www.treccani.it

ECONOMIA

Indice

ECONOMIA

Economia

di Alessandro Roncaglia, Paolo Sylos Labini

Sommario: 1. Le basi concettuali dell'economia: l'oggetto e il metodo. L'economia e la storia. L'economia e le altre scienze sociali. 2. Gli sviluppi della riflessione economica. a) L'economia preclassica. b) La nascita dell'economia politica: William Petty. c) L'economia politica classica: Adam Smith. d) David Ricardo. e) Karl Marx. f) Il marginalismo. g) La scuola austriaca. h) Dalla statica alla dinamica. i) Joseph Schumpeter e lo sviluppo ciclico dell'economia. l) Max Weber, la sociologia e il processo di burocratizzazione. m) John Maynard Keynes e la disoccupazione di equilibrio. 3. Il dibattito contemporaneo. a) La sintesi neoclassica. b) Monetaristi e teorici delle aspettative razionali. c) I postkeynesiani. d) L'offerta di moneta in Keynes e nei monetaristi. e) Piero Sraffa. 4. L'analisi dinamica. a) Diversi tipi di modelli; il processo di meccanizzazione. b) Le variazioni di lungo periodo dei

salari e dei prezzi. c) Il sottosviluppo. d) I problemi dell'ambiente e lo sviluppo sostenibile. 5. Considerazioni conclusive.

1. Le basi concettuali dell'economia: l'oggetto e il metodo. L'economia e la storia. L'economia e le altre scienze sociali

Su quali siano l'oggetto e il metodo dell'economia, esistono oggi e sono esistite in passato opinioni diverse. Per questo non cercheremo di fornire una definizione esatta di questo campo di studi, ma piuttosto di delineare in estrema sintesi le concezioni che si sono succedute nel tempo e quelle che sono presenti nel dibattito di oggi. Prima di imboccare questo sentiero, tuttavia, proveremo a fornire alcune indicazioni generali sul campo di lavoro degli economisti e sui suoi rapporti con le altre scienze. Gli economisti hanno di fronte una realtà complessa e in continuo movimento. Da questa isolano alcuni aspetti che assumono a oggetto del loro lavoro d'analisi; quindi procedono alla costruzione di teorie, introducendo ipotesi semplificatrici per isolare quelli che ritengono i principali elementi del problema che intendono affrontare. Le differenze tra i vari approcci dipendono dalle scelte compiute in questo processo di astrazione, e particolarmente da quelle compiute nella sua prima fase.Possiamo distinguere per quest'aspetto due impostazioni di fondo, radicalmente diverse, sulle quali torneremo in dettaglio più avanti: l'impostazione degli economisti classici e quella degli economisti marginalisti. Secondo la concezione classica, prevalente nel periodo della nascita dell'economia come scienza e fino alla seconda metà dell'Ottocento, l'economia politica riguarda l'organizzazione di una società caratterizzata dalla divisione del lavoro e dallo scambio di merci tramite il mercato; si tratta quindi di una scienza sociale, che studia questioni quali la distribuzione del reddito o il ritmo di accumulazione del capitale. Secondo la concezione marginalista, affermatasi a partire dagli ultimi decenni

dell'Ottocento, il problema economico coincide con quello del calcolo razionale: come ottenere il massimo risultato in presenza di un dato ammontare di risorse disponibili; si tratta quindi di un problema logico, in quanto tale suscettibile di trattazione matematica (come problema di massimo - o minimo - vincolato), e che come tutti i ragionamenti assiomatici risulta 'astorico', nel senso che la struttura dell'analisi - il complesso delle cosiddette leggi economiche, o più precisamente dei teoremi di base - non cambia quando ci si riferisca a società lontane fra loro nel tempo e nello spazio e radicalmente diverse fra loro quanto ad assetto istituzionale.
L'economista austriaco Joseph Schumpeter (1883-1950) distingueva tre fasi principali nel lavoro dell'economista: la fase della 'visione preanalitica', che suggerisce i problemi da studiare e, almeno nelle grandi linee, il modo in cui iniziare ad affrontarli; la fase della 'concettualizzazione', in cui si tenta di 'razionalizzare' la complessa realtà che ci si trova di fronte creando categorie mentali che permettono di eliminare dalla scena - di astrarre da - gli aspetti secondari, concentrando l'attenzione sulle caratteristiche distintive considerate più rilevanti (come accade, appunto, con i concetti di salario, rendita, profitto); infine, la fase della 'teorizzazione' vera e propria, in cui si collegano in strutture logiche - in modelli - gli elementi identificati nella fase precedente.
Le differenze tra approcci economici come quello classico e quello marginalista cui si accennava sopra vengono spesso discusse considerando solo l'ultima delle tre fasi di lavoro di cui parla Schumpeter, cioè i modelli teorici che costituiscono il risultato ultimo della ricerca economica; in realtà, tuttavia, queste differenze dipendono principalmente dal procedimento di astrazione-concettualizzazione, cioè dalla prima delle due fasi, e possono essere identificate in modo più preciso confrontando le diverse concezioni del problema economico e il diverso 'contenuto' dei concetti utilizzati.
Il rapporto tra l'economia e le altre scienze sociali, e in particolare tra economia

e storia, si pone in modo diverso a seconda di come viene concepita l'economia. Consideriamo ad esempio la semplice dicotomia tra impostazione classica e impostazione marginalista proposta sopra, cioè tra la concezione dell'economia come scienza sociale e storicamente determinata, e la concezione dell'economia come scienza assiomatica del comportamento razionale. Vediamo subito che nel primo caso tutta la struttura teorica - la base concettuale prima ancora che i sistemi di modelli analitici - va sviluppata e modificata nel tempo in una continua interazione tra ricerca economica e ricerca storica. Nel secondo caso, invece, quelli che mutano nel tempo sono considerati dati esogeni del problema economico: la tecnologia, le preferenze del consumatore, o altri più specifici vincoli al funzionamento del mercato, che l'economista marginalista assume come dati il cui studio è demandato a campi di ricerca separati da quello dell'economia, mentre la struttura del problema economico permane invariata nel tempo. Allo stesso modo, la concezione dell'economia come studio dei fattori che determinano l'evoluzione nel tempo della divisione del lavoro e le sue conseguenze (principalmente in termini di accumulazione, occupazione, distribuzione del reddito) implica una stretta interrelazione tra il lavoro dell'economista e quello, ad esempio, del sociologo che analizza l'evolversi della struttura sociale. Viceversa, la concezione dell'economia come scienza del comportamento razionale - ogni soggetto è un homo œconomicus che fonda ogni sua azione su un calcolo razionale - implica l'esclusione dal campo dell'economia di problemi quali il contrasto tra passioni e interessi nel determinare la condotta umana di cui tanto si è discusso dal XVI al XVIII secolo (v. ad es. Hirschman 1977), mentre le determinanti delle preferenze del consumatore, ad esempio, vengono considerate come problema di pertinenza della psicologia sociale, il cui campo d'analisi viene nettamente distinto da quello dell'economia.Un rapporto di condizionamento reciproco può essere infine individuato fra diritto ed economia. Sul piano dell'analisi è bene ricordare

che il fondatore dell'economia moderna, Adam Smith, si è occupato sistematicamente di diritto e che, per un altro verso, il sociologo Max Weber si è occupato, sia pure marginalmente, oltre che di economia, anche di diritto. Se ci poniamo dal punto di vista dello sviluppo economico, possiamo rilevare che le stesse innovazioni tecnologiche sono talvolta favorite e talvolta contrastate dal sistema giuridico, e che quest'ultimo muta nel tempo anche per effetto di impulsi generati dalle innovazioni tecnologiche: basti pensare all'elettricità, all'automobile, all'aeroplano, all'elettronica. Sui problemi dell'oggetto e del metodo dell'economia e del suo rapporto con le altre scienze sociali avremo più volte occasione di tornare nel seguito, esaminando lo sviluppo del pensiero economico. Qui ci limitiamo a sottolineare che l'economia, come tutte le scienze che studiano la società, è storicamente condizionata, giacché la società stessa cambia in modo irreversibile nel tempo storico: le scienze sociali vanno viste come cerchi che in parte si sovrappongono e che si muovono tutti nella storia.

2. Gli sviluppi della riflessione economica

a) L'economia preclassica

L'economia politica inizia a essere riconosciuta come disciplina distinta dalle altre scienze sociali assai gradualmente, a partire dalla seconda metà del XVII secolo; ma solo nella seconda metà del secolo scorso, con l'istituzione delle prime cattedre di economia nelle università, l'economista viene identificato come una figura professionale autonoma.Naturalmente, cenni più o meno sparsi a problemi oggi comunemente ritenuti di competenza degli economisti appaiono già nell'antichità e nel Medioevo. Autori come Diodoro Siculo, Senofonte o Aristotele considerano ad esempio gli aspetti economici della divisione del lavoro, sostenendo fra l'altro che essa permette di raggiungere una

migliore qualità del prodotto. Ma nel complesso per lungo tempo - almeno fino al XVII secolo - i problemi economici sono stati affrontati in modo sostanzialmente diverso da come li si affronta oggi. I filosofi e i teologi medievali in particolare, più che tentar di descrivere e interpretare il modo di funzionamento del sistema economico, si proponevano il compito di fornire indicazioni sul comportamento moralmente più giusto da tenere nel campo dei rapporti economici. Così il problema dell'usura non era quello di fornire la spiegazione del livello del tasso d'interesse, ma quello di giustificare la condanna morale del prestito a interesse e di individuare le eccezioni a questo rigoroso precetto morale. Analogamente, il problema del giusto prezzo riguardava non il tentativo di spiegare perché sul mercato i vari beni venivano scambiati a certi prezzi piuttosto che a prezzi più alti o più bassi, ma il tentativo di fornire ai mercanti criteri di condotta. (In questo senso, e non come anticipazioni delle successive teorie del valore classica e marginalista, vanno interpretati i riferimenti ai costi di produzione dei beni scambiati, o alla loro scarsità e utilità per l'acquirente, come criteri di riferimento da utilizzare per fissare rapporti di scambio moralmente corretti.).

Per comprendere il cambiamento intervenuto attorno al XVII secolo nel modo di considerare i problemi economici, occorre ricordare i radicali mutamenti verificatisi nell'organizzazione della vita economica e sociale. Il 'mercato' - inteso come scambio di beni contro denaro - esisteva anche nell'Atene di Pericle o nella Roma di Cesare; ma gli scambi coprivano una quota relativamente limitata della produzione complessiva della società, e le condizioni in cui si svolgevano erano caratterizzate dalla massima irregolarità, a causa di fattori quali l'incidenza dei fenomeni meteorologici sui raccolti, le difficoltà dei trasporti, la diffusa insicurezza rispetto alla criminalità privata e agli interventi arbitrari delle autorità. (Per il primo aspetto - portata limitata degli scambi - possiamo ricordare, ad esempio, che nel Medioevo, secondo Kula - v., 1970 -,

gli scambi sul mercato interessano tipicamente solo i beni che rientrano nel sovrappiù, cioè quella parte del prodotto che non è necessaria come mezzo di produzione o di sussistenza per continuare l'attività produttiva. In particolare esisteva già una rete di scambi di prodotti 'di lusso' - spezie, merletti, metalli preziosi - che collegava fra loro aree geografiche anche assai distanti; accanto a essa, era gradualmente sorta una rete di rapporti finanziari tra i maggiori centri commerciali, basata soprattutto su 'lettere di cambio'. Ma l'autoproduzione - cioè la produzione per il consumo diretto dei produttori - che caratterizza le piccole comunità agricole perde terreno rispetto alla produzione per il mercato solo con l'affermarsi del sistema di fabbrica, quando i lavoratori non sono proprietari dei beni che producono, e comunque questi ultimi sono beni diversi da quelli che essi consumano, e quando diviene comune l'utilizzo di mezzi di produzione specializzati - macchinari e impianti - prodotti da altre imprese. Per il secondo aspetto - irregolarità degli scambi - richiamiamo solo una delle manifestazioni più caratteristiche dell'assenza di uniformità nelle condizioni di scambio: la molteplicità e continua variabilità delle unità di misura delle merci - unità di peso, di lunghezza, di volume -, superata solo gradualmente a partire, non a caso, dal XVII secolo).

Proprio l'assenza di regolarità e uniformità nell'attività economica è probabilmente alla base dei generici rinvii che gli autori del periodo fanno alle condizioni di domanda e di offerta come determinanti dei prezzi sul mercato. In presenza di una forte variabilità della domanda e dell'offerta, e in assenza di indicazioni sui fattori che le determinano, questi generici rinvii non possono essere considerati una teoria del prezzo, né tantomeno un'anticipazione della teoria marginalista. Quest'ultima, infatti, come vedremo meglio più avanti, spiega il prezzo di equilibrio dei vari beni come corrispondente al punto d'incontro tra domanda e offerta del bene in questione, considerate come funzioni (continue e differenziabili) - la prima decrescente e la seconda

crescente - del prezzo del bene stesso, ed eventualmente di altre variabili come i prezzi degli altri beni e il reddito dei consumatori; mentre nei generici rinvii alle condizioni di domanda e offerta cui si accennava sopra non è riscontrabile - del tutto comprensibilmente, date le condizioni dell'epoca - alcuna idea di una ben specificata e stabile relazione funzionale tra la domanda e l'offerta, e altre variabili come il prezzo del bene in questione.

b) La nascita dell'economia politica: William Petty

Un punto di svolta fondamentale è costituito dalla nascita e dal consolidarsi, attorno al XV-XVI secolo, degli Stati nazionali, i quali vengono posti al centro della riflessione degli scienziati sociali; a essi corrisponderà, in campo economico, la nozione di sistema economico che fin dal XVI secolo costituisce un concetto-base per la nascente scienza dell'economia politica. In breve: ai tempi dell'antica Grecia e di Roma gli scambi sul mercato sono massimamente irregolari; poi, via via, diventano meno irregolari; ma è solo nel secolo XVII che cominciano ad assumere caratteristiche di relativa regolarità e uniformità, dopo i cambiamenti politici che si sono affermati nei secoli XV, XVI e XVII. Nello stesso periodo, cambiamenti decisivi si verificano anche nel campo delle scienze naturali, nelle quali si afferma l'idea che compito dello scienziato è quello di 'scoprire', al di sotto della superficie complessa e variegata dei fenomeni sperimentati dai nostri sensi, le 'leggi' - cioè relazioni quantitative precise e invariabili - che costituiscono l'intelaiatura del mondo naturale. La ricerca di leggi quantitative corrisponde, nel pensiero di Hobbes (1588-1679), a una concezione materialistico-meccanica dell'uomo e del mondo: il metodo d'indagine - la logica delle quantità (logica sive computatio, come dice Hobbes) - corrisponde alla natura stessa dell'oggetto indagato. A parere di vari studiosi dell'epoca, la perfezione matematica delle leggi naturali è sanzionata dal fatto che in essa si rispecchia la mano del Creatore. Caratteristica in questo senso è

l'affermazione di Galileo (1564-1642), secondo il quale "questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo) [...] è scritto [da Dio] in lingua matematica" (cfr. G. Galilei, Il saggiatore, 1623, in Opere, 1890-1909, a cura di A. Favaro, vol. VI, Firenze 1896, p.232). Così fra il XVII e il XVIII secolo la vecchia fisica e la vecchia chimica intese come descrizione delle qualità sensibili degli oggetti lasciano il campo alla scienza moderna di Newton (1642-1727) e Lavoisier (1743-1794). Il metodo induttivo baconiano, fusione di empirismo e razionalismo, è ripreso alla metà del XVII secolo dai fondatori della Royal Society inglese, fra i quali appare una figura chiave per la nascita dell'economia politica, William Petty (1623-1687).
Fra gli autori che, nel XVI e XVII secolo, affrontano questioni economiche, sono ancora numerosi quelli che ne discutono considerandole innanzitutto un problema morale, come accade negli innumerevoli trattati sull'usura pubblicati in quel periodo. Ma sono molti anche quanti, sulla scia di Machiavelli (1469-1527), indirizzano le loro opere ai sovrani consigliandoli su come mantenere e accrescere la potenza economica dei loro Stati: i cosiddetti cameralisti, spesso osservatori intelligenti della realtà in cui operano, ma nel complesso non molto sensibili a quella esigenza di sistematicità che costituisce un requisito fondamentale della scienza moderna. Di frequente i cameralisti vengono inclusi, assieme a vari altri autori di scritti sull'economia del periodo che va dal XVI alla seconda metà del XVIII secolo, sotto l'etichetta di mercantilisti (un'etichetta che in realtà vari storici del pensiero economico considerano troppo generica, riferita ad autori molto diversi l'uno dall'altro). Il termine 'mercantilisti' designa tradizionalmente i fautori di uno Stato interventista, in particolare attraverso dazi sul commercio e divieti alle importazioni, al fine di assicurare una bilancia commerciale attiva e quindi l'afflusso di metalli preziosi, ma anche al fine di favorire lo sviluppo delle manifatture nazionali proteggendole dalla concorrenza estera.Lo spirito 'laico'

di Niccolò Machiavelli e dei cameralisti - per fare solo due esempi che illustrano la nuova concezione del mondo che comincia ad affermarsi - si fonde con la nuova metodologia scientifica dell'induttivismo baconiano e con la concezione materialistico-meccanica di Hobbes nel caso di William Petty, il fondatore dell' 'aritmetica politica'. "Il metodo che intendo seguire è tuttora inconsueto: invece di usare solo comparativi e superlativi, e argomenti intellettuali, ho deciso di esprimermi in termini di numero, peso e misura (come esempio dell'aritmetica politica che ho tanto perseguito); di usare solo argomenti fondati sulla sensazione, e di considerare unicamente quelle cause che hanno fondamenta visibili nella Natura; lasciando quelle che dipendono dalle mutevoli menti, opinioni, appetiti e passioni di singoli uomini, alla considerazione di altri." (v. Petty 1963, p.244).
L'aritmetica politica, come si vede, si contrappone al metodo logico-deduttivo proprio della Scolastica: Petty non intende semplicemente rilevare e descrivere la realtà "in termini di numero, peso e misura", ma piuttosto sceglie di esprimersi in quei termini nel tentativo di interpretare la realtà individuandone le caratteristiche principali. In questo senso, e per il suo contributo alla formazione del sistema dei concetti (prezzo naturale, merce e mercato, sovrappiù, ecc.) che costituiscono le fondamenta della scienza economica, Petty può essere considerato - come afferma Marx - il fondatore dell'economia politica. Prima di Adam Smith, al quale spesso viene attribuito questo ruolo, dobbiamo ancora ricordare almeno François Quesnay, medico di Madame de Pompadour e fondatore della scuola dei fisiocrati, come più tardi furono chiamati i suoi seguaci (allora noti come les économistes): sostenitori dell'esistenza di un 'ordine naturale' che, pur non implicando un rifiuto assoluto dell'intervento dello Stato nell'economia, avrebbe dovuto comportare una maggiore libertà da vincoli per gli sviluppi spontanei dell'economia. Quesnay (1694-1774) fornisce, con i suoi celebri Tableaux économiques, una

rappresentazione analitica del funzionamento del sistema economico, come società divisa in classi sociali tra le quali ha luogo un circuito di scambi che realizza contemporaneamente la distribuzione del reddito tra le diverse classi e la condizione per il ripetersi regolare del processo produttivo: l'attività economica è concepita come un 'flusso circolare' di produzione e di scambio, una concezione che caratterizzerà tutta l'economia politica classica fino a Marx e che, come vedremo, è stata riproposta in epoca recente da Piero Sraffa.

c) L'economia politica classica: Adam Smith

Adam Smith (1723-1790) non può essere considerato il fondatore dell'economia politica, ma è comunque assai più di un semplice sistematizzatore di teorie sviluppate dai suoi predecessori. A lui infatti dobbiamo una visione moderna dei compiti e del metodo dell'economia politica, e alcuni importanti contributi analitici attuali per vari aspetti ancora oggi.Il problema al centro della riflessione di Smith è quello della divisione del lavoro: come può funzionare (sopravvivere e prosperare) una società basata sulla divisione del lavoro? Nell'affrontare questo problema Smith raccoglie in un sistema teorico ben strutturato varie linee d'analisi proposte e sviluppate da autori precedenti. Così Smith riprende dai fisiocrati la concezione di produzione, distribuzione e consumo del reddito come processo circolare; e dal filone dell'illuminismo sociologico scozzese una visione della società in cui 'passioni' e 'interessi' si intrecciano e possono integrarsi, anziché entrare necessariamente in conflitto; e su queste basi sviluppa la sua teoria della 'natura e cause della ricchezza delle nazioni', basata sul legame tra sviluppo della divisione del lavoro e allargamento degli scambi di mercato.

Vediamo meglio, sia pur solo per brevi cenni, i principali aspetti di questo sistema teorico, esposto da Smith in un ampio trattato organico di economia politica, Indagine sulla natura e le cause della ricchezza delle nazioni (1776),

che costituirà la base dei successivi sviluppi dell'economia politica classica fino a Marx. Innanzitutto, come si è appena accennato, l'approccio classico rappresenta il funzionamento del sistema economico come un processo circolare (o, meglio ancora, a spirale). All'inizio del processo produttivo abbiamo determinate quantità di varie merci, che vengono utilizzate come mezzi di produzione e come mezzi di sussistenza per i lavoratori impiegati nel processo produttivo. Al termine del processo produttivo otteniamo un prodotto che è costituito dalle stesse merci, anche se in quantità diverse. Possiamo dire che il sistema produttivo è 'vitale' se la quantità prodotta di ogni merce è eguale o superiore alla quantità della stessa merce che era stata utilizzata nei vari settori come mezzo di produzione o di sussistenza. Il sovrappiù (un concetto centrale nelle analisi della scuola classica, prima e dopo Smith) è costituito dall'eccesso del prodotto rispetto alle quantità utilizzate nel processo produttivo; il sovrappiù è quindi composto di varie merci.In un sistema economico basato sulla divisione del lavoro, il prodotto delle imprese appartenenti a ciascun settore non corrisponde al fabbisogno di mezzi di produzione del settore stesso (inclusi i mezzi di sussistenza per i lavoratori impiegati nel settore); perciò ciascun settore preso isolatamente non è in grado di continuare la propria attività, ma deve entrare in contatto con gli altri settori dell'economia per ottenere da essi i propri mezzi di produzione in cambio di una parte almeno del proprio prodotto. Si ha così quella rete di scambi che caratterizza le economie 'di mercato'.

I rapporti di scambio tra i vari settori, e quindi tra le varie merci, sono vincolati dalla necessità di permettere la 'riproduzione' del sistema: ogni settore deve ottenere, grazie alla cessione dei propri prodotti, almeno i mezzi di produzione e di sussistenza necessari per continuare il processo produttivo. Ma, data l'esistenza del sovrappiù, questa condizione da sola non è in grado di permettere la determinazione univoca dei rapporti di scambio tra le varie merci: questi

rapporti di scambio infatti, oltre a garantire a ciascun settore la disponibilità di mezzi di produzione e di sussistenza per continuare la produzione, determinano la ripartizione del sovrappiù tra i vari settori e tra le varie classi sociali. Quesnay assumeva rapporti di scambio tali che di tutto il sovrappiù, generato dalla classe 'produttiva' (cioè nel settore agricolo), si appropriavano le classi proprietarie (nobili e clero), mentre la classe 'sterile' (imprenditori e lavoratori delle manifatture) non faceva altro che trasmettere nel prodotto il valore dei mezzi di produzione e di sussistenza impiegati nel processo produttivo. Smith invece distingue tra suddivisione in settori del sistema economico e suddivisione in classi sociali della popolazione; e propone quella che dopo di lui è divenuta la partizione standard in tre classi sociali: lavoratori, capitalisti e proprietari terrieri, cui corrispondono tre tipi di reddito: salari, profitti e rendite.Nell'analizzare il funzionamento del sistema sociale così schematizzato, Smith abbandona la vecchia identificazione tra ricchezza delle nazioni e produzione economica complessiva del paese, tipica del punto di vista degli economisti 'consiglieri del principe'; e adotta una concezione moderna, identificando la ricchezza di un paese con il grado di sviluppo raggiunto dal sistema economico, misurabile attraverso il reddito pro capite. Di conseguenza, per Smith spiegare la ricchezza delle nazioni significa essenzialmente spiegare il livello del reddito pro capite.

Tale livello corrisponde al prodotto di due variabili: la produttività per lavoratore e la quota dei lavoratori produttivi sul totale della popolazione. La seconda variabile dipende sia da fattori economici sia dalle consuetudini sociali; Smith contrappone ripetutamente le consuetudini di una società feudale a quelle della nascente società capitalistica a tutto vantaggio di quest'ultima. La produttività per lavoratore dipende dallo sviluppo raggiunto dalla divisione del lavoro (grazie, dice Smith, "a tre diverse circostanze: primo, all'aumento di destrezza di ogni singolo operaio; secondo, al risparmio del tempo che di solito

si perde per passare da una specie di lavoro a un'altra; e infine all'invenzione di un gran numero di macchine che facilitano e abbreviano il lavoro e permettono a un solo uomo di fare il lavoro di molti"). Ma la divisione del lavoro progredisce solo se c'è un allargamento dei mercati, che è necessario per assorbire il maggior prodotto di un'impresa in cui cresce il numero dei lavoratori e simultaneamente cresce la produttività di ciascuno di essi. Di qui la posizione liberista di Smith. Infatti egli è favorevole all'abbattimento degli ostacoli al libero commercio, che può mettere in moto una spirale virtuosa: l'allargamento dei mercati favorisce una crescente divisione del lavoro, e quindi un aumento della produttività, che permette un aumento del reddito pro capite, e quindi un ulteriore allargamento dei mercati. L'espansione dell'area del mercato nella società, inoltre, stimola un aumento della quota dei lavoratori produttivi sul totale della popolazione.

Il coordinamento delle attività individuali è assicurato dal mercato, che 'premia' quanti producono ciò che è maggiormente richiesto, e 'punisce' quanti si dedicano ad attività che gli altri considerano inutili. L'interesse individuale agisce quindi da molla per il conseguimento dell'interesse collettivo.

Tuttavia, come si accennava sopra, ciò non significa che Smith sposi le tesi del medico e filosofo olandese Bernard de Mandeville (1670-1733): tesi sintetizzate nel detto "vizi privati, pubbliche virtù", e centrate sul riconoscimento dell'egoismo come fondamento del benessere collettivo. Nella sua opera La teoria dei sentimenti morali, infatti, Smith sottolinea il ruolo di una 'morale della simpatia' (nel senso etimologico di 'sentire insieme') nel guidare le azioni individuali. La ricerca dell'interesse personale risulta automaticamente vincolata da tale morale, escludendo quelle azioni (come la sofisticazione degli alimenti da parte dei commercianti) che potrebbero danneggiare altri. La diffusa adesione dei cittadini alla 'morale della simpatia', affiancata e sostenuta dalle istituzioni pubbliche per la difesa dell'ordine e l'amministrazione della giustizia,

costituisce dunque un presupposto necessario per la tesi liberista, sostenuta da Smith nella Ricchezza delle nazioni, secondo cui il benessere collettivo è meglio conseguito affidandosi al libero perseguimento degli interessi individuali nell'ambito di un'economia di mercato concorrenziale. Conviene sottolineare qui che Smith propone questa tesi, senza sviluppare un concetto preciso di massimizzazione del benessere collettivo e prescindendo dalla distribuzione del reddito tra le diverse classi sociali, essenzialmente in base al confronto fra le istituzioni feudali ancora persistenti nella sua epoca e le nuove istituzioni dell'economia di mercato, ancora in via di affermazione. I meriti della costruzione smithiana, come si vede, risiedono principalmente nell'aver fornito un quadro concettuale organico e solido per l'interpretazione dello sviluppo dei sistemi economici. Per vari aspetti tuttavia l'analisi di Smith lasciava ampio spazio per ulteriori progressi, specie nel senso di una maggiore coerenza formale; e ciò anche su questioni centrali come la teoria del valore e della distribuzione.

d) David Ricardo

La costruzione di una solida struttura analitica per l'economia politica classica costituisce il contributo principale dell'opera dell'inglese David Ricardo (1772-1823). Punto di partenza della sua riflessione analitica è il contrasto d'interessi tra i proprietari terrieri, politicamente dominanti, e la nascente borghesia industriale, contrasto che si manifestava nel dibattito sui dazi all'importazione dei cereali. A parere di Ricardo, i dazi spingevano verso l'alto i prezzi dei cereali e quindi, data la loro importanza nei consumi dei lavoratori, anche i salari; ciò spingeva verso il basso i profitti, rallentando il processo di accumulazione e frenando lo sviluppo del sistema economico.Alla base dell'analisi di Ricardo è il concetto di sovrappiù, definibile - come si è già accennato - come la parte del prodotto che resta disponibile una volta

ricostituite le scorte iniziali di mezzi di produzione e di mezzi di sussistenza per i lavoratori impiegati. Ricardo, seguendo la teoria della popolazione di Thomas R. Malthus (1766-1834), assume che il salario sia pari al livello di sussistenza (inteso non in senso biologico, ma come minimo tenore di vita accettabile da parte dei lavoratori di una data società in un dato periodo storico), e in tal modo considera i salari come corrispondenti ai consumi necessari dei lavoratori. Perciò, accogliendo la tripartizione smithiana della società nelle classi dei lavoratori, proprietari terrieri e capitalisti, il sovrappiù risulta ripartito tra rendite - che vengono utilizzate essenzialmente in consumi di lusso - e profitti, largamente destinati agli investimenti.Il problema della rendita viene poi superato grazie alla teoria della rendita differenziale; secondo tale teoria, la rendita viene pagata solo sulle terre più fertili e corrisponde alla differenza tra i costi di produzione relativi alla meno fertile tra le terre in coltivazione, e i costi relativi alle terre più fertili.
Il profitto risulta così definito come grandezza residuale, cioè come quella parte del sovrappiù che non viene assorbita dalla rendita. Il saggio di profitto è allora determinabile come rapporto tra profitti e capitale anticipato. Ma ciò richiede che profitti e capitale anticipato siano espressi in termini di grandezze omogenee: o interpretandoli come quantità diverse di una stessa merce, il 'grano' (come, secondo l'interpretazione di Sraffa, Ricardo fa in un saggio del 1815), o misurando prodotto e mezzi di produzione in valore, come Ricardo fa nei Principî dell'economia politica e della tassazione del 1817. In quest'opera, infatti, egli - come vari altri economisti classici prima e dopo di lui - ricorre a una teoria del valore-lavoro contenuto, secondo la quale le merci si scambiano in proporzione alle quantità di lavoro direttamente o indirettamente necessario alla loro produzione; ma occorre sottolineare che Ricardo è consapevole delle carenze di questa teoria. (Tali carenze consistono essenzialmente nel fatto che beni prodotti con le stesse quantità di lavoro ma su un arco di tempo più o meno

lungo - ad esempio vino 'giovane' e vino 'invecchiato' - dovranno avere prezzi diversi).Ricardo è anche noto per la sua 'teoria dei costi comparati', con cui mostra i vantaggi della divisione internazionale del lavoro permessa dal commercio internazionale.

e) Karl Marx

La struttura analitica sviluppata dalla scuola classica fino alle opere di Ricardo viene ripresa dal tedesco Karl Marx (1818-1883), che pone però al centro della sua attenzione il contrasto d'interessi tra borghesia e proletariato.Con la sua 'teoria dello sfruttamento', Marx cerca di mostrare che i profitti derivano da 'lavoro non pagato' pur in un sistema che rispetti il criterio capitalistico di scambio tra valori eguali. A tale scopo egli riprende la teoria del valore-lavoro contenuto - dei cui limiti, peraltro, è ben consapevole. (Marx tentò anche - ma senza successo - di mostrare che i risultati ottenuti in base a tale teoria restano validi quando ci si basa piuttosto sui 'prezzi di produzione', che rispettano la condizione di uniformità del saggio di profitto nei vari settori). Marx, dunque, assume che le merci si scambino sul mercato in proporzione al lavoro contenuto in esse, cioè alle quantità di lavoro direttamente e indirettamente necessario a produrle. Ciò vale anche per la forza-lavoro, che ha un valore pari al lavoro contenuto nei mezzi di sussistenza che costituiscono il consumo necessario dei lavoratori. Il valore d'uso della forza-lavoro (che è cosa diversa dal valore di scambio, così come il valore di scambio del carbone è ben distinto dal suo valore d'uso, che consiste nel fornire calore) sta nel fornire lavoro al capitalista che l'ha acquistata: tanto lavoro quanto il capitalista riesce a ottenere e quindi, date le consuetudini sociali che regolano la lunghezza della giornata lavorativa, di regola un numero di ore di lavoro superiore a quelle 'contenute' nella forza-lavoro stessa.

Questo eccesso di lavoro prestato rispetto a quello richiesto per la semplice

riproduzione dei beni necessari al consumo dei lavoratori, o 'pluslavoro', corrisponde al valore del 'plusprodotto' o sovrappiù, e costituisce quindi la fonte dei profitti e delle rendite.Questa situazione di sfruttamento è celata agli occhi dell'osservatore superficiale dal 'feticismo delle merci', cioè dal fatto che i rapporti di cooperazione per il funzionamento del sistema economico tra i lavoratori appartenenti ai vari settori appaiono in un'economia capitalistica come rapporti di scambio tra merci sul mercato. Occorre rilevare che nel Marx del Capitale il concetto di sfruttamento sostituisce quello giovanile, di derivazione smithiana ed hegeliana, di alienazione, corrispondente al fatto che sono sottratti al controllo del lavoratore, e costituiscono quindi 'altro da sé' (alius), sia i mezzi di lavoro, sia il prodotto del lavoro, sia il processo produttivo, e che, di conseguenza, il lavoro risulta per il lavoratore come un mezzo per un fine distinto - procurarsi il salario, e quindi i mezzi di sussistenza -, anziché come diretta autorealizzazione dell'individuo nella società.
Lo sfruttamento che caratterizza il modo di produzione capitalistico (e, in forma ancor più diretta, i modi di produzione precedenti, come il feudalesimo e lo schiavismo) è a parere di Marx superabile con il passaggio a un modo di produzione più avanzato, il comunismo. La transizione dal capitalismo al socialismo, caratterizzato dalla proprietà collettiva dei mezzi di produzione e che sembra costituire nella visione di Marx una fase di preparazione al comunismo, sarebbe inevitabile conseguenza di alcune 'leggi di movimento del capitalismo', in particolare la crescente bipolarizzazione tra un proletariato sempre più vasto ('processo di proletarizzazione') e sempre più povero ('legge dell'immiserimento crescente') e una borghesia sempre più ricca e più forte ('legge della concentrazione capitalistica'), che sarebbe alla fine sfociata in una rivoluzione. Di fatto, questa parte almeno della visione di Marx è stata direttamente contraddetta dall'evoluzione concreta delle società capitalistiche, nelle quali si sono verificati un notevole miglioramento del tenore di vita dei

lavoratori e la crescita di una classe media impiegatizia e professionale, con interessi distinti sia da quelli dei grandi capitalisti industriali e finanziari sia da quelli dei lavoratori non qualificati.Marx prospetta due schemi di analisi generale, quello della riproduzione semplice e quello della riproduzione su scala allargata, o dell'accumulazione. Il primo rientra nella categoria degli schemi circolari, il secondo nella categoria degli schemi a spirale, in cui una parte almeno del sovrappiù è impiegata produttivamente.
Entrambi gli schemi debbono rispettare la condizione della riproducibilità: le quantità delle merci prodotte che entrano fra i mezzi di produzione debbono essere eguali o superiori alle quantità delle stesse merci impiegate nel processo produttivo. Inoltre, se si fa l'ipotesi che in tutti i mercati viga la concorrenza, deve verificarsi una seconda condizione: l'unicità del saggio del profitto. In una successiva approssimazione si può fare l'ipotesi che certi mercati siano non concorrenziali, con profitti più alti della norma; tuttavia, prima d'introdurre una tale ipotesi occorre assicurarsi della coerenza logica dell'analisi nel caso più semplice della concorrenza generalizzata, dimostrando che in una tale ipotesi il saggio del profitto è unico. Ora, già nello schema di riproduzione semplice di Marx, la prima condizione - quella della riproducibilità - è rispettata, ma non lo è la seconda, quella relativa all'unicità del saggio del profitto. È questo il famoso problema della trasformazione dei valori-lavoro in prezzi di produzione cui si è già accennato; per esso Marx indica, con un breve accenno, una soluzione puramente intuitiva, proponendosi di fornirne in seguito una trattazione analitica, ciò che tuttavia poi non fa effettivamente. Seguendo l'analisi di Sraffa - di cui parleremo più avanti - l'erroneità della teoria del valore-lavoro risulta in modo ancora più compiuto: fra i 'valori' espressi dal lavoro contenuto e i prezzi vi è proporzionalità solo nell'ipotesi, non rilevante in un'economia capitalistica, di un saggio dell'interesse, e quindi del profitto, pari a zero.

La teoria del valore-lavoro si muove a un alto livello di astrazione. A un livello molto più vicino alla realtà concreta si muove invece una tesi che in Marx è appena accennata, ma che assume poi grande rilievo per essere stata sostanzialmente accolta anche da economisti non marxisti come Joseph Schumpeter e lo statunitense John Kenneth Galbraith (n. 1908): è la tesi della tendenza alla progressiva concentrazione delle unità produttive, un processo che troverebbe il suo principale impulso nella diffusione delle economie di scala, a loro volta imputabili al progresso scientifico e organizzativo. Un tale processo sembrava trovare ampio riscontro nel fenomeno delle fusioni e poi nella comparsa e nella crescita di grandi e grandissime imprese, nazionali e transnazionali, un fenomeno che assume particolare rilievo tra la fine del secolo scorso e il principio del nostro e quindi dopo la seconda guerra mondiale. Secondo diversi economisti questo processo avrebbe spinto le economie capitalistiche nella direzione di una sorta di collettivismo privato, preludio di un collettivismo pubblico. Tuttavia, negli ultimi due decenni il processo di concentrazione ha subito un netto rallentamento o addirittura un rovesciamento, in quanto, grazie anche a grandi innovazioni, come quelle connesse con l'elettronica, si sono creati nuovi spazi per le piccole imprese. Inoltre, la tesi del processo di concentrazione trascura il fatto che non di rado le piccole imprese avviano innovazioni poi sviluppate dalle grandi imprese: in tal modo, le piccole imprese possono contrastare gli effetti negativi sull'efficienza e sulla capacità d'innovare che le grandi imprese potrebbero subire con la loro burocratizzazione.

f) Il marginalismo

Il momento cruciale per la 'rivoluzione marginalista' è comunemente indicato negli anni tra il 1871 e il 1874, quando escono le opere principali dei capifila della scuola marginalista austriaca, Carl Menger (1840-1921), di quella inglese,

William Stanley Jevons (1835-1882), e di quella francese, Léon Walras (1834-1910). Infatti nel 1871 escono sia i Principî di economia pura di Menger sia la Teoria dell'economia politica di Jevons, nel 1874 gli Elementi di economia politica pura di Walras. Va ricordato comunque che la rivoluzione marginalista aveva avuto importanti precursori, quali il francese Antoine-Augustin Cournot (1801-1877), e che tra i suoi cofondatori, per quanto riguarda il ramo inglese, Alfred Marshall (1842-1924; i suoi Principî di economia appaiono nel 1890) riveste probabilmente un'importanza superiore a quella di Jevons; in Italia, le idee di Marshall conseguono una discreta diffusione grazie soprattutto agli scritti di Maffeo Pantaleoni (1857-1924). Le differenze fra l'approccio austriaco dell'imputazione, quello francese dell'equilibrio economico generale, e quello marshalliano degli equilibri parziali sono notevolissime, sia per quanto riguarda il metodo sia per quanto riguarda la visione di fondo del funzionamento del sistema economico. Ma prima di considerare più da vicino i tre principali filoni di ricerca tradizionalmente inclusi sotto l'etichetta del marginalismo conviene indicare alcuni elementi di fondo che li accomunano, contrapponendoli all'approccio classico illustrato sopra.

La contrapposizione è stata sintetizzata da Sraffa individuando nell'approccio classico la "presentazione del sistema della produzione e del consumo come processo circolare", "in netto contrasto con l'immagine offerta dalla teoria moderna di un corso a senso unico che porta dai 'fattori della produzione' ai 'beni di consumo'" (v. Sraffa, 1960, p. 121). In questo modo, Sraffa richiama sinteticamente le differenze di fondo tra l'approccio classico e quello marginalista, relative alla visione del problema economico e del funzionamento dell'economia, e alla struttura dell'analisi, particolarmente nel campo della teoria del valore e della distribuzione che è quello dove più direttamente si esprime l'impostazione di fondo dei diversi approcci economici. Come si

accennava sopra presentando l'approccio classico, esso ha al centro della sua analisi il problema del modo in cui può riprodursi, e crescere nel tempo, una società basata sulla divisione del lavoro; l'approccio marginalista invece considera il problema economico come consistente nell'individuare l'ottima distribuzione delle risorse scarse tra gli usi alternativi possibili, date le preferenze dei soggetti economici.
Pertanto, l'impostazione marginalista tradizionale concepisce il problema del valore come relativo alla determinazione di prezzi e quantità 'di equilibrio', cioè tali da assicurare l'eguaglianza tra domanda e offerta, derivati dal confronto fra le dotazioni iniziali di risorse produttive e le preferenze degli agenti economici. (L'impostazione non muta nella sostanza, quando dai modelli di puro scambio si passa ai modelli di scambio e produzione, di modo che il rapporto tra le dotazioni iniziali e la soddisfazione dei desideri è mediato dall'attività produttiva, oltre che dall'attività di scambio e di consumo; o quando fra le risorse produttive sono inclusi anche mezzi di produzione prodotti).
Di fronte al problema del valore così concepito, la tesi centrale delle teorie marginaliste tradizionali (quelle che Keynes chiama teorie classiche, e tra le quali prende ad esempio l'analisi di Arthur Cecil Pigou, 1877-1959, successore di Marshall sulla cattedra di economia politica dell'Università di Cambridge) è che un sistema economico in cui prevalga la concorrenza perfetta e che non sia soggetto a continui disturbi esogeni tende a raggiungere una posizione di equilibrio dotata di caratteristiche di ottimalità, nel senso che non è possibile migliorare la posizione di qualche soggetto economico senza peggiorare la posizione di qualcun altro. In particolare, le teorie marginaliste tradizionali sostengono che in regime di concorrenza pura il salario reale si muove, assieme a tutti gli altri prezzi relativi, verso un livello tale da assicurare l'eguaglianza tra domanda e offerta di lavoro, cioè la piena occupazione.Fra i meccanismi riequilibratori automatici che portano il sistema economico verso la piena

occupazione, alcuni teorici marginalisti hanno messo in rilievo la flessibilità del rapporto tra capitale e lavoro: se sotto la pressione della disoccupazione il salario reale diminuisce, le imprese troveranno più conveniente utilizzare tecniche produttive che sostituiscono lavoratori a capitale, di modo che il rapporto capitale-lavoro diminuisce, e una data dotazione originaria di 'capitale' permette d'impiegare un numero di lavoratori via via crescente, fino a giungere alla piena occupazione. Questa tesi e in generale l'impostazione stessa della teoria del valore e della distribuzione, come si accennava, hanno assunto forme diverse in autori appartenenti a filoni diversi dell'approccio marginalista tradizionale. Ad esempio il filone francese dell'equilibrio economico generale inaugurato da Walras, ripreso e sviluppato al principio del secolo dall'italiano Vilfredo Pareto (1848-1923) e poi, negli ultimi trent'anni, da autori quali lo statunitense Kenneth Arrow (n. 1921, premio Nobel nel 1972) e il francese Gerard Debreu (n. 1921, premio Nobel nel 1983), è basato sull'ipotesi che siano considerate date, da un lato, le preferenze dei soggetti, dall'altro le quantità disponibili, specificate in termini fisici, delle risorse iniziali (diversi tipi di capacità lavorative, di terre, di beni capitali).
Il filone inglese di Jevons e Marshall, a differenza del filone francese, tende a considerare come variabili da determinare all'interno della teoria anche le quantità disponibili delle varie risorse, utilizzando come dati esogeni le funzioni di utilità e disutilità dei vari soggetti economici (cioè considerando l'offerta di lavoro, ad esempio, come determinata dal confronto fra il salario, e quindi l'utilità dei beni acquistabili con esso, e il sacrificio che il lavoratore compie, cioè la disutilità del lavoro). Inoltre Marshall presenta la sua analisi come tentativo di sintesi tra l'approccio classico e quello marginalista 'puro', sostenendo che l'analisi classica commette l'errore di concentrarsi 'sul lato dell'offerta', mentre quella marginalista nella versione più estrema commette l'errore speculare di concentrarsi 'sul lato della domanda'. Marshall, viceversa,

concentrando l'attenzione sulla singola impresa o sulla singola industria (con il metodo 'dell'analisi parziale', cioè considerando offerta e domanda di ciascun bene come indipendenti da quanto contemporaneamente avviene sugli altri mercati), individua prezzo e quantità 'normali'. Le variabili 'normali' corrispondono a una situazione di equilibrio del mercato considerato, e sono determinate dall'intersezione di una curva di offerta, che esprime l'andamento dei costi di produzione al variare della quantità prodotta, e una curva di domanda, che esprime le preferenze del consumatore. Le due curve rappresentano due funzioni che legano la quantità offerta e quella domandata all'andamento del prezzo del singolo bene. L'offerta è considerata funzione crescente del prezzo, almeno a partire da un certo livello di produzione in poi, in quanto, secondo il cosiddetto 'postulato della produttività marginale decrescente', il costo necessario a ottenere una unità addizionale di prodotto cresce al crescere della quantità prodotta, e pertanto gli imprenditori sono disposti ad aumentare la quantità offerta solo se il prezzo aumenta compensando l'aumento dei costi.
Le varie giustificazioni addotte per il postulato della produttività marginale decrescente sono comunque risultate o irrealistiche, o incompatibili con l'ipotesi di concorrenza e/o con l'ipotesi di 'isolabilità' delle vicende del mercato considerato rispetto agli altri mercati, ipotesi che è alla base del metodo degli equilibri parziali. Di fatto, il postulato della produttività marginale decrescente è motivato essenzialmente da un'esigenza analitica di simmetria con il postulato dell'utilità marginale decrescente. Secondo quest'ultimo, l'utilità di una dose addizionale di bene consumato diminuisce al crescere della quantità consumata; perciò il consumatore sarà disposto ad acquistare dosi addizionali del bene solo a prezzi man mano decrescenti, e la domanda risulterà funzione decrescente del prezzo.
Abbiamo infine la scuola austriaca di Menger e dei suoi allievi, delle cui

caratteristiche distintive ci occuperemo nel prossimo paragrafo. È possibile attribuire una concezione di fondo sostanzialmente unitaria ai vari filoni dell'approccio marginalista: quella cui si accennava sopra, di un 'corso a senso unico', caratterizzato da un punto di partenza (sia esso costituito dalle risorse inizialmente disponibili, o dalle risorse originarie, o dalle disutilità) e da un punto di arrivo, costituito dalla soddisfazione dei gusti dei consumatori. Ma questa comune concezione di fondo assume poi una varietà notevole di connotati nelle diverse scuole che si ricollegano ad essa.

Comunque, la comune concezione di fondo facilita una valutazione sintetica del contributo teorico delle scuole marginaliste, in quanto essa implica l'esistenza di aspetti centrali comuni anche nella struttura analitica delle diverse scuole.

Intanto, per quanto riguarda le preferenze dei singoli consumatori è comune l'ipotesi di indipendenza del sistema di preferenze di ciascun soggetto economico dalle preferenze degli altri: l'unica influenza riconosciuta delle scelte altrui sulle scelte di ciascun singolo consumatore è quella indiretta, che passa attraverso i prezzi che si determinano sul mercato. Le 'esternalità', cioè il fatto che il comportamento degli altri ha un'influenza decisiva sul mio sistema di preferenze - come nel caso degli 'effetti di dimostrazione', per cui se tutti i miei vicini comprano la televisione a colori anch'io sarò indotto a comprarla -, sono escluse dalla struttura analitica della teoria marginalista tradizionale.

Questa concezione delle scelte di consumo è caratterizzata da un individualismo estremo: ciascun individuo decide per sé, sulla base di preferenze che la teoria assume come date e che quindi si suppone non interagiscano con il comportamento altrui. In tal modo essa appare contrapposta all'assunzione classica, forse valida soprattutto come prima approssimazione rudimentale, ma probabilmente più 'concreta' di quella marginalista, di un comportamento di consumo sostanzialmente simile all'interno di ciascuna classe sociale. In quella che è tradizionalmente chiamata 'ipotesi classica',

infatti, i profitti sono destinati all'accumulazione di capitale, cioè agli investimenti, e le rendite dei proprietari terrieri sono destinate essenzialmente a consumi di lusso; quanto ai salari, per i classici - che avevano ben presenti le condizioni del tempo in cui vivevano - il comportamento di consumo è reso omogeneo da ragioni obiettive, cioè dal fatto che il salario si colloca al livello di sussistenza (un livello che tuttavia, come si è già accennato, dipende dalle condizioni storico-sociali).

Un secondo elemento comune alla struttura analitica dei vari filoni dell'approccio marginalista è costituito dal ruolo centrale che in essi gioca il 'principio di sostituzione', nella produzione come nel consumo: al variare dei prezzi relativi dei diversi beni di consumo, la quantità consumata dei beni il cui prezzo è cresciuto si riduce, mentre viceversa cresce quella dei beni il cui prezzo è diminuito; analogamente, al variare dei prezzi relativi dei 'fattori di produzione', quelli il cui prezzo è aumentato vengono sostituiti da quelli il cui prezzo è diminuito. Un terzo elemento comune, che appare una conseguenza logica della struttura analitica dell'approccio marginalista, e che lo differenzia nettamente dall'approccio classico, è dato dalla simmetria nella trattazione delle variabili distributive. Infatti salario e saggio del profitto (o, nelle trattazioni 'disaggregate', i prezzi dei servizi dei diversi fattori produttivi) sono posti su uno stesso piano, come determinati (assieme alle quantità utilizzate di capitale e lavoro) dalla realizzazione sui rispettivi mercati dell'equilibrio tra domanda e offerta. Così salario e saggio di profitto corrispondono, dal lato della domanda, alla produttività marginale di lavoro e capitale. Simultaneamente, dal lato dell'offerta, salario e saggio del profitto corrispondono al 'costo reale' dei rispettivi fattori di produzione: la penosità marginale del lavoro e il sacrificio dell'astinenza dal consumo nel caso del capitale. In altre formulazioni teoriche dell'approccio marginalista, specie nelle moderne riformulazioni assiomatiche della teoria dell'equilibrio economico generale, il saggio d'interesse (che spesso

nella terminologia marginalista sostituisce il saggio del profitto come 'prezzo' dell'uso del capitale) è collegato alle 'preferenze intertemporali' dei soggetti economici tra consumo presente e consumo futuro.
Le critiche al marginalismo, che nel passato assumevano carattere occasionale o frammentario, in tempi recenti si sono intensificate, soprattutto dopo la pubblicazione dell'opera di Sraffa nel 1960. Torneremo più oltre su tali critiche; qui ci limitiamo a ricordare alcuni aspetti particolarmente problematici della teoria dell'equilibrio economico generale formulata da Léon Walras e poi ripresa e ampliata da Vilfredo Pareto. Walras rappresenta il sistema economico con un sistema di equazioni elaborate originariamente dai cultori della meccanica razionale (Walras aveva studiato ingegneria, Pareto era ingegnere). Walras sviluppa la sua teoria attraverso tre approssimazioni successive, rappresentate rispettivamente da: 1) equazioni dello scambio; 2) equazioni dello scambio e della produzione e, infine, 3) equazioni dello scambio, della produzione, dell'accumulazione e del credito. L'assunzione di partenza - che resta valida anche nelle successive approssimazioni - è che esista una dotazione iniziale di beni capitali fisici (che includono le risorse naturali) e di capitali personali (capacità di lavoro delle persone). Inoltre sono incluse fra i dati del problema le preferenze - le funzioni di utilità - dei singoli soggetti e la tecnologia; le imprese operano tutte in condizioni di concorrenza. Sulla base di queste premesse Walras dimostra che il numero delle equazioni eguaglia il numero delle incognite - prezzi e quantità dei beni e dei servizi - e che il sistema di equazioni, nei tre casi che egli considera, è determinato.
Sulla scia di Walras, numerosi economisti matematici affrontano i problemi dell'esistenza, unicità e stabilità delle soluzioni per il modello di puro scambio; i risultati cui è giunto questo filone di ricerca possono essere sintetizzati come segue: a) sotto condizioni sufficientemente generali, il sistema ha soluzioni economicamente significative (con valori non negativi per i prezzi e le quantità

di equilibrio delle varie merci); b) in generale è possibile una molteplicità di equilibri; c) la stabilità dell'equilibrio è dimostrabile solo ricorrendo a ipotesi particolari, troppo specifiche perché a tali dimostrazioni possa essere attribuita validità generale. (Questi risultati sono stati elaborati in modo rigoroso in riferimento a modelli assiomatici di equilibrio economico generale proposti negli anni cinquanta da Arrow e Debreu; tali modelli considerano anche 'equilibri intertemporali', trattando come merci diverse una stessa merce in momenti diversi del tempo e prendendo come date le preferenze dei soggetti economici tra diverse quantità di beni non solo in un dato momento nel tempo, ma anche tra momenti diversi: meno uova e più galline oggi e più uova e meno galline domani). Comunque, la fragilità - secondo alcuni critici l'insostenibilità - della costruzione di Walras e Pareto appare in piena luce nella terza approssimazione. Walras infatti assume che all'inizio di un certo periodo vi sia una dotazione "data a caso" ("criée au hasard") di beni capitali, intesi in senso lato; nella terza approssimazione egli ammette che si producano nuovi beni capitali, particolarmente quelli che risultano scarsi e che forniscono il più alto rendimento. Ma, nella logica della sua costruzione, suppone che tali beni entrino in funzione, non nello stesso periodo, bensì nel periodo successivo; nulla però assicura che, in tale periodo successivo, avrà luogo l'unicità dei rendimenti dei beni capitali. Mentre nel caso dei beni di consumo l'aggiustamento dei prezzi e delle quantità avviene in uno stesso periodo e quindi le correzioni potranno dar luogo a movimenti convergenti verso certi livelli normali, così non è nel caso dei beni capitali, in particolare di quelli fisici. Pertanto, nel sistema di Walras non si ha in generale un unico tasso di rendimento in tutti i settori produttivi; ciò è una logica conseguenza del fatto che tale teoria segue 'l'ottica della scarsità' anziché 'l'ottica della riproducibilità', in quanto essa si fonda non sulla concezione classica dell'attività economica come processo circolare o a spirale, ma sulla concezione di un corso a senso

unico, che parte da una certa dotazione di fattori produttivi e si conclude con l'offerta di un certo gruppo di beni di consumo: le risorse "date a caso" portano a rendimenti "dati a caso".

g) La scuola austriaca

I teorici della scuola austriaca (oltre a Menger, vanno ricordati almeno i suoi allievi Friedrich von Wieser, 1851-1926, ed Eugen von Böhm-Bawerk, 18511914) adottano un'ottica soggettiva radicale, in base alla quale il valore di ciascun bene o servizio viene dedotto dall'utilità che esso ha per il consumatore finale, direttamente nel caso dei beni di consumo o indirettamente nel caso dei beni di produzione; in quest'ultimo caso si 'imputa' al mezzo di produzione una quota dell'utilità che il bene prodotto ha per il consumatore, calcolando tale quota in proporzione al contributo del bene o servizio considerato al processo produttivo (da cui l'espressione 'teoria dell'imputazione').Le prime fasi di sviluppo della scuola austriaca furono comunque caratterizzate dall'aspro scontro metodologico con la cosiddetta 'scuola storica tedesca', tra i cui rappresentanti ricordiamo Wilhelm Roscher (1817-1894) e Gustav von Schmoller (1838-1917). La scuola storica attribuiva importanza centrale allo studio dell'evolversi delle istituzioni, criticando l'eccesso di astrazione e di astoricismo dell'impostazione marginalista; ma finiva in realtà col ridurre l'economia alla semplice descrizione di vicende e situazioni specifiche. Contro questa posizione, la scuola austriaca sostenne la necessità di ragionamenti deduttivi rigorosi basati su premesse generali, riportando nel complesso una vittoria culturale schiacciante. Una influenza della scuola storica tedesca può comunque essere individuata nella cosiddetta scuola istituzionalista, attiva ancora oggi soprattutto negli Stati Uniti, dove il suo principale esponente è stato Thornstein Veblen (1857-1929). La scuola austriaca 'moderna', in particolare con Friedrich von Hayek (1899-1992), si caratterizza per un deciso sostegno a

posizioni radicalmente liberiste, e si differenzia dal marginalismo tradizionale nella visione del funzionamento dell'economia: in particolare, l'atto di scelta compiuto dal soggetto economico è visto come esperimento in condizioni di incertezza, il cui risultato modifica le aspettative e le conoscenze iniziali in un processo continuo; rispetto all'approccio marginalista francese e anglosassone, il concetto di equilibrio perde così il suo tradizionale ruolo analitico centrale come soluzione del problema economico.
Tra la fine degli anni venti e i primi anni trenta, inoltre, Hayek, riprendendo e sviluppando una teoria proposta da un altro esponente di rilievo della scuola austriaca, Ludwig von Mises (1881-1973), sostenne in vari scritti una teoria monetaria del ciclo economico (la cosiddetta 'teoria del risparmio forzato'), che dopo un discreto successo iniziale fu relegata in secondo piano dall'affermazione della teoria keynesiana. Secondo la teoria del risparmio forzato, mentre il livello di equilibrio di lungo periodo della produzione corrisponde alla piena occupazione dei lavoratori e al pieno utilizzo della capacità produttiva disponibile, nel breve periodo è possibile un'accelerazione degli investimenti finanziati da un'espansione dell'offerta di moneta da parte delle banche, a scapito di una riduzione non desiderata dei consumi, cioè di un risparmio non desiderato - o 'forzato' - prodotto dall'aumento dei prezzi che riduce il potere d'acquisto dei lavoratori salariati; in un secondo momento, quando entra in funzione la nuova capacità produttiva corrispondente agli investimenti addizionali, l'accresciuta offerta di prodotti genera una spinta deflazionistica, e quindi un aumento del potere d'acquisto dei lavoratori e una ripresa dei consumi, con una distruzione di risparmio che riporta il sistema economico alla situazione di partenza. Questa teoria permette di conciliare il riconoscimento di un'influenza dei fenomeni monetari sui livelli di produzione con la concezione marginalista tradizionale secondo la quale prezzi relativi e quantità di equilibrio (inclusi i livelli di produzione e di occupazione) sono

determinati esclusivamente dai fattori reali: disponibilità di risorse, tecnologia, preferenze dei consumatori. In questo senso, nonostante le caratteristiche specifiche che abbiamo appena richiamato, la scuola austriaca può essere a buona ragione collocata all'interno della tradizione marginalista.

h) Dalla statica alla dinamica

Come si è accennato sopra, l'approccio marginalista in tutte le sue varie formulazioni imposta il problema economico come centrato sul confronto fra risorse disponibili e preferenze dei consumatori finali, e risolve tale problema determinando prezzi e quantità di equilibrio che assicurino l'eguaglianza tra quantità domandate e offerte dei vari beni e servizi. Tramite l'operare del principio di sostituzione, i mercati assicurano l'eliminazione di scarsità ed eccedenze, e quindi l'ottima utilizzazione delle scarse risorse disponibili. In particolare, in condizioni di concorrenza il normale funzionamento del mercato del lavoro fissa il salario a un livello tale da assicurare la piena occupazione della forza-lavoro disponibile. L'impostazione marginalista è dunque essenzialmente statica: dati l'ammontare di risorse disponibili, la tecnologia e i gusti dei consumatori prevalenti in un dato momento del tempo, la teoria determina prezzi e quantità di equilibrio. L'evoluzione di un sistema economico nel tempo è legata al cambiamento dei dati: risorse disponibili, tecnologia e gusti dei consumatori. Così essa può essere studiata in due modi: o tramite analisi di statica comparata, cioè determinando prezzi e quantità di equilibrio prima con l'iniziale insieme di dati e poi con quello nuovo, e confrontando le due soluzioni per individuare l'effetto del cambiamento dei dati; oppure tramite analisi dinamiche, cioè introducendo una precisa ipotesi sull'evoluzione nel tempo di uno o più fra i dati del problema, e determinando così un sentiero regolare per lo sviluppo del sistema economico. In entrambi i casi, comunque, restano generalmente esclusi dall'analisi sia i problemi di transizione da un

equilibrio all'altro, sia quelli di spiegare all'interno della scienza economica i cambiamenti nella tecnologia o nelle abitudini di consumo.
L'aspetto sorprendente della 'rivoluzione marginalista', se consideriamo la sua natura essenzialmente statica, è che essa ha luogo proprio in un tempo in cui i cambiamenti tecnologici, che hanno caratterizzato il capitalismo industriale sin dal suo apparire, tendono a crescere, non a diminuire, d'importanza, come tendono a crescere d'importanza i cambiamenti nelle abitudini di consumo (e spesso questo secondo fenomeno è condizionato dal primo). D'altra parte, i cambiamenti tecnologici sono all'origine del processo di sviluppo economico, che costituiva il problema teorico fondamentale per economisti classici come Smith, ma che scompare del tutto dall'orizzonte teorico dei marginalisti. Per spiegare tale svolta paradossale sono state indicate diverse ragioni, tre in particolare, che qui ci limitiamo a ricordare e a commentare molto concisamente. La prima ragione si ricollega all'aspirazione a un maggior rigore. Ciò implica l'applicazione sistematica dei metodi matematici, cosa agevole se si assumono costanti la tecnologia e le abitudini di consumo, ma ben più difficile nel caso di problemi dinamici. Quella statica era considerata solo un'analisi di prima approssimazione; ma i modelli di norma erano concepiti in modo da rendere impossibile il passaggio all'analisi dinamica: per far questo occorreva ricominciare da capo.
La seconda ragione è legata a motivi di natura ideologica. La teoria classica era sfociata nei cosiddetti socialisti ricardiani e poi nell'analisi di Marx, che aveva finalità dichiaratamente rivoluzionarie. Per esorcizzare gli sviluppi di tipo marxista, e quindi rivoluzionari, si diffuse la tendenza a rifiutare l'impostazione stessa delle teorie classiche, abbandonando l'esame oggettivo dei rapporti sociali di produzione e rivolgendo lo studio verso le caratteristiche soggettive dei bisogni e quindi dell'individuo astrattamente considerato. La terza ragione si ricollega alla crescita del reddito individuale di strati sempre più ampi della

popolazione. Mentre al tempo dei classici la gran massa della popolazione viveva al livello della sussistenza e il problema delle preferenze dei consumatori si poneva solo per un'esigua minoranza di privilegiati, nelle nuove condizioni quel problema si poneva invece per una maggioranza sempre più ampia; e le teorie marginaliste davano il massimo rilievo alle preferenze dei consumatori.
Le tre ragioni vanno forse considerate congiuntamente, ricordando che i classici non erano riusciti a risolvere in modo soddisfacente il problema del valore.
Un'ulteriore spinta all'affermazione del marginalismo fu data dalla sua vittoria nello scontro con la nuova scuola storica, che tendeva a risolvere la teoria nella storia economica.
Non sarebbe esatto, tuttavia, affermare che la costruzione marginalista sia integralmente statica. In effetti, vari economisti tentano di affrontare, all'interno dei diversi filoni dell'approccio marginalista, il problema delle situazioni 'di squilibrio', generalmente considerate come oscillazioni di breve periodo attorno alle situazioni 'di equilibrio' implicanti, come si è detto, il pieno utilizzo delle risorse disponibili. In questo quadro viene affrontato il problema del ciclo economico. Esso è attribuito a shock, cioè a variazioni improvvise e impreviste dei dati del problema che spingono il sistema economico fuori dell'equilibrio per qualche tempo prima che l'operare delle forze di mercato ve lo riporti; oppure a fenomeni monetari che si sovrappongono alle forze 'reali' determinanti l'equilibrio di mercato. Si tratta, in ogni caso, di eventi che rendono non più di equilibrio i prezzi relativi prevalenti nella situazione precedente; il protrarsi dello squilibrio e la sua trasformazione in ciclo economico sono ricollegati all'incapacità del mercato di condurre istantaneamente i prezzi relativi ai valori coerenti con il nuovo equilibrio. Da questa base comune, le specifiche teorie del ciclo si differenziano per l'accento posto su prezzi relativi particolari, quali salario reale, tasso d'interesse e tasso del profitto. In particolare Marshall e vari economisti americani della Scuola di Chicago mettono l'accento sui movimenti

del salario e del tasso d'interesse confrontati con le aspettative sulle variazioni dei prezzi; lo svedese Knut Wicksell (1851-1926) sugli scostamenti del tasso d'interesse monetario da quello 'naturale' corrispondente al tasso di rendimento degli investimenti; l'inglese Ralph Hawthrey (1879-1975) sui movimenti dei tassi d'interesse sui crediti a breve termine, e quindi sul 'ciclo delle scorte'; l'inglese Dennis Robertson (1890-1963) su cause monetarie per quanto riguarda i cicli brevi, e sull'addensarsi degli acquisti per il rinnovo di impianti e beni di consumo durevoli determinato dall'alternarsi di 'sciami' di innovazioni con periodi di relativo ristagno tecnologico per quanto riguarda i cicli di media durata. Alle teorie austriache del risparmio forzato che almeno in parte derivano da quella di Wicksell, si è già accennato nel paragrafo precedente.
Da queste tradizioni provengono anche due economisti che tuttavia si scostano dalle fondamenta tradizionali dell'approccio marginalista in misura sufficiente da essere considerati 'eretici', fondatori di nuove scuole di pensiero: Schumpeter e Keynes.

i) Joseph Schumpeter e lo sviluppo ciclico dell'economia
L'economista austriaco Joseph Schumpeter (18831950) si contrappone alla tradizione marginalista (di cui pure dichiara di accettare le principali elaborazioni teoriche, come soluzione però di un problema specifico, quello 'statico', concernente la raffigurazione di un sistema economico che non cambia nel tempo) per il suo tentativo di costruire una teoria dello sviluppo. Tale teoria è basata sulle innovazioni, introdotte dagli imprenditori e finanziate dai banchieri; imprenditori e banchieri risultano così le figure attive, le cui scelte determinano l'evoluzione dell'economia, mentre le scelte di consumo e di risparmio, che risultavano centrali nella teoria tradizionale, appaiono qui come secondarie. Tutte le principali nozioni della teoria tradizionale risultano modificate da questo cambiamento d'ottica; ad esempio, la nozione di

concorrenza della teoria statica risulta 'travolta' dalla forza della concorrenza dinamica, che viene dagli imprenditori che introducono nuovi processi produttivi o nuovi prodotti. I profitti degli imprenditori-innovatori derivano da beni nuovi, che riescono ad affermarsi nel mercato, e dai minori costi rispetto alle imprese che utilizzano i vecchi metodi produttivi; e, dato che la domanda di finanziamenti alle banche proviene essenzialmente dagli imprenditori-innovatori, sono le opportunità d'investimento aperte dalle innovazioni e la loro redditività che determinano i tassi d'interesse.
Il ciclo economico è spiegato dall'andamento irregolare del flusso delle innovazioni. Queste, d'altra parte, sono d'importanza molto diversa: le innovazioni di portata storica - la macchina a vapore per usi fissi e per il trasporto per terra o per mare, l'elettricità, l'automobile, l'aereo, oggi la microelettronica - danno origine a 'cicli di lunga durata' (i 'cicli Kondrat'v' di durata approssimativamente cinquantennale); le innovazioni di minor rilievo, che spesso s'innestano in quelle maggiori e ne costituiscono sviluppi specifici, danno origine a cicli di durata più breve: i 'Juglar', di circa nove anni, e i 'Kitchin', di circa tre anni (dal nome degli economisti che li hanno studiati sistematicamente, anche sul piano empirico). Il modello teorico di Schumpeter è originariamente presentato nella Teoria dello sviluppo economico del 1912 ed è poi riproposto, con diverse importanti modificazioni, nel trattato su I cicli economici del 1939, nel quale l'analisi teorica si combina con quella empirica (storica e statistica). Schumpeter parte dall'analisi del 'flusso circolare', ossia del processo economico che riproduce uniformemente se stesso: un concetto simile, se pure non identico, a quello dei classici e di Sraffa. La rottura di tale flusso è provocata dagli imprenditori tramite le innovazioni, ossia nuove e più efficienti combinazioni di fattori produttivi, dalle quali emerge il profitto. Gli imprenditori-innovatori per finanziare le innovazioni domandano prestiti alle banche, le quali creano mezzi di pagamento addizionali. Compare quindi una

schiera di imitatori, crescono gli investimenti e, di conseguenza, la domanda di beni di consumo; si sviluppa così la fase di prosperità. Dalla prosperità si passa alla flessione, man mano che vengono a maturazione i frutti delle innovazioni e man mano che crescono le produzioni delle imprese che non s'innovano ma si avvantaggiano della generale prosperità. Alla fine del ciclo, il sistema dei prezzi risulta cambiato e i redditi reali accresciuti. Pertanto, il nuovo ciclo parte da un livello più alto del reddito complessivo e di quasi tutti i redditi individuali: cicli e sviluppo risultano le manifestazioni di un unico processo, il processo dello sviluppo ciclico.

Secondo Schumpeter, tale processo è caratterizzato da una tendenza di lungo periodo verso la concentrazione delle imprese, una tendenza condizionata dalla crescente importanza delle economie di scala nelle principali attività economiche. In tal modo sono emerse quelle grandi o grandissime imprese, specialmente nell'industria e nella finanza, che in diversi casi hanno assunto rilevanza internazionale. Schumpeter sostiene che nelle grandi imprese, le quali organizzano grandi laboratori di ricerca, l'innovazione tende a essere trasformata in un'attività di routine. La crescita assoluta e relativa delle grandi imprese e la burocratizzazione del processo innovativo renderebbero superflua la figura centrale del capitalismo moderno, l'imprenditore, e renderebbero pertanto sempre più difficile la sopravvivenza del capitalismo, favorendo l'avvento del socialismo centralizzato: il collettivismo pubblico verrebbe a sostituire un sistema che tende ad assumere sempre più le caratteristiche di un collettivismo privato. In tal modo Schumpeter, che ideologicamente era un conservatore, analiticamente si avvicina a Marx che aveva già considerato, sia pure per brevi cenni, la tendenza alla concentrazione; Marx, tuttavia, sosteneva l'ineluttabilità dell'avvento del socialismo su basi in gran parte diverse: il ruolo principale era svolto dal crescente immiserimento delle masse.

Oggi appare chiaro che la tendenza alla concentrazione, che è stata osservata in

diverse importanti attività economiche, è stata interpretata in termini troppo schematici sia da Marx sia da Schumpeter. In particolare, Schumpeter ha sottovalutato gravemente il ruolo che tuttora svolgono gli inventori individuali e il ruolo dei ricercatori e degli inventori che operano in istituti universitari e in altri organismi autonomi. Il flusso delle nuove idee e delle invenzioni che hanno una tale origine riveste importanza essenziale anche per le grandi imprese, i cui laboratori spesso sviluppano invenzioni di provenienza esterna. Schumpeter ha inoltre sottovalutato il ruolo che piccole e piccolissime imprese possono avere e spesso hanno avuto nell'avviare importanti innovazioni, che in seguito sono riprese e sviluppate dalle grandi. Per di più, anche come conseguenza di certe innovazioni, come quelle connesse all'elettronica, sono sorti nuovi spazi per le piccole imprese, il cui peso, nonostante la tendenza verso la concentrazione, era sempre rimasto considerevole e negli ultimi due decenni è progressivamente cresciuto in tutti i paesi industrializzati.

l) Max Weber, la sociologia e il processo di burocratizzazione

Joseph Schumpeter è un economista, ma è anche un sociologo e un politologo: possono essere classificate con queste ultime due etichette alcune sue opere, come L'imperialismo e le classi sociali e Capitalismo, socialismo e democrazia. Un altro grande studioso, quasi contemporaneo di Schumpeter, Max Weber (1864-1920), viene generalmente considerato come un sociologo, ma è stato professore di economia politica (prima a Friburgo, poi a Heidelberg) e si è occupato fra l'altro anche di temi strettamente economici (quali i fondamenti della teoria marginalista del valore, in una recensione del 1908 a un libro di Brentano). La sua opera principale comunque resta Economia e società, pubblicata postuma nel 1922. Weber e Schumpeter sono due cospicui esempi dell'ampia sovrapposizione che spesso si crea fra le scienze sociali, in particolare fra economia e sociologia. Un altro esempio dello stretto

collegamento tra economia e sociologia è fornito dall'opera dell'italiano Vilfredo Pareto, anche se, come si è accennato sopra, il contributo di Pareto alla teoria economica rientra nel filone marginalista dell'equilibrio economico generale e fa capo quindi a una concezione che ha contribuito più a separare che ad avvicinare le due discipline.

L'opera di Weber riguarda l'interpretazione del capitalismo moderno, concepito come un grandioso processo di trasformazione, che, a sua volta, è in primo luogo l'espressione di un processo di razionalizzazione non solo dell'attività economica, ma dell'organizzazione della società intera. Tale processo s'incontra e per certi versi si scontra con le convinzioni religiose, che possono essere più o meno favorevoli allo 'spirito del capitalismo'; tale processo s'incarna, fra l'altro, in una progressiva burocratizzazione dell'organizzazione statale e del sistema produttivo. Il processo di cui parla Weber si contrappone sotto diversi aspetti, se non altro sotto l'aspetto quantitativo, al processo di crescente proletarizzazione delineato da Marx. L'evoluzione storica ha dato ragione a Weber, non a Marx, giacché la burocrazia pubblica e privata è cresciuta enormemente nei paesi sviluppati, mentre la quota dei proletari - operai salariati- è cresciuta fin verso la fine del secolo scorso, e in seguito, durante questo secolo, ha mostrato tendenza a declinare. (Negli ultimi cento anni la massa degli impiegati è cresciuta dal 5 al 30% della popolazione attiva in Italia, dal 12 al 45% in Inghilterra e dal 14 al 49% negli Stati Uniti, con l'avvertenza che oggi in Italia la burocrazia pubblica rappresenta oltre la metà del totale, in Inghilterra poco meno della metà e negli Stati Uniti un terzo. La quota del 'proletariato', dopo aver toccato o superato il 50% della popolazione attiva, negli ultimi decenni è scesa nettamente sotto tale livello. Com'è noto, oltre che dagli impiegati e dagli operai, la popolazione attiva è costituita da lavoratori autonomi).

L'economista deve considerare il processo di burocratizzazione congiuntamente all'espansione degli interventi pubblici nella vita economica e, nei tempi più

vicini a noi, alla crescita dello Stato sociale e degli apparati fiscale e parafiscale. Le stesse teorie economiche vanno considerate con riferimento a queste tendenze. Così, le politiche economiche ispirate da Keynes non sarebbero state concepibili se il processo di burocratizzazione nel senso di Weber non fosse andato molto avanti. Non pochi modelli teorici vanno visti nella stessa prospettiva. Ad esempio, una teoria delle retribuzioni - salari e stipendi - che miri ad avere efficacia interpretativa non può non riconoscere che nel nostro tempo le retribuzioni nominali sono rigide verso il basso: una rigidità che si ricollega in vari modi alle forme organizzative prevalenti nella burocrazia pubblica e privata e ai sindacati, che per diversi aspetti rientrano nel processo weberiano di burocratizzazione.

m) John Maynard Keynes e la disoccupazione di equilibrio

In opposizione alle tradizioni precedenti (classica e marginalista), che considerano la moneta un semplice 'velo', cioè priva di influenza sui livelli di equilibrio delle variabili 'reali' (quantità prodotte, prezzi relativi di equilibrio e variabili distributive), Keynes (1883-1946) insiste sull'influenza persistente, e non solo come fattore di disturbo ciclico, che le vicende monetarie possono avere su quelle reali. La formulazione matura della teoria keynesiana - quella della Teoria generale del 1936 - può essere illustrata a partire da tre elementi fondamentali: il concetto di domanda effettiva, la teoria degli investimenti e del moltiplicatore del reddito e una teoria monetaria basata sulla 'preferenza per la liquidità'. Consideriamo rapidamente questi tre elementi. La domanda effettiva esprime, per ogni livello di occupazione, l'ammontare di prodotto che gli imprenditori si attendono di vendere sul mercato; ed è considerata da Keynes come funzione crescente - ma a velocità decrescente - del livello di occupazione. Gli imprenditori decidono quanto produrre, e quindi in quale misura utilizzare la capacità produttiva disponibile e quanti lavoratori

impiegare, confrontando la curva di domanda effettiva con la curva di offerta. Quest'ultima indica, in corrispondenza di ciascun livello di occupazione, l'ammontare di vendite che gli imprenditori ritengono necessario per recuperare i costi e ottenere un profitto appena sufficiente a indurli a continuare la produzione. La curva di offerta è quindi una funzione crescente del livello di occupazione; inoltre, nella Teoria generale, Keynes afferma che essa tende a crescere man mano più rapidamente a causa dell'aumento progressivo del costo unitario di produzione (postulato della produttività marginale decrescente). Pertanto, per livelli sufficientemente bassi dell'occupazione la curva di domanda effettiva risulterà superiore alla curva di offerta, mentre a un certo punto (detto 'punto di domanda effettiva') le due curve si incontreranno, e per livelli di occupazione superiori la curva di offerta risulterà superiore a quella di domanda.

Il livello di occupazione scelto dagli imprenditori corrisponde al punto di domanda effettiva. In questo modo, osserva Keynes, tutto dipende dalle aspettative degli imprenditori, e più precisamente dalle 'aspettative di breve periodo', relative alle possibilità di smercio dei prodotti. Inoltre, non vi è alcuna ragione per cui il livello di occupazione così determinato debba corrispondere a quello che assicura il pieno utilizzo della capacità produttiva disponibile e, soprattutto, la piena occupazione della forza-lavoro. Le aspettative degli imprenditori risultano realizzate quando la domanda per beni di consumo e d'investimento corrisponde al livello di produzione da loro prescelto; in questo caso, rileva Keynes, anche in presenza di disoccupazione e di capacità produttiva inutilizzata gli imprenditori non hanno alcuno stimolo a espandere produzione e occupazione. A sostenere quest'ultima tesi concorrono gli altri elementi costitutivi della teoria keynesiana, in particolare la teoria dell'investimento e quella della preferenza per la liquidità che illustreremo tra poco. In sintesi, Keynes ricorda che i consumi dipendono essenzialmente dal

reddito e quindi dai livelli di produzione, mentre lo stimolo esterno alle variazioni della domanda aggregata viene dagli investimenti; questi, a loro volta, dipendono dalle aspettative degli imprenditori e dal tasso d'interesse, determinato sul mercato monetario: non vi è motivo per cui questi elementi debbano rispondere alla presenza di disoccupazione in modo da stimolarne il riassorbimento. Di qui la tesi della possibilità di 'equilibri di sottoccupazione', cioè di situazioni di disoccupazione persistente.

Il secondo elemento costitutivo della teoria keynesiana, come si è accennato, è la sua teoria dell'investimento. Anche in questo caso Keynes si pone dal punto di vista dell'imprenditore, che decide se e quanto investire confrontando i costi dell'investimento con i ricavi che ci si attendono da esso. Le aspettative dell'imprenditore - che in questo caso sono 'aspettative di lungo periodo', in quanto riguardano un arco di tempo piuttosto lungo, corrispondente alla vita attiva degli impianti in cui si concreta l'investimento - sono secondo Keynes piuttosto 'volatili', in quanto riguardano un futuro incerto, e sono quindi estremamente sensibili a variazioni del 'clima generale di opinioni'. Per confrontare i costi dell'investimento con i ricavi che ci si attende di ottenere da esso negli anni di funzionamento dell'impianto occorre scontare - cioè ridurre a valori attuali, in lire di oggi - il valore atteso dei ricavi, utilizzando il tasso d'interesse; per questa via le vicende monetarie, che come vedremo fra poco determinano il livello dei tassi d'interesse, influiscono sulle vicende 'reali' dell'economia, contribuendo a determinare il livello degli investimenti e, tramite esso, il livello della produzione e dell'occupazione. Infatti il livello della produzione è collegato agli investimenti tramite il cosiddetto 'moltiplicatore': un concetto originariamente elaborato da un allievo di Keynes, Richard Kahn (1905-1989), in un articolo del 1931, per indicare l'impulso espansivo prodotto da un investimento (o da una spesa pubblica) addizionale sul reddito in condizioni di diffusa disoccupazione; l'effetto espansivo è superiore al volume

dell'investimento addizionale a causa dell'espansione della spesa per consumi da parte dei lavoratori precedentemente disoccupati e che ora trovano un impiego.
Il terzo elemento fondamentale della teoria keynesiana è costituito dalla teoria della 'preferenza per la liquidità'. Keynes respinge sia la 'teoria quantitativa della moneta', sia la teoria del tasso d'interesse basata sul confronto fra domanda e offerta di 'fondi disponibili per i prestiti'. Consideriamo per prima la teoria quantitativa. Essa si basa sull'equazione degli scambi, MV = PQ, che indica semplicemente che il valore dei beni e servizi scambiati sul mercato (pari a quantità Q moltiplicata prezzo P) è pari al valore della moneta che passa di mano in senso inverso (pari a M, quantità di moneta in circolazione, moltiplicata per V, velocità di circolazione, cioè numero di volte in cui viene utilizzata in media ciascuna unità di moneta in circolazione). La teoria quantitativa sostiene che il livello generale dei prezzi varia in proporzione alle variazioni dell'offerta di moneta (cioè che P varia in proporzione a M), in quanto suppone che sia possibile considerare relativamente stabili sia la velocità di circolazione della moneta V, determinata da fattori istituzionali che cambiano solo molto lentamente, sia il prodotto Q, che le forze di mercato tendono a mantenere al livello di piena occupazione. Viceversa, Keynes sostiene che né l'una né l'altra ipotesi possono essere considerate valide; infatti la velocità di circolazione della moneta cambia anche nel breve periodo, in conseguenza di spostamenti da moneta a titoli e viceversa nelle preferenze dei soggetti economici; il prodotto Q, come si è appena accennato, può variare in quanto per Keynes è possibile che gli imprenditori decidano di non utilizzare appieno la capacità produttiva e la forza-lavoro disponibili.
Come si è accennato sopra, Keynes respinge anche la teoria del tasso d'interesse dominante nella tradizione marginalista, secondo cui esso è determinato sul mercato dei fondi prestabili dal confronto tra l'offerta di tali fondi (i risparmi, che sono funzione crescente del tasso d'interesse) e la domanda, che viene

essenzialmente dagli imprenditori. Questi ultimi infatti chiedono denaro in prestito per finanziare gli investimenti, che sono funzione decrescente del tasso d'interesse. Keynes replica che sia i risparmi sia gli investimenti dipendono essenzialmente da altre variabili: i risparmi dal livello del reddito, gli investimenti dalle aspettative degli imprenditori. Il tasso d'interesse risulta invece determinato sul mercato monetario dalla 'preferenza per la liquidità', cioè dalle scelte dei soggetti economici sulla forma in cui tenere la ricchezza accumulata (che è uno stock, e quindi generalmente ha dimensioni ben maggiori del flusso annuo di risparmi e investimenti). Tale scelta riguarda essenzialmente titoli (azioni e obbligazioni) e moneta (legale o bancaria); e dipende sia dai rendimenti, sia soprattutto dalle aspettative di variazioni dei prezzi delle attività finanziarie, che continuamente determinano guadagni e perdite in conto capitale. Le attività più 'liquide' sono quelle che all'occorrenza è più facile trasformare in moneta legale, con probabilità minime di ottenere guadagni o subire perdite in conto capitale: dopo le banconote vengono i depositi di conto corrente, e poi man mano, in ordine di liquidità decrescente, i buoni del Tesoro, le obbligazioni, le azioni, beni rifugio come l'oro e i diamanti, gli immobili.

Il tasso d'interesse dipende dall'offerta di attività liquide (moneta a corso legale e moneta bancaria), cioè dalle scelte delle autorità monetarie, e dalla domanda di moneta, cioè dalla 'preferenza per la liquidità' dei privati. Pertanto il tasso d'interesse risulta determinato sul mercato monetario, e non rappresenta una variabile 'reale' determinata dalle scelte di risparmio e investimento. In questo modo, come si accennava, le vicende monetarie influiscono su investimenti, reddito e occupazione. L'obiettivo fondamentale di Keynes consiste dunque nel mostrare che un'economia di mercato non tende automaticamente alla piena occupazione. Ma il ristagno produttivo e la disoccupazione costituiscono una minaccia per l'organizzazione civile della società. Perciò la stessa sopravvivenza

del capitalismo, che sarebbe messa in discussione da condizioni di disoccupazione elevata e persistente, richiede, secondo Keynes, un vigoroso intervento pubblico per stimolare l'attività economica e sostenere l'occupazione. Negli ultimi cinquant'anni, e in particolare negli anni cinquanta e sessanta, numerosi economisti 'keynesiani' hanno identificato tale intervento pubblico in politiche fiscali e monetarie espansive, cioè in politiche di spesa pubblica in disavanzo per sostenere la domanda aggregata di beni e servizi, e in politiche di espansione della liquidità dirette a ridurre i tassi d'interesse e quindi a favorire gli investimenti. In realtà, Keynes proponeva soprattutto di creare un ambiente di consuetudini e istituzioni, anche internazionali, tale da stimolare sia un clima di aspettative imprenditoriali favorevoli agli investimenti, sia una riduzione dell'incertezza che condiziona le decisioni sui livelli di produzione delle imprese; in quest'ambito rientrano proposte quali la predisposizione di programmi d'investimento pubblici da realizzare nei momenti di ristagno, o la riforma delle istituzioni monetarie e finanziarie internazionali diretta a favorire i commerci e lo sviluppo economico.
L'analisi di Keynes, concentrando l'attenzione su alcuni aspetti del funzionamento del sistema economico, fa passare in secondo piano altri aspetti non meno importanti di quelli considerati; in gran parte degli economisti postkeynesiani, ciò si traduce in una visione eccessivamente semplificata dell'economia. In particolare, l'assunzione della tecnologia data, fatta da Keynes nella Teoria generale per focalizzare l'attenzione sugli elementi al centro della sua analisi, ha spesso indotto gli economisti keynesiani a trascurare il ruolo del cambiamento tecnologico, e quindi a sottovalutare l'importanza dei fattori, anche 'di breve periodo', che lo favoriscono, come dei suoi effetti su reddito e occupazione, ma anche sulla struttura stessa del sistema economico. L'ottica 'aggregata' dell'analisi keynesiana, evidente nell'uso di un indice unico per il 'livello dei prezzi' (un uso che Keynes stesso aveva criticato nel suo Trattato

della moneta del 1930), ha portato non solo a trascurare il rapporto tra mutamenti strutturali dell'economia da un lato e andamento del reddito e dell'occupazione dall'altro, ma anche a una perniciosa separazione tra analisi microeconomica (teoria dei prezzi relativi, della struttura dei consumi, delle forme di mercato e della distribuzione del reddito) e analisi macroeconomica (teoria della moneta, del reddito e dell'occupazione), già criticata dallo stesso Keynes ma ormai cristallizzata nella pratica di molti corsi d'insegnamento universitari. Anche da questo probabilmente deriva la scarsa attenzione spesso prestata nella concreta attuazione di politiche fiscali e monetarie espansive alla distinzione tra spese produttive e improduttive.

3. Il dibattito contemporaneo

Nel variegato panorama contemporaneo sono presenti diversi gruppi di economisti, i principali fra i quali si richiamano più o meno direttamente alle maggiori scuole economiche illustrate sopra. Distinguiamo quattro gruppi principali: gli economisti della 'sintesi neoclassica', dominanti per oltre trent'anni dopo la conclusione della seconda guerra mondiale, caratterizzati dall'innesto di elementi keynesiani - particolarmente per quel che riguarda la politica economica - sul tronco della tradizione marginalista; gli economisti monetaristi e la scuola delle aspettative razionali, caratterizzati dal rifiuto più o meno radicale dell'intervento pubblico nell'economia e, sul piano più strettamente teorico, da un rifiuto della teoria keynesiana in quanto contraddittoria con la struttura analitica dell'approccio marginalista; gli economisti postkeynesiani, che tendono a sviluppare gli elementi della teoria keynesiana più eterodossi rispetto alla tradizione marginalista; Sraffa e gli economisti che condividono il suo progetto culturale di un ritorno all'impostazione della scuola classica.

a) La sintesi neoclassica

Di fronte all'esperienza della grande depressione degli anni trenta, molti economisti sono indotti a prestare orecchio alle idee di Keynes sui rimedi di politica economica alla disoccupazione, pur senza abbandonare la teoria marginalista del valore e della distribuzione che costituisce la base della loro formazione professionale. Questi economisti perciò cercano di reinterpretare la teoria di Keynes introducendo alcune ipotesi per rendere l'esistenza di disoccupazione compatibile con la teoria marginalista. Lungo questa strada procede in particolare John Hicks (1904-1989, premio Nobel nel 1972). In un articolo del 1939, Hicks propone il cosiddetto schema IS-LM, che traduce la teoria keynesiana nei termini più tradizionali di un modello di equilibrio economico generale semplificato, caratterizzato dalla presenza di tre mercati: il mercato dei beni, quello della moneta e quello dei titoli (quest'ultimo però gioca un ruolo puramente passivo, mentre l'attenzione si concentra sui primi due). Il mercato dei beni è in equilibrio quando l'offerta, cioè la produzione, è eguale alla domanda aggregata (che nell'ipotesi semplificata di un sistema senza rapporti con l'estero, senza spesa pubblica e senza prelievo fiscale corrisponde alla somma della domanda per beni di consumo e di quella per beni d'investimento); e ciò si verifica quando i risparmi, che sono funzione crescente del reddito, sono eguali agli investimenti, considerati funzione decrescente del tasso d'interesse. Il mercato della moneta è in equilibrio quando offerta e domanda di moneta sono eguali; secondo l'ipotesi della moneta esogena, l'offerta di moneta è determinata dalle autorità monetarie che controllano l'emissione di moneta a corso legale e, indirettamente, la quantità di moneta creditizia che può essere creata dalle banche; la domanda di moneta è pari alla somma della domanda di moneta a scopo di transazione, che è funzione crescente del reddito, e della domanda di moneta a scopo speculativo - quella su

cui Keynes aveva concentrato l'attenzione, e che esprime la scelta sulla forma, moneta o titoli, in cui tenere la propria ricchezza -, che è considerata funzione decrescente del tasso d'interesse.
Lungo la strada intrapresa da Hicks procede anche Franco Modigliani (n. 1918, premio Nobel nel 1985), economista statunitense di origine italiana, emigrato negli Stati Uniti per sfuggire alle persecuzioni razziali. Per certi aspetti Modigliani accetta lo schema IS-LM; accanto al mercato dei beni e a quello della moneta (e a quello dei titoli finanziari, che come si è detto resta sullo sfondo), egli però considera esplicitamente anche il mercato del lavoro. Come per gli altri mercati, le variazioni del prezzo tendono ad assicurare l'equilibrio tra domanda e offerta: nel nostro caso, dunque, le variazioni del salario reale - che è il prezzo dei servizi dei lavoratori -, portando in equilibrio domanda e offerta di lavoro, tendono ad assicurare la piena occupazione. Per ottenere il risultato keynesiano, cioè la possibilità di una situazione di disoccupazione persistente, occorre allora introdurre qualche ostacolo che impedisca il libero funzionamento del mercato del lavoro. Quest'ostacolo è individuato nella natura non concorrenziale del mercato del lavoro dovuta alla forza contrattuale dei sindacati.In questo modo la teoria keynesiana viene presentata come un caso particolare della teoria marginalista: quel caso in cui l'equilibrio di piena occupazione non può essere raggiunto, perché il mercato del lavoro non è concorrenziale, per cui la presenza di disoccupazione non influisce sui salari monetari e, tramite questi, sui salari reali. Si ha così la sintesi neoclassica, che in questo dopoguerra, e fino a un'epoca relativamente recente, ha dominato l'insegnamento della macroeconomia nelle università di tutto il mondo (grazie anche al successo del testo Economia di Paul Samuelson - n. 1912, premio Nobel nel 1970 -, che ha avuto oltre 10 milioni di lettori dal 1948 a oggi).
La sintesi neoclassica riassorbe la tesi keynesiana della possibilità di equilibri di sottoccupazione nell'ambito della concezione marginalista tradizionale, legata

all'idea di mercati in cui le variazioni del prezzo tendono ad assicurare l'equilibrio tra domanda e offerta. La natura non concorrenziale del mercato del lavoro spiega la disoccupazione, causata come si è detto dalla rigidità dei salari verso il basso; ciò apre la strada a riconoscere l'utilità dell'intervento pubblico nell'economia, perché la disoccupazione può essere combattuta tramite l'utilizzo della politica fiscale e monetaria, utili in generale per regolare l'andamento dell'economia evitandone o riducendone le oscillazioni cicliche. Naturalmente, in presenza di un qualche potere di mercato da parte dei sindacati, l'intervento pubblico diretto a favorire la riduzione della disoccupazione può contemporaneamente favorire un aumento del tasso di crescita dei salari monetari, e quindi dell'inflazione. Il trade-off (relazione inversa) tra disoccupazione e tasso d'inflazione è stato riproposto in un celebre articolo del 1958 dall'economista neozelandese A.W. Phillips (1914-1975); la curva decrescente che rappresenta tale relazione inversa (detta 'curva di Phillips') costituisce, per gli economisti della 'sintesi neoclassica', l'insieme delle possibili scelte di politica economica. Ma, come si accennerà più avanti, tale concezione è stata oggetto di varie critiche negli ultimi venticinque anni.

b) Monetaristi e teorici delle aspettative razionali

All'interno della tradizione marginalista si apre, a partire dagli anni cinquanta, un vivace dibattito sulla plausibilità delle ipotesi necessarie per assicurare il risultato 'keynesiano' di una disoccupazione persistente. Questo dibattito in sostanza riguarda la forza dei meccanismi di mercato nel ristabilire l'equilibrio tra domanda e offerta nel caso del lavoro, e l'opportunità dell'intervento pubblico in campo economico. Fra quanti nutrono fiducia nei meccanismi riequilibratori del mercato e ostilità verso l'intervento pubblico nell'economia, particolare rilievo ha la Scuola di Chicago. Milton Friedman (n. 1912, premio Nobel nel 1976) è considerato il massimo esponente di questa scuola. Egli

sviluppa una teoria della moneta diversa da quella di Keynes, riprendendo e sviluppando le tesi della vecchia teoria quantitativa. In particolare, nel lungo se non nel breve periodo, il livello di equilibrio del reddito dipende da fattori 'reali' come le dotazioni di risorse, la tecnologia e le preferenze dei soggetti economici; la velocità di circolazione della moneta è considerata come funzione stabile dei tassi di rendimento dei vari tipi di attività (moneta, titoli, beni, 'capitale umano').

Friedman sostiene quindi che le vicende monetarie, in particolare l'offerta di moneta (che è considerata esogena), possono influire sul reddito e sull'occupazione solo nel breve periodo; nel lungo periodo le variazioni dell'offerta di moneta influiscono solo sul livello generale dei prezzi (in altri termini, la 'curva di Phillips' risulta inclinata negativamente solo nel breve periodo, ma diventa verticale nel lungo periodo).Inoltre Friedman condanna gli interventi di politica monetaria e fiscale diretti a sostenere la domanda globale e quindi il reddito e l'occupazione; non solo perché l'efficacia di questi interventi è limitata al breve periodo, ma anche perché gli stessi effetti di breve periodo sono incerti, e possono anzi risultare negativi. Infatti, ricorda Friedman, gli interventi di politica economica sono soggetti a tre tipi di ritardi e di incertezze: 1) quelli inerenti alla valutazione della situazione su cui intervenire; 2) quelli inerenti al passaggio dalla valutazione alla decisione dell'intervento e alla sua attuazione; e, infine, 3) quelli relativi al passaggio dall'attuazione dell'intervento al momento in cui si esplicano i suoi effetti. A causa di questi ritardi e incertezze è possibile, ad esempio, che gli interventi esercitino gli effetti previsti ma in una situazione diversa, in cui sarebbero necessari interventi di segno opposto. Quindi gli interventi di politica economica possono avere un effetto destabilizzante, cioè ampliare, anziché ridurre, le fluttuazioni del reddito.

Una tesi ancora più estrema viene sostenuta dai teorici delle 'aspettative razionali' (gli statunitensi J.F. Muth, R.E. Lucas, T.J. Sargent). Secondo questi

ultimi, i soggetti economici imparano a tener conto dell'intervento pubblico nell'economia, scontandone gli effetti in anticipo; così, ad esempio, una spesa pubblica in disavanzo, cioè non finanziata con un aumento contemporaneo delle tasse, decisa per stimolare la domanda globale, viene controbilanciata da una riduzione dei consumi privati, decisa per accantonare i risparmi con cui pagare le tasse che prima o poi dovranno venire imposte per far fronte agli oneri del debito pubblico con cui viene finanziata la spesa pubblica. Pertanto, la 'curva di Phillips' risulta verticale anche nel breve periodo: gli interventi di politica monetaria e fiscale espansivi possono produrre solo aumenti del tasso d'inflazione, e non del livello di disoccupazione. (Si può notare, per inciso, che queste ipotesi presuppongono che tutti i soggetti economici condividano uno stesso modello di funzionamento dell'economia, e siano dotati di una cultura economica e di una capacità di previsione che sarebbe un eufemismo definire irrealistiche).L'unico tipo di politica economica ammessa dai teorici delle aspettative razionali è quello diretto a ridurre le frizioni nel funzionamento del mercato: le cosiddette 'politiche dell'offerta', consistenti ad esempio nel facilitare la mobilità dei lavoratori da un posto di lavoro a un altro, o nell'assicurare che le qualifiche di cui la forza-lavoro del paese viene dotata corrispondano agli sbocchi professionali offerti dal mercato.

c) I postkeynesiani

Di fronte alla reinterpretazione della teoria di Keynes proposta dalla sintesi neoclassica e alle critiche monetariste, si è avuta una decisa reazione da parte di un gruppo di economisti noti come postkeynesiani (come gli inglesi Joan Robinson, 1903-1983, e Richard Kahn; l'ungherese, naturalizzato inglese, Nicholas Kaldor, 1908-1986 ; gli statunitensi Sidney Weintraub, 1914-1983, e Hyman Minsky, n. 1920).Questi economisti sottolineano che lo schema IS-LM proposto da Hicks e condiviso dagli economisti della sintesi neoclassica relega

in secondo piano l'elemento più caratteristico della concezione keynesiana: l'incertezza sul futuro che domina le decisioni degli operatori economici. Nel caso della funzione degli investimenti, ben più importanti del tasso d'interesse sono le aspettative degli imprenditori sulla redditività dei vari progetti d'investimento: aspettative che Keynes considera 'volatili', nel senso che cambiano continuamente, a seconda ad esempio del clima politico e delle condizioni economiche generali.
Nel caso della domanda di moneta, Keynes considera le aspettative sul futuro (per la precisione, sull'andamento futuro del tasso d'interesse) essenziali nel determinare la domanda di moneta a scopo speculativo, che egli considera la componente principale della domanda di moneta, e soprattutto quella che ne determina la continua variabilità. Di fronte alla rilevanza dell'incertezza sul futuro, alla volatilità delle aspettative, e alla conseguente variabilità delle relazioni che legano gli investimenti e la domanda di moneta a scopo speculativo al tasso d'interesse, gli economisti postkeynesiani considerano fuorviante la rappresentazione di mercati in equilibrio, per i beni come per la moneta, cioè la concezione che è alla base dello schema IS-LM. In luogo dell'equilibrio simultaneo dei vari mercati, tipico della tradizione marginalista e ripreso nello schema IS-LM, gli economisti postkeynesiani propongono una caratterizzazione del sistema economico basata su una sequenza di nessi di causa ed effetto: la domanda speculativa di moneta, assieme alla politica monetaria della banca centrale, influisce sul tasso d'interesse; questo a sua volta, assieme alle aspettative, influisce sul livello degli investimenti; gli investimenti, tramite il moltiplicatore, determinano il livello del reddito e dell'occupazione. In questo modo si sottolinea l'influenza che le vicende dei mercati monetari e finanziari esercitano sul reddito e sull'occupazione, in contrapposizione alle tesi della tradizione classica e marginalista sulla 'neutralità' della moneta. Inoltre, vari economisti postkeynesiani sostengono che

l'offerta di moneta è endogena: cioè che la quantità di moneta in circolazione (particolarmente la moneta bancaria) non è controllata in modo rigido dalle autorità monetarie, ma dipende, almeno in parte, dalle decisioni di altri soggetti.

d) L'offerta di moneta in Keynes e nei monetaristi

Data l'importanza non solo teorica ma anche pratica che riveste la questione del modo di concepire l'offerta di moneta, conviene ritornare sui punti di vista espressi da Keynes e dai monetaristi.Si è già ricordato che vari economisti postkeynesiani hanno sostenuto che l'offerta di moneta è endogena; hanno sostenuto tale tesi in chiave critica sia dello schema IS-LM di Hicks sia delle teorie dei monetaristi. Ora, è necessario ricordare che la massa monetaria si compone di diversi tipi di mezzi monetari, fra cui i principali sono i biglietti della banca centrale e i depositi presso le banche ordinarie. Se per i biglietti il controllo dell'autorità monetaria è diretto, pur se neppure qui totale, per i depositi è solo indiretto. Le variazioni dei depositi dipendono infatti principalmente dalle decisioni delle imprese; tuttavia, le loro iniziative assumono rilevanza monetaria solo se le banche ordinarie accolgono le richieste di prestiti concedendo aperture di credito o trasformando titoli di credito che non hanno funzioni monetarie, come le cambiali, in depositi, che costituiscono moneta bancaria. A loro volta, le decisioni delle banche ordinarie nel creare depositi possono incontrare un limite superiore nella politica della banca centrale; tuttavia, l'affermazione che la banca centrale è in grado di porre un limite all'espansione della moneta bancaria è ben diversa dall'affermazione che la massa dei mezzi monetari sia determinata dalla banca centrale. Questi due flussi di mezzi monetari dipendono dunque da impulsi significativamente, anche se solo parzialmente, diversi: come è dimostrato dal fatto che in certe circostanze, quali quelle che si verificarono all'insorgere della grande depressione nel 1929, la massa dei biglietti non diminuì e anzi ebbe un sia pur

limitato aumento, mentre la massa dei depositi subì una forte contrazione. Fra l'altro, ciò indica che anche nelle fluttuazioni cicliche l'influenza della politica monetaria adottata dalla banca centrale è importante ma non dominante, almeno di norma.

Né la tesi del carattere endogeno né quella del carattere esogeno della moneta sono dunque pienamente valide; sembra tuttavia più vicina al vero la prima tesi. Sia Keynes nella Teoria generale, sia Friedman assumono come esogena la moneta; sotto questo aspetto Hicks nel suo modello IS-LM non tradisce il pensiero di Keynes. Tuttavia, mentre nel caso della teoria keynesiana l'assunzione di moneta esogena può essere abbandonata senza gravi conseguenze (le conseguenze possono essere gravi per lo schema di Hicks), ciò non vale per la teoria monetarista, la quale non può fare a meno di quell'assunzione: soltanto con essa, infatti, la moneta può apparire come motore e non come cinghia di trasmissione. La teoria monetarista tuttavia - l'abbiamo già accennato - considera la moneta come motore non delle quantità reali (che possono variare per effetto della politica monetaria solo in via transitoria), ma solo delle quantità nominali, in particolare dei prezzi. La grave limitazione che i moderni monetaristi hanno in comune con i sostenitori dell'antica teoria quantitativa della moneta - tra cui emerge lo statunitense Irving Fisher (1867-1947) - sta proprio in ciò, che essi non riescono a concepire variazioni generalizzate dei prezzi che non siano da attribuire a variazioni della quantità di moneta, mentre in realtà hanno luogo sia diminuzioni sia aumenti generalizzati dei prezzi non provocati da impulsi monetari. Ad esempio, la moneta non gioca un ruolo attivo quando i prezzi variano perché i salari monetari crescono più rapidamente della produttività del lavoro, o perché variano i prezzi delle materie prime e delle fonti di energia, o per misure di politica tributaria capaci d'influire sui prezzi attraverso i costi; e gli esempi potrebbero continuare. Si verificano indubbiamente circostanze in cui variazioni generalizzate dei prezzi dipendono

da quelle della massa dei mezzi monetari; così, per esempio, quando durante una guerra o in tempo di pace la banca centrale aderisce alla richiesta del governo di finanziare il disavanzo del bilancio statale con la creazione di biglietti, può aver luogo un processo inflazionistico.

In breve, la formula con cui Fisher aveva sintetizzato l'antica teoria quantitativa della moneta, rielaborata ma non radicalmente modificata da Friedman, e cioè MV = PQ ovvero P = MV/Q, ha la natura di una tautologia e non può non essere sempre vera; per trasformarla in una formula esplicativa, occorre fare delle ipotesi su quali fra le quattro variabili che compaiono nella formula vadano considerate date, cioè determinate da fattori esterni alla teoria (o 'esogeni'). Se si assumono come dati la velocità di circolazione, V, e la quantità di beni, Q, lasciando liberi di variare M e P; se inoltre si assume che M vari per effetto del finanziamento di un deficit pubblico; allora è vero che le variazioni del livello dei prezzi dipendono da quelle di M. Ma se si ipotizza che il livello dei prezzi vari per impulsi esterni alla formula, come quelli provenienti dai costi cui si è fatto riferimento poco fa, allora la massa dei mezzi monetari e la velocità di circolazione giocano un ruolo passivo e non attivo nelle variazioni dei prezzi. Anzi, se l'autorità monetaria adotta una politica accomodante e accresce la quantità di moneta, man mano che i prezzi aumentano per effetto dell'aumento di certi costi, allora all'aumento di P si accompagna l'aumento di M, ma il secondo aumento è essenzialmente effetto e non causa del primo aumento.

Se invece l'autorità monetaria adotta una politica non accomodante ma restrittiva, può aver luogo una recessione senza che essa blocchi l'aumento dei prezzi. (In un secondo momento, tuttavia, la crescita dei prezzi potrà rallentare o arrestarsi, se proviene da un aumento dei salari, e se questo viene frenato dall'aumento della disoccupazione che si accompagna alla recessione).Il fatto che la relazione P = MV/Q è sempre vera non autorizza dunque ad attribuire

sempre alla massa monetaria la responsabilità di aumenti generalizzati dei prezzi. Osservazioni analoghe valgono per flessioni generalizzate dei prezzi, come quelle che avvennero ripetutamente nel secolo scorso: queste flessioni vanno collegate principalmente ad aumenti della produttività del lavoro accompagnati da aumenti più lenti nei salari nominali. In contrasto con certe spiegazioni che s'ispiravano a qualche variante della teoria quantitativa, il ruolo giocato dalla moneta - nel secolo scorso nei paesi più sviluppati la moneta di base era l'oro - fu, se non proprio nullo, decisamente secondario.

e) Piero Sraffa

Il disegno culturale perseguito da Piero Sraffa (1898-1983) è decisamente di vasta portata: operare un capovolgimento delle linee di ricerca della scienza economica, detronizzando l'approccio marginalista dominante e proponendo in suo luogo l'impostazione originaria degli economisti classici.In analogia con la linea d'indagine seguita dagli economisti classici, in Produzione di merci a mezzo di merci (pubblicato nel 1960) Sraffa pone al centro della sua analisi i rapporti che intercorrono in condizioni 'normali' tra i vari settori, o industrie, in cui si articola un sistema economico basato sulla divisione del lavoro. Come si è già accennato a proposito dell'economia politica classica, ciascun settore deve entrare in contatto con gli altri settori dell'economia per ottenere da essi i propri mezzi di produzione in cambio di una parte almeno del proprio prodotto. Si ha così quella rete di scambi che caratterizza le economie basate sulla divisione intersettoriale del lavoro. Come mostra Sraffa, il problema della determinazione dei rapporti di scambio che si stabiliscono tra i vari settori va affrontato, in un'economia capitalistica, simultaneamente al problema della distribuzione del reddito tra le classi sociali dei lavoratori, dei capitalisti e dei proprietari terrieri. L'intersezione tra questi due problemi costituisce ciò che nella tradizione classica è indicato come problema del valore.

La critica dell'approccio marginalista proposta da Sraffa pone in rilievo il fatto che il 'capitale' è in realtà un insieme di mezzi di produzione prodotti, i cui prezzi variano in modo non univoco al variare della distribuzione del reddito, di modo che non è possibile affermare a priori che una riduzione del salario provochi una riduzione nell'utilizzo di 'capitale' rispetto al lavoro.
Senza entrare nei dettagli analitici del dibattito, ci limitiamo a osservare che la critica di Sraffa colpisce una tesi vitale per la tradizione marginalista: l'idea che la riduzione del salario reale causata dalla disoccupazione nel caso di un mercato del lavoro concorrenziale porti a un riassorbimento della disoccupazione stessa, inducendo gli imprenditori a scegliere tecniche a maggiore intensità di lavoro (e a minore intensità di capitale). Questo meccanismo è essenziale per sostenere la tesi della capacità autoregolatrice del mercato, e la visione dell'economia come scienza che studia, appunto, i meccanismi equilibratori del mercato.Inoltre, Sraffa ripropone l'approccio classico (di cui con la sua edizione critica delle opere di Ricardo aveva riscoperto le fondamenta concettuali e la struttura analitica), risolvendo il problema centrale del valore lasciato aperto dagli economisti classici e da Marx. La soluzione di Sraffa consiste nel determinare simultaneamente prezzi relativi e una variabile distributiva 'residuale' (salario o saggio del profitto), data la tecnologia corrispondente a un determinato insieme di livelli di produzione e una variabile distributiva 'esogena'. Sulla scia del lavoro di Sraffa, numerosi economisti - fra i quali diversi italiani - hanno sviluppato i vari aspetti del progetto culturale di ripresa della concezione classica: con analisi di storia del pensiero economico, dirette a chiarire le fondamenta concettuali dell'approccio classico e a distinguerle da quelle proprie dell'approccio marginalista; con lavori di critica a specifiche teorie marginaliste, in particolare relative alla teoria della distribuzione, dell'occupazione, dell'accumulazione, del commercio internazionale, e così via; con sviluppi analitici diretti ad approfondire su punti

specifici (come la produzione congiunta, il capitale fisso, la scelta delle tecniche) il contributo offerto nel libro di Sraffa; infine, con lavori meno direttamente connessi all'analisi sraffiana, ma diretti a proporre teorie d'impostazione classica sui diversi problemi dell'economia politica, dalla teoria della distribuzione e dello sviluppo alla teoria delle forme di mercato. Attualmente fra le forme di mercato sembrano preminenti quelle di tipo oligopolistico, il cui studio può essere utilizzato sia in analisi parziali (analisi di singoli mercati), sia in analisi di carattere generale (come quella consentita dal sistema di Sraffa), sia nell'analisi delle variazioni nel tempo dei prezzi e delle quote distributive. (Un'analisi di questo tipo è stata elaborata dal polacco Michal Kalecki, 1899-1970, noto soprattutto per aver precorso e poi sviluppato in modo originale alcuni aspetti centrali della teoria keynesiana).

4. L'analisi dinamica

a) Diversi tipi di modelli; il processo di meccanizzazione

Le variazioni su cui si concentra l'analisi marginalista sono variazioni istantanee e ipotetiche, fuori dal tempo. In realtà, l'impostazione di quest'analisi è essenzialmente statica: i pochi modelli teorici dinamici che sono stati elaborati nell'ambito del marginalismo comportano spostamenti ipotetici delle funzioni, un metodo che è dubbio rientri nella dinamica. I modelli dell'analisi marginalista non soltanto sono statici, ma soprattutto non sembrano suscettibili, neppure attraverso successive approssimazioni, di sviluppi dinamici. È un limite grave, se si considera che la nostra epoca è dominata dai mutamenti tecnologici e dal processo di sviluppo economico.Gli economisti classici, segnatamente Adam Smith e David Ricardo, attribuivano invece importanza fondamentale a entrambi i fenomeni. Questo è manifestamente vero per Smith (il processo della crescente divisione del lavoro consiste appunto in una serie ininterrotta di

mutamenti grandi e piccoli dei metodi produttivi da cui consegue un aumento sistematico della produttività del lavoro), ma è vero anche per Ricardo, non solo e non tanto per la sua famosa analisi riguardante i possibili effetti negativi sull'occupazione derivanti dall'introduzione di macchine, quanto per il fatto che l'analisi delle tendenze delle quote distributive (salari, profitti e rendite) è posta al centro della sua costruzione teorica proprio per la sua importanza fondamentale in relazione al processo di accumulazione e di sviluppo. In particolare Ricardo pensava che la tendenza, da lui presunta, all'aumento progressivo della quota del reddito nazionale destinata alle rendite fondiarie fosse preoccupante proprio perché avrebbe comportato la progressiva compressione della quota destinata ai profitti, da cui dipende il processo di accumulazione.

Fra i modelli teorici di sviluppo non formalizzati, oltre quelli degli economisti classici, troviamo il modello dello sviluppo ciclico di Schumpeter, di cui si è detto, e il modello di Marco Fanno (1878-1965), che riguarda principalmente il ciclo economico ma, subordinatamente, anche il processo di sviluppo. Fra i modelli formalizzati troviamo quello proposto dal grande matematico John von Neumann (1903-1957), e i modelli di derivazione keynesiana elaborati in Gran Bretagna da Roy Harrod (1900-1978) e negli Stati Uniti da Evsey Domar (n. 1914). L'originaria teoria keynesiana aveva essenzialmente carattere statico, ma si è rivelata suscettibile di sviluppi dinamici. Ci sono poi modelli formali di crescita collegati più alla lontana con la teoria keynesiana, come quelli di Nicholas Kaldor e di Luigi Pasinetti (n. 1930). In questi modelli, come già nelle analisi di Smith, si attribuisce il massimo rilievo al progresso tecnico, visto come originato all'esterno del sistema economico.

Quest'ultimo punto merita riflessione. Infatti, mentre certe innovazioni possono essere considerate indipendenti da impulsi economici, altre invece provengono essenzialmente da impulsi di tal genere, come l'espansione della domanda e

l'aumento dei costi, cosicché è compito dell'economista studiarli. È da notare che Smith, secondo il quale la divisione del lavoro è condizionata dall'estensione del mercato, aveva già ben compreso che la crescita della domanda influisce in modo significativo sul ritmo del progresso tecnico. Di norma, possono essere considerate in gran parte indipendenti da impulsi economici le innovazioni di grande rilievo, originate da invenzioni scientifiche fuori dall'ordinario, mentre le innovazioni più frequenti, di rilievo più modesto, spesso semplici perfezionamenti di grandi innovazioni, sono indotte principalmente da impulsi economici, fra cui sono appunto l'espansione della domanda e l'aumento dei costi. Particolare importanza assumono le innovazioni determinate da aumenti del costo relativo del lavoro, ossia da aumenti del rapporto fra salari e prezzi delle macchine. In effetti, lo sviluppo del capitalismo industriale moderno è stato caratterizzato da un processo di progressiva meccanizzazione (e in tempi recenti di progressiva automazione) dei processi produttivi. L'aumento della produttività del lavoro che accompagna la meccanizzazione può aver luogo in presenza di salari monetari tendenzialmente stabili, o crescenti più lentamente della produttività; ovvero può aver luogo in presenza di salari monetari che crescono con la stessa velocità o più rapidamente della produttività. In tutti i casi, cresce il rapporto fra salari e prezzi (compresi i prezzi delle macchine); e l'aumento di questo rapporto stimola la sostituzione di macchine a lavoro. Il processo si autoalimenta, nel senso che l'aumento di quel rapporto, che stimola l'introduzione di nuove macchine, fa crescere la produttività; a sua volta, tale aumento consente un aumento dei salari rispetto a tutti i prezzi, compresi i prezzi degli stessi beni acquistati dai lavoratori; di conseguenza, quel rapporto subisce un nuovo aumento. Il processo comporta un aumento sistematico dei salari reali.
Nei periodi in cui l'aumento della domanda globale di beni è stato più lento dell'aumento di produttività è emersa una disoccupazione, che può essere vista

come disoccupazione tecnologica in quanto l'aumento di produttività di norma trae origine da innovazioni tecnologiche. Non è affatto necessario, tuttavia, che le innovazioni determinino disoccupazione: spesso la domanda aggregata aumenta alla stessa velocità e anche più rapidamente della produttività, cosicché la disoccupazione non compare o, al contrario, ha luogo un aumento dell'occupazione. D'altra parte, in certe condizioni la domanda aggregata diminuisce: compare allora una disoccupazione particolare, di tipo keynesiano.

b) Le variazioni di lungo periodo dei salari e dei prezzi

L'aumento dei salari reali può aver luogo quando i prezzi dei beni acquistati dai lavoratori diminuiscono mentre i salari monetari restano stabili o aumentano limitatamente; oppure quando i salari monetari aumentano mentre i prezzi restano stabili o crescono più lentamente dei salari. La prima tendenza ha avuto luogo nel secolo scorso (in Inghilterra i prezzi dei prodotti finiti sono diminuiti di oltre il 70%, negli Stati Uniti di quasi la metà, mentre i salari monetari sono aumentati, rispettivamente, del 70% e del 90%, ossia, in media, dello 0,5 o 0,6 % l'anno). La seconda tendenza - salari che aumentano più rapidamente dei prezzi, anch'essi in aumento - ha avuto luogo dopo la fine della seconda guerra mondiale. (Nel periodo compreso fra le due guerre mondiali salari e prezzi hanno subito violente oscillazioni; questo periodo è stato dominato dalla grande depressione e richiede un'analisi a sé stante). Non è indifferente che l'aumento di produttività dia luogo a un aumento dei salari reali attraverso una flessione dei prezzi mentre i salari monetari restano stabili o in moderato aumento, ovvero attraverso un aumento dei salari monetari mentre i prezzi restano stabili o in moderato aumento. Il primo meccanismo infatti stimola il processo di sviluppo, tramite una catena di diminuzioni di costi che si verifica quando l'aumento di produttività ha luogo in un settore che produce mezzi di produzione: in regime di concorrenza, la diminuzione dei costi crea extra-

profitti che inducono le imprese a espandere la produzione, provocando diminuzioni di prezzo del prodotto (e quindi ulteriori diminuzioni dei costi nei settori che lo utilizzano come mezzo di produzione) fino a quando gli extra-profitti non vengono riassorbiti.
Questa specifica sequenza, favorevole al processo di sviluppo, viene meno quando - come accade se prevalgono forme di mercato oligopolistiche - i prezzi non diminuiscono e i lavoratori partecipano ai frutti del progresso tecnico tramite aumenti dei salari monetari. In quest'ultimo caso restano, come stimoli allo sviluppo già presenti nel periodo precedente (ma con importanza minore), le innovazioni e gli aumenti della domanda proveniente dall'estero o dal settore pubblico.
La transizione dalla prima alla seconda delle due tendenze alternative, che contrassegnano due successivi periodi storici, è collegata a profonde trasformazioni nella struttura delle moderne economie capitalistiche. Le principali trasformazioni sono la conseguenza del processo di concentrazione, che ha condotto alla formazione di grandi e grandissime imprese, spesso di dimensioni internazionali, e del processo di differenziazione dei prodotti, consentito da un crescente livello del reddito individuale anche nelle fasce relativamente più povere della popolazione e favorito dalla diffusione della pubblicità. I due processi si sono verificati nei mercati dei prodotti industriali e, con caratteristiche particolari, nelle attività finanziarie e creditizie. Al tempo stesso, si sono rafforzati i sindacati dei lavoratori e si sono diffuse, nel mercato del lavoro, varie forme di contrattazione collettiva, ciò che ha contribuito a determinare una crescente rigidità verso il basso dei salari. Nel secolo scorso e ancora fino alla seconda guerra mondiale accadeva che i salari monetari diminuissero; nei moderni paesi capitalistici ciò non si è più verificato dopo la seconda guerra mondiale, anzi, in questi paesi, i salari aumentano almeno in proporzione alla produttività. Corrispondentemente, i prezzi all'ingrosso dei

prodotti finiti ben di rado diminuiscono e, se ciò accade, la diminuzione è minima. Restano flessibili verso l'alto come verso il basso i prezzi delle materie prime, nei cui mercati hanno avuto scarsa rilevanza i due processi, sopra ricordati, di concentrazione delle imprese e di differenziazione dei prodotti. Con salari rigidi verso il basso e flessibilità minima dei prezzi dei prodotti finiti, quando si verifica una crisi o una depressione, la ripresa, che nel passato era pressoché automatica, oggi incontra difficoltà molto maggiori. Infatti, l'aumento della domanda reale non è più stimolato dal meccanismo concorrenziale di diffusione a catena di riduzioni dei costi, extra-profitti, aumenti di produzione e riduzioni dei prezzi. Nelle nuove condizioni, la ripresa può avere luogo o per effetto d'investimenti stimolati da innovazioni, o per un aumento della domanda estera, o per un'azione del governo. Nel caso che le prime due spinte siano insufficienti, è il governo che deve intervenire: la sua azione può consistere non solo in un aumento delle spese pubbliche, ma anche in una politica creditizia attiva. Gli stessi sindacati, spingendo in alto i salari, possono contribuire all'aumento della domanda.Le considerazioni appena svolte si collegano alla storia economica più che alla teoria economica, e costituiscono elementi di un'interpretazione che non è generalmente accolta. Ciò nonostante le abbiamo proposte, perché i più recenti sviluppi della teoria economica (di cui si è discusso sopra, nel cap. 3) ben difficilmente possono essere compresi se non si fa riferimento alle trasformazioni strutturali verificatesi nelle moderne economie capitalistiche. Inoltre, quelle considerazioni possono servire a illustrare un tema fondamentale già ricordato al principio della nostra trattazione, cioè i rapporti fra teoria economica e storia. Gli economisti che ignorano la necessità di tali rapporti continuano imperterriti a costruire modelli fondati sulle ipotesi, del tutto irrealistiche nel mondo di oggi, della concorrenza atomistica e della flessibilità verso il basso come verso l'alto dei prezzi e dei salari.

Pur non analizzandole esplicitamente, Keynes aveva correttamente interpretato le conseguenze di quelle trasformazioni, raccomandando, in particolari condizioni, una politica attiva del credito e una vigorosa espansione delle spese pubbliche, anche in deficit, per promuovere la ripresa economica. Sulla scia di queste raccomandazioni per l'assunzione di un ruolo attivo dello Stato nell'economia, e sulla scia del crescente peso politico dei lavoratori, un programma di spese sociali fu poi sistematicamente elaborato da lord Beveridge (1879-1963), la cui opera Pieno impiego in una società libera (1946), preparata con la collaborazione di Kaldor e ricollegandosi esplicitamente alle teorie keynesiane, ebbe grande influenza sulle politiche sociali di tutti i paesi capitalistici.

Già in passato lo Stato era intervenuto per proteggere le fasce più deboli della popolazione in diversi campi, segnatamente in quelli delle pensioni, della sanità e della disoccupazione. Ma solo dopo la seconda guerra mondiale l'azione pubblica ha assunto le dimensioni che tutti conosciamo. Ciò è stato reso possibile, fra l'altro, dall'accresciuto reddito individuale medio. Tuttavia in alcuni paesi una spesa pubblica eccessiva (anche per il suo utilizzo come 'ammortizzatore sociale' dopo la crisi petrolifera del 1973-1974) ha contribuito negli anni settanta e ottanta alla formazione e alla crescita dei disavanzi pubblici non di breve ma di lungo periodo. A sua volta, ciò ha contribuito - assieme alla diffusione, soprattutto negli Stati Uniti e in Inghilterra, di dottrine monetariste favorevoli all'uso di politiche monetarie restrittive come strumento di lotta all'inflazione - a spingere verso l'alto il tasso dell'interesse, frenando così gli investimenti e la crescita produttiva (non solo nei paesi sviluppati, ma anche e soprattutto nei paesi del Terzo Mondo, spesso appesantiti da enormi debiti esteri). Al tempo stesso, in vari paesi sviluppati si è avuta una reazione dell'opinione pubblica contro l'aumento della pressione fiscale (la cosiddetta 'rivolta fiscale'), e contro l'espansione delle spese pubbliche o, più in generale,

contro gl'interventi pubblici nell'economia. La diffusione del monetarismo o di teorie come quella delle aspettative razionali, alle quali si è accennato sopra, costituiscono espressioni teoriche di questa reazione.
L'importanza dell'azione pubblica nella vita economica, comunque, non deve essere misurata semplicemente considerando il peso delle spese pubbliche rispetto al prodotto interno lordo: un peso che in certi paesi giunge al 50% e che nella patria del capitalismo privato, gli Stati Uniti, si aggira sul 35%. Occorre anche considerare il carattere dell'azione pubblica nei diversi settori, specialmente nel fondamentale settore del credito, il cui vertice - la banca centrale - è ormai in tutti i paesi un'istituzione pubblica, pur dotata di autonomia, e la cui base - le aziende di credito - è in vari modi controllata dall'autorità monetaria o è addirittura, in parte, posseduta dallo Stato o da enti pubblici (in Italia la quota dei depositi che fa capo ad aziende di credito pubbliche raggiunge il 70%; in altri paesi sviluppati la quota è minore, ma non è mai trascurabile). D'altra parte, il gran peso assunto dai titoli pubblici nei mercati finanziari ha reso possibile un controllo pubblico della politica creditizia impensabile nel secolo scorso. Anche questi interventi di carattere istituzionale hanno dato luogo ad abusi molto gravi, cosicché anche in questo caso le reazioni contro gl'interventi pubblici sono ben comprensibili. Tuttavia, il rimedio non sta nella condanna globale e acritica di tali interventi; il rimedio sta nell'introduzione di cambiamenti organizzativi tendenti a eliminare gli abusi, promuovendo con decisione, non la riprivatizzazione generalizzata delle aziende - creditizie e non creditizie - controllate dallo Stato, il che non sarebbe possibile, ma una privatizzazione ampia e differenziata secondo un ben meditato ordine di priorità.Si deve osservare tuttavia che l'idea, condivisa da parecchi economisti, che lo Stato da un lato e i sindacati dall'altro costituiscano un male in sé per l'economia, è un'idea erronea. Il fatto è che pur in presenza di sindacati relativamente forti e di un accresciuto ruolo dello Stato nell'economia,

dopo la fine della seconda guerra mondiale lo sviluppo economico e civile è stato più, e non meno, sostenuto che nel passato: nonostante errori, eccessi e sprechi di ogni genere, la somma algebrica è stata positiva.

c) Il sottosviluppo

Fra i modelli dinamici ve ne sono alcuni che riguardano lo sviluppo dei paesi arretrati. Quello del sottosviluppo è forse il maggiore problema della nostra epoca, per i suoi riflessi umani, civili, ambientali. Qui possiamo fornire solo alcuni brevissimi cenni.In via preliminare, si può osservare che per diversi importanti aspetti l'analisi dei paesi sottosviluppati può trovare un valido punto di partenza nelle opere degli economisti classici, a cominciare dall'opera di Adam Smith, giacché la situazione odierna dei diversi paesi del Terzo Mondo, pur con profonde differenze, presenta non poche analogie con i paesi europei del Settecento. Gli economisti classici insistevano sull'impiego, produttivo o improduttivo, del sovrappiù, che allora era costituito da tutti i redditi non da lavoro (profitti, interessi e rendite), i quali potevano essere in parte risparmiati o tassati. Oggi la distinzione fra impieghi produttivi e improduttivi non è considerata rilevante dalla teoria moderna, mentre lo è se si considerano i paesi del Terzo Mondo, nei quali pertanto acquista importanza essenziale esaminare sia il reimpiego produttivo del sovrappiù da parte delle stesse unità produttive, sia l'apparato per il trasferimento volontario di una parte del sovrappiù (sistema creditizio), sia l'apparato per il trasferimento coattivo (sistema tributario).
Al tempo stesso, acquista importanza essenziale distinguere il sovrappiù che è tale per l'intera economia dal 'sovrappiù' che è tale solo per singoli privati: il primo comporta una crescita del reddito, il secondo una sua redistribuzione.
Con riferimento ai profitti, questa distinzione corrisponde a quella proposta, sulla scia degli economisti classici, da Alberto Breglia (1900-1955), fra profitti 'sterili' e profitto 'fecondo'. I primi sono profitti 'da sottrazione', in quanto

corrispondono a un trasferimento di risorse da un soggetto a un altro, e non a un aumento delle risorse complessive. I profitti 'da sottrazione' possono essere imputati a posizioni di monopolio o a operazioni puramente speculative in periodi di inflazione ovvero - ma qui i profitti sarebbero 'distruttivi' e non soltanto sterili - possono provenire dalla produzione e dal commercio di sostanze stupefacenti e attività consimili. Il profitto 'fecondo' o 'da addizione', invece, è quello proveniente da una crescita della produttività e dalla conseguente riduzione dei costi. Se si ammette che i soggetti economici tendono a ripetere le operazioni alle quali sono abituati, si può presumere che di regola coloro che ottengono un profitto 'fecondo' tendono a reimpiegarlo produttivamente, cosicché la spirale produttiva tende a perpetuarsi, originando un processo di sviluppo che è tale sia per il singolo sia per la società. Questi concetti sono tutti presenti, esplicitamente o implicitamente, nell'impostazione stessa delle analisi elaborate dai classici, e riacquistano tutta la loro importanza nello studio dei paesi sottosviluppati.

Tra i modelli teorici relativi ai paesi sottosviluppati quello di Arthur Lewis (n. 1915, premio Nobel nel 1979) - che riguarda specialmente i paesi delle fasce tropicali e subtropicali - si ricollega, sotto importanti aspetti, alle analisi degli economisti classici, e comunque si situa fuori dalla tradizione marginalista. Il modello di Lewis concentra l'attenzione sulle condizioni dell'offerta di lavoro, che in quei paesi è economicamente illimitata (com'era nella prima fase dell'industrializzazione dei paesi oggi sviluppati), non solo per la crescita demografica, ma anche per la possibilità, per le imprese capitalistiche, di reclutare manodopera sottraendola ad attività premoderne, come quelle svolte nelle tribù. Il modello considera inoltre le condizioni delle produzioni di beni alimentari di base, nelle quali, per la bassa produttività, è assai limitato il sovrappiù che può essere investito sia nella stessa agricoltura sia in attività extra-agricole: come già avevano messo in rilievo i classici, infatti, un sovrappiù

limitato frena l'accumulazione. D'altra parte, l'offerta economicamente illimitata di lavoro - anche questa è una caratteristica che nella sostanza troviamo già nelle analisi dei classici - comporta un livello dei salari basso, vicino al livello di sussistenza, e relativamente stazionario, cosicché ogni aumento di produttività nelle attività basate sul lavoro salariato, come quelle svolte nelle piantagioni e nelle miniere, tende a tradursi in una flessione dei prezzi relativi (che nelle relazioni internazionali sono denominati 'ragioni di scambio'), in direzione sfavorevole ai paesi produttori, cioè in genere ai paesi in via di sviluppo.

Da notare che durante gran parte del secolo scorso, un periodo in cui gli aumenti di produttività si traducevano in flessioni dei prezzi dei beni prodotti dai paesi sviluppati, le 'ragioni di scambio' variavano non contro ma a favore dei paesi in via di sviluppo, giacché nell'industria manifatturiera dei paesi sviluppati la produttività cresceva e i prezzi diminuivano a ritmi più rapidi di quanto accadeva nei paesi arretrati.Il modello di Lewis riguarda in modo particolare i paesi sottosviluppati tropicali che producono materie prime agrarie e minerarie: paesi che si trovano in larghe zone dell'Africa e dell'Asia, e in zone più ristrette dell'America Latina. Per numerosi paesi sottosviluppati dell'America Latina e dell'Asia, tuttavia, valgono modelli interpretativi alquanto diversi. Occorre rilevare che in un numero ancora piccolo ma in continua crescita di paesi asiatici ha avuto luogo un processo di sviluppo economico e, in particolare, industriale, relativamente vigoroso: sono i 'paesi di nuova industrializzazione', fra i quali troviamo la Corea del Sud e Taiwan; l'Indonesia sta entrando in una fase di crescita sostenuta. L'India, oltre a un non trascurabile sviluppo industriale, è riuscita, grazie anche a innovazioni di tipo agrario ('rivoluzione verde'), a ottenere una crescita della produzione di beni alimentari un po' più rapida della pur ragguardevole crescita demografica, cosicché le frazioni della popolazione colpite dalla fame si stanno decisamente

restringendo, mentre permangono estese le fasce di popolazione che soffrono di malnutrizione.
Non solo per l'India, ma in generale per molti paesi in via di sviluppo è importante la questione del rapporto fra crescita delle produzioni di beni alimentari e crescita demografica. Tale questione può essere risolta da paesi arretrati relativamente piccoli procurandosi i beni alimentari di cui hanno bisogno attraverso gli scambi internazionali piuttosto che attraverso la produzione diretta; ma per paesi grandi e popolosi come l'India tale via può rappresentare solo un contributo parziale. Accanto alla questione dello sviluppo produttivo c'è dunque un problema di crescita demografica. Dopo la seconda guerra mondiale nei paesi del Terzo Mondo tale crescita si è accelerata, non per un aumento della natalità ma per una rapida diminuzione della mortalità, imputabile alla costruzione di strutture igieniche e alla diffusione dei farmaci moderni e dei servizi sanitari. In larga misura, il problema della miseria del Terzo Mondo è da attribuire proprio alla rapida crescita demografica. È bene osservare che agli inizi della scienza economica moderna i problemi dello sviluppo produttivo e quelli della crescita demografica erano considerati congiuntamente e non separatamente, come oggi accade. Il problema demografico chiama direttamente in causa i problemi dell'evoluzione culturale e del grado dell'istruzione. Infatti per i demografi è ormai un dato acquisito che il grado dell'istruzione, in particolare quello delle donne (che nei paesi arretrati di regola hanno un'istruzione inferiore, non di rado assai inferiore, a quella degli uomini), condiziona la velocità del declino del saggio di natalità: a parità di reddito individuale, maggiore è il grado d'istruzione delle donne, più rapida è la flessione del saggio di natalità. (La flessione della mortalità prima o poi porta con sé quella della natalità, ma la velocità relativa delle due flessioni ha importanza fondamentale per l'andamento del reddito pro capite.) Una politica demografica tendente ad accelerare la flessione della natalità deve pertanto

collegarsi alla politica di diffusione dell'istruzione; si possono poi utilizzare anche incentivi e disincentivi di carattere economico.
Questi temi vengono tutti trattati, spesso separatamente, da demografi e da economisti. Anche questi temi rientrano nell'ambito della dinamica economica, intesa in senso ampio; e anche per questi temi c'è una ripresa d'interesse da parte degli economisti, alcuni dei quali si rifanno esplicitamente, per questo come per altri aspetti, agli economisti classici come Adam Smith.

d) I problemi dell'ambiente e lo sviluppo sostenibile

Il processo di sviluppo ha portato con sé, nei paesi in cui ha avuto luogo, cospicui benefici, ma ha avuto e sta avendo altresì costi rilevanti sotto l'aspetto dei valori morali e umani e sotto l'aspetto economico. Già al principio di questo secolo, alcuni economisti mettevano in rilievo i costi economici che lo sviluppo economico può comportare, per esempio, per via dell'inquinamento. Negli ultimi decenni gli effetti della crescita esplosiva delle produzioni industriali si sono manifestati in forme sempre più allarmanti. Si può stimare che nei paesi sviluppati la produzione industriale sia cresciuta di oltre venti volte negli ultimi cento anni; se si ammette che le esalazioni, i fumi, i rifiuti provenienti dai processi produttivi e i rifiuti provenienti dai consumatori siano cresciuti in una proporzione simile, ci si rende conto delle dimensioni gigantesche assunte dai problemi che oggi vengono definiti ambientali. Questi problemi inoltre sono stati fortemente aggravati dal fatto che certe produzioni sono risultate non semplicemente inquinanti ma addirittura tossiche, con effetti che si propagano attraverso l'aria, le acque e i terreni, e quindi attraverso le produzioni alimentari.
Di fronte a questi problemi, molti studiosi - economisti e non economisti - si sono chiesti in quale misura e in quale modo lo sviluppo economico sia 'sostenibile', cioè tale da non danneggiare l'ambiente naturale in cui viviamo. Il concetto di 'sviluppo sostenibile' ha attratto sempre più l'attenzione in questi

ultimi anni; dal nostro punto di vista, il problema principale non riguarda la sua concreta definizione, che è affidata a campi scientifici in rapido sviluppo genericamente indicati con il termine 'ecologia', ma il modo in cui ci si può assicurare che le scelte degli operatori economici si muovano nella direzione desiderata. Le questioni rilevanti, da questo punto di vista, riguardano il conflitto tra interessi privati e interesse pubblico, e tra mercato e Stato, ma anche il rapporto tra economia e morale. Si tratta di questioni che sono state al centro del dibattito fin dalla nascita dell'economia politica, alle quali in parte abbiamo già accennato sopra parlando della concezione di Adam Smith (v. § 2b), e che di recente sono tornate a costituire oggetto di riflessione (ricordiamo ad esempio le ricerche di Amartya Sen). Così come non potrebbe funzionare un'economia di mercato in cui il macellaio e il fornaio fossero liberi di adulterare la loro merce, non sarebbe possibile evitare il degrado dell'ambiente naturale senza una coscienza civica che abbia interiorizzato la sua importanza per il benessere sociale, e senza istituzioni capaci di intervenire per imporre il rispetto dei vincoli ambientali nei casi di violazione della norma morale. Il problema del rapporto tra intervento pubblico e libera iniziativa privata nell'ambito di un'economia di mercato appare così come una questione di complementarità, piuttosto che di opposizione. Ma questo non vale solo per le macrostrutture giuridiche e amministrative: vale anche per gli interventi più specifici di politica economica. A titolo esemplificativo, consideriamo alcuni problemi relativi al settore energetico.

La crescita del settore energetico e i cambiamenti nella sua struttura interna sono collegati da complesse relazioni di causa ed effetto all'evoluzione dell'economia nel suo complesso. Così è evidente che i consumi energetici complessivi dipendono strettamente dall'andamento della produzione e del reddito; ed è altrettanto evidente che una crescente disponibilità di energia costituisce un prerequisito per lo sviluppo economico. In altri termini, lo

sviluppo economico è condizionato dall'offerta di energia, ma allo stesso tempo ne determina la domanda. Non dobbiamo trascurare poi il ruolo del cambiamento tecnologico: da un lato la crescente meccanizzazione e l'aumento del prodotto pro capite spingono nella direzione di un'espansione dei consumi energetici; ma dall'altro lato il progresso tecnico, nella costante ricerca della riduzione dei costi di produzione, è anche la fonte di una riduzione dei fabbisogni energetici per unità di prodotto, e di una maggiore efficienza nell'uso di energia in generale. Il risultato netto di queste due spinte contrastanti dipende in misura probabilmente decisiva dall'andamento dei prezzi delle varie forme di energia: nei periodi di crescita di tali prezzi, si ha un processo di 'sostituzione dinamica', in cui imprese e famiglie dedicano maggiore attenzione allo sviluppo e all'applicazione pratica di nuove tecnologie che consentono risparmi energetici; mentre nei periodi di prezzi calanti dell'energia (non necessariamente in assoluto, ma rispetto ai prezzi degli altri mezzi di produzione e di consumo) la spinta a una riduzione dei consumi energetici viene meno, e questi ultimi tendono a seguire da vicino l'andamento della produzione e del reddito. Anzi, i consumi energetici possono crescere più rapidamente della produzione e del reddito, sia per i motivi indicati sopra (crescente meccanizzazione, e quindi crescente 'intensità energetica', della produzione), sia perché i consumi energetici delle famiglie corrispondono in misura notevole alla domanda di beni e servizi per soddisfare bisogni di ordine superiore, che per loro natura assorbono una quota crescente del reddito.
Il progresso tecnico, che procede a velocità diseguale nei vari campi, è il fattore principale anche nel determinare i cambiamenti nella struttura interna del settore energetico. La sua importanza è confermata da due circostanze. In primo luogo, la sequenza legna-carbone-petrolio-fissione nucleare e gas naturale-fusione nucleare ed energia solare, che indica la successione delle fonti di energia dominanti (dove l'ultimo anello della catena indica lo scenario più

verosimile, anche se non l'unico possibile, per la metà del prossimo secolo), appare come una sequenza di miglioramenti nella capacità tecnologica dell'uomo di estrarre energia dalla natura, caratterizzata da fortissimi aumenti dell'offerta di energia a costi mediamente decrescenti. In secondo luogo c'è da registrare la crescente penetrazione dell'elettricità, cioè l'aumento della quota dei consumi di energia soddisfatti dall'elettricità. Tale tendenza assicura maggiore flessibilità all'offerta di energia, dato che l'elettricità può essere prodotta usando diverse fonti primarie, e quindi assicura una maggiore autonomia di politica energetica ai vari paesi, che possono compiere scelte diverse a seconda delle proprie dotazioni di risorse naturali; inoltre, nei sistemi industriali moderni l'elettricità permette un uso più flessibile dell'energia, oltre a costituire il supporto necessario per la diffusione dell'informatica nelle imprese manifatturiere (automazione) e nei servizi.
Lo sviluppo dei consumi di energia ha posto problemi gravissimi per la salvaguardia dell'ambiente naturale che, come si accennava sopra, vanno affrontati dalle autorità pubbliche. Infatti gli effetti sull'ambiente della produzione e dell'utilizzo delle diverse fonti di energia sono un caso classico di 'esternalità', cioè di effetti dell'attività di uno specifico gruppo di produttori o consumatori che non costituiscono costi o benefici per il singolo produttore o consumatore, ma vantaggi o svantaggi per un gruppo più ampio di agenti economici, talvolta per la società nel suo complesso. Nel caso di una 'esternalità negativa' (ad esempio le emissioni di sostanze inquinanti), l'impresa o il consumatore che ne sono responsabili non hanno alcun incentivo economico a limitarne la portata. Perciò, tradizionalmente, la teoria economica suggerisce di controbilanciare le 'esternalità negative' tramite apposite tasse o tramite specifiche norme che impongano limiti o interventi di depurazione. Tuttavia, in pratica, la difficoltà di individuare gli effetti ambientali delle varie attività umane, e in particolare di quelle connesse alla produzione e all'utilizzo di

energia, e poi la difficoltà di determinarne con precisione la portata, hanno favorito in passato un atteggiamento lassista. Solo negli ultimi anni è diventato evidente che le conseguenze ambientali della produzione e del consumo di energia sono state sottovalutate, se non completamente ignorate, nei decenni successivi alla rivoluzione industriale e fino a un'epoca molto recente. Questioni come l'effetto serra, completamente ignorate fino a pochi anni fa, sono ora al centro dell'attenzione. Possiamo prevedere, dunque, che i problemi ambientali avranno un peso crescente nelle scelte strategiche nel campo dell'energia.

La necessità di favorire scelte compatibili con il rispetto dell'ambiente naturale implica sia una normativa sempre più precisa e vincolante sulle diverse fonti di energia (ad esempio sulla sicurezza delle centrali nucleari, o sulle emissioni inquinanti derivanti dall'utilizzo di combustibili fossili), sia un deciso stimolo a ricerche tecnologiche finalizzate a migliorare l'impatto ambientale delle diverse fonti di energia, sia una politica di imposte specifiche anche assai elevate (come quella sui consumi di benzina). Una politica decisa in questo senso può contribuire a ridurre l'elasticità rispetto al reddito dei consumi di energia, e in casi estremi a renderla negativa, permettendo una crescita del reddito accompagnata da una riduzione dei consumi di energia. Questa riduzione, tuttavia, potrà riguardare i paesi sviluppati, non quelli in via di sviluppo, molti dei quali presentano oggi consumi energetici pro capite bassissimi, destinati a crescere se appena - come tutti desiderano - il loro reddito pro capite tenderà a salire verso quello attuale dei paesi oggi industrializzati. La politica energetica dovrà perciò assicurare che la disponibilità di fonti di energia non costituisca un ostacolo per lo sviluppo economico; la compatibilità dello sviluppo con la difesa dell'ambiente dovrà essere assicurata, oltre che frenando i consumi di energia per unità di prodotto, anche favorendo le scelte più opportune tra le varie fonti di energia disponibili, e assicurando nell'utilizzo di ciascuna di esse

il rispetto dei vincoli ambientali.Accanto ai contrasti tra interessi privati e pubblici, un altro tipo di contrasti emerge sul piano internazionale: se in un certo paese le imprese vengono obbligate ad adottare costosi accorgimenti per ridurre o eliminare l'inquinamento, tali imprese possono trovarsi in condizioni svantaggiose, nella concorrenza internazionale, rispetto alle imprese di altri paesi in cui questi obblighi non siano stati introdotti. I contrasti di questo tipo possono essere ridotti attraverso accordi internazionali che stabiliscano obblighi comuni. Lo sviluppo sostenibile non può essere conseguito da un solo paese, ma solo da una cooperazione internazionale su vasta scala: lo stesso tipo di cooperazione che è necessaria per affrontare i problemi drammatici del sottosviluppo.

5. Considerazioni conclusive

Mentre nelle scienze chiamate sperimentali la costante preoccupazione degli studiosi è di verificare empiricamente i loro modelli teorici, in economia una simile preoccupazione è più l'eccezione che la regola. Ciò dipende solo limitatamente dal fatto che in questa disciplina, come nelle altre discipline sociali, non ci sono e non possono esserci laboratori: dipende soprattutto dalle caratteristiche assunte nel nostro tempo dalla teoria economica dominante, che ha privilegiato i ragionamenti assiomatici, nei quali ciò che conta è essenzialmente il rigore logico, mentre la rilevanza empirica conta poco o nulla. In effetti, nelle analisi economiche si assiste a una sorta di polarizzazione: da un lato troviamo modelli puramente astratti; dall'altro lato, indagini essenzialmente empiriche. Sono relativamente rari i lavori che mirano a combinare la riflessione teorica con l'analisi empirica: da un lato ci si preoccupa essenzialmente del rigore, dall'altro lato essenzialmente della rilevanza, mentre in qualsiasi disciplina scientifica entrambi i requisiti sono importanti.

Ciò non significa affatto sostenere che non siano apprezzabili e anzi raccomandabili i modelli astratti; né che non si debba far ricorso a metodi formali, particolarmente a metodi matematici; significa invece sostenere che quando si elaborano modelli astratti ci si deve domandare se potenzialmente siano suscettibili di successive approssimazioni che consentano di avvicinarsi progressivamente alla realtà economica e di interpretare i fenomeni concreti. Se gli economisti non dispongono di laboratori, dispongono tuttavia di una serie di strumenti analitici ausiliari, come quelli forniti dalla statistica e dall'econometria che, pur non consentendo controlli paragonabili agli esperimenti dei fisici e dei chimici, rendono possibili verifiche di carattere empirico. Si tratta di verifiche assai meno robuste di quelle compiute dagli scienziati sperimentali, che tuttavia, se usate con prudenza, possono suggerire ipotesi e problemi interessanti, stimolare dubbi e riflessioni critiche, e, più in generale, ridurre quelle incertezze e quella indeterminazione che sono connaturate a tutte le scienze, ma che sono particolarmente estese nelle discipline sociali.

In conclusione, le pecche più gravi della teoria economica moderna sono tre. In primo luogo, sono relativamente scarsi i lavori che combinano la riflessione teorica con l'analisi empirica. In secondo luogo, dominano ancora i modelli statici, che prescindono dal tempo e quindi ignorano in via di principio i più importanti fenomeni dell'epoca in cui viviamo, cioè il progresso tecnico e lo sviluppo economico. Infine, c'è una sorta di spaccatura fra microeconomia e macroeconomia, ossia, da un lato l'analisi dei prezzi e di tutti quei fenomeni che si collegano ai singoli soggetti, come le imprese e i consumatori, e dall'altro lato l'analisi dei grandi aggregati economici, come il reddito nazionale e l'occupazione. La divisione del lavoro, di cui Adam Smith parlava in senso concreto, è andata crescendo anche nelle diverse discipline, e quindi fra le discipline sociali e, in particolare, nell'economia. In questa esposizione abbiamo

cercato di fornire un ragguaglio estremamente conciso dello stato e delle tendenze osservabili nell'analisi economica; la menzione delle tendenze non poteva non comportare l'indicazione di alcuni problemi e dibattiti critici oggi in corso. In trattazioni specifiche di questa Enciclopedia vengono illustrate, non meno concisamente, le linee essenziali sia delle discipline ausiliarie, come l'econometria, sia dei diversi rami in cui, a questo stadio della sua evoluzione, l'economia si è suddivisa (economia agraria, industriale, internazionale, monetaria, pubblica). Analogamente, resta affidata a trattazioni specifiche l'illustrazione di alcune tendenze recenti che qui non è stato possibile considerare: come il neoistituzionalismo, che spiega le istituzioni economiche e sociali mediante modelli contrattualistici; o come l'utilizzo di modelli di disequilibrio economico per spiegare la disoccupazione; o come la diffusione della teoria dei giochi nelle analisi dell'equilibrio economico generale da un lato e nelle moderne teorie dell'organizzazione industriale dall'altro lato, per considerare la possibilità di 'ragionamenti strategici' dei soggetti economici, che nelle loro decisioni tengono conto delle possibili reazioni degli altri alle loro scelte.

Un ultimo aspetto al quale è necessario dedicare almeno un cenno è costituito dai rapporti fra conoscenza e azione, ossia fra clima culturale e modelli teorici da un lato, e strategie politiche generali e linee di politica economica dall'altro lato.

A titolo illustrativo conviene considerare tre fra gli economisti precedentemente ricordati, e cioè Adam Smith, Karl Marx e John Maynard Keynes. Con la sua grande opera, Smith ha certamente contribuito a determinare un mutamento radicale nella cultura politica del suo tempo e, ancora di più, del tempo successivo. Sul piano pratico, Smith raccomandava la progressiva eliminazione delle barriere e dei vincoli all'attività economica che provenivano dall'epoca feudale e dalle politiche mercantilistiche. Il liberismo di Smith va inteso

appunto in questo senso e non, come spesso si sostiene, nel senso di un atteggiamento passivo o inerte della pubblica autorità - governo e parlamento - nell'attività economica. (Così Smith era favorevole all'istruzione elementare pubblica, una posizione assolutamente minoritaria ai suoi tempi; era favorevole a interventi dello Stato per diverse opere pubbliche; raccomandava una riforma dei contratti agrari per favorire lo sviluppo agricolo: e questi sono solo tre esempi). L'analisi di Marx deve essere valutata con riferimento alle condizioni osservabili nel primo stadio del capitalismo, una fase in cui i salari erano ancora assai vicini al livello di sussistenza e molte donne e molti bambini erano costretti a svolgere lavori pesanti nelle fabbriche. È nota la grande influenza che le idee di Marx hanno esercitato su intellettuali - non solo economisti -, su partiti politici e su sindacati, soprattutto in Europa e in Asia: un'influenza che è stata enorme in certi paesi a regime dittatoriale, in cui la dottrina marxista era divenuta addirittura la dottrina ufficiale.

Il crollo dei regimi che si richiamavano al marxismo, verificatosi in diversi paesi a partire dal 1989, è almeno in parte legato al peggioramento delle condizioni economiche di tali paesi, imputabile principalmente alla loro incapacità d'introdurre innovazioni. Questa incapacità non si è manifestata in una prima fase, fino a quando è stato possibile concentrare lo sforzo economico sulle fondamentali infrastrutture, né in una seconda fase immediatamente successiva, quando è stato possibile acquistare dai paesi a economia di mercato impianti 'chiavi in mano' per l'industria di base; si è invece manifestata in modo drammatico quando dalle grandi unità industriali utilizzate per la produzione di beni di base qualitativamente omogenei è stato necessario passare alle medie e piccole unità produttive, che spesso impiegano tecnologie relativamente più sofisticate. Infatti, in un'economia pianificata, i dirigenti delle aziende monopolistiche di Stato potevano eseguire più o meno efficacemente gli ordini dell'ufficio centrale di pianificazione, ma non avevano alcun incentivo a

introdurre innovazioni: un'attività che necessariamente presuppone l'iniziativa individuale e la disponibilità a correre rischi. Ciò vale non solo per le grandi innovazioni, ma anche per le piccole, quasi sempre scientificamente irrilevanti, ma molto importanti per lo sviluppo economico. Innovazioni di questo genere sono particolarmente rilevanti in agricoltura, a condizione che i contadini abbiano la proprietà della terra, ovvero operino nell'ambito di contratti agrari che riconoscano i miglioramenti introdotti da chi coltiva la terra.
Oggi nei paesi dell'ex Unione Sovietica e dell'Europa orientale si discute sulla necessità di passare da un'economia pianificata a un'economia di mercato, e non di rado si ragiona come se l'economia di mercato fosse sinonimo di puro laissez faire; ma non è così. Il mercato è l'espressione di un sistema di contratti e, più in generale, di un complesso sistema istituzionale. Il passaggio da un'economia pianificata a un'economia di mercato implica il passaggio a un nuovo sistema istituzionale, in cui sia consentita la proprietà privata dei mezzi di produzione (pur con correttivi e limitazioni, come ad esempio una legislazione antimonopolistica), siano incoraggiate le innovazioni grandi e piccole, e in cui lo Stato abbia un suo ruolo, non onnicomprensivo, ma neppure insignificante. Una tesi di questo tipo, per quanto riguarda il ruolo dello Stato, è stata sostenuta da Keynes, la cui influenza è stata grande soprattutto nei paesi occidentali. La principale opera teorica di Keynes è nata quando, a causa della grande depressione (1929-1939), la disoccupazione aveva raggiunto proporzioni straordinariamente ampie. In generale, per combattere la disoccupazione e sostenere la crescita dell'economia, Keynes e i suoi discepoli raccomandano, oltre a politiche monetarie e fiscali espansive, anche il "controllo sociale degli investimenti". Gli interventi pubblici nell'area della sicurezza sociale hanno origini assai antiche, ma la teoria keynesiana ha dato alla loro crescita un nuovo vigoroso impulso, senza il quale il moderno Stato sociale, almeno in Europa, non avrebbe probabilmente assunto l'importanza che conosciamo. L'influenza

di Keynes, tuttavia, ha fortemente indebolito le resistenze presenti nella fase precedente, almeno nei periodi di pace, alla diffusione dei controlli pubblici sull'attività produttiva e alla crescita delle spese pubbliche.
In tempi recenti gli eccessi hanno dato luogo a reazioni; tali reazioni hanno trovato una giustificazione teorica in certi modelli, come il monetarismo di Friedman e il modello delle aspettative razionali.
Da queste teorie emergevano precettistiche di carattere ultraliberistico, che negli ultimi vent'anni hanno influenzato in modo significativo le politiche economiche di diversi governi. Le misure di riduzione dei controlli - interventi di privatizzazione delle imprese pubbliche e di deregolamentazione - in vari casi hanno avuto effetti tutto sommato positivi. La stessa cosa non si può dire però per gli effetti delle nuove politiche monetarie e fiscali; fra l'altro, le politiche fiscali hanno accentuato la diseguaglianza nella distribuzione dei redditi e, quanto alle spese pubbliche, hanno determinato un freno alla loro espansione, ma non una loro riduzione. Inoltre, le politiche monetarie restrittive adottate nei paesi più sviluppati hanno avuto un pesantissimo effetto negativo a livello internazionale, frenando la crescita dei paesi in via di sviluppo, particolarmente attraverso l'aumento dei tassi d'interesse e quindi dell'onere per i debiti internazionali.Conoscenza e azione sono i due termini riscontrabili in ogni ramo della cultura e non solo nelle scienze sociali. Le forme sono tuttavia diverse: nelle scienze sociali, e soprattutto in economia, l'aspetto dell'intervento pubblico assume un rilievo tutto particolare, sia al livello delle manovre contingenti di politica economica, sia al livello delle grandi strategie.

(V. anche Banca e sistema bancario; Benessere, Stato del; Cambio; Capitale; Capitalismo; Cicli economici; Credito; Debito pubblico; Diritto dell'economia; Diritto ed economia; Disoccupazione; Distribuzione della ricchezza e del reddito; Divisione del lavoro; Ecologia; Econometria; Economia e politica

agraria; Economia e politica del lavoro; Economia e politica industriale; Economia e società; Economia internazionale; Economia pubblica; Equilibrio economico; Finanza pubblica; Finanziari, mercati; Fisco e sistemi fiscali; Moneta; Politica economica e finanziaria; Sottosviluppo; Sviluppo economico).

Mercato
di Mario Deaglio

Sommario: 1. Definizione e tassonomia del mercato. a) Concorrenza perfetta e imperfetta. b) Monopolio, cartello, monopsonio, monopolio bilaterale. c) Alcuni casi particolari di mercato. 2. La complessità dei mercati reali. a) Soggetti, motivazioni, intervento pubblico. b) Estensioni recenti del mercato. c) Mercati irregolari e malavitosi. 3. Le interpretazioni del mercato. a) Il 'filone centrale' del pensiero economico. b) Marx e il filone marxista. c) La rivoluzione keynesiana e la reazione successiva. d) Schumpeter e Hayek: due interpretazioni eterodosse. e) Altre interpretazioni del mercato. 4. Lo sviluppo storico del mercato. a) Il problema dell'origine del mercato. b) L'evoluzione del mercato tra tecnologie e regole. c) Le tendenze attuali. 5. I limiti del mercato. □ Bibliografia.

1. Definizione e tassonomia del mercato

Dicesi mercato un insieme di scambi di natura economica, o comunque a essa riconducibile, aventi per oggetto un diritto reale (proprietà, uso, ecc.) su beni, materiali o immateriali, già esistenti o di futura realizzazione, oppure una prestazione (lavoro, ecc.). Tradizionalmente si dice mercato anche il luogo in cui avvengono gli scambi; con i moderni mezzi di comunicazione tale luogo può essere anche soltanto virtuale, essendo le transazioni possibili senza contiguità fisica tra le parti.
Per l'esistenza di un mercato sono necessarie le seguenti condizioni.
1. La libertà d'azione dei partecipanti, i quali possono essere soggetti individuali o collettivi. Tale libertà si concreta in un potere autonomo di scelta tra comportamenti diversi; essa è, peraltro, variamente limitata da restrizioni di tipo giuridico e dalle risorse di cui i partecipanti

hanno la disponibilità.
2. Un complesso di regole per lo svolgimento degli scambi, alle quali i partecipanti devono uniformarsi, e un'organizzazione che ne cura il rispetto.
3. Un prezzo, o un insieme di prezzi, relativo all'oggetto dello scambio, che si formi, almeno parzialmente, a seguito delle preferenze dei partecipanti e non già per semplice imposizione dal centro.
4. Un mezzo di pagamento accettato da tutti i partecipanti.
L'esistenza stessa del mercato amplia la gamma delle scelte economiche non solo di chi effettivamente vi partecipa ma anche di chi, pur decidendo di non farlo, può parteciparvi. Il mercato è quindi un bene collettivo e aumenta il benessere sociale, pur in presenza di numerosi limiti ed effetti negativi (v. cap. 5). Le economie in cui sono presenti tutte le condizioni sopra elencate, e in cui sono altresì libere e decentrate le scelte di produzione per la maggior parte dei beni e servizi, possono essere definite economie di mercato.
Lo studio delle caratteristiche dei mercati e delle loro conseguenze è argomento precipuo dell'economia politica, ma interessa anche le discipline giuridiche, la storia economica e la sociologia. Pur nella diversità delle interpretazioni (v. cap. 3), l'analisi economica ha sviluppato una tassonomia del mercato ormai sufficientemente consolidata, sia pure con qualche variazione terminologica e concettuale.Essa si basa sul carattere paradigmatico del mercato di concorrenza perfetta (v. § 1a), adottato come modello di riferimento e talora indicato come il mercato tout court, mentre le altre forme di mercato sono considerate come varianti o scostamenti da tale modello.

a) Concorrenza perfetta e imperfetta
Il mercato di concorrenza perfetta richiede condizioni molto restrittive, quali l'omogeneità dell'oggetto dello scambio, la pluralità dei venditori (normalmente identificati con i produttori del bene venduto) e dei compratori - gli uni e gli altri animati dalla volontà di rendere massimo il proprio vantaggio -, l'esiguità delle quote di mercato dei singoli, tali che le decisioni di ciascun operatore non possano influenzare il prezzo. Il prezzo si forma pertanto in maniera impersonale, per tentativi, oppure mediante l'opera di un agente esterno (il 'banditore'). Il mercato di concorrenza perfetta richiede inoltre l'istantaneità e la disponibilità generale delle informazioni relative ai prezzi e alle quantità scambiate e la possibilità dei partecipanti di entrare e uscire dal mercato stesso senza costi né barriere di

altro tipo (v. Walras, 1896[3]).
Tali condizioni possono sussistere anche solo allo stato potenziale: perché si realizzi un comportamento concorrenziale è sufficiente la presenza virtuale, non quella reale, di un gran numero di partecipanti, ossia l'istantanea possibilità di accesso al mercato di qualsiasi soggetto. Non è necessario che i mercati siano 'contesi', è sufficiente che essi siano 'contendibili' (contestable markets) (v. Baumol, Panzar e Willig, 1982).
In queste condizioni si determina un prezzo di equilibrio al quale tutte le quantità offerte vengono acquistate e si realizza un ottimo paretiano, ossia la condizione in cui è impossibile migliorare la posizione di un qualsiasi partecipante al mercato senza peggiorare la posizione di almeno un altro partecipante (v. Pareto, 1906). È questo il fondamento dell'utilità sociale del mercato.
La concorrenza perfetta è assai rara, per non dire inesistente, nella realtà. La telematica, tuttavia, rendendo possibile la sostanziale istantaneità e l'universale disponibilità a basso costo delle informazioni di mercato, ha contribuito, unitamente alla riduzione delle barriere legali all'entrata, a determinare condizioni non troppo distanti dalla concorrenza perfetta in alcuni mercati finanziari (v. § 1c).
La formazione impersonale del prezzo non conduce a un'effettiva competizione tra i partecipanti, e quindi all'interdipendenza delle loro decisioni, quanto piuttosto a una sequenza di eventi che tende a escludere dal mercato i produttori meno efficienti, per una sorta di selezione naturale (v. § 3c). I partecipanti non hanno alcun potere sul mercato; i profitti, al di là della remunerazione del capitale che equilibra i mercati finanziari, hanno carattere transitorio.
Potere e profitti compaiono invece quando si abbandonano alcune delle condizioni definitorie di cui sopra e si passa alle varie forme della concorrenza imperfetta (v. Robinson, 1933; v. Chamberlin, 1933). Se si segue questo percorso intellettuale, un primo caso di scostamento dalla concorrenza perfetta si ha con l'attenuazione della condizione di numerosità dei venditori, permanendo quelle della numerosità dei compratori e dell'omogeneità del bene scambiato. Si è allora in regime di oligopolio, un mercato in cui, dal lato dei venditori, il basso numero fa cadere la condizione di irrilevanza della quota di mercato del singolo, mentre, dal lato degli acquirenti, la persistente numerosità esclude qualsiasi influenza sul prezzo. I venditori sono quindi in effettiva competizione tra loro e le decisioni di uno di essi sulle quantità da vendere e sul prezzo da richiedere possono essere influenzate da, e a loro

volta influenzare, quelle degli altri venditori. È frequente nell'oligopolio la presenza di un'impresa dominante, generalmente di dimensioni maggiori, la quale fissa il prezzo (oppure la quantità che intende produrre) senza tener conto delle altre; queste ultime regolano le proprie decisioni sul comportamento della prima.
Ciascun produttore cerca allora di introdurre elementi migliorativi o distintivi nel suo prodotto, differenziando ciò che vende anche solo per piccoli particolari, come la marca, e trasferendo la competizione sul piano qualitativo. Si perde così il requisito della omogeneità del prodotto e il mercato che ne consegue è detto di concorrenza monopolistica. Nella concorrenza monopolistica ciascun venditore è monopolista (v. § 1b) del proprio prodotto, ma i prodotti sono tra loro sostituibili, sia pure imperfettamente, nella scelta degli acquirenti. Tra oligopolio e concorrenza monopolistica esiste un continuum logico che ne rende difficile l'esatta differenziazione. Qualora sussistano differenze secondarie tra i prodotti, entrambi gli schemi possono di fatto essere utilizzati per l'analisi di un particolare mercato. Nelle moderne economie di mercato la grande maggioranza dei beni di consumo e di investimento è scambiata in condizioni di oligopolio-concorrenza monopolistica, al punto che con il termine 'mercato' si intende spesso non già la concorrenza perfetta bensì l'insieme degli scambi in tali condizioni.
La vera differenziazione dei mercati non deriva tanto dalla numerosità dei venditori quanto dall'effettivo livello delle barriere all'ingresso (v. § 2a; v. Bain, 1956; v. Sylos Labini, 1956), che possono essere dovute a vincoli tecnologici, distanze geografiche, abitudini, norme legislative. La natura concorrenziale di un mercato non ha quindi carattere assoluto, nel senso che è presente oppure manca del tutto; esistono piuttosto diversi livelli di concorrenzialità, che, come si vedrà (v. cap. 4), caratterizzano i vari periodi storici.

b) Monopolio, cartello, monopsonio, monopolio bilaterale

Se si annulla, invece che attenuare, la condizione di numerosità dei produttori-venditori, si ha il monopolio. Nel monopolio esiste un solo venditore che fronteggia una pluralità di compratori. Il monopolista fissa il prezzo del proprio prodotto in base alle sue conoscenze delle caratteristiche della domanda aggregata degli acquirenti, la quale non deve essere direttamente influenzata dal prezzo di altri prodotti, in quanto, in tal caso, si avrebbe concorrenza monopolistica (v. § 1a). Si immagina quindi quale quantità sarà complessivamente acquistata a ciascuno dei possibili prezzi e tra questi fisserà quello che

meglio si confà ai suoi obiettivi.
A parità di struttura produttiva e di condizioni di domanda, il monopolista, se intende conseguire il massimo profitto, produce una quantità inferiore a quella che si realizzerebbe in un mercato di concorrenza perfetta e la vende a un prezzo superiore. Spesso può aumentare i propri profitti tenendo un comportamento discriminante, ossia praticando prezzi differenziati a diverse categorie di acquirenti oppure per diverse quantità vendute. Il profitto del monopolista è superiore a quello (corrispondente alla remunerazione di equilibrio del capitale) che i produttori complessivamente realizzano sul mercato di concorrenza perfetta e riduce quindi, in maniera corrispondente, il vantaggio (surplus) che gli acquirenti derivano dal mercato. L'aumento del profitto del venditore e la riduzione del surplus degli acquirenti danno una somma negativa che viene detta perdita netta di monopolio.Da quali fattori può trarre origine una situazione di monopolio, dal momento che essa comporta una perdita netta di tipo sociale? Si possono individuare quattro casi.
1. Un'innovazione tecnologica tale da conferire all'innovatore un'esclusiva, talvolta protetta da un'apposita legislazione (brevetto) che ne impedisca la diffusione. È il caso oggi più frequente e può produrre effetti positivi sulla crescita, in quanto alla perdita netta per la collettività, derivante dall'esistenza del monopolio stesso, fa da contropartita una migliore soddisfazione dei bisogni. Questo monopolio ha carattere temporaneo per l'invecchiamento fisiologico dell'innovazione o la scadenza del brevetto, e lascia in eredità un incremento del patrimonio tecnologico. Il venire meno delle barriere monopolistiche fa allora evolvere il mercato verso forme di concorrenza monopolistica-oligopolistica (v. § 1a).
2. Il 'monopolio naturale', che si verifica quando un determinato volume produttivo può essere raggiunto in maniera efficiente, con le tecnologie correnti, soltanto se esiste un unico produttore-venditore (come accade, per esempio, nel settore della produzione-distribuzione di elettricità o in quello dei trasporti ferroviari). In questo caso la contropartita del monopolio è la più favorevole struttura dei costi; il monopolista naturale è spesso un'azienda pubblica che non si propone l'obiettivo del massimo profitto, ma è anzi tenuta a praticare prezzi 'politici' con il solo vincolo di coprire i costi o addirittura di contenere le perdite entro un livello prefissato, finanziato mediante la tassazione.
3. Il monopolio di Stato, normalmente con finalità fiscali (vendita di sigarette e tabacchi). In questo caso il prezzo è, anche sensibilmente, superiore al costo.
4. Il monopolio sancito per legge a favore di privati con scopi di politica industriale (sviluppo

in esclusiva di determinate attività) o anche solo di potere (per esempio il monopolio del commercio con l'Oriente da parte delle Compagnie delle Indie inglese e olandese).
Una forma particolare di monopolio è il cartello, un accordo tra più produttori che stabiliscono di agire congiuntamente predeterminando il prezzo di vendita della merce e spartendosene, mediante l'attribuzione di quote di produzione, la quantità da produrre, in genere allo scopo di rendere massimi i profitti. Nel cartello il medesimo volume produttivo di un monopolista viene realizzato da un consorzio di produttori tra loro indipendenti, con la conseguente assenza dei benefici derivanti dalle economie di scala di un'analoga produzione monopolistica. Per la sua scarsissima utilità sociale, il cartello è, di regola, considerato negativamente dalle legislazioni della maggioranza dei paesi. Un esempio di cartello è rappresentato dall'OPEC, l'organizzazione dei paesi esportatori di petrolio, fondata nel 1960.
Speculare al monopolio è il monopsonio, un mercato caratterizzato dalla presenza di un solo acquirente, fronteggiato da una pluralità di venditori. Condizioni di monopsonio si verificano, non di rado, nei mercati dei fattori produttivi e, in particolare, nel mercato del lavoro, quando, per esempio, una sola grande impresa si confronta con lavoratori disorganizzati; quando però i lavoratori si organizzano e si danno una rappresentanza unitaria, si ha un caso di monopolio bilaterale, in cui un monopsonista (spesso un'organizzazione che raggruppa tutte le imprese di un settore o di un'area geografica) è fronteggiato da un monopolista (in questo caso un'organizzazione sindacale che ha il mandato di negoziare per tutti i lavoratori un contratto collettivo di lavoro).

c) Alcuni casi particolari di mercato
1. Il mercato del lavoro è normalmente caratterizzato da una situazione di quasi monopolio bilaterale, che vede contrapposti i rappresentanti dei lavoratori e quelli delle imprese, e da un considerevole intervento regolatore pubblico, che si concreta anche in una funzione istituzionale di mediazione e di indirizzo. Entrambe le parti dispongono di strumenti di pressione: i lavoratori possono far ricorso a varie forme di sciopero, le imprese alla serrata. Il risultato delle trattative, affrontate con questi strumenti, dipende dall'interazione tra le strategie adottate dalle parti e l'azione pubblica, ed è efficacemente descrivibile mediante la teoria dei giochi.
Negli anni novanta questa rappresentazione schematica subisce numerose modifiche. La maggiore libertà di iniziativa consente alle imprese di usare come strumento di pressione

anche la prospettiva di chiudere o trasferire gli impianti. D'altro canto il maggior contenuto di 'capitale umano' della prestazione lavorativa sta di fatto 'personalizzando' il lavoro; sulla contrattazione collettiva sopra delineata si innesta così, in misura molto maggiore che in passato, una contrattazione individuale o di piccoli gruppi, riguardante, oltre che la retribuzione, le modalità di esecuzione del lavoro (flessibile, a tempo parziale, ecc.), le prospettive future (opportunità di partecipare a corsi che comportino un arricchimento professionale, sviluppo della carriera), i fringe benefits, o retribuzione in natura (uso gratuito di un'autovettura, di un telefono portatile, ecc.).

2. Il mercato dei capitali finanziari, tradizionalmente incentrato nelle borse valori e nelle banche (in passato sovente di proprietà pubblica), ha subito modificazioni fondamentali nel corso dell'ultimo decennio. La funzione bancaria di intermediazione dei capitali è stata ridotta dalla possibilità dei grandi operatori di emettere direttamente titoli di debito (securitization) scambiati nei principali centri finanziari, in stretto collegamento telematico. Nel corso degli anni ottanta i movimenti di capitale a livello mondiale sono stati liberalizzati, si è ampliato il numero delle società quotate nelle borse, sono comparsi nuovi operatori quali i fondi comuni di investimento, i fondi pensione e i fondi chiusi. All'innovazione tecnologica si sono aggiunte importanti innovazioni normative, spesso coordinate a livello mondiale, tendenti a impedire l'uso di informazioni riservate (insider trading), a ridurre le occasioni di frode, a garantire in vario modo una maggiore trasparenza (offerte pubbliche di acquisto e di vendita).Sono stati introdotti numerosi nuovi strumenti finanziari (certificati di deposito, contratti 'pronti contro termine', contratti futures, ecc.) che consentono impieghi fortemente differenziati per quantità, rischio e durata. La maggior mobilità dei capitali ha grandemente ridotto il potere delle banche centrali di indirizzare e controllare i mercati.

3. Il mercato delle abitazioni è caratterizzato da un fattore produttivo non riproducibile (il terreno) con prezzo derivante non già dal costo di produzione bensì dalla scarsità, e fortemente influenzato dalle decisioni pubbliche riguardanti l'edificabilità, gli oneri di urbanizzazione, ecc. La specificità di ciascuna abitazione fa sì che sia carente il requisito dell'omogeneità del bene scambiato, e che le informazioni sui prezzi abbiano un minor grado di generalità. L'offerta è potenzialmente costituita da tutte le abitazioni esistenti e non solo da quelle nuove. Le variazioni del prezzo influenzano, pertanto, la ricchezza oltre che il reddito, interessando quindi tutti i proprietari e non solo i partecipanti al mercato.

A un'estrema libertà di prezzi sul mercato della proprietà ha fatto tradizionalmente riscontro,

in Europa, l'intervento pubblico nel mercato degli affitti, sia sul piano normativo (equo canone, blocco dei fitti, ecc.) sia mediante l'affitto a prezzi moderati di immobili di proprietà di enti pubblici. Ciò ha consentito di raggiungere fini sociali immediati, ma ha anche disincentivato la costruzione di case da parte dei privati a scopo di affitto, e reso molto realistico il rischio di ingerenze politiche nell'assegnazione delle case di proprietà pubblica. In tempi recenti si osservano la graduale liberalizzazione degli affitti, la vendita di parte degli alloggi pubblici agli inquilini, nuove regole sui mutui (aumento della quota anticipata dalle banche), che consentono una maggior diffusione della proprietà immobiliare. I nuovi fondi immobiliari consentono l'impiego in abitazioni anche di capitali finanziari molto piccoli, riducendo la barriera all'ingresso; funzione in qualche modo analoga ha, nel caso degli alloggi per le vacanze, la multiproprietà, una frammentazione temporale del diritto di proprietà.

2. La complessità dei mercati reali

a) Soggetti, motivazioni, intervento pubblico

Una moderna economia di mercato è caratterizzata da situazioni ben più complesse di quelle sopra schematicamente descritte. A tale maggiore complessità concorrono i seguenti fattori.

1. L'identificazione del produttore-venditore è oggi resa difficile dalla presenza, accanto all'impresa tradizionale, di 'gruppi', di joint ventures, di accordi produttivi o di vendita dei tipi più vari nonché di imprese di proprietà pubblica (come quelle a partecipazione statale in Italia).
2. Le motivazioni dell'impresa sono ben diverse dal semplice perseguimento del massimo profitto. Gli obiettivi risultano normalmente molto articolati e difficilmente riconducibili a comportamenti massimizzanti (v. Simon, 1992): possono, per esempio, comprendere il raggiungimento di un determinato livello di profitti e di una determinata quota di mercato, la loro stabilità o crescita nel tempo, la continuità del gruppo dirigente, la sconfitta di un concorrente monopolistico. Per imprese di proprietà pubblica la ricerca del profitto può essere temperata da considerazioni politico-sociali (come l'obbligo di investire prioritariamente in certi settori o aree geografiche). Sempre maggior peso (negli Stati Uniti il 7% dell'occupazione complessiva), infine, assumono i produttori privati senza fini di lucro (università, organizzazioni sanitarie, cooperative di lavoro, ecc.) che perseguono obiettivi di

natura non economica (v. Salamon e Anheier, 1993).
3. Sulla natura dei mercati incide fortemente la regolazione pubblica. Essa può, tra l'altro, imporre un determinato prezzo oppure fissarne i livelli massimi (tipici in Italia, fino a tempi recenti, quelli dei prodotti petroliferi, nonché l'equo canone per gli affitti) e minimi (salari contrattuali), oppure dettare regole che stabiliscano caratteristiche merceologiche dei beni o modalità per lo svolgimento dei servizi. Costituendo vere e proprie barriere all'ingresso, tali regole contribuiscono a determinare il grado di concorrenzialità del mercato. I pubblici poteri possono poi decidere imposte, sussidi diretti o indiretti, incentivi e disincentivi di natura fiscale o di altro tipo, nell'intento di influenzare le scelte degli operatori, orientandone gli investimenti verso regioni non sufficientemente sviluppate o verso attività particolari, come la ricerca scientifica. Possono infine agire in prima persona con varie finalità, costituendo aziende pubbliche.
La realtà odierna delle economie di mercato si configura quindi come un complicato intreccio delle forme teoriche originarie: una sorta di grande, sperimentale, non assestato 'brodo di coltura', che le discipline economico-aziendali, oltre a quelle economiche 'pure', cercano di esplorare e descrivere per poi poterne catalogare con sicurezza tutte le componenti.

b) Estensioni recenti del mercato
Nel corso degli anni ottanta e novanta si è verificata una forte estensione del mercato come strumento regolatore di attività in precedenza prevalentemente gratuite o soggette a regolamentazione amministrativa; questo passaggio è particolarmente rilevante nello sport, nella cultura, nella sanità. Il meccanismo del mercato comincia inoltre a essere applicato anche alla soluzione dei problemi ecologici.Nello sport si è fortemente esteso l'ambito delle prestazioni di eccellenza a pagamento rispetto a quelle a carattere gratuito e dilettantesco. Si sono formate vere e proprie 'imprese sportive', il cui scopo è il conseguimento di profitti mediante il coordinamento delle prestazioni dei singoli atleti e la 'vendita' di un gioco di squadra direttamente agli spettatori oppure a organizzazioni (sponsors) che ne fanno un uso promozionale o pubblicitario. La caratteristica principale di queste imprese è che sono prevalentemente dotate di capitale umano. Negli sport individuali (sci, tennis, golf) l'impresa sportiva è spesso costituita dallo sportivo stesso che amministra il proprio capitale con l'ausilio di professionisti stipendiati (agenti, allenatori, ecc.).

L'organizzazione dei mercati sportivi implica momenti formalizzati, ormai a livello internazionale, di scambio (per esempio, il 'mercato' dei giocatori di calcio) e di produzione (tornei e manifestazioni sportive) nonché un sistema di autoregolazione, affidato a organismi autonomi di categoria aventi il potere di comminare sanzioni.
Per quanto riguarda la cultura, il mercato dei diritti d'autore e delle opere dell'ingegno, già sviluppato a livello nazionale, si è esteso a livello internazionale con occasioni formalizzate, come le 'fiere del libro' e le grandi aste periodiche di opere d'arte gestite da organizzazioni apposite, che garantiscono l'autenticità del prodotto ed emettono veri e propri listini dei prezzi. Un altro sviluppo riguarda l'attività universitaria, per la quale alla regolamentazione amministrativa comincia a sostituirsi, a partire dal Nordamerica, un vero e proprio 'mercato degli accademici', con salari differenziati in base a giudizi di valore. Infine, l'organizzazione dei musei sta evolvendo in senso imprenditoriale, sia pure senza fini di lucro, con l'inserimento, su un'offerta di base tendenzialmente gratuita o a basso prezzo, di attività, come mostre e altre occasioni culturali, il cui costo è totalmente sopportato da visitatori e sponsors.
Nel settore sanitario, accanto agli ospedali pubblici, stanno comparendo, a partire dagli Stati Uniti, grandi organizzazioni ospedaliere aventi forma societaria e fini di lucro. In Europa l'evoluzione dei sistemi sanitari pubblici tende, a cominciare dalla Gran Bretagna, a fare degli ambulatori medici vere e proprie imprese, soggette alle pressioni competitive del mercato, sia pure all'interno del settore pubblico. In molti paesi esiste un mercato del sangue umano; in forme più o meno ufficiali e legali esistono anche mercati degli organi.
Negli Stati Uniti è in atto un tentativo per estendere il mercato ai problemi ecologici - un tempo ritenuti estranei al suo ambito oggettivo - per quanto riguarda l'inquinamento industriale dovuto a specifiche e misurabili emissioni di fumi e gas nocivi. Si obbligano i produttori a munirsi, acquistandoli sul mercato, di speciali 'diritti a inquinare', emessi dall'amministrazione pubblica. I danni all'ambiente entrano così a far parte dei costi di produzione. Tali diritti possono ugualmente essere acquistati da organizzazioni ecologiste che rendono così più caro l'inquinamento. Il prezzo di questi diritti fornisce una misura congiunta della necessità tecnologica e della sensibilità sociale all'inquinamento, non necessariamente dell'effettiva gravità del fenomeno.

c) Mercati irregolari e malavitosi

I mercati irregolari possono sorgere quando la regolamentazione del mercato non corrisponde ai desideri generali ma il costo politico della sua modifica è troppo elevato. Fu così per i 'mercatini' dei paesi dell'Est europeo nell'ultima fase del socialismo reale, tollerati dalle autorità ma non resi ufficiali nel timore di sollevare gravi conflitti interni e internazionali. I mercati irregolari possono altresì sorgere quando l'interesse a modificare le regole è proprio di una minoranza che non ha alcuna speranza di vederlo ufficialmente sanzionato, ma la natura dell'infrazione non suscita gravi reazioni repressive. In entrambi i casi alle regole 'ufficiali' del mercato si sostituiscono regole di fatto (per esempio comportamenti diffusi di evasione fiscale).
Il più tradizionale mercato irregolare è quello del cosiddetto 'lavoro nero', che vede un produttore monopsonista infrangere le regole ufficiali della contrattazione e confrontarsi con una pluralità di lavoratori non organizzati. L'esito è solitamente un salario inferiore a quello minimo dei mercati ufficiali, e spesso non vengono rispettate altre modalità della prestazione (norme di sicurezza). Esistono però altre forme di mercato irregolare del lavoro in cui un lavoratore autonomo specializzato (per esempio un falegname, un idraulico, ecc.) è in condizioni di monopolio sul mercato locale e vende la sua prestazione alle famiglie imponendo prezzo e modalità (per esempio l'evasione d'imposta).
Un mercato irregolare che infrange le leggi penali può essere definito mercato malavitoso. Perché si dia luogo a un tale mercato è indispensabile la libera decisione di acquirenti e venditori di parteciparvi, come avviene per i mercati della droga, i più importanti mercati malavitosi per volume d'affari.La loro articolazione è complessa e non interamente nota. Al livello della droga grezza ci si trova prevalentemente in presenza di un monopsonio, con una pluralità di produttori agricoli che vendono a un unico acquirente. Per il passaggio successivo, ossia la lavorazione di base e la vendita internazionale ai paesi consumatori, prevale l'organizzazione 'a cartello' dei venditori (v. § 1b). Nei paesi consumatori la forma più normale di vendita all'acquirente finale è costituita da una serie di monopoli locali in cui un unico venditore rifornisce un'area delimitata; i monopolisti mostrano spesso comportamenti collusivi, con prezzi non troppo differenti. In alcuni casi, però, soprattutto in presenza di tipi alternativi di droga, si può parlare di concorrenza monopolistica anomala, condotta sovente con metodi violenti.

3. Le interpretazioni del mercato

a) Il 'filone centrale' del pensiero economico

Il cosiddetto 'filone centrale' del pensiero economico (mainstream economics) tradizionalmente attribuisce la massima importanza al mercato quale meccanismo di regolazione sociale e non solamente economica. Al suo interno si sono succedute, nel corso di circa duecento anni, interpretazioni diverse, ma non contrastanti, del mercato, generalmente collegate ai paradigmi di volta in volta dominanti del pensiero scientifico e filosofico.

La prima trattazione sistematica del mercato, dovuta agli economisti classici, è direttamente tributaria dell'empirismo inglese e del giusnaturalismo francese, dai quali acquisisce i concetti di 'ordine naturale' e di 'stato di natura'. Questi presuppongono la naturale capacità del sistema economico di riprodursi e di restare in equilibrio (v. Quesnay, 1758), la libera estrinsecazione delle attività umane nell'ambito delle leggi (v. Locke, 1689), una naturale 'simpatia' tra i singoli (v. Smith, 1776).

Nella potente teorizzazione smithiana esiste, per ciascun prodotto, un 'prezzo naturale' che consente di remunerare in maniera 'normale' i fattori della produzione. Il prezzo effettivo di mercato tende a quello naturale, pur discostandosene per perturbazioni di ogni tipo. Essenziale non è quindi l'equilibrio, bensì la tendenza a raggiungerlo: condizioni generali di libertà, con l'assenza di barriere all'ingresso e all'uscita, la mobilità del capitale e (in misura minore per Smith) del lavoro, la disponibilità delle informazioni determinano spostamenti dei fattori produttivi da un mercato all'altro, operando nel senso di rendere uguali il saggio di profitto e il saggio salariale e di aumentare l'efficienza della produzione.Il mercato è perciò inteso soprattutto come un meccanismo attraverso il quale la società si autoregola, una 'mano invisibile' che realizza un vantaggio generale mediante il perseguimento dell'interesse individuale, peraltro espresso nell'ambito di norme morali condivise e di un apposito contesto giuridico-istituzionale.

La concezione smithiana, che fa da sfondo anche al sistema economico di Ricardo, fu variamente sviluppata nel corso della prima metà dell'Ottocento, tra l'altro da Say (v., 1803), il quale, partendo dall'osservazione che il valore della produzione complessiva è uguale a quello dei redditi distribuiti, concluse che la produzione crea potere d'acquisto ed enunciò la sua celebre e controversa 'legge dei mercati', secondo cui l'offerta crea la propria domanda.

Nella seconda metà dell'Ottocento l'influsso del positivismo induce gli economisti

neoclassici ad accentuare, accanto all'importanza della libertà, quella della razionalità individuale. L'influenza della fisica li porta ad affinare il concetto di equilibrio con l'uso di strumenti logici e matematici.
Il pensiero neoclassico tende così a sviluppare leggi 'scientifiche' dell'economia per le quali le contingenze storiche hanno scarsa rilevanza. Il modello del mercato viene formalizzato mediante l'introduzione delle curve collettive di domanda e di offerta, mentre viene abbandonato il concetto di prezzo naturale. Il prezzo di mercato e la quantità scambiata sul mercato diventano così le risultanti matematiche dell'incontro di queste curve.
Il prezzo che così si realizza pone i mercati in condizioni di equilibrio, nel senso che nulla resta invenduto, mentre ad altri prezzi si manifestano sovrabbondanze e scarsità. Si dimostra in tal modo che l'equilibrio, nella formulazione neoclassica, è unico e stabile e che il mercato naturalmente evolve verso tale equilibrio, dal quale non si allontana senza modificazioni esterne.
Secondo i neoclassici, il mercato di concorrenza perfetta è in grado di assicurare l'allocazione efficiente delle risorse e l'assenza di disoccupazione involontaria, sostanzialmente senza interventi dall'esterno. Questa teoria lascia poco spazio, forse per reazione alla loro importanza nel pensiero di Marx (v. § 3b), agli effetti sociali del mercato, che trovano una trattazione estesa soltanto nell'opera di Marshall. Il liberismo economico dei neoclassici diverge gradualmente dal liberalismo, che vede nei mercati null'altro che un mezzo per giungere a più generali forme di libertà (v. Croce, 1928).

b) Marx e il filone marxista

La concezione del mercato di Marx si caratterizza per il ruolo dell'analisi del potere all'interno del mercato. Marx, il cui pensiero ha radici classiche, attribuisce un contenuto sociologico ai fattori della produzione identificandoli con le classi sociali che li detengono (il capitale con la borghesia e il lavoro salariato con il proletariato).
Per Marx (v., 1867-1894) tutto il potere è concentrato nel capitale: con il mercato è il capitale, e non il lavoro, a essere posto in condizioni di libertà. Il mercato, sovrastruttura funzionale all'economia capitalista e sostanzialmente esistente solo dentro di essa, è il luogo dello scontro tra le due classi sociali - la borghesia e il proletariato -, e le ricorrenti crisi cicliche di tale economia sfoceranno nella crisi finale del capitalismo.
Marx attribuisce altresì un valore negativo al concetto di 'merce', in quanto bene scambiato

sul mercato. La merce, infatti, configura un particolare rapporto tra l'uomo e le cose, derivante dal suo essere scambiata al suo valore-lavoro. Tale valore sintetizza determinati rapporti di produzione, l'esistenza del profitto, il dominio dell'uomo sull'uomo.Il problema del calcolo economico nel socialismo, ossia dei valori da attribuire a beni e fattori produttivi in assenza di mercato, si era già posto in particolare con Enrico Barone (v., 1908). Dopo la Rivoluzione russa del 1917, la realizzazione di una società senza mercato divenne un problema pratico da risolversi mediante prezzi amministrati, imposti da un ente centrale di governo dell'economia. Come fissarli, però, senza rinunciare a quell'efficienza allocativa che è una conseguenza naturale del mercato?
Una soluzione di questo problema è stata fornita da Lange e Lerner. L'ente centrale di programmazione svolge le funzioni istituzionali del mercato; fissa, cioè, prezzi, salari e tassi di interesse in modo da uguagliare domanda e offerta; simula i processi di mercato decentrando varie decisioni, ma riservandone alcune (come il tasso di accumulazione) al pianificatore centrale (v. Lange, 1936).
In direzione opposta si colloca la soluzione implicita nel cosiddetto 'esperimento iugoslavo', basato sul concetto di autogestione delle unità produttive da parte dei lavoratori; esso riserva al pianificatore centrale, almeno in via di principio, soltanto compiti di coordinamento, informazione e concertazione volontaria.

c) La rivoluzione keynesiana e la reazione successiva
La grande crisi del capitalismo degli anni trenta costituisce il clima culturale entro cui si sviluppa la concezione del mercato di J.M. Keynes. Per Keynes, il mercato può non raggiungere naturalmente le condizioni di piena occupazione, previste dalla 'legge di Say' e formalizzate dagli economisti neoclassici, perché la domanda effettiva può rivelarsi inferiore all'offerta complessiva, determinando situazioni di crisi e di disoccupazione involontaria. La domanda deve, in tal caso, essere stimolata con interventi pubblici espansivi diretti (programmi di spesa pubblica) o indiretti (stimoli fiscali alla spesa di individui e imprese).
Su queste linee la scuola keynesiana sviluppò l'idea di un intervento attivo dei pubblici poteri, teorizzando la compatibilità del meccanismo del mercato con la fissazione di obiettivi nazionali e settoriali di sviluppo nell'ambito di una programmazione 'flessibile' diretta dal centro (v. Shonfield, 1965) e la necessità, in questo contesto, di una concertazione tra le parti sociali ('politica dei redditi') promossa dal governo per predeterminare il tasso dei salari e dei

profitti.
Questa concezione del mercato, che oggi appare riduttiva, fu a suo tempo intesa come l'ancora di salvezza di un capitalismo minacciato, oltre che dalla crisi interna, dallo scontro con il socialismo reale. Negli anni cinquanta Keynes fu attaccato da sinistra come il rappresentante del neocapitalismo che cercava di salvare il sistema di mercato; due decenni più tardi venne attaccato come esponente di un pensiero sostanzialmente contrario al mercato.
L'azione di governo dei mercati teorizzata dal keynesismo comportava interventi restrittivi oltre che espansivi della domanda globale. Dopo una lunga fase espansiva in cui le ricette keynesiane si rivelarono efficaci, i livelli di spesa si consolidarono e si rivelò sempre più difficile ridimensionare la domanda anche per il peso relativamente scarso attribuito alla politica monetaria. Si aprì così la via all'inflazione degli anni settanta.
Queste difficoltà contribuirono all'affermarsi di una concezione del mercato che si richiama al pensiero neoclassico e pone soprattutto l'accento sui caratteri di razionalità del comportamento individuale. Secondo tale concezione, neppure le aspettative dei partecipanti al mercato sono plasmate dall'esperienza storica, ma sono aspettative razionali; essendo razionali, gli individui effettuano da sé e per sé quell'attività di programmazione che il governo non appare in grado di svolgere con successo. L'azione di politica economica viene giudicata inutile, d'importanza secondaria o addirittura dannosa eccetto che per il controllo della crescita della massa monetaria (da effettuarsi, in ogni caso, sulla base di parametri oggettivi). Da concetto tributario di discipline filosofiche e scientifiche, il mercato diventa, con i nuovi economisti classici, esso stesso un paradigma forte: una società di uomini liberi può essere definita come un insieme di mercati, purché si intenda come mercato, in senso più ampio di quello tradizionale nell'analisi economica, qualsiasi tipo di scambi volontari, anche di natura non pecuniaria. L'allargamento del concetto di mercato a (quasi) tutti i fenomeni economico-sociali fa sì che il mercato stesso non rivesta soltanto la funzione di distributore ottimale di risorse economiche, ma tenda altresì a diventare il principale, se non l'unico, meccanismo di regolazione dei rapporti e di coordinamento tra gli individui (v. Bosanquet, 1983). I meccanismi selettivi della concorrenza (v. § 1a) vengono interpretati talora in senso darwiniano (v. Nelson e Winter, 1982), giustificando in tal modo il comportamento 'spietato' e comunque non collaborativo verso i più deboli.
Un esempio importante di estensione del concetto di mercato è dato dalla sua applicazione

alla famiglia, giustificata e spiegata in chiave razionale quale risultato della mutua convenienza dei suoi membri (v. Schultz, 1979; v. Becker, 1981). In quest'impostazione allo Stato e a tutte le istituzioni collettive non può non essere attribuita un'origine contrattualistica. Il processo politico di una democrazia viene interpretato come 'mercato politico' (v. Buchanan e Tollison, 1972). L'organo politico rappresentativo viene inteso come produttore di decisioni che gli elettori 'acquistano' mediante il voto, scegliendo tra le offerte dei candidati quella meno lontana dai loro desideri, su un mercato con caratteri concorrenziali.

d) Schumpeter e Hayek: due interpretazioni eterodosse

Due interpretazioni originali del mercato, dense di sviluppi e di conseguenze, sono dovute a due economisti austriaci quasi contemporanei: Schumpeter (nato nel 1883) e Hayek (nato nel 1899).

Per Schumpeter l'imprenditore non 'partecipa' ai mercati, li 'crea', in quanto vi si presenta con prodotti nuovi, in grado di generare profitti monopolistici transitori. Il mercato è il luogo in cui la validità di tale 'creazione' viene verificata attraverso la 'distruzione creatrice' dei prodotti esistenti. Artefici di questa distruzione creatrice sono, per Schumpeter, soprattutto le grandi imprese, tendenti al monopolio, al quale, peraltro, giungono solo per periodi limitati di tempo, data la transitorietà del loro vantaggio creativo. Il profitto monopolistico associato a questa fase è un compenso per l'attività innovativa (v. Schumpeter, 1942).

Si realizza così una concorrenza giocata a colpi di innovazioni, portatrice non già di equilibrio, bensì di squilibri, e come tale generatrice di progresso. Le innovazioni si verificano generalmente a grappoli o 'sciami' e sono quindi concentrate nel tempo; esse danno in tal modo origine a 'onde lunghe', di durata pluridecennale, con alternanza di diverse velocità di espansione, che caratterizzano l'esperienza storica delle economie di mercato.

Hayek è un esponente di primo piano del liberismo, ma la sua idea del mercato si differenzia nettamente da quella del 'filone centrale' dell'economia. Per Hayek il sistema di mercato deve intendersi essenzialmente come un meccanismo di trasmissione diretta di conoscenze attraverso i segnali rappresentati dai prezzi. Mentre però la concorrenza perfetta ipotizza conoscenza perfetta, nel modello di mercato hayekiano (che egli chiama 'catallassi' appositamente per distinguerlo dalla concorrenza) le conoscenze sono necessariamente parziali e imperfette e solo attraverso il mercato si coordinano. Mentre la concorrenza

perfetta conduce a un equilibrio stazionario, l'equilibrio della catallassi hayekiana ha carattere evolutivo e vi è in esso lo spazio per l'innovazione imprenditoriale (v. Hayek, 1973-1979).
Per Hayek la società progredisce facendo uso di conoscenze disperse, che non possono essere riunite e sintetizzate da un pianificatore centrale o anche solo da un organismo di coordinamento del mercato. L'ordine di mercato non deve essere quindi rozzamente inteso come un mezzo che consente di giungere a un fine predeterminato, bensì semplicemente come un processo di scambio che consente agli individui di agire in maniera sufficientemente prevedibile in un mondo mutevole e incerto.

e) Altre interpretazioni del mercato
In quanto elemento di importanza crescente nell'ordine sociale, il mercato ha attirato l'attenzione delle Chiese, che, in genere, lo giudicano favorevolmente, pur con limitazioni varie. Ciò è evidente nella Bibbia e nel Corano, il quale ammette esplicitamente il profitto e implicitamente assegna un ruolo allo Stato nel caso di monopoli naturali. Tuttavia il buon imprenditore musulmano deve pagare prezzi 'equi' e retribuzioni 'ragionevoli'; sussiste la proibizione dell'interesse, che viene però superata da un'associazione in partecipazione del creditore all'attività del debitore.Il pensiero protestante, in particolare, considera il mercato luogo di estrinsecazione della libertà individuale. Autorevoli interpretazioni (Weber, Sombart, Tawney) vedono un collegamento diretto tra l'etica protestante e la diffusione del capitalismo.
Per la Chiesa cattolica il limite del mercato è l'"originale destinazione comune dei beni creati". Nella Centesimus annus di Giovanni Paolo II si opera una distinzione tra bisogni 'solvibili' e risorse 'vendibili', da un lato, per i quali il libero mercato è riconosciuto come strumento efficace, e gli altri bisogni. Prima ancora dello scambio "qualcosa è dovuto all'uomo perché uomo" (Centesimus annus, 34). Un ulteriore limite del mercato è individuato nella necessità di difendere i beni collettivi (Centesimus annus, 40).

4. Lo sviluppo storico del mercato

a) Il problema dell'origine del mercato
È questione ampiamente dibattuta tra gli storici e gli antropologi se il mercato sia una

componente essenziale delle comunità umane oppure se sorga solo in tempi relativamente recenti. Su questa seconda posizione si colloca un filone di pensiero che ha le sue radici nella concezione marxiana del mercato (v. § 3b) e in Polanyi un autorevole esponente. Polanyi (v., 1944) sostiene che l'economia di mercato sarebbe stata pressoché inesistente prima del IV secolo a.C. e fa risalire alla rivoluzione industriale il momento della 'grande trasformazione', quando essa divenne il paradigma dominante della società occidentale. Prima di allora non il profitto bensì il dovere, l'onore e il prestigio avrebbero dominato i rapporti economici. I prezzi sarebbero, sì, esistiti, ma sarebbero stati determinati dalla tradizione, scarsamente e lentamente sensibili al mutare della domanda e dell'offerta. Da posizioni ideologicamente molto distanti anche Hayek (v. § 3d) afferma l'esistenza di uno stadio 'tribale' dello sviluppo umano, anteriore al mercato (v. Hayek, 1973-1979).
Recenti scoperte archeologiche hanno rivelato l'esistenza, tra la Mesopotamia e la Palestina, nel II millennio a.C., di una sofisticata rete commerciale, e rendono quindi difficile sostenere in pieno questa tesi; appare però evidente che rapporti 'tradizionali' nel senso di Polanyi e di Hayek dovessero essere ampiamente presenti in società in cui il costo delle comunicazioni era altissimo e quindi il commercio era comunque limitato a beni di elevato valore.
Fu solo con l'Impero romano che si svilupparono mercati e commerci che, per certi versi, possono essere definiti 'di massa', come dimostra il vettovagliamento di grandi concentrazioni urbane, a cominciare da Roma stessa, con grano proveniente dalla Sicilia e dall'Africa. La ricerca storica non ha chiarito appieno il contesto istituzionale e creditizio del mercato nell'Impero romano, ma è ben documentata la tendenza secolare all'aumento dell'ingerenza pubblica che giunge fino all'imposizione estesa di prezzi amministrati, limitando comunque il grado di concorrenzialità. Il frantumarsi dell'unità geoeconomica del Mediterraneo, con l'invasione araba e lo spostamento a nord del centro politico dell'Europa, provocò una sensibile riduzione degli scambi - e quindi dell'importanza dei mercati - in tutta l'Europa occidentale. Nell'economia curtense dell'alto Medioevo il feudo tende all'autarchia e il mercato ha carattere occasionale o sporadico con una limitatissima circolazione monetaria.

b) L'evoluzione del mercato tra tecnologie e regole

Dopo il Mille, il mercato, pur riguardando una piccola parte della produzione, prevalentemente incentrata sulle attività tessili e su alcuni beni di lusso, rinacque con forme giuridiche specifiche (nelle fiere veniva sospesa l'efficacia delle leggi normali) e specifici

strumenti e istituzioni finanziarie (le 'lettere di cambio' e le banche). A causa dell'elevatissima incidenza del costo dei trasporti, l'organizzazione dei commerci necessitava di una concentrazione di capitali spesso più rilevante di quella necessaria per la produzione, realizzata con tecnologie preindustriali. Si formò così una classe 'capitalista' di natura mercantile e di respiro internazionale operante in condizioni di tipo concorrenziale; parallelamente, per i prodotti di uso locale, il sistema delle corporazioni diede origine a un mercato strettamente regolato con prezzi rigidi ed elevatissime barriere all'entrata.
Tra il Quattrocento e il Settecento, mentre permaneva la rigida regolazione dei mercati locali, specialmente urbani, il 'grande commercio' si organizzò a livello mondiale; si affermarono lentamente e in maniera spesso contrastata i mercati di alcuni prodotti agricoli (il tè, le spezie, il cotone e simili), con le grandi compagnie commerciali, l'attività bancaria e le borse (v. Wallerstein, 1974-1980; v. Braudel, 1979). Contemporaneamente, con le leggi inglesi sulle enclosures e il graduale passaggio, nell'Europa occidentale, dai beni feudali ai beni allodiali, si affermò il diritto di proprietà privata sulla terra e con esso la possibilità di alienarla (v. Deane, 1965). Nacque un mercato dei fondi agricoli, premessa a una riorganizzazione produttiva dell'agricoltura, resa possibile da nuove tecniche di coltivazione e allevamento, che la orientò maggiormente allo scambio. Il lavoro salariato con retribuzione monetaria sostituì gradualmente il lavoro servile e i compensi in natura.
Lo sviluppo del mercato appare quindi fortemente influenzato da due tipi di fattori, entrambi necessari: le tecnologie e le regole. Dalle prime dipendono la realizzabilità dei prodotti, le condizioni di efficienza produttiva, il livello di concentrazione necessaria e le barriere tecniche all'ingresso sui mercati. Le seconde, essenzialmente stabilite da leggi, determinano invece altri tipi di barriere, relative ai soggetti che possono partecipare ai mercati e alle loro possibilità operative; determinano altresì il ruolo del settore pubblico in quanto controllore, stimolatore e penalizzatore dell'attività economica - tramite politiche di incentivazione e di disincentivazione - e anche produttore.
Questa chiave interpretativa consente di dar ragione dell'estendersi e del rafforzarsi del mercato a partire dalla seconda metà del Settecento con la rapida diffusione delle tecnologie industriali e il parallelo sviluppo istituzionale dei mercati (v. North, 1990). Appare peraltro difficile stabilire un prius storico e logico tra tecnologie e regole, elementi tra i quali si stabilisce una dialettica complessa. Secondo le concezioni liberiste (v. §§ 3a e 3d) tale dialettica è conoscibile solo sperimentalmente attraverso i segnali dei prezzi, mentre secondo

le concezioni socialiste essa è governabile da un'autorità pianificatrice centrale (v. § 3b).
È in ogni caso possibile stabilire che dalle tecnologie industriali vennero due spinte opposte, con fasi alterne di prevalenza. Le tecnologie, infatti, richiesero da un lato crescenti concentrazioni di capitale, con tendenza alla riduzione del numero dei produttori-venditori; parallelamente, però, il diffondersi delle ferrovie, delle navi a vapore, del telegrafo ridusse il costo dei trasporti, attenuò le barriere tecniche all'ingresso, allargò l'ambito geografico dei mercati, favorendo l'aumento del numero dei produttori-venditori. L'allargamento fu facilitato da potenti ondate di innovazione tecnologica, che crearono mercati nuovi, nei quali a condizioni temporanee di monopolio seguirono rapidamente condizioni di oligopolio e concorrenza monopolistica.
Si può ugualmente stabilire che lo sviluppo delle regole procedette nel senso dell'estensione e della specificazione dei diritti di proprietà, soprattutto sui prodotti dell'ingegno, della regolamentazione del mercato del lavoro, con contrattazione collettiva in regime di quasi monopolio bilaterale, e di quello del capitale (v. § 1c). Si restrinsero, con la società per azioni, i limiti della responsabilità personale del proprietario, fino a fare del capitale un'entità anonima, separata dalle persone dei possessori e pertanto aggregabile in concentrazioni superiori alle disponibilità dei singoli. Queste spinte regolatrici contribuirono anch'esse al sorgere di un proletariato industriale e di una borghesia detentrice del capitale produttivo.
Nelle ricorrenti crisi cicliche prevalse la tendenza a politiche industriali interventiste, di tipo protezionista, con elevate tariffe doganali, sussidi, salvataggi ed estensione dell'area produttiva pubblica. Nei periodi di espansione si osserva invece la tendenza all'abbattimento delle barriere, all'allargamento del mercato, al contenimento dell'intervento pubblico.L'ampliamento del mercato ha dato origine, nel corso degli anni cinquanta e sessanta di questo secolo, ad alcune centinaia di grandi gruppi industriali e finanziari - le cosiddette 'società multinazionali' - con attività in ambito mondiale, generalmente non limitate a un solo settore merceologico, ma con controllo fortemente accentrato nei singoli paesi d'origine. Queste società hanno operato sovente in condizioni di accentuato oligopolio e con un forte potere di orientamento del mercato.

c) Le tendenze attuali

Si può datare dalla crisi petrolifera (1973) l'inizio di un periodo di profonda trasformazione del sistema di mercato. L'emergenza energetica, infatti, stimolò la sostituzione con tecnologie

elettroniche delle tecnologie meccaniche, più dispendiose in termini di energia. Le tecnologie elettroniche, a loro volta, determinarono la riduzione sia delle barriere tecniche di ingresso sia - in molti ma non in tutti i settori - del capitale necessario per una produzione efficiente. Esse favorirono così il ridimensionamento delle unità produttive, la loro localizzazione diffusa, il sorgere di imprese di dimensioni medie o medio-piccole con interessi estesi ben al di là del loro tradizionale ambito, locale o nazionale.
L'effetto congiunto di queste due tendenze ha scardinato, in molti settori, gli antichi oligopoli (ne è un esempio eloquente lo stesso settore dei computer), obbligando sovente le società multinazionali a ridurre le proprie dimensioni relative e assolute e a concentrarsi in uno specifico settore di attività, con l'attenuazione o la perdita della leadership dei mercati.
Contemporaneamente nuove regole commerciali creano aree di libero scambio sempre più estese. Tale tendenza, iniziata con la costituzione della Comunità Europea (1957) e con gli accordi del GATT (1948 e 1961), è culminata, attraverso un iter tortuoso, nella costituzione, all'inizio del 1995, della World Trade Organisation (WTO), che ha gettato le basi per la libera circolazione non solo delle merci ma anche dei servizi, dei brevetti, del lavoro e del capitale.
Alle regole di origine pubblica si aggiungono quelle, spesso elaborate da organismi autonomi, che sovrintendono ai singoli mercati. In ambedue i casi vi è la tendenza a intervenire ampiamente in difesa degli acquirenti sui mercati a concorrenza monopolistica (controllo dei messaggi pubblicitari, norme a garanzia della qualità) contro i comportamenti collusivi tra produttori e le conseguenze negative indirette derivanti dalla produzione (diseconomie esterne: v. cap. 5). Privatizzazioni e riduzione dei poteri pubblici di intervento e controllo (deregulation) attenuano la presenza, diretta e indiretta, dello Stato nei mercati.
Quest'evoluzione ha fatto sì che, verso la metà degli anni novanta, i mercati siano divenuti maggiormente 'contendibili' (v. § 1a) e che il livello di concorrenzialità delle economie di mercato sia oggi molto elevato.

5. I limiti del mercato

Non vi è mercato se i soggetti non hanno potere di scelta (v. cap. 1); l'assenza di tale potere si verifica sempre in un sistema totalmente pianificato dal centro, mentre nelle economie di mercato una simile situazione, che comporta prezzi amministrati e razionamento, si verifica in maniera generalizzata solo in caso di guerra, di calamità naturali e di altre gravi

emergenze.
Ugualmente non è possibile il mercato in assenza di regole che lo governino, oppure quando non esistono mezzi di pagamento adeguati. Tutto ciò esclude dal mercato gli scambi di servizi personali che avvengono all'interno delle famiglie: nulla può 'pagare' l'assistenza gratuita di una madre a un figlio infermo.
La natura del bene o della prestazione può essere tale da escludere che sia oggetto di scambio, e quindi di mercato, anche al di fuori dell'ambito familiare. È il caso dei beni collettivi, per l'impossibilità di formulare una domanda autonoma (se non eventualmente, nella formulazione iperliberista, mediante il mercato politico; v. § 3c): nessun privato può 'comprare' la manutenzione stradale, l'illuminazione pubblica, la sentenza di un giudice (se ciò formalmente avviene, l'atto cambia natura) o il servizio della polizia (se ciò avviene, anche la prestazione della polizia cambia natura e il poliziotto si trasforma in guardia del corpo). Vi è quindi un fallimento del mercato, mentre è invece possibile che i servizi pubblici, forniti in risposta a una domanda non di mercato, vengano prodotti a seguito di aste competitive o che diverse unità produttive dell'amministrazione pubblica competano tra loro per soddisfare la domanda, ottenendo, in caso di successo, non già maggiori profitti ma dotazioni migliori, ecc. È questo il cosiddetto mercato interno all'amministrazione pubblica, che si ha, per esempio, quando gli utenti possono spendere presso diversi fornitori pubblici speciali 'buoni' (vouchers) loro assegnati che danno diritto alla prestazione gratuita.
Vi è inoltre una forte difficoltà a far rientrare in maniera efficiente nell'ambito degli scambi di mercato i beni non riproducibili oppure riproducibili in quantità limitata e in tempi lunghi, specie quando, attraverso il loro sfruttamento, subiscono trasformazioni, spesso irreversibili, con forti svantaggi generali (diseconomie esterne). La scarsità sociale sposta il confine del mercato annettendo una parte della sfera del legittimo interesse individuale a quella delle obbligazioni sociali. Il trascurare questa sfera corrode la base stessa delle economie di mercato (v. Hirsch, 1976). Per allocare questi beni è possibile procedere con un razionamento casuale (per esempio nel caso del parcheggio dell'auto in una zona congestionata) o con un razionamento consapevole (effettuato in base a criteri esterni al mercato, ad esempio riservando un certo numero di posti nei parcheggi alle auto dei medici, della polizia, ecc.). La realtà mostra spesso la coesistenza pragmatica di questi tre criteri (una parte dei posti-auto assegnata a categorie speciali; una parte affidata al razionamento casuale; una parte lasciata al mercato, spesso grazie all'opera di un posteggiatore abusivo).

Un altro limite è dato dalla spinta all'uniformità. Secondo una massima attribuita a Ludwig von Mises, il mercato è "produzione di massa per le masse". In quanto meccanismo di regolazione sociale, esso tende a ottenere risultati migliori là dove maggiori sono le uniformità dei beni oggetto del mercato e là dove la domanda è meno differenziata. Il mercato esercita, perciò, una forte spinta all'uniformità dell'oggetto, solo in parte contrastata dalle onde lunghe dell'innovazione.
Per l'azione delle economie di scala, i gusti delle minoranze troveranno la loro soddisfazione a un prezzo più elevato di quelli delle maggioranze; pur avendo la tecnologia elettronica spesso contribuito ad abbassare la soglia di produzione minima per poter restare sul mercato, una domanda al di sotto di tale soglia non trova soddisfazione sul mercato. Se un quotidiano, per poter sopravvivere, deve vendere ogni giorno almeno centomila copie, qualsiasi gruppo politico, religioso o simile che non riesca ad assicurare il volume di vendita minimo deve rinunciare a vedere le proprie convinzioni espresse da un quotidiano in grado di stare sul mercato.Il mercato tende inoltre a rappresentare una soluzione 'totalizzante', nel senso che la presenza di mercati a monte è spesso una condizione indispensabile per il funzionamento ottimale di un determinato mercato. Esso si propone quindi come soluzione globale dei problemi della società e tende a espandersi all'intera società. Ciò può causare l'eliminazione di tutto quanto non è in grado di stare sul mercato. Si rischia così l'eliminazione del 'diverso'. Questo significa che il mercato può ampliare enormemente la gamma delle scelte in alcuni settori, fino al punto di rendere impossibile una decisione razionale, ma ridurla contemporaneamente in altri, eliminando le opzioni più 'radicali' o minoritarie.
Per conservare il 'diverso' appare quindi necessario un deliberato intervento pubblico - di tipo non assistenziale ma promozionale - che, al limite, costituisce la vera salvaguardia della molteplicità e della varietà. Per il funzionamento accettabile di un'economia di mercato è indispensabile, anche al di là della fornitura dei servizi pubblici, la presenza di attività non di mercato, finanziate volontariamente dalla collettività oppure mediante la tassazione.

Moneta

di Paolo Sylos Labini

sommario: 1. L'evoluzione storica: brevi cenni. a) Vari tipi di moneta. b) I biglietti e i depositi. c) L'oro e la convertibilità dei biglietti. d) Sistema aureo e sistema bimetallico. e) L'inconvertibilità: da eccezione a regola. 2. La moneta e le relazioni economiche internazionali. a) La moneta e i cambi. b) La manovra dello sconto in regime aureo. c) Cambi fissi e cambi flessibili. d) La manovra dello sconto in regime di inconvertibilità e altre manovre della banca centrale. 3. Moneta e prezzi. a) La teoria quantitativa della moneta. b) Le tendenze dei prezzi nel XIX e nel XX secolo. c) I diversi ruoli dei diversi mezzi monetari. 4. Il processo inflazionistico. a) I molteplici impulsi. b) I differenziali di inflazione. c) L'inflazione nei paesi del Terzo Mondo. 5. Moneta e sviluppo economico. a) La creazione della moneta bancaria. b) Profitto, interesse e sviluppo. c) Le variazioni di breve periodo del saggio dell'interesse. d) La moneta e il ciclo economico. 6. Una riflessione conclusiva. □ Bibliografia.

1. L'evoluzione storica: brevi cenni

Vari tipi di monetaMan mano che le società progrediscono, superando lo stadio selvaggio e primitivo, cresce l'esigenza di andare oltre la produzione per il consumo diretto e oltre il baratto, ossia lo scambio di un bene contro un altro bene: occorre che emerga un bene generalmente accettato come intermediario negli scambi e come misura dei rapporti di scambio - dei valori - fra i diversi beni. Quel bene può avere una sua specifica utilità ovvero può essere accettato come intermediario negli scambi per convenzione, anche senza un'utilità intrinseca, proprio perché risponde all'esigenza dell'intermediazione. Tale convenzione, tuttavia, non può non essere fondata sulla ragionevole aspettativa che quel bene non sia disponibile in quantità illimitata o non sia riproducibile a volontà, giacché solo in queste condizioni si può confidare che i suoi rapporti di scambio con gli altri beni restino tollerabilmente stabili nel corso del tempo. Tale relativa stabilità rende idoneo quel bene non solo a svolgere la funzione di intermediario negli scambi, ma anche a fungere da serbatoio di valore, dove 'valore' è da intendersi come la capacità del bene considerato di acquistare altri beni. I due ruoli - intermediario negli scambi e serbatoio di valore - sono quelli tradizionalmente attribuiti al bene di cui si discute e che viene definito come moneta, un bene che nei tempi più vicini a noi è stato man mano sostituito da segni cartacei. A questo punto cade l'utilità intrinseca della moneta e resta solo l'elemento convenzionale.

Coloro che hanno studiato l'evoluzione delle società umane in diverse parti del mondo nei tempi antichi hanno individuato i beni più diversi prescelti come moneta: conchiglie (in zone lontane dal mare), bestiame, pelli, e infine metalli. Per un lungo perio do storico le società che progredivano a un ritmo relativamente più rapido hanno adottato come moneta determinati metalli. E qui compare una bipartizione: da un lato troviamo i metalli nobili, non facilmente riproducibili, quasi inattaccabili da agenti esterni, come l'oro e l'argento; dall'altro troviamo altri metalli e leghe, come il rame e il bronzo. I metalli nobili costituiscono la moneta fondamentale, usata negli scambi di rilevante entità; gli altri metalli vengono usati nei piccoli pagamenti e costituiscono la moneta detta divisionaria o sussidiaria. Sia i metalli nobili sia gli altri vengono suddivisi in piccoli pezzi - in dischi - coniati dall'autorità centrale della società per dare la garanzia del peso e del contenuto.

Durante gli ultimi secoli in Europa, nelle società più evolute, accanto alla moneta costituita da metalli nobili (e oltre la moneta divisionaria) compare e man mano si diffonde la moneta cartacea; per un certo periodo i due tipi di moneta coesistono, la moneta cartacea essendo subordinata a quella metallica. A un certo punto la moneta vera e propria scompare, mentre sopravvive la moneta divisionaria, quella usata per gli scambi di piccola entità. I banchi, che sono i progenitori delle banche moderne e si trovano già nell'antica Grecia e poi, più tardi, nell'antica Roma, compiono principalmente il servizio di custodire le monete depositate da singoli individui, che in quel tempo sono monete d'oro, d'argento e di rame; servono anche a cambiare le monete dei diversi metalli; possono inoltre far prestiti, con monete proprie o, raramente, depositate. Nel tardo Medioevo in certe zone dell'Europa, e in particolare nelle parti centrali e settentrionali della nostra penisola, i commercianti che svolgono traffici a lunga distanza cominciano a usare lettere di cambio - quelle che poi diventano cambiali - con cui un soggetto s'impegna a pagare in moneta sonante un altro soggetto che opera in un luogo diverso: si usano lettere di cambio ovvero promesse di pagamento in luogo di monete coniate d'oro o d'argento per evitare i rischi, allora molto elevati, connessi col trasferimento di queste monete. In origine le lettere di cambio erano titoli di credito; cominciarono ad acquisire una funzione di tipo monetario quando, sul finire del Medioevo, si introdusse la 'girata', che poteva essere ripetuta più volte, fino a quando chi riceveva la lettera di cambio non decideva di riscuotere in moneta sonante la somma promessa. Le lettere di cambio potevano essere

accettate in luogo delle monete vere e proprie dalle banche, che davano al commerciante che gliele consegnava moneta sonante, trattenendosi per il servizio una percentuale - un interesse anticipato, uno 'sconto'. Contro le lettere di cambio, che erano promesse di pagamento, ossia erano titoli di credito ma non mezzi monetari, le banche, quando raggiungevano una notorietà sufficientemente ampia, potevano offrire le proprie promesse di pagamento, variamente configurate, che erano largamente accettate come moneta, mentre questo non succedeva con le promesse di pagamento dei commercianti. Si aveva allora uno scambio fra titoli di credito di tipo diverso: fra cambiali, rilasciate da commercianti ad altri commercianti, e promesse di pagamento delle banche, che, dato il credito di cui queste godevano, circolavano come moneta in ampie cerchie di persone. L'operazione di sconto - scambio fra cambiali e mezzi di pagamento bancari - risultò via via conveniente non solo fra soggetti operanti in luoghi diversi, ma anche fra soggetti operanti nello stesso luogo, che tuttavia avevano rapporti che implicavano tempi diversi: le operazioni di credito, che rientrano nel genere più ampio delle operazioni finanziarie, hanno significato proprio se si considerano gli intervalli temporali - finanza viene da finis che significa 'termine'. Nel Medioevo i mezzi di pagamento bancari non sono propriamente moneta, ma sono usati in sostituzione della moneta e cioè svolgono funzioni monetarie.

I biglietti e i depositiCol tempo emerge via via una bipartizione fra gli stessi mezzi di pagamento bancari: da un lato quelli che fanno capo alle banche che prevalgono sulle altre poiché hanno raggiunto notorietà e credito in ampie zone e presso un gran numero di soggetti; dall'altro i mezzi di pagamento che fanno capo alle banche minori. La bipartizione dà luogo da un lato ai biglietti, che fanno capo alle banche dette poi 'di emissione', e dall'altro ai depositi, che fanno capo alle banche dette poi 'ordinarie' o 'di deposito e sconto'. Le banche che via via superano le altre per importanza e dimensioni possono ricevere un forte impulso per un ulteriore sviluppo se l'autorità politica attribuisce ai loro titoli il potere liberatorio per il pagamento dei tributi o di particolari obblighi contrattuali; in contropartita di tale privilegio l'autorità politica si fa dare, come anticipazioni, somme di biglietti: sono prestiti concessi a condizioni di favore e non di rado rinnovati più volte o addirittura mai rimborsati. In certi casi le banche di emissione sono create direttamente dall'autorità politica, come banche pubbliche, senza la precedente evoluzione di cui si è appena detto. Anche nei paesi in cui operano diverse banche autorizzate a emettere biglietti si afferma man mano la tendenza

verso l'unificazione, cosicché alla fine emerge una sola banca di emissione. Oggi questa è la regola ed è per tale motivo che si parla al singolare di banca di emissione o di banca centrale. Già prima i biglietti erano generalmente accettati come moneta perché un gran numero di soggetti aveva la certezza che in qualsiasi momento essi potevano essere scambiati in moneta sonante secondo un rapporto fisso - dischi coniati d'oro o d'argento. L'espressione 'pagabili a vista al portatore', che ancora si legge sui biglietti della banca centrale, aveva un tempo un significato sostanziale, mentre oggi è un'espressione priva di significato concreto giacché i biglietti non sono più convertibili in monete d'oro o d'argento. Ormai alla moneta di base, che non ha più un valore intrinseco ed è costituita dai biglietti della banca centrale, resta solo l'elemento convenzionale, emerso dall'esigenza sociale di un intermediario negli scambi e suffragato da regole stabilite dall'autorità politica.

Quanto agli assegni, che fanno capo alle banche ordinarie e sono usati come mezzi monetari sul fondamento dei depositi, vale per essi una relazione simile a quella che sussisteva fra biglietti e monete auree: come i biglietti circolavano sulla fiducia di poterli convertire in qualsiasi momento in monete metalliche, così gli assegni circolano sulla fiducia di poterli convertire in qualsiasi momento in biglietti. Le due convertibilità possono essere assicurate se da un lato la banca di emissione e dall'altro le banche ordinarie dispongono di riserve adeguate: rispettivamente, di monete auree o argentee e di biglietti. Tuttavia, per mantenere la convertibilità non era necessario che le riserve di monete metalliche rappresentassero il 100% dei biglietti e quelle di biglietti il 100% dei depositi. Non era necessario, poiché, se c'erano soggetti che facevano valere il loro diritto alla conversione, c'erano altri soggetti che portavano alla banca i biglietti o gli assegni senza pretendere la conversione in quanto dovevano restituire i prestiti ricevuti. Del resto, di norma bastava la fiducia che la conversione era possibile senza difficoltà per far sì che essa non venisse richiesta. Man mano che cresceva il numero dei soggetti che accettavano normalmente in pagamento i biglietti e gli assegni (nella prima fase dello sviluppo delle banche moderne tale numero era assai ristretto), diminuivano le richieste di conversione e poteva quindi diminuire la quota delle riserve. (Al principio del Novecento la quota delle riserve auree delle banche di emissione si aggirava sul 10%, oggi la riserva in biglietti delle banche ordinarie si aggira sul 20-25%).

L'oro e la convertibilità dei bigliettiDi norma, dunque, una quota modesta di riserve poteva

essere sufficiente ad assicurare la duplice convertibilità. Ai circuiti del credito corrispondevano i circuiti della produzione: quando questi si concludevano positivamente i prestiti potevano essere restituiti; in tali condizioni la fiducia nei mezzi monetari bancari non veniva in alcun modo incrinata. Ciò valeva anche quando alcune imprese fallivano e non restituivano i prestiti, se la quota dei fallimenti sul totale dei prestiti era modesta; quando diventava elevata, il circuito del credito si rompeva, i mezzi monetari restavano in circolazione come peso morto e la fiducia nella stabilità del loro potere d'acquisto si incrinava. A questo punto poteva esplodere il panico, che provocava una corsa alla conversione e rendeva inadeguate le riserve. Se ciò accadeva per le banche ordinarie queste fallivano, a meno che l'autorità pubblica o la banca centrale non intervenissero per salvarle; se il panico investiva i biglietti, la banca di emissione era costretta a sospendere la convertibilità in monete auree. Una tale catena di eventi poteva essere avviata da una serie di fallimenti di imprese e, di riflesso, di banche; più spesso, però, la convertibilità doveva essere sospesa perché l'autorità politica imponeva - di solito per esigenze connesse a una guerra - la stampa e la cessione di cospicue e crescenti somme di biglietti che non avevano una contropartita nell'attività produttiva: quella militare è, per definizione, un'attività distruttiva. La convertibilità poteva poi essere ripristinata bloccando per un certo periodo l'emissione dei biglietti: quando la ripresa della produzione arrivava al punto di controbilanciare la massa dei biglietti, questi recuperavano l'antico potere d'acquisto e scompariva il divario fra valore della moneta metallica e valore dei biglietti - un divario chiamato 'aggio'.

Sistema aureo e sistema bimetallicoIn certi periodi e in certi paesi è stato usato, come moneta, un solo metallo, l'oro o l'argento. In altri periodi e in altri paesi sono stati usati simultaneamente due metalli, l'oro e l'argento. In tempi antichi si trattava solo di dischi coniati; in tempi relativamente recenti nei paesi che adottavano il bimetallismo i biglietti della banca di emissione erano convertibili in moneta d'oro o d'argento secondo un rapporto fisso (spesso 1 a 15). Dal bimetallismo ci si attendeva una stabilità dei prezzi maggiore di quella ottenibile col monometallismo: era questo il vantaggio sperato che poteva indurre a quella opzione; per i paesi economicamente deboli che avevano rapporti molto stretti con un paese forte che aveva adottato il bimetallismo, un altro vantaggio era costituito da una maggiore facilità nei rapporti economici col paese più forte. In effetti, fino a quando non avevano luogo scoperte rilevanti di nuove miniere, d'oro o d'argento, il vantaggio riguardante

i prezzi era reale. Naturalmente, in economie in crescita, di tanto in tanto non poteva non emergere una scarsezza relativa dell'uno o dell'altro metallo; ma grazie alla convertibilità in entrambi i metalli le variazioni dei prezzi risultavano più attenuate di quanto sarebbero state col monometallismo. Oltre certi limiti, però, in seguito a scoperte rilevanti di nuove miniere, il sistema bimetallico entrava in crisi e diventava, almeno di fatto, monometallismo. In effetti s'imponeva la cosiddetta 'legge di Gresham', secondo cui la moneta cattiva caccia la buona: quando vi sono più mezzi monetari con cui si possono fare pagamenti, la gente preferisce pagare col mezzo di minor valore intrinseco rispetto all'altro, che viene tesoreggiato, o impiegato in usi non monetari o nei pagamenti esteri, quando i soggetti di altri paesi chiedono di essere pagati con la moneta 'buona'. Giova ricordare che, storicamente, la Francia istituì un sistema bimetallico nel 1803; quel sistema fu poi adottato nel 1865 dall'Unione monetaria latina, di cui faceva parte anche l'Italia, ma fu travolto nel 1878, come conseguenza delle grandi scoperte di miniere d'argento nel continente americano.

L'inconvertibilità: da eccezione a regolaIn diversi paesi evoluti nel XIX secolo e fino alla prima guerra mondiale la convertibilità dei biglietti era la regola e l'inconvertibilità l'eccezione - si parlava di 'corso forzoso' e questa espressione indicava una situazione di emergenza. Oggi non si parla più di corso forzoso e l'inconvertibilità è diventata la regola che non soffre eccezioni. Il sistema aureo è stato sostanzialmente abbandonato nel periodo fra le due guerre.Quando la moneta era puramente metallica l'autorità politica interferiva assai poco sulla circolazione monetaria: garantiva il peso e la qualità dei dischi metallici per mezzo del conio, ma talvolta alterava il peso o la qualità dei dischi coniati e imponeva alla gente di accettare in pagamento le monete come se l'alterazione non fosse avvenuta. La conseguenza era una perdita del potere d'acquisto delle monete, ossia un aumento dei prezzi. In ogni modo si trattava di periodi relativamente brevi, anche se non infrequenti. Tutto sommato, quando circolava la moneta metallica l'autorità politica non era in grado di commettere abusi troppo gravi. Le cose cambiano quando si diffonde la moneta cartacea. Già cambiano quando vige la convertibilità dei biglietti in moneta metallica, e cambiano perché l'autorità politica può imporre alla banca di emissione di 'anticipare' certe somme di biglietti a vantaggio dell'erario; quando tali anticipazioni diventano molto rilevanti, la proclamazione del corso forzoso diviene inevitabile. Le cose cambiano poi radicalmente quando il corso forzoso diviene la norma: in queste condizioni gli abusi possono diventare gravi e frequenti. Contro

questi abusi i paesi evoluti hanno cercato di difendersi rendendo pubblica, quando non lo era già, la banca di emissione e introducendo leggi che stabiliscono regole rigorose circa l'emissione dei biglietti, anche se poi, in caso di necessità vera o presunta, il governo trova il modo di non rispettarle. Su queste regole ci sono stati e ci sono dibattiti fra economisti e politici. Con l'inconvertibilità dei biglietti divenuta norma permanente, il cambiamento nel sistema monetario è talmente profondo da indurre un economista di spicco come Alberto Breglia a parlare di una vera e propria nazionalizzazione della moneta.Come si è accennato, la fine del sistema aureo ha avuto luogo nel periodo compreso fra le due guerre mondiali, dopo la grande crisi cominciata nel 1929, ma il sistema aveva dato segni di cattivo funzionamento da quando, a partire dal principio del nostro secolo, prezzi e salari erano divenuti sempre più rigidi verso il basso: la flessibilità verso il basso dei prezzi, come si cercherà di chiarire, era uno dei requisiti essenziali per il buon funzionamento del sistema aureo.

2. La moneta e le relazioni economiche internazionali

La moneta e i cambiNel tempo in cui le monete erano esclusivamente o prevalentemente metalliche le stesse monete che circolavano all'interno di un dato paese erano usate per i pagamenti a soggetti di altri paesi. Considerati i diversi comportamenti dei diversi Stati - alcuni intervenivano piuttosto spesso per alterare a proprio vantaggio il peso e la qualità dei coni, altri raramente -, nel corso del tempo erano le monete coniate dagli Stati più corretti a prevalere. Così, il tallero d'argento di Maria Teresa è stato usato a lungo nei paesi più diversi e tuttora è usato, sia pure marginalmente, in certi paesi arabi e africani. Questo stato di cose, tuttavia, poteva permanere fino a quando gli scambi fra paesi diversi restavano quasi stazionari ovvero crescevano lentamente. Quando la crescita è divenuta vigorosa, è aumentato parallelamente il bisogno di mezzi di pagamento. Per questo motivo e per ridurre i rischi inerenti al trasferimento di monete auree o argentee, si è sviluppato l'uso, nei pagamenti internazionali, di pagare con lettere di cambio stilate nella moneta cui appartiene il debitore e cambiate poi in moneta nazionale nel paese dove risiede il creditore. Compare, così, un vero e proprio mercato delle lettere di cambio, o cambiali estere, o divise. Quando vigeva la convertibilità dei biglietti in oro, le oscillazioni dei prezzi dei cambi - o, brevemente, dei cambi - incontravano due limiti molto ristretti, uno verso l'alto, l'altro verso il

basso, rispetto al livello normale, o parità, dei cambi, determinato dal peso in oro delle unità monetarie dei due paesi considerati. I limiti di oscillazione si aggiravano sull'1-2% della parità e dipendevano dalle spese di trasporto e di assicurazione delle monete auree; solo quando il prezzo delle divise superava l'uno o l'altro limite - i 'punti dell'oro' - le monete auree venivano trasferite (esportate o importate). In regime aureo dunque il livello dei cambi, intesi come prezzi delle divise, era dato dal rapporto fra i pesi in oro con due ristretti margini di oscillazione.

La manovra dello sconto in regime aureoIn qualsiasi regime monetario i prezzi delle divise - i cambi - dipendono dalla domanda e dall'offerta delle divise, la domanda essendo determinata dal complesso delle voci passive, l'offerta dal complesso delle voci attive della bilancia dei pagamenti (compravendita di merci e di servizi, trasferimenti di capitali), cosicché un aumento dei cambi - considero solo questo caso - indica un eccesso della domanda sull'offerta, ossia un deficit nei conti con l'estero. Se il deficit è consistente, in regime aureo le monete d'oro vengono esportate e diminuiscono le riserve della banca centrale, la quale, per difenderle, eleva lo sconto. Tale aumento attira capitali dall'estero, giacché gli investimenti finanziari a scadenza relativamente breve nel paese che si considera diventano più vantaggiosi. Già questo è un contributo al riequilibrio della bilancia dei pagamenti; è tuttavia un contributo temporaneo, poiché i prestiti esteri, che immediatamente rientrano fra le voci attive, danno poi luogo a voci passive, per il pagamento degli interessi e per la restituzione. L'aumento dello sconto della banca centrale - definito anche 'saggio ufficiale' - ha anche un altro effetto: costringe le banche ordinarie a elevare il loro sconto, poiché per rifornirsi di biglietti esse debbono ricorrere alla banca centrale, riscontando le cambiali ricevute dalle imprese. Le imprese chiedono allora minori prestiti e riducono sia gli investimenti sia la domanda di servizi dei lavoratori; pertanto diminuisce, o cessa di crescere, la massa dei mezzi monetari bancari e diminuisce anche la domanda di beni. Nel secolo scorso e fino alla prima guerra mondiale, sia i salari che i prezzi erano flessibili verso il basso e la flessione dei prezzi, determinata dalla diminuzione della domanda, stimolava la crescita delle esportazioni dal paese considerato, il che tendeva a riequilibrare in modo più stabile i conti con l'estero. La diminuzione della domanda, d'altra parte, riduceva il livello o il saggio di aumento delle importazioni e anche ciò contribuiva a eliminare il deficit estero, che era all'origine del deflusso d'oro.

Cambi fissi e cambi flessibiliSi è fatto riferimento al tempo in cui nei paesi più evoluti vigeva il sistema aureo, giacché in quegli anni si è via via articolata la manovra dello sconto attuata dalla banca centrale. Allora la manovra dello sconto serviva soprattutto a difendere la riserva aurea, che aveva una funzione propriamente monetaria. Anche oggi, nel tempo dell'inconvertibilità, la banca centrale detiene una riserva d'oro, come anche di divise estere; ma oggi l'oro non ha più funzioni propriamente monetarie: se mai, quando si profila un deficit consistente nei conti con l'estero, l'oro può essere ceduto ai paesi creditori, che non hanno difficoltà ad accettarlo, per quegli stessi requisiti che per tanto tempo hanno indotto i paesi più diversi a sceglierlo come moneta unica o, nel periodo della convertibilità, come moneta di base. L'oro aveva allora una funzione propriamente monetaria, poiché i soggetti erano certi che l'unità monetaria fondamentale aveva un determinato peso e un determinato titolo; oggi non è così e il prezzo dell'oro oscilla come quello di qualsiasi merce. Mette il conto di soffermarsi sulla manovra dello sconto, giacché quella manovra viene attuata ancora oggi anche se per fini alquanto diversi; certi fini tuttavia coincidono e il confronto fra il modo in cui operava quella manovra nel passato e il modo in cui opera oggi è un confronto istruttivo. Oggi non si tratta più di difendere le riserve auree in quanto riserve monetarie, ma si tratta pur sempre di difendere le riserve d'oro e di divise, e la manovra dello sconto può essere il mezzo adatto per gli effetti che tende ad avere sui movimenti di capitali e sui cambi.

Riguardo ai cambi, attualmente bisogna distinguere il regime dei cambi fissi e quello dei cambi flessibili. Il regime dei cambi fissi rassomiglia al regime della convertibilità: anche allora i cambi erano fissi o, più propriamente, quasi fissi, e il livello normale era costituito dal rapporto fra i pesi in oro; anche allora vi erano, per effetto di un meccanismo spontaneo, margini ristretti di oscillazione - i 'punti dell'oro'. In regime di inconvertibilità sia il livello normale sia i margini di oscillazione dipendono non da meccanismi spontanei, ma da accordi internazionali: i margini, che imitano i punti dell'oro, servono al buon funzionamento del sistema. Con i cambi flessibili, invece, non ci sono limiti né verso il basso né verso l'alto. Le oscillazioni dipendono essenzialmente dall'andamento dei conti con l'estero e dalla stabilità politica - l'instabilità tende a generare sfiducia nei mercati finanziari - ed entro certi limiti possono essere controllate dai responsabili della politica economica e, in particolare, della politica monetaria.

A questo punto è bene ricordare che le relazioni fra cambi e prezzi sono bidirezionali: l'aumento dei cambi tende a far crescere i prezzi interni attraverso l'aumento di quelli dei beni importati; d'altra parte l'aumento dei prezzi interni imprime una spinta verso l'alto ai cambi, poiché scoraggia le esportazioni e stimola le importazioni e quindi fa diminuire le voci attive e fa crescere le voci passive dei conti con l'estero.

I vantaggi dei cambi fissi consistono, in primo luogo, nel fatto che quando i salari aumentano la concorrenza internazionale costringe le imprese a concentrare i loro sforzi sull'aumento della produttività poiché non possono elevare i prezzi. In secondo luogo, i cambi fissi riducono al minimo i rischi d'inflazione importata. In terzo luogo, essi riducono i rischi di movimenti speculativi di capitali, che possono provocare dannose alterazioni nei movimenti di merci e nei flussi finanziari. Con i cambi flessibili si hanno svantaggi speculari: debole incentivo ad accrescere la produttività, rischio di inflazione importata, movimenti dei capitali provocati da operazioni speculative sui cambi. Con i cambi flessibili si hanno, però, adattamenti vantaggiosi alle condizioni che cambiano e l'incentivo a espandere le esportazioni anche quando i prezzi interni aumentano, dato che la speculazione sui cambi opera con maggior forza della speculazione sui beni; pertanto l'aumento dei cambi tende ad anticipare quello dei prezzi, che a loro volta ricevono, con ritardo, dai cambi una spinta verso l'alto.

Tutto considerato, i vantaggi dei cambi fissi (o quasi fissi) superano quelli dei cambi flessibili: è questo il motivo per cui i paesi più evoluti nell'ultimo mezzo secolo hanno aderito ad accordi che tendevano a ricostituire un sistema di cambi quasi fissi, come gli accordi internazionali di Bretton Woods del 1944 e gli accordi del 1979 fra un certo numero di paesi europei. Gli accordi di Bretton Woods comportavano un sia pur parziale legame con l'oro attraverso il dollaro, che a certe condizioni poteva essere convertito in oro, ma solo su richiesta dei governi; gli accordi del 1979 che dettero vita al Sistema Monetario Europeo (SME) non comportavano, invece, alcun legame prestabilito col dollaro o con l'oro, ma impegnavano i governi a mantenere i cambi a determinati livelli. Il sistema di Bretton Woods entrò in crisi nel 1971, quando gli Stati Uniti abbandonarono ogni legame fisso con l'oro. Dal Sistema Monetario Europeo sono usciti, in anni recenti, alcuni paesi fra cui Inghilterra e

Italia; gli altri sono rimasti, ma con margini di oscillazione più ampi. Ciò, tuttavia, è avvenuto per cause che si possono definire di forza maggiore, tanto che tutti i paesi europei hanno manifestato, con gli accordi di Maastricht del 1991, l'intenzione di arrivare a un'unica banca centrale europea e a una moneta unica, si potrebbe dire a un sistema a cambi rigidamente fissi, con margini di oscillazione nulli; per entrare nel sistema previsto da quegli accordi, tuttavia, i paesi interessati dovranno soddisfare certe condizioni, riguardanti principalmente l'ammontare relativo del debito pubblico e il ritmo della pressione inflazionistica.

La manovra dello sconto in regime di inconvertibilità e altre manovre della banca centraleIl rialzo dello sconto, come si è ricordato, può sostenere il cambio della moneta nazionale attirando capitali esteri e frenando il deflusso di capitali interni; e ciò vale sia in regime di convertibilità sia in regime di biglietti inconvertibili. Quel rialzo può essere anche attuato, nell'uno e nell'altro regime, per fini principalmente interni, per esempio per frenare un aumento dei prezzi o per provocarne una flessione. Questo secondo obiettivo, tuttavia, poteva essere perseguito con successo in passato, quando sia i prezzi sia i salari erano flessibili verso il basso. Oggi, come meglio vedremo, solo i prezzi dei prodotti grezzi, agricoli e minerari, dipendono, nel breve periodo, dalle variazioni della domanda e dell'offerta: i prezzi dei prodotti finiti, che in un'economia moderna sono i più importanti, variano soprattutto per effetto delle variazioni dei costi, in primo luogo dei salari, che tuttavia vanno messi in rapporto alla produttività. La restrizione del credito e della moneta ha effetti sulla domanda di tutti i beni, a cominciare dai beni d'investimento, ma non ha effetti diretti di rilievo sui prezzi dei prodotti finiti. Può avere un effetto indiretto: frenando gli investimenti il rialzo dello sconto tende a far crescere la disoccupazione e in questo modo indebolisce il potere contrattuale dei lavoratori e frena l'aumento dei salari, cosa che, a sua volta, frena un aumento dei prezzi. L'aumento dello sconto può essere usato come monito: ai sindacati, affinché moderino le loro richieste di aumenti salariali; agli industriali, affinché non usino subito e pienamente il loro potere di mercato sui prezzi. Può essere usato come monito anche nei riguardi del governo, qualora ci sia un consistente deficit di bilancio: l'aumento dello sconto rende più pesante l'onere per interessi e può indurre il governo a contenerlo.

Alternativamente, o congiuntamente, la banca centrale dispone di altri strumenti di

intervento, oltre lo sconto; anzi, coi cambi fissi è meno acuto l'assillo dei conti con l'estero e dei movimenti speculativi di capitali, e lo sconto viene usato più raramente. Le altre manovre - mi limito alle principali - sono le seguenti: 1) la compravendita di titoli pubblici; 2) la manovra sulle riserve delle banche ordinarie; 3) indicazioni non coercitive, che fanno appello alla persuasione (moral suasion), rivolte alle banche ordinarie e agli intermediari finanziari. Le diverse manovre - che vanno poste in relazione ai diversi obiettivi della banca centrale: difesa delle riserve, stabilità dei cambi, stabilità dei prezzi, stimolo agli investimenti e allo sviluppo - sono tutte storicamente condizionate, nel senso che tutte divengono possibili solo in seguito a determinate trasformazioni nella struttura dell'economia e del sistema creditizio. In generale, possono avere efficacia solo se c'è un'unica banca di emissione, il che storicamente non si è verificato né in tutti i tempi né in tutti i paesi (in Italia l'unificazione dei diversi istituti di emissione ha avuto luogo solo nel 1927). La stessa manovra dello sconto può avere piena efficacia solo dopo che i traffici e i movimenti internazionali di capitali hanno superato certe dimensioni ed è divenuto sufficientemente ampio l'uso dei biglietti e delle divise. La compravendita di titoli di Stato (in Inghilterra si parla di open market operations) viene attuata dalla banca centrale per far variare la massa dei biglietti in circolazione - con l'acquisto di titoli tale massa aumenta, con la vendita diminuisce - senza modificare il saggio di sconto e gli altri saggi dell'interesse. La compravendita di titoli di Stato ha potuto assumere carattere sistematico solo dopo che il loro mercato è divenuto ampio, e ciò è avvenuto col progressivo ampliamento dell'intervento pubblico nell'economia (nel passato solo durante e subito dopo le guerre quel mercato assumeva dimensioni ragguardevoli). La manovra sulle riserve delle banche ordinarie consiste nel variare la quota, rispetto al volume dei prestiti, che quelle banche debbono tenere sotto forma di biglietti per assicurare la convertibilità dei depositi: variando quella quota varia in senso inverso la massa dei prestiti e quindi dei depositi. Tale manovra presuppone che la banca di emissione non solo sia unica ma sia anche dotata di poteri coercitivi, che un'istituzione privata non può avere. Anche gli interventi volti a favorire ovvero a scoraggiare i rapporti di debito e di credito con l'estero delle banche ordinarie e le raccomandazioni che a tali banche rivolge la banca centrale in tanto sono concepibili in quanto questa banca ha poteri conferiti dalla legge: le banche ordinarie sanno che quei poteri esistono, anche se non vengono usati, e la 'persuasione' può avere efficacia.

In definitiva, la politica della banca centrale influisce sull'economia principalmente in due modi: variando il saggio ufficiale di sconto e variando la massa dei biglietti in circolazione e quindi la liquidità complessiva del sistema. Il rialzo dello sconto è efficace soprattutto sui conti con l'estero e sui cambi, e viene usato più in regime di cambi flessibili che in regime di cambi fissi. Ma il rialzo dello sconto rende inevitabile anche il rialzo dei saggi sui prestiti concessi dalle banche ordinarie alle imprese e ciò frena gli investimenti, non solo reali ma anche finanziari. Le variazioni della liquidità complessiva possono essere attuate anche senza variare lo sconto ufficiale; ma poiché di quelle variazioni risentono soprattutto le banche, esse si rifletteranno, in senso inverso, sui saggi dei prestiti concessi dalle banche alle imprese. 3. Moneta e prezzi

La teoria quantitativa della monetaNel descrivere la manovra dello sconto tradizionalmente si pone l'accento sugli effetti che questa manovra ha sui prezzi; in particolare, un rialzo dello sconto provocherebbe una flessione dei prezzi o, quantomeno, una diminuzione del loro saggio di aumento. Come si è accennato, da un rialzo dello sconto era lecito attendersi, nel XIX secolo e, in minor misura, al principio del XX, una flessione di tutti i prezzi, quelli delle materie prime e quelli dei prodotti finiti. Dopo la prima guerra mondiale e, in misura più accentuata, dopo la seconda, i prezzi dei prodotti finiti hanno mostrato una netta rigidità verso il basso, mentre è rimasta immutata la flessibilità dei prezzi delle materie prime in entrambe le direzioni; anzi, per i prezzi delle materie prime la doppia flessibilità è divenuta perfino più netta dopo il totale sganciamento del dollaro dall'oro, avvenuto nel 1971. Pertanto nel nostro tempo, salvo situazioni eccezionali, il rialzo non provoca una flessione nei prezzi dei prodotti finiti, ma ci si può invece attendere un rallentamento nell'aumento dei prezzi. Si tratta di chiarire meglio perché.

Il meccanismo tradizionalmente indicato per spiegare la flessione dei prezzi come conseguenza di un rialzo dello sconto è il seguente. L'aumento dello sconto da parte della banca centrale porta con sé l'aumento dello sconto delle banche ordinarie; i prestiti che le banche fanno alle imprese, con lo sconto di cambiali o con operazioni analoghe, diventano più cari. Diminuiscono quindi la massa dei prestiti e quella della moneta bancaria, giacché una parte cospicua dei depositi proviene dallo sconto di cambiali (un'altra parte proviene da versamenti di biglietti fatti dalle famiglie). Secondo l'argomentazione tradizionale, la

diminuzione dei mezzi di pagamento bancari comporta una diminuzione dei prezzi; ciò appare dalla cosiddetta equazione degli scambi, P=MV/Q, dove P indica il livello generale dei prezzi, M la massa di tutti i mezzi monetari, V la velocità di circolazione, ossia il numero delle volte che in media ogni unità monetaria è usata negli scambi nell'unità di tempo, e Q la quantità dei beni prodotti e scambiati. Posto che V e Q nel breve periodo siano relativamente stabili, risulta che P, il livello dei prezzi, varia al variare di M, la quantità di moneta. La cosiddetta equazione degli scambi rappresenta la formulazione più semplice della teoria quantitativa della moneta. Ci sono diverse altre formulazioni, meno semplici - una delle più recenti è quella proposta da Milton Friedman, il padre del monetarismo - ma le variabili fondamentali sono quelle indicate nell'equazione degli scambi.

A questo punto conviene riflettere sulla determinazione e sulle variazioni delle diverse categorie di prezzi tenendo conto delle diverse condizioni prevalenti nei diversi mercati. Nel XIX secolo nella maggior parte dei mercati prevalevano le condizioni di concorrenza: le imprese erano relativamente piccole, ed erano modeste sia le barriere all'entrata imputabili alle economie di scala sia quelle connesse con le sempre più rilevanti spese per la pubblicità. Il processo di concentrazione che ha avuto luogo nell'industria e in particolari servizi, come quelli del credito, è da porre in relazione con la crescente importanza, in alcuni settori, delle economie di scala, mentre il processo di differenziazione dei prodotti è entrato in un rapporto di interazione con la crescita della pubblicità; i due processi, che in certe attività si sono combinati, hanno modificato le condizioni dei mercati industriali e dei servizi, mentre nell'agricoltura e in parte nell'attività mineraria sono rimaste condizioni per diversi aspetti simili a quelle prevalenti nel XIX secolo, di modo che in queste attività permangono condizioni non lontane dalla concorrenza; nelle altre oggi prevalgono condizioni di oligopolio concentrato, differenziato o misto. Ora, in concorrenza i prezzi dipendono, nel breve periodo, dal gioco della domanda e dell'offerta e sono flessibili sia verso l'alto che verso il basso; nel lungo periodo dipendono dai costi. In oligopolio, invece, i prezzi dipendono dai costi anche nel breve periodo: se la domanda varia, variano le quantità prodotte piuttosto che i prezzi, giacché le imprese hanno un certo potere di mercato e preferiscono variare l'offerta piuttosto che i prezzi (ciò vale non solo per la riduzione dell'offerta, ma anche per l'aumento, giacché di norma le imprese dispongono di capacità inutilizzata; inoltre in mercati aperti, se la domanda aumenta, possono crescere le

importazioni senza aumento dei prezzi). Il risultato complessivo consiste in ciò, che oggi, nei mercati più importanti, che sono quelli dei prodotti industriali e dei servizi, i prezzi sono divenuti rigidi verso il basso rispetto alle variazioni della domanda; resta la flessibilità verso il basso nel caso dei prezzi delle materie prime agrarie e minerarie, ma tale flessibilità nelle condizioni attuali ha effetti diversi da quelli che aveva quando riguardava la maggior parte dei mercati, compresi quelli dell'industria e dei servizi, e si esplicava non solo verso l'alto ma anche verso il basso.

Posto che in prima approssimazione la massa dei mezzi di pagamento in circolazione esprima la domanda totale, nel secolo scorso la cosiddetta equazione degli scambi poteva avere un certo potere interpretativo, con riferimento a un periodo relativamente breve. Oggi ciò può valere solo quando la quantità dei mezzi monetari e specialmente dei biglietti aumenta a un ritmo talmente più rapido della produzione da prevalere su tutte le altre variabili; in tali condizioni la crescita della moneta tende a spingere rapidamente in alto tutti i prezzi, indipendentemente dall'andamento dei costi. Sono condizioni che si verificano in guerra o dopo una guerra; nei paesi evoluti raramente hanno luogo in tempi di pace. Oggi, in condizioni di relativa normalità, nel breve periodo le variazioni dei prezzi dei prodotti finiti e dei servizi dipendono dalle variazioni dei costi e in particolare dei costi diretti, in cui ha un ruolo dominante il costo del lavoro, dato dal rapporto fra salari e produttività. Sebbene ciò appaia evidente sul piano empirico, molti economisti sono restii ad abbandonare la teoria quantitativa della moneta. A un tale atteggiamento contribuisce il fatto che, in certe condizioni, quando i prezzi aumentano la banca centrale restringe l'emissione dei biglietti col risultato, sia pure non immediato, di un rallentamento dell'aumento dei prezzi. Tuttavia se, come spesso accade, l'aumento dei prezzi è spinto da un aumento del costo del lavoro, il risultato positivo della politica restrittiva non dipende da un rallentamento della domanda di beni originato da una crescita più lenta della massa monetaria, ma dipende invece dal fatto che la politica restrittiva frena gli investimenti e quindi la domanda di lavoro e quindi ancora l'aumento dei salari: il rallentamento dell'aumento dei prezzi è la conseguenza della crescita più lenta dei salari. Una tale politica, tuttavia, frena il processo di sviluppo e in certi casi dà luogo a una vera e propria recessione: la politica restrittiva, dunque, non influisce direttamente sui prezzi ma passa attraverso un aumento della disoccupazione.

Le tendenze dei prezzi nel XIX e nel XX secoloIn pieno contrasto con quanto accade oggi, nel XIX secolo i prezzi hanno avuto una nettissima tendenza a scendere. Fatto eguale a 100 l'indice dei prezzi in Inghilterra nel 1800, nel 1897, anno in cui bruscamente termina la flessione, l'indice è sceso a 25. Negli Stati Uniti le cifre corrispondenti sono 100 e 40: la flessione è meno accentuata per il fatto (così pare) che l'inflazione determinata dalla guerra civile ha spinto l'indice stabilmente in alto, anche se poi la flessione è riapparsa. Di un tale andamento abbiamo due ipotesi esplicative, una fondata sulla moneta, l'altra sui costi, avvertendo che allora la convertibilità era la regola e ricordando che nel lungo periodo i prezzi si adeguano ai costi anche in condizioni di concorrenza, allora prevalenti in quasi tutti i mercati.

Chi adotta l'interpretazione monetaristica della flessione dei prezzi assume come data la crescita della produzione - che va attribuita a fattori 'reali', da esaminare separatamente - e studia l'andamento dello stock di moneta reale - l'oro o l'argento. Così, l'economista svedese Gustav Cassel riteneva che nel XIX secolo nei principali paesi la produzione dei beni tendeva a crescere a un saggio medio annuale del 3%; da vari indizi egli pensava che lo stock d'oro tendesse a crescere a un saggio maggiore dopo le scoperte di nuove miniere, e a un saggio minore in altri periodi, contrassegnati dall'assenza di quelle scoperte, in quanto la produzione cresceva solo estendendo lo sfruttamento delle miniere già note. C'erano dunque periodi di relativa abbondanza d'oro, con una corrispondente tendenza dei prezzi a salire, e periodi di relativa scarsità, con la tendenza dei prezzi a scendere; nel complesso, nel XIX secolo avrebbero avuto la prevalenza i periodi di scarsità. L'ipotesi esplicativa proposta da Cassel, che dopo la prima guerra mondiale ebbe una notevole diffusione, è stata criticata da diversi economisti fra cui Schumpeter (v., 1939, p. 473). Le critiche sono di tre ordini: 1) l'ipotesi di Cassel presupponeva che in tutti i principali paesi avesse vigore la convertibilità in monete auree, mentre in alcuni di tali paesi vigeva la convertibilità in monete auree e/o argentee; 2) essa presupponeva anche una quota relativamente stabile fra monete auree e biglietti in circolazione, mentre così non è stato: più si diffondeva l'uso dei biglietti (per un lungo periodo i contadini erano stati restii ad accettarli in pagamento dei loro prodotti) più la quota delle riserve metalliche poteva scendere, senza che questo mettesse a repentaglio la convertibilità; 3) analogamente tendeva a scendere la quota delle riserve in biglietti detenuti dalle banche ordinarie, cosicché la supposta scarsezza d'oro poteva essere agevolmente

aggirata.

Se dunque l'ipotesi monetarista non regge, occorre considerare la seconda ipotesi esplicativa, quella dei costi, che in termini embrionali si trova già in Adam Smith. Secondo Smith, con lo sviluppo economico i salari tendono a crescere ma, grazie alla progressiva divisione del lavoro, la 'capacità produttiva' dei lavoratori cresce anche di più e ciò comporta una flessione dei costi e, man mano, dei prezzi. Per molte materie prime vale lo stesso ragionamento, anche se nel complesso, secondo Smith, qui la produttività del lavoro cresce meno rapidamente. (In effetti, nel secolo scorso i prezzi dei prodotti industriali sono diminuiti più dei prezzi delle materie prime).

Se questa ipotesi esplicativa è fondata, perché - ci possiamo chiedere - la flessione dei prezzi si arresta bruscamente nel 1897? E perché da allora in poi i prezzi, pur attraverso ampie oscillazioni, tendono a salire?La risposta alla prima domanda non è difficile: nel 1897, dopo una crisi ciclica, si profila una ripresa che, com'era la regola, portò con sé un aumento dei prezzi. Il quesito più rilevante è l'altro: come mai da allora in poi e specialmente dopo la seconda guerra mondiale si è imposta la tendenza verso l'aumento? Di nuovo l'attenzione va concentrata sui salari e sulla produttività. Col rafforzarsi dei sindacati operai, che si accompagna al processo di concentrazione delle imprese, e con la sempre più accentuata differenziazione dei prodotti e delle capacità dei lavoratori, cresce il potere di mercato dei lavoratori: i salari diventano sempre più rigidi verso il basso e tendono sistematicamente a salire, non di rado, anche più della produttività. Per le materie prime, soprattutto per quelle prodotte da paesi del Terzo Mondo, i due processi (concentrazione e differenziazione) non hanno luogo o hanno luogo in forma assai più limitata, i sindacati operai appaiono tardivamente e raramente raggiungono una rilevante forza contrattuale. Di conseguenza, i prezzi delle materie prime nel lungo periodo salgono meno (l'aumento può essere attribuito principalmente all'aumento nei prezzi dei mezzi di produzione venduti dai paesi industrializzati). Così, nel lungo periodo il rapporto fra prezzi dei prodotti industriali e prezzi delle materie prime, che nel secolo scorso ha avuto la tendenza a diminuire, nell'ultima parte di quel secolo e poi nel nostro ha mostrato la tendenza ad aumentare, mentre nel breve periodo i prezzi delle materie prime sono flessibili sia verso l'alto che verso il basso. Se il costo del lavoro non varia, ossia se i salari aumentano come la produttività e i prezzi delle

materie prime flettono, i prezzi dei prodotti finiti possono diminuire. Tale diminuzione, però, è assai limitata, poiché di regola la quota dei costi diretti imputabile alle materie prime è modesta. Solo quando ha luogo una spettacolare caduta dei prezzi delle materie prime, tale da determinare una sensibile flessione dei costi diretti, ci si può attendere una diminuzione sensibile dei prezzi dei prodotti finiti. (Una tale situazione si verificò durante la grande depressione negli Stati Uniti: i prezzi delle materie prime caddero di oltre il 50% e i prezzi dei prodotti finiti diminuirono, per questo motivo, di circa il 20%).

I diversi ruoli dei diversi mezzi monetariL'analisi fin qui svolta mette in evidenza la necessità di tenere ben distinti i diversi tipi di mezzi monetari, per motivi non solo formali ma anche sostanziali: nel passato le monete metalliche, i biglietti, i depositi; oggi i biglietti e i depositi. È bene avvertire che la moneta divisionaria non compare quasi mai nelle analisi giacché svolge un ruolo puramente sussidiario, specialmente rispetto ai biglietti. D'altra parte, in questo dopoguerra si sono progressivamente diffuse le carte di credito, che svolgono funzioni monetarie; e poiché queste carte debbono essere coperte da depositi, in prima approssimazione è possibile limitare l'analisi all'andamento dei depositi. Tuttavia, la diffusione delle carte di credito crea nuovi problemi di analisi e di politica monetaria, che sono tutti da esplorare a causa dell'origine relativamente recente del fenomeno. È certo che stanno rapidamente crescendo i tipi degli strumenti monetari e di intermediazione finanziaria, mentre al tempo stesso le moderne attrezzature elettroniche hanno consentito uno sviluppo straordinario delle transazioni finanziarie al livello mondiale. Tutto ciò rende incerta la nozione stessa di quantità di moneta e aggrava la labilità delle aspettative e i rischi di crisi finanziarie (v. Minsky, 1986); la nozione di quantità di moneta appare ulteriormente incerta se si ricorda che fra i mezzi monetari si considerano anche i depositi a risparmio e i titoli pubblici a breve scadenza, che possono svolgere alcune funzioni monetarie; a rigore sono da considerare solo come quasi-moneta.

Oggi conviene concentrare l'attenzione su due grandi aggregati monetari, quello dei biglietti e quello costituito dai depositi. Spesso nelle analisi monetarie, e in particolare in quelle dei monetaristi, in modo esplicito o implicito si assume che sussista un rapporto stabile fra i due aggregati. Ora, le banche ordinarie debbono tenere in biglietti (o in mezzi equivalenti) una riserva pari a una certa quota dei depositi, ma questo comporta solo un limite massimo alla

creazione dei depositi: non comporta affatto un rapporto stabile fra biglietti e depositi. In effetti, i due aggregati hanno andamenti diversi e in certi periodi di crisi perfino contrastanti, e ciò perché ne sono diversi la genesi e il ruolo.

È diversa la genesi: la banca centrale è un'istituzione pubblica e nell'emettere i biglietti ha molteplici obiettivi. Non necessariamente sorgono contraddizioni fra tali obiettivi; quando sorgono, però, la banca centrale è costretta a stabilire una scala di priorità: assicurare una relativa stabilità dei prezzi o dei cambi, assecondare lo sviluppo produttivo, favorire la crescita dell'occupazione, assecondare il collocamento dei titoli pubblici. I depositi, che fanno capo alle banche ordinarie, nascono fondamentalmente in due modi: provengono dalle entrate che le famiglie consegnano alle banche e che periodicamente ritirano e ricostituiscono; vengono 'creati' dalle banche sulla base di garanzie reali o, più spesso, di cambiali portate dalle imprese per lo sconto. È bene ricordare che in quest'ultimo caso ha luogo una trasformazione di titoli di credito che fanno capo a soggetti privati, e che non sono moneta, in titoli di credito che fanno capo alle banche e che sono moneta. La natura della moneta creditizia e particolarmente di quella che viene 'creata' dalle banche sulla base delle cambiali è stata vista esattamente da Schumpeter (v., 1912), il quale ha fatto della creazione della moneta bancaria una colonna portante della sua teoria dello sviluppo; già prima, però, il fenomeno era stato visto da Marx (v., 1867-1894, vol. II) e da De Viti De Marco (v., 1898).

Le concezioni del ruolo da attribuire alla moneta sono assai diverse. Mentre sono pochi gli economisti che si sono resi pienamente conto del ruolo della moneta nel processo di sviluppo, sono numerosi quelli che hanno elaborato analisi su ruoli degni di considerazione, ma meno rilevanti; tali analisi sono cresciute di numero dopo lo sviluppo delle dottrine di tipo monetarista. (Per una pregevole rassegna v. Bianchi, 1982).Tornando ai due fondamentali aggregati monetari, e cioè biglietti e depositi, è bene notare che fra i due aggregati sussistono differenze di rilievo anche nel loro ruolo di intermediari negli scambi. I biglietti sono usati prevalentemente dai consumatori, soprattutto da coloro che acquistano beni di consumo non durevoli; i depositi sono usati dagli uomini d'affari nei loro rapporti interni, per acquistare mezzi di produzione diversi dai servizi dei lavoratori; sono anche usati dalle famiglie per acquistare beni durevoli di consumo, come case e automobili. Per pagare i servizi dei lavoratori gli uomini d'affari convertono i depositi in biglietti, che i lavoratori

usano nei loro acquisti quotidiani. I commercianti man mano restituiscono alle banche i biglietti, se avevano avuto prestiti, altrimenti li depositano; nell'uno e nell'altro caso contribuiscono a ricostituire le riserve delle banche ordinarie. Pertanto il circuito del credito, che prende corpo nei due grandi aggregati monetari, si salda col circuito della produzione attraverso il commercio.

Giova sottolineare che i biglietti sono usati prevalentemente nelle transazioni riguardanti i beni che entrano nel costo della vita, i depositi in quelle dei prodotti intermedi e dei prodotti finiti venduti ai commercianti. Di tali diversi ruoli l'economista che intende analizzare l'andamento dei prezzi deve tener conto, cominciando col distinguere fra prezzi al consumo - per i quali conviene concentrare l'attenzione sui biglietti - e prezzi all'ingrosso - per i quali è particolarmente interessante l'esame dell'andamento dei depositi.

Nell'analisi dei prezzi sono dunque rilevanti due distinzioni, una orizzontale, l'altra verticale. La distinzione orizzontale è fra prezzi delle materie prime e prezzi dei prodotti finiti, quella verticale è fra prezzi al minuto e prezzi all'ingrosso. Le due distinzioni riguardano la logica della formazione e delle variazioni delle diverse categorie di prezzi nonché il ruolo dei prezzi nel processo di sviluppo, come si vedrà brevemente nel cap. 5.

4. Il processo inflazionistico

I molteplici impulsiSono numerosi, anche fra gli economisti che non accettano la dottrina monetarista, quelli che considerano l'inflazione come un fenomeno sempre essenzialmente monetario. In un certo senso ciò è ovvio: l'inflazione consiste in un aumento generalizzato dei prezzi e i prezzi sono ammontari monetari riferiti alle unità dei beni e dei servizi. Se però con quell'affermazione si vuol dire che l'impulso principale, e comunque l'impulso iniziale, parte sempre dalla moneta, l'affermazione è inaccettabile, come già può apparire dall'analisi precedente e come sarà meglio chiarito in questo paragrafo.Il cenno alle tendenze dei prezzi di lungo periodo ha messo in evidenza che l'inflazione è tipicamente un problema del nostro tempo e, in misure molto diverse, è oggi un problema che riguarda tutti i paesi del mondo. È interessante ricordare che nel rapporto annuale della Banca Mondiale nella prima tabella, che fornisce una serie di dati caratteristici dei 132 più importanti paesi del mondo, troviamo una

colonna con l'indicazione 'tasso medio annuale di inflazione' per due periodi pluriennali. Se per ipotesi fosse stato pubblicato un rapporto simile nel XIX secolo, si sarebbe invece trovata una colonna con i dati sul 'tasso medio annuale di deflazione', giacché allora la tendenza prevalente dei prezzi era verso la deflazione.

In prima approssimazione l'interpretazione monetarista può aver fondamento per le inflazioni e le iperinflazioni connesse con periodi bellici; non ha fondamento, invece, in periodi relativamente normali, durante i quali, sempre in prima approssimazione, vale lo schema dell'inflazione da costi. Tuttavia anche nel caso delle iperinflazioni - come quelle osservate, non in periodi bellici, in certi paesi dell'America Latina - non sempre il primo impulso parte dalla moneta. D'altra parte nei paesi sviluppati, durante i periodi di relativa normalità, solo eccezionalmente l'inflazione raggiunge tassi molto elevati, superiori al 20%; di solito i tassi variano dal 2 al 6-7% l'anno, ma va ricordato che tassi del genere riguardano i prezzi al minuto, mentre i prezzi all'ingrosso o aumentano di meno o non aumentano affatto. Con riferimento a tassi del 2-3-4% l'anno, si parla di 'pressione inflazionistica' proprio per indicare un'inflazione ininterrotta ma contenuta.

Tutto considerato, i principali impulsi inflazionistici provengono da sei direzioni: dalla quantità di moneta che alimenta la domanda complessiva; dalla domanda di certe categorie di beni, indipendentemente dall'andamento della massa monetaria; dai costi diretti; dai prezzi dei prodotti importati; dai cambi; dal fisco. C'è poi un caso speciale, che può aver luogo durante guerre o sconvolgimenti sociali: una scarsezza generalizzata di beni, che può originare o aggravare un processo inflazionistico.

Sulle relazioni fra quantità di moneta e inflazione si è già detto. La domanda influisce direttamente sui prezzi delle materie prime, agricole e minerarie, nei cui mercati di regola tuttora prevalgono condizioni concorrenziali. I costi e in particolare i costi diretti influiscono sui prezzi dei prodotti non agricoli. L'aumento nei prezzi dei prodotti importati può dipendere dall'andamento dei prezzi nei paesi d'origine, ma può dipendere anche da un aumento dei cambi esteri; in entrambi i casi conviene distinguere fra prezzi delle materie prime e dei mezzi di produzione e prezzi dei prodotti finiti importati, che entrano in concorrenza con quelli prodotti all'interno: nel primo caso si rientra nell'inflazione da costi,

nel secondo la spinta inflazionistica dipende dal fatto che si alza il limite che frena il potere delle imprese nazionali di spingere in alto i prezzi. Quanto alla politica fiscale, essa può generare inflazione in almeno due modi: attraverso un aumento delle imposte indirette, le quali, a differenza delle imposte dirette, normalmente si trasferiscono sui prezzi, o attraverso un aumento delle tariffe dei servizi pubblici, che entrano nel costo della vita.

Prima di procedere conviene fare qualche osservazione sul tentativo, compiuto in tempi recenti in Inghilterra, di applicare la dottrina monetarista per combattere l'inflazione, poiché tale tentativo ha fornito la prova empirica che in tempi normali la dottrina monetarista non funziona. Margaret Thatcher, che ha posto la lotta all'inflazione come uno degli obiettivi prioritari del suo governo e ha adottato una politica di rigido contenimento della quantità di moneta attraverso alti interessi, andò al potere nel 1979, quando i prezzi al consumo aumentavano alla velocità del 13%; quando il suo governo cadde, nel 1990, quei prezzi aumentavano del 10%. È vero che in certi anni l'aumento dei prezzi era sceso al 3-5%, ma ciò era avvenuto soprattutto per la flessione nell'aumento del costo del lavoro; la ripresa dell'inflazione è imputabile a una politica fiscale sbagliata (specialmente per gli aumenti di imposte indirette e di tariffe), adottata nella convinzione che l'andamento dei prezzi dipenda soltanto dall'andamento degli aggregati monetari. Secondo i monetaristi la politica fiscale può generare inflazione solo se c'è un deficit di bilancio e se questo deficit è finanziato con stampa di biglietti; ma non è così.

Il processo inflazionistico può essere amplificato dalle aspettative: se è ampia la schiera di coloro che si attendono un aumento dei prezzi e si comportano di conseguenza, i prezzi tendono effettivamente ad aumentare. Per evitare di ragionare in circolo, tuttavia, si possono chiamare in causa le aspettative solo dopo aver chiarito quali sono gl'impulsi che le generano. Inoltre, nei diversi mercati l'effetto di amplificazione può essere più o meno forte secondo le caratteristiche dei mercati medesimi. Così, gli effetti delle aspettative e quindi delle speculazioni nei mercati delle divise estere sono spesso particolarmente rilevanti proprio perché in quei mercati operano soggetti che per ragioni professionali cercano senza sosta di anticipare il futuro: ciò facendo accelerano gli eventi e talvolta con la loro condotta, sia pure temporaneamente, addirittura li provocano. Il fattore speculativo ha sempre avuto importanza nei mercati mondiali delle materie prime, che di regola vengono prodotte e offerte in

condizioni vicine alla concorrenza e vengono richieste dalle imprese industriali di tutto il mondo. In effetti, da tempo immemorabile nei mercati internazionali delle materie prime hanno luogo, oltre ai contratti 'a pronti', anche contratti 'a termine', nei quali convergono le operazioni speculative, che sono condizionate dalla politica degli intermediari finanziari e che si riflettono, in primo luogo, nelle variazioni delle scorte.

I differenziali di inflazioneDal 1971 l'inflazione si è accentuata in quasi tutti i paesi, sia pure con intensità diverse. Per chiarire questo fatto, conviene far riferimento, da un lato, alla crisi del sistema monetario internazionale, creato dagli accordi di Bretton Woods del 1944 e all'adozione, per un certo numero di anni, dei cambi flessibili in sostituzione dei cambi fissi, e dall'altro, ai mercati internazionali delle materie prime e di quella essenziale fonte di energia che è il petrolio.

Dopo la crisi del sistema monetario internazionale la flessibilità dei prezzi delle materie prime, che è sempre stata più rilevante di quella relativa ai prezzi dei prodotti finiti, è divenuta ancora più accentuata: il campo di variazione prima del 1971 era di circa il 5-6%, in aumento e in diminuzione, dopo quella data si è più che raddoppiato. Ciò perché, senza più l'ancoraggio delle principali monete al dollaro e, indirettamente, all'oro, nei mercati delle materie prime il fattore speculativo è divenuto nettamente più vigoroso che nel passato. D'altra parte, quando i prezzi delle materie prime aumentano, i prezzi dei prodotti finiti aumentano in proporzione al peso che le materie prime e le fonti di energia hanno nel costo diretto, mentre, quando i prezzi delle materie prime diminuiscono, i prezzi dei prodotti finiti diminuiscono di meno o non diminuiscono affatto, anche perché spesso lo spazio che si crea viene riempito da un aumento dei salari. Il risultato è che la maggiore variabilità, verso l'alto e verso il basso, dei prezzi delle materie prime e dell'energia tende ad avere, sui prezzi dei prodotti finiti, una spinta netta verso l'alto.

Dunque, dopo il 1971 il processo inflazionistico mondiale si è accentuato, ma in misure molto diverse, secondo i paesi. Mettendo per ora da parte i paesi dell'America Latina, i differenziali d'inflazione fra i paesi sviluppati sono riconducibili principalmente a tre ragioni.

In primo luogo, al diverso andamento del costo del lavoro per unità di prodotto. In secondo luogo, all'andamento dei cambi: i cambi fissi circoscrivono ai soli aumenti dei prezzi che hanno luogo nei mercati d'origine gli impulsi inflazionistici provenienti dall'estero, mentre coi cambi flessibili tali impulsi possono essere provocati anche da aumenti dei cambi. In terzo luogo, i differenziali d'inflazione dipendono dalla composizione della produzione. Beninteso, quei differenziali dipendono in primo luogo dalle politiche seguite dai diversi paesi.

I prezzi al consumo, che generalmente sono usati per misurare l'inflazione, tendono ad aumentare più dei prezzi all'ingrosso; anzi, in certi periodi e in certi paesi tendono ad aumentare anche quando i prezzi all'ingrosso sono stabili o diminuiscono limitatamente (ciò è accaduto più volte in Italia negli anni cinquanta). Il fatto è che i prezzi al consumo comprendono i servizi, dove la produttività aumenta più lentamente della produttività e dei salari industriali, che normalmente guidano l'aumento di tutti i salari, compresi quelli del commercio; inoltre, i prezzi al consumo comprendono anche i fitti, il cui andamento risente della scarsità di carattere monopolistico delle aree fabbricabili.Dopo l'introduzione del Sistema Monetario Europeo, che per un importante gruppo di paesi ha portato con sé il ritorno ai cambi quasi fissi, il processo inflazionistico mondiale ha subito una certa attenuazione. Tuttavia da qualche anno quel sistema è entrato in crisi.

L'inflazione nei paesi del Terzo MondoNel paragrafo precedente sono stati considerati prevalentemente i paesi industrializzati. Ben più ampi sono i differenziali di inflazione fra i paesi industrializzati considerati nel loro complesso e diversi paesi del Terzo Mondo. Particolarmente rapidi sono i processi inflazionistici nei paesi dell'America Latina, dove un saggio annuo del 70 o del 100% e anche più non è affatto eccezionale. L'esame di un tale fenomeno può mettere in luce gli impulsi inflazionistici che esprimono tensioni e conflitti sociali e politici. Queste tensioni e questi conflitti stanno dietro agli impulsi di tipo economico prima ricordati. La differenza fra i paesi più sviluppati e i paesi del Terzo Mondo afflitti da inflazioni galoppanti sta in ciò, che nei paesi più sviluppati di norma quei conflitti sono meno gravi, mentre nei paesi sottosviluppati possono essere e in certi periodi sono nettamente più gravi. La conflittualità va vista con riferimento al mercato del lavoro e al bilancio pubblico: sono queste le due aree in cui si manifestano e in un modo o nell'altro si

compongono i conflitti sociali e politici (uno dei modi di composizione è costituito dai trasferimenti sociali). In diversi paesi latinoamericani il sistema tributario funziona molto male, non solo per motivi organizzativi, ma anche per ragioni che si possono definire politiche, cosicché le entrate tendono a crescere più lentamente delle spese, che crescono a un ritmo sostenuto o per motivi militari (spesso dovuti a conflitti interni) o per motivi sociali o di politica economica. Il deficit pubblico, di conseguenza, cresce rapidamente e non di rado viene finanziato con la stampa di biglietti: si profila allora una situazione simile a quella che si ha in periodi di guerra e l'economia subisce un forte impulso inflazionistico. D'altra parte, il pessimo funzionamento del sistema tributario riguarda soprattutto le imposte dirette, che colpiscono le famiglie e le società e che esigono adeguati strumenti di accertamento o di controllo; le imposte indirette, che colpiscono i beni e i servizi, possono essere amministrate con relativa efficacia anche attraverso apparati tributari mal funzionanti. Per contenere il deficit il governo è allora indotto a ricorrere alle imposte indirette, cosa che a volte è la conseguenza di clausole poste da istituzioni finanziarie per concedere prestiti. Ma le imposte indirette, a differenza di quelle dirette, sono inflazionistiche: un aumento ampio e generalizzato di tali imposte può innescare un processo di inflazione, che viene via via aggravato dalle rivendicazioni dei sindacati, i quali cercano di recuperare la perdita di potere d'acquisto dei salari; si mette così in moto la spirale salari-prezzi. In alcuni casi nel processo inflazionistico hanno un ruolo di rilievo i cambi: certi governi hanno praticato una politica di successive svalutazioni, piccole ma frequenti, che annulla l'effetto di freno che i cambi possono avere sui prezzi interni e, dato che di norma i cambi aumentano prima e più rapidamente dei prezzi, favorisce le esportazioni, il che costituisce il motivo principale di una tale manovra; in tal modo però l'inflazione viene accelerata.

Nei paesi latinoamericani sono state messe in evidenza alcune cause strutturali dell'inflazione. Una di tali cause risiede nell'agricoltura e dipende dalla distribuzione relativamente concentrata della proprietà: da un lato grandi proprietari, spesso assenteisti, dall'altro contadini poverissimi che producono principalmente per l'autoconsumo e sono comunque incapaci di rispondere attivamente agli stimoli del mercato. La conseguenza è che l'offerta di prodotti agricoli per il consumo interno risulta relativamente rigida, cosicché una domanda in crescita, anche per la rapida urbanizzazione, tende a tradursi più in prezzi crescenti che in offerta crescente. Va anche considerata l'organizzazione del commercio dei

prodotti agricoli: pochi grandi commercianti all'ingrosso, spesso collegati ai grandi proprietari, trovano più profittevole puntare su maggiori margini d'intermediazione piuttosto che sull'espansione delle vendite.

In breve, se l'inflazione è un fenomeno complesso nei paesi sviluppati, esso è ancora più complesso nei paesi semisviluppati dell'America Latina. Nei paesi decisamente arretrati dell'Africa, specialmente dell'Africa subsahariana, l'inflazione è per certi aspetti un fenomeno meno complesso; certamente è diverso. In tali paesi la quota dei lavoratori dipendenti sul totale costituisce una minoranza, in certi casi non supera il 10-15%; gli altri sono lavoratori indipendenti: piccoli commercianti, modesti artigiani e, spesso, membri di tribù nelle quali prevale la produzione per l'autoconsumo. In queste condizioni i salari e gli stipendi hanno ben poco spazio in un processo inflazionistico che, quando ha luogo (e in alcuni paesi assume un'intensità rilevante), si manifesta nelle città, nelle quali il mercato, pur con caratteristiche elementari, assume un peso non trascurabile e articolazioni differenziate. Nelle capitali, dove risiede l'amministrazione centrale, troviamo il maggior numero di pubblici funzionari e i centri delle spese statali, che sono costituite principalmente dalle spese per investimenti e dalle spese militari, connesse con conflitti interni che non di rado sono conflitti etnici e conflitti sociali, solo limitatamente paragonabili a quelli che hanno luogo nei paesi sviluppati. 5. Moneta e sviluppo economico

La creazione della moneta bancariaCon la transizione dalle economie quasi stazionarie dell'epoca premoderna - economie prevalentemente agrarie, di tipo feudale o quasi feudale - alle economie moderne, caratterizzate prima dal capitalismo commerciale e in seguito, in certi paesi, dal capitalismo industriale, lo sviluppo economico diventa, da episodico e saltuario, un processo normale. In tali condizioni la moneta è sempre più legata al credito; anzi la separazione man mano si attenua e la moneta creditizia diviene il principale mezzo di pagamento nelle sue due configurazioni - biglietti e depositi -, con l'avvertenza che quando l'inconvertibilità diviene la regola i biglietti non sono più titoli di credito, ma costituiscono la moneta di ultima istanza.

Un'economia stazionaria può essere paragonata a un circolo, mentre un'economia che si sviluppa può essere vista come una spirale; Joseph Schumpeter distingue tra flusso circolare

e sviluppo, Karl Marx tra riproduzione semplice e riproduzione su scala allargata o accumulazione. Nel capitalismo industriale il processo reale è l'accumulazione, che si svolge attraverso periodi di crescita accelerata e periodi di temporaneo ristagno e di crisi. Per Schumpeter il processo di sviluppo è spinto da innovazioni tecnologiche e organizzative ed è attuato da imprese nuove o da imprese già esistenti che s'innovano: in tale processo il ruolo essenziale spetta alla creazione di moneta creditizia, mentre quello del risparmio è del tutto secondario. Vi sono tuttavia anche imprese che espandono la produzione su una base tecnica sostanzialmente invariata: lo stimolo all'espansione può venire da un aumento della domanda generato dalle imprese che s'innovano o da spese pubbliche o dalla domanda estera, o da una combinazione di queste spinte. Infine, vi sono imprese che in ciascun periodo non mutano né il volume della produzione né la base tecnica: sono quelle che, idealmente, rientrano in un processo astratto di riproduzione semplice. Per tali imprese il credito non è indispensabile, giacché la produzione può essere finanziata con le entrate correnti; il credito può servire a far combaciare nel tempo i cicli di produzione coi cicli delle uscite e delle entrate monetarie, una funzione che può essere svolta anche dal risparmio.

Il flusso dei mezzi monetari direttamente connessi col processo di sviluppo è quello costituito dai depositi in conto corrente creati dalle banche per le imprese che si espandono, con o senza innovazioni. Indirettamente collegati con lo sviluppo sono i biglietti emessi dalla banca centrale per sostenere l'azione delle banche ordinarie, che debbono mantenere certe riserve in biglietti e si rivolgono alla banca centrale per riscontare i titoli di credito che hanno scontato alle imprese. La banca centrale emette biglietti anche per sostenere l'azione del governo, particolarmente attraverso l'acquisto e la vendita di titoli pubblici; infine, la banca centrale emette (o ritira) biglietti per ragioni connesse con le relazioni economiche internazionali (merci e capitali). La banca centrale ha quindi un panorama ampio nella sua politica di emissione di biglietti, i quali pertanto non possono essere posti in rapporto diretto col processo di sviluppo economico portato avanti dalle imprese. La banca centrale regola il flusso netto dei biglietti attraverso i tre canali ora richiamati (banche ordinarie, Tesoro, estero) con gli obiettivi plurimi di cui si è detto.

Il volume dei biglietti può certo essere visto come una quantità esogena, ossia determinata da decisioni esterne al sistema economico, anche se si tratta di decisioni che condizionano

l'economia e dall'economia sono condizionate. Il volume dei depositi in conto corrente creati dalle banche ordinarie per le imprese costituisce invece una quantità endogena, nel senso che dipende dalla richiesta di prestiti bancari, a sua volta determinata dai profitti attesi, le cui variazioni dipendono dall'andamento dell'economia (v. Sylos Labini, 1948).

Oltre ai depositi in conto corrente creati dalle banche per le imprese, ci sono i depositi in conto corrente di cui sono creditrici le famiglie che hanno compiuto versamenti in biglietti: qui si deve parlare non di creazione, ma di trasformazione di mezzi monetari-depositi in luogo di biglietti. Ci sono poi i depositi a risparmio delle imprese e delle famiglie, che a rigore non sono mezzi monetari.I depositi in conto corrente delle famiglie sono da collegare all'andamento dell'occupazione e degli investimenti e quindi variano insieme ai depositi creati per le imprese.

Le forme di risparmio sono dunque diverse. In primo luogo occorre distinguere fra risparmio offerto e risparmio usato dagli stessi soggetti che lo compiono, di regola imprese. Il risparmio offerto può prendere diverse vie. Può andare alle banche sotto forma di depositi veri e propri e le banche lo usano per alimentare le riserve sulla cui base fanno prestiti alle imprese per un ammontare ben superiore a quello dei depositi ottenuti. Può essere usato per acquistare titoli, che servono a finanziare le imprese o lo Stato. Il risparmio impiegato direttamente dalle imprese varia in relazione alle aspettative di profitto: qui risparmio e investimenti addirittura coincidono (v. Kalecki, 1939). L'interesse sul risparmio che viene offerto da alcuni soggetti ad altri serve a indurre coloro che risparmiano a compiere questa offerta, che ha un ruolo importante nel processo di sviluppo sia perché è di necessario ausilio per le banche, sia perché contribuisce direttamente al finanziamento dell'attività produttiva. L'interesse sui prestiti bancari ha invece la funzione di procurare alle banche il guadagno che le induce a svolgere la loro attività. Per l'autofinanziamento c'è da considerare solo il profitto, da cui va sottratto l'interesse su un piano puramente contabile.

Nel nostro tempo il problema del risparmio va inquadrato nel più ampio problema del finanziamento dello sviluppo. In pieno contrasto con questo punto di vista, che si pone sul piano dinamico, è quello della tradizione neoclassica, che si pone sul piano statico e prescinde dallo sviluppo. Per i neoclassici l'interesse è il prezzo che equilibra la domanda di

risparmio, espressa dagli imprenditori che debbono fare investimenti, con l'offerta di risparmio, espressa appunto dai risparmiatori.

Secondo la concezione qui esposta, l'interesse sui prestiti alle imprese è legato alla crescita del reddito; nei prestiti al consumo, invece, l'interesse comporta una redistribuzione del reddito. Ciò è vero sia per i prestiti ai privati sia per quelli contratti dallo Stato al fine di coprire spese improduttive, come sono le spese militari e quelle occorrenti per pagare gl'interessi sul debito pubblico. Nel nostro tempo, tuttavia, la dinamica dell'intero sistema dei saggi dell'interesse dipende dalla dinamica delle imprese: conviene perciò concentrare l'attenzione sull'interesse produttivo.

Da quanto precede appare che l'andamento del risparmio usato per fini produttivi e l'andamento degli investimenti vanno insieme. Per l'autofinanziamento ciò è addirittura ovvio; è bene avvertire che l'autofinanziamento comprende non solo quello delle imprese vere e proprie ma anche quello dei lavoratori autonomi, che in un paese come l'Italia hanno importanza assai notevole (i lavoratori autonomi da noi rappresentano un terzo del totale dei lavoratori). Per il risparmio delle famiglie vale l'argomentazione di Keynes: un aumento del reddito complessivo determinato da un aumento degli investimenti dà luogo a un aumento del reddito delle famiglie e, data la loro propensione al risparmio, a un aumento del risparmio. Sia nel caso dell'autofinanziamento, sia in questo caso sono gli investimenti che generano il risparmio e non il contrario, come la teoria tradizionale induce a credere.

Profitto, interesse e sviluppoPer mettere in evidenza il nesso, nel capitalismo moderno, tra profitto e sviluppo possiamo ipotizzare, in prima approssimazione, che gli investimenti siano finanziati completamente con profitti, senza fondi esterni. Diviene poi agevole fare l'ipotesi, più realistica, di un finanziamento alimentato sia dai profitti non distribuiti sia da prestiti bancari, che comportano creazione di depositi. Facendo dunque l'ipotesi di completo autofinanziamento e indicando con I gli investimenti e con G i profitti totali, abbiamo

$I=aG$, (1)

dove $a=1$ se i profitti sono integralmente investiti e $a<1$ se sono investiti solo in parte. In

luogo della (1) possiamo scrivere $I/K=aG/K$ dove K è lo stock di capitale fisso. Dato che gli investimenti possono essere visti come l'incremento dello stock di capitale, ossia $I=\Delta K$, e supponendo che il rapporto fra stock di capitale e reddito rimanga stabile quando lo stock viene accresciuto, abbiamo $K/Y=\Delta K/\Delta Y$, ovvero $\Delta Y/Y=I/K$ o anche, indicando con $\hat{Y}$ il saggio di aumento del reddito e tenendo conto della (1),

formula (2)

Ciò significa che il saggio di aumento del reddito, $\hat{Y}$, coincide col saggio del profitto, G/K, quando $a=1$; se $a<1$, ne rappresenta una quota.

Qui possiamo introdurre la creazione di moneta bancaria.Se l'imprenditore svolge la sua attività con mezzi propri, può tenere il profitto tutto per sé. Se invece, com'è la regola nel tempo moderno, svolge la sua attività in tutto o in parte con mezzi monetari ottenuti in prestito, deve cedere al creditore, che spesso è una banca, una parte del profitto complessivo sotto forma d'interesse. Il profitto netto è dunque dato dalla differenza tra profitto e interesse, che, come sostiene Schumpeter, è un'imposta sull'utile dell'imprenditore.

Dunque, quando i profitti non sono integralmente investiti vale la relazione $\hat{Y}=aG/K$, dove $a<1$. D'altra parte gli imprenditori cercano di ottenere un profitto maggiore dell'interesse, i. Se così accade, si ha $i=bG/K$, dove $b<a$. Pertanto, $a/b\ i=\hat{Y}$, cosicché, se $a=b$,

formula (3)

ossia il saggio dell'interesse tende a eguagliare il saggio di incremento del reddito. D'altra parte, per la relazione $i=bG/K$, il saggio dell'interesse tende a livellarsi sul saggio del profitto; ammesso che $b\langle 1$, resta una differenza a favore del profitto anche se l'imprenditore finanzia la sua attività solo con prestiti esterni - un caso eccezionale, come il caso opposto, di completo autofinanziamento.

Le variazioni di breve periodo del saggio dell'interesseLa tesi secondo cui l'interesse tende a livellarsi sul saggio del profitto, a meno di una differenza, fa capo addirittura agli economisti

classici, Smith e Ricardo in particolare.Le relazioni considerate nel paragrafo precedente si riferiscono al lungo periodo. Se in un tale contesto c'è una tendenziale corrispondenza fra interesse e profitto, nel breve periodo spesso si osserva una correlazione inversa; per lo meno c'è correlazione inversa fra interesse sui prestiti bancari e profitti. Ciò perché, quando il profitto aumenta, aumenta l'autofinanziamento e diminuisce la pressione delle imprese sulle banche, le quali, d'altra parte, preferiscono accrescere i loro guadagni allargando il più possibile gli impieghi (i profitti crescono in una fase di congiuntura favorevole) e sono disposte a ridurre il margine fra interessi sui prestiti (che sono saggi attivi per le banche) e interessi sui depositi (saggi passivi).

Dunque, nel lungo periodo il saggio dell'interesse tende ad avvicinarsi al saggio del profitto; nel breve periodo, però, le variazioni dell'interesse dipendono da altri fattori: oltre che, inversamente, dai profitti, le variazioni dell'interesse dipendono dall'andamento dei prezzi, da quello dei cambi, dal fabbisogno di bilancio che lo Stato intende coprire con la vendita dei titoli, dalle esigenze di liquidità delle famiglie e delle imprese; in breve, si tratta di una funzione non di una ma di più variabili.

I saggi dell'interesse sono molteplici. La distinzione fondamentale è fra l'interesse a lungo e l'interesse a breve termine. Per l'interesse a breve conviene considerare il saggio di sconto della banca centrale e il saggio dell'interesse sui titoli pubblici - i due saggi si muovono insieme.Consideriamo dunque, uno per uno, i fattori che influiscono sugli interessi a breve. L'andamento dei prezzi condiziona questi interessi in vari modi. I prezzi in aumento scoraggiano l'offerta di risparmio, giacché diminuisce l'interesse reale (interesse nominale meno saggio di aumento dei prezzi): un aumento dell'interesse nominale può impedire una tale diminuzione. D'altra parte, se l'aumento dei prezzi dipende da un aumento troppo rapido dei salari, la banca centrale può elevare lo sconto per ammonire i sindacati e le imprese. Se invece i prezzi aumentano come conseguenza di un aumento dei cambi determinato da un deficit estero, la banca centrale eleva lo sconto per riequilibrare i conti con l'estero - l'aumento dello sconto attira capitali di altri paesi e scoraggia l'esportazione di capitali nazionali. In condizioni di instabilità politica i cambi possono aumentare anche indipendentemente da un deficit estero, giacché i soggetti che operano nei mercati finanziari hanno aspettative genericamente negative; anche in tali condizioni la banca centrale può

elevare lo sconto, per contrastare i movimenti speculativi sulla moneta nazionale.

L'andamento dei prezzi influisce sugli interessi a breve termine in periodi brevi; tuttavia, se hanno luogo un aumento o una diminuzione molto prolungati dei prezzi, anche gli interessi a lungo termine ne risentono. Così, nella seconda metà del secolo scorso anche l'interesse a lungo termine mostrò una tendenza sistematica a diminuire, insieme con la maggior parte dei prezzi, tanto che alcuni economisti ritennero che tale tendenza fosse l'espressione della caduta tendenziale del saggio del profitto e rappresentasse quindi la conferma della tesi proposta da un lato da Ricardo e dall'altro da Marx. In realtà si trattava, più semplicemente, di una tendenza che esprimeva il nesso fra interesse e prezzi, che allora si manifestava in un periodo lungo. Si può ipotizzare che oggi, in un'epoca di alti interessi sia a breve che a lungo termine e di pressione inflazionistica generalizzata, ricompaia lo stesso nesso, col segno cambiato.

Procedendo nell'esame dei fattori che influiscono sugli interessi a breve, consideriamo ora il deficit pubblico. Per coprire un tale deficit senza ricorrere all'emissione dei biglietti, lo Stato vende titoli pubblici: maggiore è il volume dei titoli offerti, più alto è l'interesse sui titoli necessario a persuadere i risparmiatori ad acquistarli tutti. Imprese e famiglie debbono tenere una parte delle loro entrate in forma liquida per far fronte a pagamenti scaglionati nel tempo; possono tuttavia tenere una certa liquidità per approfittare di acquisti vantaggiosi, per esempio, di titoli pubblici; così, se i soggetti si attendono, nel futuro, un interesse più alto di quello corrente, il valore dei titoli tende a salire e conviene quindi comprarli, rinunciando a conservare la quota della liquidità non necessaria per i pagamenti normali. Viceversa, man mano che si riduce il divario fra interesse atteso e interesse corrente, tende a diminuire la liquidità tenuta da parte per motivi speculativi. Sotto questo aspetto fra l'interesse corrente e il volume della liquidità complessiva, espressa in biglietti, c'è una relazione inversa. Tuttavia in una situazione di crisi le esigenze di liquidità delle imprese aumentano rapidamente, giacché per pagare i debiti ed evitare il fallimento molte imprese hanno bisogno di danaro liquido per rimborsare i prestiti, senza poter aprire nuovi cicli produttivi da cui ricavare nuove entrate, finché la crisi dura. In tali condizioni l'interesse a breve può salire a livelli molto elevati, anche se la liquidità che fa capo alla banca di emissione cresce sensibilmente. Queste osservazioni mostrano che le relazioni tra liquidità complessiva e interessi a breve

sono molto complesse; di norma prevale una relazione inversa, ma durante una crisi ciò non è più vero.Un'ultima osservazione. La recente esperienza italiana mostra che l'instabilità politica influisce non solo sui cambi, ma anche sugli interessi, a breve e a lungo termine: cambi e interessi possono essere visti come indici del grado di fiducia di cui gode un paese nella comunità finanziaria internazionale. Dopo aver considerato alcuni fattori quantificabili che influiscono su cambi e prezzi, bisogna fare i conti con quel problematico fattore qualitativo, quantificabile solo nei suoi effetti, che definiamo grado di fiducia (v. Ciocca e Nardozzi, 1996).

La moneta e il ciclo economicoNel capitalismo moderno il processo di sviluppo procede attraverso alti e bassi, ovvero attraverso accelerazioni e decelerazioni o riflessioni.Riconosciuto che il processo di sviluppo è sostenuto dal credito e, in primo luogo, dalla moneta bancaria, dobbiamo attenderci una stretta corrispondenza fra andamento della produzione e andamento della moneta bancaria, particolarmente dei depositi in conto corrente. Una tale corrispondenza non implica che la moneta bancaria abbia un ruolo determinante nell'andamento ciclico. A rigore, non è ipotizzabile neppure il nesso opposto, cioè che l'attività produttiva determina l'andamento della moneta bancaria; ma certo questa ipotesi è meno lontana dalla verità, giacché sono le aspettative di profitto che influiscono sulle decisioni di investimento, che a loro volta regolano l'andamento dell'attività produttiva e quindi della moneta bancaria e quindi ancora, in via subordinata, quello del risparmio.

Economisti di grande rilievo, come Knut Wicksell, Ralph Hawtrey, John Maynard Keynes e Marco Fanno hanno elaborato interpretazioni del ciclo economico che assegnano alla moneta un ruolo importante nel ciclo, anche se non necessariamente un ruolo causale. In un modo o nell'altro questi economisti concentrano l'attenzione sulle variazioni del rapporto fra profitto e interesse: nell'ascesa il primo tende a sopravanzare il secondo, e quindi il profitto netto è positivo ed è crescente; l'opposto accade durante una flessione, o almeno al principio di una flessione. La moneta - meglio precisare: la moneta bancaria - tende a espandersi nell'ascesa a un ritmo sostenuto, mentre nella flessione diminuisce ovvero, se la flessione è blanda, aumenta a un saggio calante. Nel passato, quando il gioco della domanda e dell'offerta influiva nel breve periodo sui prezzi nella maggior parte dei mercati, i prezzi medesimi tendevano ad avere un andamento ciclico, come la produzione; oggi una vera e propria

regolarità non si nota più.

Fra andamento della produzione e andamento della moneta bancaria nel ciclo la corrispondenza è stretta, mentre non c'è una corrispondenza altrettanto stretta - anzi in certi periodi c'è una netta divergenza - fra attività produttiva e base monetaria-biglietti e riserve bancarie. Sono stati già più volte indicati i motivi di tale possibile divergenza. È sembrato opportuno mettere nella massima evidenza questo fatto perché di norma esso viene trascurato dagli economisti, specialmente dai monetaristi i quali, dopo aver riconosciuto il ruolo preminente svolto dalla base monetaria, ragionano come se esistesse un rapporto relativamente stabile fra la stessa base monetaria e il volume dei depositi. Che le cose non stiano così è dimostrato in modo lampante dall'andamento dei due aggregati monetari negli anni in cui ha inizio negli Stati Uniti la 'grande depressione', un evento straordinario e drammatico che ha segnato profondamente non solo la storia economica ma anche la storia politica del nostro secolo. Gli andamenti dei due aggregati e della produzione industriale, che rappresenta l'indice più significativo dell'attività economica, risultano dalla figura.

Appare evidente che i depositi variano insieme con la produzione industriale mentre l'andamento della base monetaria non corrisponde affatto alla congiuntura, anzi, il suo volume rimane pressoché stazionario dal 1924 al 1930, anni in cui ha luogo una diminuzione appena percettibile, e invece cresce dal 1930 al 1933, mentre il volume dei depositi addirittura precipita insieme con la produzione. La depressione quindi esplode per motivi non imputabili alla banca centrale, in pieno contrasto con quanto afferma il fondatore del monetarismo, Milton Friedman (fu Nicholas Kaldor che per primo mise in evidenza questo punto; gli scritti dei due autori si trovano in Bellone: v., 1970). È tuttavia possibile muovere una critica severa alla banca centrale americana, la quale non avrebbe dovuto, particolarmente all'inizio della crisi, nel 1929, limitarsi a tenere stabile il volume della base monetaria: avrebbe dovuto invece accrescerlo, come fece poi in modo tardivo e inadeguato, concedendo crediti assai più larghi alle banche e bloccando così quella catena di fallimenti bancari che contribuì notevolmente alla gravità e alla durata della grande depressione. Più recentemente, negli ultimi mesi del 1987, quando si era delineata una grave crisi finanziaria, la banca centrale americana ha effettivamente esercitato un'azione controbilanciante per impedire che la crisi finanziaria degenerasse in una crisi della produzione e dell'occupazione,

come molti temevano, dentro e fuori degli Stati Uniti, e quell'azione ha avuto successo.

Se non esiste un rapporto neppure approssimativamente stabile fra base monetaria e moneta bancaria, ne seguono almeno tre corollari, utili per la politica monetaria. Il primo lo abbiamo visto proprio ora: quando occorre, la banca centrale deve esercitare un'azione di contrappeso rispetto alle forze che spingono verso il basso il volume della moneta bancaria e l'attività creditizia. Il secondo: non è affatto saggia la regola automatica, a prima vista seducente, di far crescere la base monetaria secondo obiettivi predeterminati. Infine - terzo corollario - non è affatto consigliabile togliere alla banca centrale la vigilanza sulle banche ordinarie, con l'argomento che essa serve solo a evitare abusi che possono danneggiare i risparmiatori, ma non è rilevante per la politica monetaria; la banca centrale deve vigilare non solo per evitare quegli abusi, ma anche per essere in grado di individuare subito i motivi per cui in certi periodi la solvibilità di molte banche diviene rischiosa, anche senza colpa degli amministratori; solo così può intervenire in modo tempestivo ed efficace. 6. Una riflessione conclusiva

Tutti i fenomeni economici sono storicamente condizionati, ma varia il tasso di storicità, se così si può dire: tra i fenomeni meno storicamente condizionati troviamo il baratto (anche se lo scambio fra beni senza l'intervento della moneta assume connotati diversi quando la moneta ha una diffusione minima e quando viceversa sono ampiamente diffuse diverse forme di mezzi monetari); all'opposto, fra i fenomeni più storicamente condizionati troviamo proprio la moneta: lo stesso sistema monetario internazionale può cambiare una o due volte in modo significativo nel corso di una generazione. In un lontano passato la moneta era soltanto un intermediario negli scambi e un serbatoio, oltre che una misura, di valore, come ancora si legge nei libri di testo. Nel nostro tempo la moneta ha assunto nuovi ruoli e, in quanto moneta creditizia, è divenuta la leva essenziale dello sviluppo economico.

Assai pochi fra gli economisti teorici - ricordo solo Schumpeter e Breglia - riconoscono che i fenomeni economici sono storicamente condizionati; eppure riconoscere ciò non significa affatto negare la possibilità e l'utilità di elaborare modelli teorici, usando anche, se occorre, gli strumenti matematici. Il fatto è che tali modelli possono avere efficacia interpretativa solo se si fondano su ipotesi ricavate dalla realtà economica, la quale non è immutabile, ma

cambia anche radicalmente nel tempo storico.

Capitale
di Giorgio Lunghini

sommario: 1. Introduzione. 2. L'epoca classica e la critica marxiana. a) Il capitale come anticipazioni: François Quesnay. b) Adam Smith: il lavoro diviso. c) David Ricardo: saggio dei profitti, macchine e accumulazione. d) John Stuart Mill: il capitale come fattore della produzione. e) Karl Marx: la critica della formula trinitaria. f) Marx: lavoro salariato e capitale. g) Marx: il processo di produzione-riproduzione. h) Marx: plusvalore e profitto. 3. L'epoca neoclassica. a) L'equilibrio economico come ordine naturale e necessario. b) Il capitale nell'equilibrio economico generale. c) La teoria 'austriaca': capitale e tempo. d) Irving Fisher: impazienza e volontà. e) Domanda di capitale e produttività marginale. f) Knut Wicksell: una curiosa divergenza. 4. Capitale e moneta. a) Joseph A. Schumpeter: il capitale monetario nel processo capitalistico. b) John M. Keynes: capitale, investimenti e animal spirits. c) Keynes e il saggio di interesse: il distacco dalla tradizione. d) Keynes: capitale e (dis)occupazione. 5. Controversie recenti sulla misura del capitale. a) Il capitale nella funzione di produzione. b) Le 'parabole' neoclassiche. c) La critica sraffiana: l'armonia neoclassica come caso particolarmente astratto. d) Capitale e società: la scomparsa della categoria. □ Bibliografia.

1. Introduzione

'Capitale' è un termine tratto dal linguaggio comune, con cui si indica normalmente qualsiasi forma di ricchezza accumulata e accantonata, che consenta di ottenere dal lavoro proprio o altrui, passato o futuro, un 'sovrappiù', o comunque prometta un godimento ovvero un 'profitto'. In questo senso sarebbe capitale qualsiasi mezzo di produzione prodotto, il quale aiuti a produrre di più e meglio (dall'amigdala al robot), o qualsiasi somma di denaro che frutti o possa fruttare un guadagno netto (dai ducati prestati da Shylock ad Antonio, ai dollari investiti da Henry Ford nella produzione della Ford T). Non avrebbe dunque importanza chi

possieda il capitale, a quale scopo e con quali tecniche lo impieghi, e come tutto ciò condizioni le modalità di produzione-riproduzione del sistema esistente. La definizione è evidentemente troppo ampia.
Come tutte le categorie dell'economia politica, anche quella di 'capitale' richiede una determinazione storica. Anche il 'lavoro' ha forme diverse nelle diverse epoche, tuttavia di esso si può parlare in generale. Il lavoro è semplicemente la principale attività materiale con la quale l'uomo si pone in rapporto con la natura, al fine di trarne valori d'uso. Esso è certamente più produttivo quando è svolto con l'ausilio di strumenti appropriati, che nulla impedisce di chiamare 'beni capitali': sarebbe però eccessivo ridurre la nozione di capitale a quella di un insieme di beni capitali, poiché in tal caso ci si limiterebbe a constatare che la produzione con capitale dà luogo a un sovrappiù maggiore di quello ottenibile senza un tale ausilio, e si rinuncerebbe a spiegare perché mai questo modo di produzione conferisca al sovrappiù la forma del profitto. Proprio dal concetto di sovrappiù converrà partire per intendere come nel modo capitalistico di produzione il capitale non sia soltanto un insieme di cose o una somma di denaro, ma un rapporto sociale.
Il sovrappiù può essere definito come quel che resta del prodotto sociale (tutto ciò che in un'economia viene prodotto in un dato periodo di tempo), una volta reintegrati i beni di consumo necessari per la sussistenza e la riproduzione dei lavoratori (produttivi) nonché i beni capitali che si sono consumati o logorati nel corso della produzione. Il sovrappiù sarà nullo, per definizione, in un'economia di mera sussistenza, ma normalmente sarà positivo, e può essere positivo in qualsiasi modo di produzione. Diversi, tuttavia, sono i modi in cui il sovrappiù viene prodotto, le persone o classi che se ne appropriano, l'uso che se ne fa e il ruolo che in tutto ciò hanno l'istituto della proprietà, il mercato e la moneta.
In un'astratta società precapitalistica - diciamo 'feudale' - il sovrappiù viene prodotto mediante il comando diretto del lavoro dei servi (la corvée, per esempio). Del sovrappiù, in natura o in denaro, il signore si appropria in virtù di un rapporto di potere strettamente politico e non di scambio economico, e lo impiega non per l'allargamento del processo produttivo, ma per quello che si può chiamare 'consumo signorile'. Al mercato si ricorre essenzialmente per gli scambi intercomunitari; la moneta ha come funzione pressoché esclusiva quella di facilitare gli scambi.
Con l'avvento del capitalismo, quali che ne siano state le cause ('accumulazione originaria' o 'grande trasformazione'), si assiste a una polarizzazione della società. Se si trascurano i

residui feudali - la classe dei rentiers e gli artigiani - sul mercato si fronteggiano due classi: i capitalisti, proprietari dei mezzi di produzione, e i lavoratori salariati, liberi ma proprietari di un'unica merce: la propria forza lavoro. Il sovrappiù (se realizzato) prende la forma di profitto, e questo - il profitto e non più l'uso - diventa lo scopo della produzione. Del sovrappiù il capitalista si appropria in quanto possiede o controlla i mezzi di produzione e dopo aver pagato al suo prezzo la forza lavoro. La destinazione del sovrappiù, d'altra parte, non è più il consumo, bensì l'allargamento della produzione. La moneta diventa essenziale al processo di produzione-riproduzione, poiché la produzione capitalistica non è produzione di merci a mezzo di merci, ma produzione di denaro a mezzo di denaro. E il mercato, infine, pervade tutta la società: tutti i rapporti fra gli uomini passano per il mercato.
Il processo economico acquista ora una sua autonomia: da finalizzato ad altro, diventa fine a se stesso, circolare. E soltanto ora l'economia politica si può costituire in disciplina autonoma e sistematica: in scienza del capitalismo. Alla nozione di capitale come categoria eterna (il capitale come mezzi di produzione prodotti) viene a contrapporsi quella di capitale come categoria propria e fondante di un dato modo di produzione: il modo di produzione capitalistico. Soltanto nel modo capitalistico di produzione la ricchezza prende la forma di capitale, in quanto rapporto che si instaura fra queste due classi: capitalisti e lavoratori salariati. Può allora dirsi 'capitale' qualsiasi proprietà (di denaro, macchine o altre forme di potere) mediante la quale sia possibile comandare lavoro salariato in vista di un profitto realizzabile vendendone il prodotto.
All'interno di queste due grandi categorie (il capitale come insieme di mezzi di produzione prodotti, oppure come nesso sociale) ci sono numerosissime definizioni di 'capitale'. Qui se ne ripercorrerà la storia per grandi epoche: l'epoca classica (che ai fini di questo articolo si farà cominciare con François Quesnay e finire con David Ricardo); Marx; l'epoca neoclassica (che dura fino a oggi ma ha radici anche nella dissoluzione della scuola ricardiana e dunque contemporanee alla critica marxiana); Schumpeter e Keynes; infine i moderni accademici. A questa successione temporale non corrisponde un'evoluzione coerente e progressiva del concetto, il cui significato oscilla periodicamente fra due grandi campi definitori. Nella vasta galleria di personaggi eminenti è infatti possibile tracciare una distinzione: da un lato vi sono coloro che, come gli economisti classici, Marx, e gli eretici contemporanei quali Schumpeter e Keynes, avevano in mente una società divisa in classi; dall'altro coloro che, come Senior, Say, i marginalisti, i neoclassici moderni, hanno in mente una società integrata. Per i primi il

capitale è un aspetto del potere che la classe proprietaria esercita sul lavoro: potere di decisione sull'uso del lavoro, sfruttamento, utilizzazione del sovrappiù. Per i secondi il capitale è un aspetto di decisioni riguardanti cose: quindi le categorie sono quelle dell'astinenza, del risparmio, della produzione di beni strumentali. Le due linee di pensiero, sempre compresenti nella storia delle dottrine economiche, danno necessariamente luogo a definizioni diverse di ogni categoria economica, e quindi anche del termine capitale (da rapporto sociale a beni prodotti e destinati a ulteriore produzione).
Poiché la storia del concetto di capitale rispecchia quella del capitalismo, sarà utile averne in mente una sinossi (come quella tracciata da Eric J. Hobsbawm) e ricordare le date fondamentali nella storia del concetto di capitale: il Tableau économique di François Quesnay è del 1758; la Ricchezza delle nazioni di Adam Smith è del 1776; il Saggio e i Principî di David Ricardo, qui rilevanti, sono del 1815 e del 1817-1821; i Principî di John Stuart Mill sono del 1848; gli scritti di Karl Marx sul capitale vanno dal 1835 al 1883; gli Elementi di economia politica pura di Léon Walras sono del 1874; la Teoria positiva del capitale di Eugen von Böhm-Bawerk è del 1884; la Teoria dello sviluppo economico di Joseph A. Schumpeter è del 1911; il Trattato della moneta di John M. Keynes è del 1930, la Teoria generale del 1936; la Produzione di merci a mezzo di merci di Piero Sraffa è del 1960.

2. L'epoca classica e la critica marxiana

Nella Francia della metà del Settecento l'industria, nel senso moderno del termine, è praticamente assente; le attività manifatturiere e commerciali hanno forma artigianale; l'economia è fondamentalmente un'economia agricola, retta da rapporti di proprietà di tipo feudale (alla corvée si rinuncia nel 1776). Nelle campagne francesi del settentrione il processo lavorativo comincia però a prendere forma capitalistica sia per l'uso che vi si fa dei mezzi di produzione, sia - e soprattutto - perché il lavoro è diventato lavoro salariato.
La rivoluzione del 1789 e la rivoluzione industriale sono ancora lontane; tuttavia i fisiocratici, per dirla con Marx, riescono a scoprire l'essenza borghese nascosta in un involucro feudale. Nel sistema fisiocratico "il feudalesimo viene riprodotto e spiegato sub specie della produzione borghese e l'agricoltura come il settore produttivo in cui si manifesta esclusivamente la produzione capitalistica, cioè la produzione del plusvalore. Così, mentre il feudalesimo si imborghesisce, la società borghese assume un aspetto feudale" (v. Marx,

1867-1894).

a) Il capitale come anticipazioni: François Quesnay

François Quesnay è il massimo esponente della scuola fisiocratica, e per quanto riguarda la teoria del capitale gli si devono intuizioni e analisi ancora oggi insuperate: la visione del processo produttivo come processo circolare, come processo di produzione-riproduzione (il Tableau économique); l'essenzialità, ai fini dell'analisi di questo processo, del concetto di classi sociali (sia pure nella loro forma precapitalistica: la classe produttiva, o degli agricoltori, la classe dei proprietari fondiari, la classe sterile o dei manifatturieri); infine il concetto di capitale non solo come insieme di beni capitali, ma come anticipazioni.

Per Quesnay le anticipazioni sono essenziali ai fini della produzione di un sovrappiù. Soltanto il capitale investito (in agricoltura) fa sì che il processo produttivo dia luogo a un produit net, e l'origine di questo va ricercata non nella sfera della circolazione, che pure è essenziale alla riproduzione del sistema, ma nella sfera della produzione. La produzione non può essere ridotta a un rapporto immediato fra uomo e natura; il processo di produzione di merci non può essere descritto semplicemente in termini di una combinazione di lavoro e terra, poiché per produrre merci occorre impiegare come mezzi di produzione anche quantità appropriate di queste stesse merci: l'ammontare del sovrappiù dipende in modo cruciale dall'entità e dalla composizione delle anticipazioni.

Quesnay distingue tre tipi di anticipazioni (fondando così la moderna distinzione fra capitale fisso e capitale circolante): primitive, fondiarie e annuali. Sono anticipazioni primitive il bestiame, gli edifici, gli attrezzi; anticipazioni fondiarie le opere idrauliche e di recinzione e in generale le opere di miglioramento permanente dei fondi; anticipazioni annuali i salari dei lavoratori, le sementi e tutte le altre spese annuali ricorrenti.

Anticipazioni primitive e anticipazioni fondiarie hanno la natura del capitale fisso e partecipano alla produzione del prodotto netto cedendo il loro valore nel corso di più periodi di produzione (e richiedendo quindi adeguati reintegri per ciascun periodo di produzione). Le anticipazioni annuali dei ricchi fittavoli hanno invece la natura di capitale circolante, in quanto trasferiscono il loro valore nei beni prodotti nel corso di un solo periodo di produzione. Tutte e tre queste forme di anticipazioni, nella loro composizione e nel loro ammontare, sono essenziali nella determinazione dell'ammontare e della composizione del sovrappiù, che si forma bensì soltanto nell'agricoltura ma che - una volta distribuito e speso -

sostiene tutti i ceti e tutte le professioni. Così il prodotto netto, reso possibile da un dato ammontare di anticipazioni annuali, dipende non soltanto da questo, ma anche dall'ammontare delle anticipazioni primitive e fondiarie. L'accumulazione di capitale in agricoltura è essenziale affinché crescano il prodotto lordo e la parte di questo costituita dal prodotto netto; in particolare occorre che le anticipazioni siano di natura e ammontare tali da consentire la 'grande coltura', poiché soltanto così sarà possibile l'introduzione di metodi sempre più produttivi (di sovrappiù). Una somma di denaro diventa capitale soltanto se si trasforma in capitale produttivo, e quindi procura un prodotto netto per l'intera società.

b) Adam Smith: il lavoro diviso

Nell'Inghilterra di Adam Smith il capitale ha ormai pervaso tutta l'attività produttiva. Ciò consente a Smith di fare un importante passo avanti rispetto ai fisiocratici e di riconoscere che in tutte le attività è possibile produrre sovrappiù, e che il sovrappiù prende non solo e non tanto la forma di rendita (per i fisiocratici il produit net prende per l'appunto la forma di rendita pagata dalla classe produttiva alla classe dei proprietari), ma anche e soprattutto quella di profitto. D'altra parte, come noterà Marx, Smith compie un passo indietro trascurando, nella determinazione quantitativa del sovrappiù, la quota del prodotto sociale necessaria al reintegro dei mezzi di produzione consumati nel processo produttivo: sarebbe capitale soltanto quello speso in salari e sarebbe vero il 'dogma veramente fantastico' per il quale il prezzo delle merci è composto di salario, profitto (interesse) e rendita fondiaria, quindi soltanto di salario e sovrappiù. Errore analogo compirà David Ricardo, e questo - come si vedrà a proposito di Marx - ha conseguenze importanti sulla definizione e l'analisi del saggio dei profitti.

In Smith la nozione di capitale è strettamente intrecciata con le sue distinzioni fra lavoro produttivo e lavoro improduttivo. Per Smith è volta a volta lavoro produttivo quello che si fissa in una merce vendibile, quello che crea valore e quello che produce un sovrappiù. La nozione qui rilevante è la terza, quella per la quale è lavoro produttivo (di sovrappiù) il lavoro che si scambia contro capitale, mentre è lavoro improduttivo quello pagato dal reddito dei capitalisti e dei rentiers. Simmetricamente si potrà dire che è capitale quella parte del fondo posseduto da un individuo che viene impiegata per mettere in attività il lavoro produttivo, e che quindi non solo consentirà di reintegrare le spese inizialmente sostenute, ma darà altresì luogo a un reddito. L'altra parte del fondo, in quanto destinata al consumo immediato, non

darà invece alcun reddito. Soltanto con l'impiego di lavoro produttivo sarà possibile un processo di accumulazione di capitale.
Il capitale, d'altra parte, può prendere la forma di capitale fisso e di capitale circolante. Come quasi tutte le distinzioni smithiane (quella fra lavoro produttivo e lavoro improduttivo, ma soprattutto quelle interne alla teoria smithiana del valore), anche questa è confusa, in quanto gli elementi del capitale produttivo vengono distinti a seconda del modo in cui compaiono nella sfera della circolazione, anziché in quella della produzione. La distinzione fra capitale fisso e capitale circolante in Smith si regge letteralmente sul fatto che il capitale circoli oppure no, che generi un profitto restando presso il capitalista, dentro al processo produttivo, oppure distaccandosene, circolando con le merci prodotte. Per Smith, ad esempio, sono capitale fisso i fondi impiegati nel miglioramento delle terre coltivate, le costruzioni destinate alla locazione, gli strumenti di produzione di qualsiasi genere, il bestiame da lavoro e le stesse "abilità acquisite e utili" dei lavoratori; mentre sono capitale circolante la moneta, le merci prodotte per la vendita, le materie prime impiegate in un processo produttivo o nel mantenimento del bestiame da lavoro e i salari dei lavoratori. (Contare come capitale fisso le abilità acquisite e utili dei lavoratori e come capitale circolante i loro salari evoca la difficile questione delle relazioni fra forze produttive e rapporti di produzione, difficile soprattutto per chi ha una concezione strettamente materiale del capitale produttivo. È difficile far coesistere l'idea che nel capitale vadano contate quelle che vengono chiamate utilità 'immateriali' e l'idea secondo la quale tutto il capitale è un prodotto del lavoro, e dunque se ne debbano escludere dal computo le forze e i doni della natura: è molto difficile distinguere e contabilizzare quali siano i doni della natura e quali i risultati del lavoro).
Per quanto riguarda l'origine del sovrappiù e in particolare del profitto, Smith la individua nella produttività del lavoro in generale. A sua volta, la produttività del lavoro dipende dalla divisione del lavoro, e questa dalla tendenza propria della natura umana al baratto e allo scambio: "E poiché è in questo modo, col baratto e con lo scambio, che noi otteniamo la maggior parte di quei reciproci buoni uffici di cui abbiamo bisogno, così è questa stessa tendenza a trafficare che in origine dà occasione al sorgere di quella divisione del lavoro sulla quale si fonda tutto il benessere delle società evolute" (v. Smith, 1776). Nella produzione capitalistica, in particolare, il capitale, riunendo insieme un gran numero di lavoratori e anticipando loro le sussistenze di cui essi non dispongono, può attuare la migliore divisione e distribuzione degli impieghi e può fornire agli operai le migliori macchine: la forma

capitalistica di produzione è destinata a diventare la forma dominante e definitiva dell'economia e della società. La storia sarebbe dunque finita.
La società borghese, secondo Smith, si regge sulla classe dei lavoratori produttivi; questi - producendo prodotto netto - sostengono se stessi e tutte le altre classi; i padroni, che avendo anticipato le sussistenze si appropriano del prodotto netto, ne trattengono per sé una parte come profitto, destinandola elettivamente all'accumulazione del capitale, e ne redistribuiscono l'altra parte ai proprietari fondiari e ai lavoratori improduttivi. Si manifesta così la duplicità del rapporto fra capitale e lavoro salariato: in quanto lavoro produttivo di sovrappiù, il lavoro produce il capitale, ma in quanto lavoro salariato viene comandato dal capitale. Come Smith scrive nell'Abbozzo della Ricchezza delle nazioni, "In un paese civile i poveri provvedono a se stessi e all'enorme lusso dei loro signori [...]. Tra i selvaggi, invece, ognuno gode dell'intero prodotto della propria attività" (v. Smith, 1763).
Occorre ricordare qui un problema che Smith, a differenza dei fisiocratici, si trova a dover affrontare: il problema di che cosa determini il valore delle merci. Poiché per i fisiocratici il prodotto netto si dava soltanto in agricoltura, di esso si poteva dare una determinazione quantitativa in termini fisici, come semplice differenza fra le quantità di beni prodotte e le quantità di beni impiegate come mezzi di produzione e mezzi di sussistenza; ma quando può darsi sovrappiù in tutte le attività produttive, come è per Smith, l'unica determinazione quantitativa logicamente possibile diventa quella in valore. Nasce di qui la centralità della teoria del valore nell'analisi dell'economia capitalistica.

c) David Ricardo: saggio dei profitti, macchine e accumulazione
Anche David Ricardo, come Smith, concepisce il capitale nella sua materialità, piuttosto che come rapporto sociale, e come Smith lo riduce alle anticipazioni salariali. Per Ricardo il problema principale dell'economia politica non è però l'indagine sulle cause della ricchezza delle nazioni, bensì la determinazione delle leggi che regolano la distribuzione del prodotto sociale, al netto della rendita, fra capitalisti e lavoratori. Vi è dunque almeno l'intuizione di un rapporto fondamentale fra capitale e lavoro salariato.
Nell'ipotesi che il capitale consista soltanto nelle anticipazioni salariali, il saggio dei profitti (che per definizione è uguale al rapporto fra l'ammontare dei profitti e il valore del capitale) risulta pari al rapporto fra profitti e salari, ed è dunque la misura ricercata della distribuzione del prodotto sociale netto di rendita fra capitalisti e lavoratori. Si tratterà perciò di individuare

le determinanti del saggio dei profitti, il modo in cui il saggio dei profitti si muove con il procedere dell'accumulazione del capitale, e se e come l'introduzione delle macchine nel processo produttivo abbia influenza sugli interessi delle diverse classi della società ("argomento questo di grande importanza che sembra non sia stato mai esaminato in modo da condurre a risultati certi e soddisfacenti").

Ricardo non dà una spiegazione dell'origine del profitto, poiché lo concepisce come un residuo: come quel che resta nelle mani dei capitalisti, una volta pagati rendita e salari; tuttavia la sua teoria del saggio dei profitti mostra che il rapporto fra capitalisti e lavoratori salariati è un rapporto conflittuale: è questa un'importante implicazione politica dell'analisi ricardiana delle relazioni fra capitale sociale e processo lavorativo sociale, quali si danno nel modo di produzione capitalistico.

Il saggio dei profitti, secondo Ricardo, dipende da due ordini di circostanze: dalle condizioni tecniche della produzione e dal saggio di salario. Date le condizioni tecniche della produzione (e in ogni dato momento esse sono date), saggio dei profitti e saggio di salario stanno fra di loro in una relazione inversa: a un alto saggio di salario corrisponde un basso saggio dei profitti e viceversa. Inoltre, e a differenza di quanto sosterrà la teoria neoclassica nella sua versione egemone, non è vero che a una data configurazione delle tecniche di produzione corrisponda una e una sola configurazione distributiva di equilibrio: in astratto è compatibile con una configurazione data delle tecniche di produzione qualsiasi configurazione distributiva compresa fra i due estremi in cui tutto il prodotto netto (di rendita) va ai salari, o in cui tutto il prodotto netto va ai profitti (è importante sottolineare che ciò è vero in astratto: nella realtà non è vero che non esista un vincolo distributivo alla determinazione del livello di attività o al processo di accumulazione del capitale, come mostrerà Marx e come è stato dimostrato dagli esiti politici delle tesi neoricardiane, fatte proprie da una parte del sindacato italiano negli anni settanta, circa il salario come 'variabile indipendente').

Il prodotto sociale che si spartiscono capitalisti e lavoratori è il prodotto sociale al netto della rendita. Sul livello del saggio dei profitti, e sulla sua dinamica nel corso del processo di accumulazione, hanno dunque influenza le determinanti della rendita stessa. Questa, secondo Ricardo (che ne mutua la spiegazione da Malthus), dipende dal fatto che il processo produttivo è caratterizzato da rendimenti decrescenti; se si suppone, come Ricardo fa nella versione più semplice della sua teoria, che l'unica attività produttiva sia quella agricola e che

le diverse terre abbiano diversa fertilità, allora la concorrenza fra capitalisti da un lato e fra proprietari terrieri dall'altro farà sì che la rendita risulti pari alla differenza fra il prodotto effettivamente ottenuto e quello che si sarebbe ottenuto se tutte le terre fossero state di fertilità pari a quella della terra meno fertile fra quelle messe a coltura. Su quest'ultima la rendita è nulla, e il suo prodotto (il prodotto marginale) basterà appena a pagare i salari e i profitti, a un saggio (dei profitti) che per effetto della concorrenza dovrà essere uniforme su tutte le terre. Il saggio di salario, d'altra parte, dovrà anch'esso essere uniforme su tutte le terre, e sarà mantenuto al livello di sussistenza dall'operare di un meccanismo demografico di tipo malthusiano.
Per Ricardo (che almeno in questo senso aderisce alla legge di Say, secondo la quale tutti i redditi sono integralmente spesi) se vi sono profitti positivi, questi saranno investiti dai capitalisti nella coltivazione di nuove terre, che avranno una produttività via via decrescente. I profitti saranno schiacciati fra rendita e salari (vi è dunque conflitto non soltanto fra capitalisti e lavoratori salariati, ma anche fra queste due classi e quella dei rentiers); il saggio dei profitti diminuirà, e il sistema - prima o poi - raggiungerà lo stato stazionario.
Il rapporto (distributivo) fra capitale e lavoro salariato non è però istituito, per Ricardo, soltanto dalle leggi 'naturali' che governano la relazione inversa fra saggio di salario e saggio dei profitti e la caduta di questo nel corso del processo di accumulazione (non tendenziale, ma necessaria), bensì anche dall'eventuale introduzione di macchine nel processo produttivo: che non è un fatto di natura, ma è il risultato di decisioni dei capitalisti.
Per Ricardo, e contro l'opinione allora e oggi dominante, l'introduzione di macchine nel processo produttivo riesce spesso dannosa agli interessi della classe dei lavoratori. L'occupazione aumenta sicuramente soltanto quando aumentano le anticipazioni salariali e il salario è al suo livello 'naturale', di sussistenza; mentre l'introduzione delle macchine può far sì che aumenti il reddito netto della società (rendite e profitti) e contemporaneamente diminuisca il reddito lordo (e dunque l'ammontare dei salari). A ciò consegue, secondo Ricardo, che la stessa causa che può aumentare il reddito netto del paese può nello stesso tempo rendere esuberante la popolazione e peggiorare le condizioni dei lavoratori.

d) John Stuart Mill: il capitale come fattore della produzione

Il tentativo di John Stuart Mill di emendare, completare e sistemare le teorie ricardiane prelude in realtà allo stravolgimento che del pensiero classico compirà la scuola neoclassica.

Ricardo non dà una spiegazione dell'origine del profitto, che concepisce come grandezza residuale; egli nega, in altre parole, che il capitale materiale abbia in sé un proprio potere produttivo. Mill si pone invece il problema di spiegare - ma principalmente di giustificare - i redditi da profitto come generati da un siffatto potere produttivo, anticipando così una concezione del capitale come di un elemento materiale che si combina con lavoro e terra nella produzione di ricchezza: una concezione del capitale, dunque, come 'fattore' della produzione. Su ciò si tornerà più avanti, ma vanno indicati almeno alcuni aspetti della concezione milliana del capitale perché rappresentativi di indirizzi di pensiero presenti anche in altri autori dell'epoca (J.B. Say, J. M. Lauderdale, J. R. McCulloch).
La teoria milliana del capitale - definito come "un fondo accumulato dei prodotti del lavoro precedente" - è riassunta dallo stesso Mill in quattro "proposizioni fondamentali":
1) l'industria è limitata dal capitale (l'occupazione può essere aumentata soltanto mediante l'accumulazione di capitale, cioè sarebbero impossibili problemi malthusiani di insufficienza della domanda);
2) il capitale è il risultato del risparmio;
3) il capitale, sebbene sia il risultato del risparmio, è tuttavia consumato (vi è dunque corrispondenza, come vuole la legge di Say, fra risparmio e spese);
4) la domanda di merci non è domanda di lavoro. (Quest'ultima e discussa proposizione probabilmente significa che la domanda di merci non è necessariamente domanda di lavoro, poiché la decisione di ricostituire il fondo salari mediante il ricavato dalla vendita delle merci prodotte spetta al capitalista. A questa proposizione si lega il quesito milliano, se la distinzione fra ciò che è capitale e ciò che non lo è non dipenda dalle intenzioni del capitalista, ovvero del proprietario del capitale potenziale).
Mill non afferma esplicitamente il potere produttivo del capitale, e anzi scrive che "il capitale, rigorosamente parlando, non ha alcun potere produttivo: l'unico potere produttivo è quello del lavoro"; ma aggiunge subito "assistito indubbiamente da utensili e operante sulle materie prime. Si può forse dire, senza grande improprietà, che la parte di capitale che consiste di utensili e di materie prime possiede un potere produttivo, poiché essi contribuiscono, insieme col lavoro, all'espletamento della produzione. [...] Il concetto appropriato di capitale è che tutto quanto una persona possiede costituisce il suo capitale, purché questa persona possa e voglia impiegarlo non nel consumo a scopo di soddisfazione, ma per procurarsi i mezzi di produzione con l'intenzione di impiegarli produttivamente. Ora i

mezzi di produzione sono lavoro, strumenti e materie prime. L'unico potere produttivo che esiste è il potere produttivo del lavoro, degli strumenti e delle materie prime" (v. Mill, 1844).

e) Karl Marx: la critica della formula trinitaria

Karl Marx intitola proprio alla "formula trinitaria" il primo capitolo dell'ultima sezione del terzo volume del Capitale (dedicata a I redditi e loro fonti; l'ultimo capitolo, là dove il manoscritto si interrompe, ha per titolo Le classi).

Scrive Marx: "Questa formula trinitaria si riduce più precisamente alla seguente: capitale-interesse, terra-rendita fondiaria, lavoro-salario, nella quale il profitto, la forma del plusvalore che caratterizza specificamente il modo di produzione capitalistico, è felicemente eliminato. Le pretese fonti della ricchezza annualmente disponibile appartengono a sfere completamente diverse e non vi è fra di esse la più piccola analogia, come non vi è analogia fra gli onorari di un notaio, le carote rosse, la musica. Capitale, terra, lavoro! Ma il capitale non è una cosa, bensì un determinato rapporto di produzione sociale, appartenente a una determinata formazione storica della società. Il capitale è costituito dai mezzi di produzione monopolizzati da una parte determinata della società, dai prodotti e dalle condizioni di attività della forza-lavoro, resi autonomi nei confronti della forza-lavoro vivente, che vengono mediante questa contrapposizione personificati nel capitale. [...] Viene poi la terra, la natura inorganica come tale, rudis indigestaque moles in tutta la sua selvaggia primitività. [...] E infine, come terzo in questa alleanza, un semplice fantasma, 'il' lavoro, che non è altro che un'astrazione, l'attività produttiva dell'uomo in generale, per mezzo della quale egli rende possibile il ricambio organico con la natura. [...] L'economia volgare non fa altro, in realtà, che interpretare, sistemare e difendere le idee di coloro che, impigliati nei rapporti di produzione borghesi, sono gli agenti di questa produzione. Non ci dobbiamo quindi meravigliare che l'economia volgare si senta particolarmente a suo agio proprio in questa forma fenomenica estraniata dai rapporti economici, in cui questi prima facie sono assurdi e del tutto contraddittori - e ogni scienza sarebbe superflua, se l'essenza delle cose e la loro forma fenomenica direttamente coincidessero - e che questi rapporti le appaiano tanto più evidenti di per sé quanto più le rimane nascosto il loro nesso interno".

f) Marx: lavoro salariato e capitale

Per Marx il processo di produzione capitalistico è una forma storicamente determinata del

processo di produzione sociale in generale. Quest'ultimo è al tempo stesso il processo di produzione delle condizioni materiali della vita umana e un processo che si sviluppa entro specifici rapporti di produzione storico-economici, producendo e riproducendo questi rapporti stessi di produzione e dunque i rappresentanti di questo processo, le loro condizioni materiali di esistenza e i loro rapporti reciproci, ossia la loro determinata forma economica e sociale. Il complesso di questi rapporti, in cui i rappresentanti di questa produzione stanno con la natura e fra di loro, costituisce precisamente la società, considerata nella sua struttura economica.

La caratteristica principale del modo di produzione capitalistico, se lo si guarda dal punto di vista della circolazione, è che il processo è del tipo Denaro-Merce-Denaro, e non Merce-Denaro-Merce. Ciò vuol dire che mentre nel ciclo M-D-M lo scopo dello scambio è quello di ottenere una merce finale atta a soddisfare bisogni diversi da quelli che possono essere soddisfatti con la merce posseduta e ceduta inizialmente (scopo dello scambio è il valore d'uso e la moneta serve soltanto all'intermediazione nello scambio delle merci), nel ciclo D-M-D si cede denaro per ottenere altro denaro; questo denaro, d'altra parte, non viene mai speso definitivamente, ma rifluisce sempre al punto di partenza. Scopo di questo processo non è ottenere valori d'uso, bensì un plusvalore; la somma ottenuta, perché l'operazione abbia un senso, dovrà essere maggiore di quella ceduta inizialmente: la forma effettiva del ciclo dovrà dunque essere D-M-D´, dove D´ sarà maggiore di D e la differenza rispetto al valore originario sarà costituita dal plusvalore. Nella circolazione, dice Marx, il valore anticipato originariamente si valorizza, e questo movimento lo trasforma in capitale.

Si tratta ora di spiegare come mai D´ possa (e debba) essere maggiore di D. La spiegazione marxiana è la seguente. Come si è visto, il processo si apre con uno scambio di potere d'acquisto contro una merce: il potere d'acquisto originario si trasforma in capitale proprio in quanto assume la forma intermedia di merce. Tuttavia non esiste nessuna risorsa o merce (salvo una, come si vedrà) che allo stesso tempo abbia valore d'uso e sia fonte di valore, così come occorre affinché D´ sia maggiore di D. Diventa dunque necessario indagare come nel modo di produzione capitalistico si svolga il processo di produzione (e riproduzione). L'idea marxiana è che l'unica merce che insieme abbia valore d'uso e funzioni come fonte di valore sia la forza lavoro. Di tale merce, che costituisce l'unica proprietà del lavoratore, il lavoratore stesso non può fare uso, poiché non possiede i mezzi di produzione; può soltanto venderla a chi - il capitalista - possiede potere d'acquisto da trasformare in capitale, la immette e utilizza

nel processo produttivo e ne trae il plusvalore che si realizzerà (se si realizzerà) nella forma di profitto con l'ulteriore trasformazione della merce prodotta in denaro.
Se le cose stanno così, non è più possibile pensare il capitale soltanto come entità materiale e come categoria distinta dal lavoro, come semplice insieme di mezzi di produzione che vengono combinati con il lavoro per produrre valori d'uso. Nel modo di produzione capitalistico il lavoratore è lavoratore salariato, e il rapporto materiale fra strumenti di lavoro e capacità lavorativa, quale si dà nel processo lavorativo, sottende un nesso sociale fra capitalista e lavoratore che condiziona tutto il processo di valorizzazione e di riproduzione.
Scrive Marx (in Lavoro salariato e capitale): "Anche il capitale è un rapporto sociale di produzione. Esso è un rapporto borghese di produzione. Il capitale non è dunque soltanto una somma di prodotti materiali; esso è una somma di merci, di valori di scambio, di grandezze sociali [...]. Come dunque una somma di merci, di valori di scambio, diventa capitale? Per il fatto che essa, come forza sociale indipendente, cioè come forza di una parte della società, si conserva e si accresce attraverso lo scambio con il lavoro vivente, immediato. L'esistenza di una classe che non possiede null'altro che la capacità di lavorare è una premessa necessaria del capitale" (v. Marx, 1849).

g) Marx: il processo di produzione-riproduzione

Che cosa avviene nello scambio fra capitale e lavoro salariato? Per Marx il valore di una merce si scinde in tre parti: capitale costante, capitale variabile, plusvalore. Il processo produttivo si apre con la spesa da parte del capitalista del suo capitale monetario nell'acquisto dei mezzi di produzione e della forza lavoro (la quale costituisce il capitale produttivo). La forza lavoro, come qualsiasi altra merce, viene pagata secondo il suo valore, che è pari al tempo di lavoro necessario per riprodurla (cioè per produrre i mezzi di sussistenza del lavoratore).
Al termine del processo di produzione il capitale produttivo è trasformato in capitale merce, in merci che hanno un valore superiore a quello del capitale produttivo iniziale. La parte di capitale monetario spesa nei mezzi di produzione (capitale costante) non cambia la sua grandezza di valore, mentre la parte spesa in forza lavoro (capitale variabile) ha aumentato il suo valore, producendo il plusvalore che viene trattenuto dal capitalista e che, una volta realizzato, può essere trasformato in nuovo capitale produttivo. Il lavoratore salariato si troverà così di fronte il valore che egli stesso ha prodotto; egli vende la propria forza lavoro

per produrre ciò che gli si contrapporrà come proprietà del capitalista. Da un lato la forza lavoro, in quanto produce plusvalore, è all'origine del profitto; dall'altro il lavoro, in quanto lavoro salariato, è incluso nel capitale.
Ciascun singolo capitale, tuttavia, può riprodursi soltanto se la merce prodotta viene venduta e il ricavato viene riconvertito in nuovo capitale produttivo. La realizzazione della merce può però essere ostacolata da tre ordini di circostanze: il bisogno che la collettività ha della merce stessa; la quantità di moneta esistente; l'effettiva trasformazione in denaro dell'intera quantità prodotta. Gli schemi di riproduzione di Marx, che riprendono l'idea del Tableau économique di Quesnay, hanno lo scopo di individuare le condizioni necessarie affinché il processo di produzione possa ripetersi, riprodursi, ed è proprio nel processo di riproduzione che può manifestarsi l'equilibrio - oppure la crisi - del capitale.
Se si suddivide il sistema economico nei due settori fondamentali (quello che produce mezzi di produzione e quello che produce merci salario), allora dalla condizione di uguaglianza fra domanda e offerta che è necessaria per ciascun settore risulterà che la domanda di mezzi di produzione da parte del settore che produce merci salario deve essere uguale alla domanda di merci salario da parte del settore che produce mezzi di produzione. In assenza di un piano, è chiaro non soltanto che il verificarsi spontaneo di questa condizione è improbabile, ma anche che, se quella uguaglianza per caso si verifica, nulla assicura che essa comporti il massimo livello di attività e di occupazione del sistema; sono cruciali, comunque, nella determinazione del livello, delle proporzioni e della regolarità e stabilità delle modalità di riproduzione del sistema, le decisioni dei capitalisti circa l'impiego dei profitti realizzati.
Da questo punto di vista, cioè dal punto di vista della riproduzione e della circolazione del capitale complessivo, Marx dimostra che l'equilibrio reale, nello scambio delle diverse parti del prodotto annuo, non necessariamente esiste, né è necessariamente unico, stabile e ottimo per tutti, come invece vogliono gli economisti ortodossi. In particolare, la funzione del denaro quale capitale monetario produce determinate condizioni dello svolgimento normale della riproduzione (sia su scala semplice sia su scala allargata), che si trasformano in altrettante condizioni di svolgimento anormale della produzione, in possibilità di crisi: "poiché l'equilibrio stesso - dato il carattere primitivo di questa produzione - è un caso".

h) Marx: plusvalore e profitto

Per Marx la soluzione del problema lasciato irrisolto da Ricardo - quale sia l'origine del

profitto - è dunque questa: all'origine del profitto sta il plusvalore, e all'origine di questo sta il pluslavoro prestato nella fabbrica dal lavoratore, dopo che questi aveva venduto sul mercato, e al suo prezzo, la propria forza lavoro. Lo scambio che ha per oggetto la merce forza lavoro è uno scambio fra equivalenti nella sfera della circolazione, mentre è uno scambio fra non equivalenti se si considera il processo capitalistico complessivo, che è allo stesso tempo processo di circolazione, produzione e riproduzione.
La giornata lavorativa si divide in due parti: una prima parte serve a ricostituire i beni di consumo necessari alla riproduzione della forza lavoro (lavoro necessario), la seconda (pluslavoro) costituisce il plusvalore, quella parte del valore del prodotto che non ritorna al lavoratore salariato e che anzi gli si contrapporrà come nuovo capitale. Conviene aggiungere che un rapporto di sfruttamento non si dà soltanto nel modo di produzione capitalistico, ma soltanto in questo esso è mediato, e celato, dallo scambio. Il profitto è il risultato della forma capitalistica del rapporto di sfruttamento.
Poiché il valore di ogni merce prodotta capitalisticamente si scinde per Marx in capitale costante (c), capitale variabile (v) e plusvalore (s), la grandezza alla quale vanno commisurati i profitti ai fini della determinazione del saggio del profitto sarà il valore dell'intero capitale speso negli elementi della produzione; cioè la somma del capitale costante e del capitale variabile, e non soltanto l'ammontare dei salari anticipati, come si ha in Ricardo. Se il plusvalore si trasforma in profitto (se cioè la merce prodotta viene realizzata), il saggio del profitto sarà dato da $s/(c + v)$, ovvero, se si dividono numeratore e denominatore per v, dal rapporto fra s/v (che si può definire saggio del plusvalore o di sfruttamento) e $(c/v + 1)$ (dove il rapporto c/v può essere definito 'composizione organica del capitale').
Mentre il saggio del plusvalore rende manifesta l'origine del plusvalore stesso, che è il capitale variabile, nel saggio del profitto tale origine viene occultata poiché, perdendosi la distinzione fra capitale costante e capitale variabile, il profitto apparirà come generato da 'qualità segrete' del capitale nel suo complesso (e non soltanto da quella sua parte che per Marx ha capacità di valorizzazione). Il capitalista anticipa il capitale complessivo senza riguardo alle diverse funzioni assolte nella produzione del plusvalore dalle singole parti costitutive del capitale, e dunque ai suoi occhi il profitto è originato dal capitale in sé, mentre in realtà esso non è altro che una forma mistificata del plusvalore, "la forma fenomenica del plusvalore". (Si pone qui il cosiddetto 'problema della trasformazione', che ci limitiamo a menzionare. Mentre in Ricardo valore e prezzo coincidono immediatamente, per Marx valore

e prezzo sono due categorie distinte: i valori si 'trasformano' in prezzi di produzione per effetto della concorrenza, in maniera tale da assicurare ai vari capitali l'eguaglianza dei saggi di profitto. I prezzi di produzione dovrebbero dunque essere determinati a partire dai valori. Sulla trasformazione in prezzi dei valori misurati in termini di lavoro incorporato è bene avvertire che l'analisi elaborata da Piero Sraffa (v., 1960) in Produzione di merci a mezzo di merci, ha messo in evidenza che, nel procedere alla trasformazione, si incontrano difficoltà gravi, secondo alcuni economisti insuperabili, secondo altri non decisive).
Un cenno, infine, agli effetti dell'accumulazione del capitale sul saggio del profitto. Si è visto come per Ricardo il saggio del profitto dipenda dalle condizioni tecniche della produzione e dal saggio di salario, e che la messa a coltura di terre via via meno fertili, dato il saggio di salario, comporta necessariamente e naturalmente una caduta del saggio dei profitti. Nella definizione marxiana (e nella contabilità capitalistica) del saggio del profitto, questo dipende invece dal rapporto fra il saggio del plusvalore e la composizione organica del capitale; cioè dalla divisione della giornata lavorativa (che è una misura della distribuzione del prodotto fra capitalisti e lavoratori) e dalle condizioni tecniche della produzione (descritte in termini di composizione organica del capitale). Ciò significa che, a parità del saggio di plusvalore (ovvero di sfruttamento), l'andamento del saggio del profitto dipenderà da quello della composizione organica del capitale. Questa, a sua volta, dipende dalle decisioni dei capitalisti circa le tecniche di produzione e circa il rapporto fra capitale costante e capitale variabile, che in generale tenderà ad aumentare, non tanto in dipendenza di un 'progresso tecnico' ineluttabile, il quale deterministicamente conduca a sostituire macchine a lavoro vivo, ma in quanto in tal modo - generando disoccupazione (esercito industriale di riserva) - si manterrà il salario al suo livello di sussistenza. Il saggio dei profitti, dunque, tenderà a diminuire, a causa delle decisioni stesse dei capitalisti; anche se alla legge della caduta tendenziale del saggio del profitto potranno opporsi, quali contraddizioni intrinseche della legge, delle cause antagonistiche.

3. L'epoca neoclassica

Nel suo splendido libro su Gli anni dell'alta teoria. Invenzione e tradizione nel pensiero economico, 1926-1939, George L. S. Shackle narra che nei quarant'anni seguenti il 1870 fu fondata una "grande teoria" o "grande sistema della scienza economica", il cui unico scopo

era di dimostrare quali fossero le implicazioni logiche di gusti o bisogni dati, combinati con "conoscenza perfetta" e messi a confronto con "scarsità e versatilità di risorse". Scarsità e versatilità di risorse, unite alla perfetta e generale conoscenza delle soddisfazioni che si possono trarre da ciascun uso di tali risorse attraverso l'intera gamma di possibilità rivelate da un determinato stadio della tecnologia, assicuravano che queste risorse sarebbero state sempre pienamente impiegate. La conoscenza perfetta implicava anche la perfezione dei mercati, così che ogni bene veniva prodotto da un gran numero di imprese di pari dimensioni fra le quali i compratori si muovevano in completa indifferenza. Inoltre la conoscenza perfetta aboliva del tutto la necessità di un qualche mezzo di conservazione del potere generale d'acquisto (distinto dalla ricchezza incorporata in forme concrete capaci di fornire direttamente la soddisfazione del consumo o di intervenire fisicamente nella produzione) e quindi non c'era moneta reale (la cui funzione come riserva di valore è di rendere possibile la posposizione di una decisione circostanziata a chi possiede una conoscenza imperfetta). La teoria evitava di prendere in considerazione lo sviluppo in qualsiasi sua forma: "La Grande Teoria era quindi la teoria dell'equilibrio generale stazionario (o, meglio, atemporale), perfettamente concorrenziale e di piena occupazione" (v. Shackle, 1967).

a) L'equilibrio economico come ordine naturale e necessario

Con l'avvento della scuola neoclassica l'economia politica si trasforma in una disciplina che di 'politico', esplicitamente, ha ben poco; la 'scienza' economica, secondo la definizione che ne darà Lionel Robbins tra il 1932 e il 1935, si riduce all''economica': "scienza che studia la condotta umana come una relazione tra scopi e mezzi scarsi, applicabili a usi alternativi". Questa riduzione (o generalizzazione, a seconda dei punti di vista) nell'ambito della teoria del capitale conduce a concepire il capitale stesso esclusivamente nella sua connotazione materiale: come un insieme di mezzi di produzione prodotti e dotati di produttività.

Già nell'epoca classica erano presenti teorie che concepivano il capitale come 'fattore produttivo'. Nassau Senior, in particolare, riconduce la produttività del capitale alla categoria dell''astinenza', cioè al "comportamento di un soggetto che o si astiene dall'uso improduttivo di ciò di cui dispone o deliberatamente preferisce il conseguimento di risultati remoti anziché quello di risultati immediati". All'origine del profitto starebbe l'astinenza, e nella definizione e nell'analisi del capitale diventa centrale la dimensione temporale del processo produttivo. Un'impostazione siffatta ha una conseguenza di grande portata sulla visione del processo

economico. La conseguenza è che se il lavoro non è l'unica fonte del valore, ma esistono anche altri 'fattori' della produzione, allora viene meno la categoria del sovrappiù, che per l'appunto presuppone che sia il lavoro produttivo (la forza lavoro, per Marx) a produrre di più di quanto occorra alla sua riproduzione, e si afferma la concezione del problema distributivo come retto dall'armonia anziché dal conflitto fra classi antagonistiche. Se si pensa ciascuna quota distributiva come il risultato di un contributo produttivo specifico di ciascun fattore, la distribuzione del prodotto sociale non è più determinata anche da un conflitto di classe, ma soltanto dalle condizioni tecniche della produzione, che sono assunte come date. Il mondo è retto da un'unica, determinata legge economica ed esiste un'unica configurazione di equilibrio, che in assenza di attriti si afferma e si mantiene naturalmente e con il massimo vantaggio per tutti. Il reddito percepito da ciascun soggetto non è altro che il prezzo per i servizi produttivi dei fattori della produzione di cui ciascun soggetto è proprietario.

b) Il capitale nell'equilibrio economico generale

Del capitale la teoria di Léon Walras vuole e può dire soltanto ciò che se ne può dire nel linguaggio matematico. Ciò rende particolarmente difficile esporla mediante il linguaggio comune, ma soprattutto le impedisce di indagare il ruolo che il capitale ha nelle diverse fasi del processo capitalistico di produzione-riproduzione, riducendo la questione al modo in cui il capitale entra (come dato) in un problema di massimizzazione sotto vincoli. Questo, nella visione di Walras, è un problema generale e universale, che va studiato con riferimento a un sistema il cui funzionamento non è condizionato da alcuna determinazione storica o istituzionale.

Il problema che Walras affronta con la sua teoria dell'equilibrio economico generale è il seguente: quali siano le quantità di beni prodotti e scambiati, e i prezzi ai quali avvengono tali scambi, in quella configurazione che vede realizzate simultaneamente le posizioni di equilibrio cui tendono i soggetti economici. I dati del problema, d'altra parte, sono le quantità iniziali di risorse produttive, la tecnica di produzione, il sistema di preferenze dei soggetti. Si suppone, inoltre, che il mercato sia concorrenziale e che i soggetti siano razionali nel senso che tendono a massimizzare la propria soddisfazione.

Lo schema analitico complessivo comprende quattro stadi: una teoria dello scambio, sotto il vincolo dell'eguaglianza fra domanda e offerta dei beni di consumo; una teoria della produzione, sotto il vincolo dell'eguaglianza fra prezzo di vendita dei prodotti e costi dei

servizi produttivi che vi sono entrati; una teoria della capitalizzazione, circa le quantità prodotte dei capitali propriamente detti, sotto il vincolo dell'uniformità del saggio di rendimento dei beni capitali e dell'eguaglianza fra reddito risparmiato e valore dei nuovi beni capitali prodotti; infine una teoria della circolazione che tenga conto della necessità (strumentale) della moneta come capitale circolante.
Per Walras il capitale in generale è costituito da "tutte le forme di ricchezza sociale che non sono consumate affatto o che vengono esaurite soltanto dopo un certo periodo di tempo". Sono invece redditi "tutti i beni non durevoli, tutte le forme di ricchezza sociale che vengono consumate immediatamente". Per intendere il modo in cui il capitale entra nella determinazione dell'equilibrio economico walrasiano, occorre dunque partire dalla nozione walrasiana di ricchezza sociale, che è "l'insieme delle cose materiali o immateriali che sono scarse, cioè che da una parte ci sono utili e dall'altra non sono disponibili che in quantità limitata". Gli elementi della ricchezza sociale sono di due tipi: i capitali, beni che servono più di una volta, e i redditi, beni che servono una sola volta. I capitali, a loro volta, sono di tre tipi: i capitali, o risorse, naturali, i capitali, o capacità, personali, e i capitali propriamente detti, quei beni capitali che non sono né terra né facoltà personali. Quanto ai redditi, essi comprendono i beni per uso di consumo (beni di consumo e servizi consumabili delle tre specie di capitali) e i beni per usi di produzione (beni intermedi e servizi produttivi delle tre specie di capitali).
La principale preoccupazione analitica di Walras riguarda la trasformazione dei servizi produttivi in prodotti. Nel processo economico vi sarebbe dunque un punto di partenza (le dotazioni iniziali di capitali) e un fine: il consumo, da parte dei soggetti, dei beni di consumo, e l'acquisto - in quanto risparmiatori - di nuovi capitali. La successione lineare che conduce dal punto di partenza al traguardo è la seguente: i soggetti proprietari dei capitali (in quantità date) ne vendono i servizi agli imprenditori; questi, comprati tali servizi, si scambiano (nel senso del baratto) i servizi intermedi fra di loro; ora dispongono dei fattori della produzione e li immettono in un processo produttivo, la cui tecnologia è data; l'esito del processo è la vendita dei prodotti, dagli imprenditori ai proprietari dei servizi produttivi, che i primi avevano acquistato dai secondi e che ora i secondi sono in grado di acquistare dai primi.
Per esprimersi in termini moderni, si può dire che nella teoria walrasiana della produzione le merci non sono prodotte (come è invece per gli economisti classici) mediante altre merci prodotte, bensì grazie ai servizi produttivi di beni capitali assunti come dati nello schema. La

funzione dell'imprenditore si riduce a quella di comprare servizi produttivi e vendere beni di consumo, senza alcuna iniziativa o reddito suoi propri; come noterà Schumpeter, "è chiaro che nel pensiero di Walras le famiglie sono realmente gli agenti che, sia come compratori che come venditori di servizi, determinano il processo economico " (v. Schumpeter, 1954).
Per quanto riguarda il valore dei beni capitali occorre distinguere due casi: un'economia in stato stazionario e un'economia in sviluppo. Nella prima non vi sarà un mercato sul quale si determinano tali valori, poiché in uno stato stazionario non vi sarà produzione di nuovi beni capitali, ma soltanto ammortamenti e rinnovi. In un'economia in sviluppo vi sarà invece domanda di nuovi beni capitali e offerta di risparmio. Gli imprenditori domanderanno nuovi beni capitali fino a quando non saranno uguali il rendimento di questi e il prezzo di offerta del risparmio, sotto la condizione che in equilibrio i prezzi dei beni capitali siano proporzionali ai loro rendimenti netti. L'offerta di risparmio, da parte delle famiglie, è spiegata da Walras in termini di utilità. A questo fine - per stabilire una relazione fra beni capitali e utilità - Walras immagina un bene ideale, il "ricavo netto perpetuo", per il quale ciascuna famiglia ha una funzione di domanda (basata sull'utilità marginale che attribuisce al bene stesso). I prezzi dei beni capitali possono ora essere espressi in termini del prezzo di un'unità di ricavo netto perpetuo per unità di tempo, che è il reciproco del saggio di interesse. In equilibrio la domanda complessiva di nuovi beni capitali (che deve essere uguale al risparmio) dovrà ripartirsi fra le industrie che producono questi nuovi beni capitali secondo la condizione di proporzionalità ricordata sopra (in modo che il rendimento netto dei beni capitali sia proporzionale ai loro prezzi). Il fattore di proporzionalità comune ai prezzi di tutti i beni capitali sarà dato dal reciproco del prezzo di un'unità di ricavo netto perpetuo e dunque - in assenza di moneta - coinciderà con il saggio di interesse.
L'edificio walrasiano è senza dubbio molto bello, come scrive Schumpeter (che ne è affascinato e però - tentato da Marx - vuol prenderne le distanze): "Il piano terreno di questa costruzione è la teoria del 'mercato' dei beni di consumo. Al primo piano troviamo la teoria della produzione e il 'mercato' dei servizi produttivi, non separato dal primo mercato ma con esso integrato. Al secondo piano abbiamo il 'mercato' dei beni capitali, analogamente integrato con gli altri due" (ibid.). (E al terzo piano c'è un altro mercato, integrato con gli altri tre, quello del 'capitale circolante' e della moneta).
Sebbene la teoria walrasiana sia spesso invocata come fondamentale, in quanto schema di riferimento indispensabile per una fondazione microeconomica delle proposizioni relative al

funzionamento del sistema economico nel complesso (fondazione che molti presumono necessaria, sebbene con ragioni dubbie), quella stessa teoria sembra a sua volta avere fondamenta discutibili. Infatti la teoria dell'equilibrio economico generale ha dato luogo a una vastissima letteratura, in accordo o critica, sulla sua coerenza formale e il suo realismo; qui ci si limita a riprendere una considerazione di Claudio Napoleoni, per il quale al fondo delle difficoltà analitiche della teoria walrasiana c'è un modo sostanzialmente contraddittorio di concepire il capitale: "Il capitale è una realtà essenzialmente unitaria proprio perché costituisce, nel suo insieme e indipendentemente dai singoli beni che lo costituiscono, il termine di riferimento rispetto al quale si determina il saggio dell'interesse. La contraddizione in cui cade Walras sta allora in ciò, che, da un lato, il capitale è frantumato nelle sue singole componenti (i tanti 'beni capitali') e, dall'altro lato, è riaffermato come unitario nel momento in cui ciò è inevitabile, ossia nel momento in cui si introduce il saggio dell'interesse (comune valore dei saggi di rendimento)" (v. Napoleoni, 1976).

c) La teoria 'austriaca': capitale e tempo

Al centro della teoria austriaca del capitale (il cui campione è Eugen von Böhm-Bawerk, ma in cui confluiscono da varie parti culturali e in tempi diversi i contributi di William S. Jevons - "Il capitale è un fatto di tempo" -, Carl Menger, Knut Wicksell e altri) sta l'idea che il capitale, in quanto fattore della produzione, possa essere misurato in termini di 'periodo di produzione'. Il capitale, per Böhm-Bawerk - "il Marx borghese", secondo Schumpeter -, non è un fattore produttivo originario, quali sono il lavoro e la terra, ma è esso stesso prodotto. I beni di consumo possono essere prodotti anche direttamente, mediante l'impiego dei soli fattori originari; normalmente, tuttavia, essi sono prodotti con metodi indiretti, che comportano l'impiego di beni intermedi, di beni capitali: tali metodi indiretti sono più produttivi, e diventano tanto più produttivi (sia pure con incrementi decrescenti) quanto più lungo è il periodo di produzione. Il problema diventa allora quello di determinare una misura dell'intensità capitalistica delle tecniche di produzione, che a sua volta consenta di determinare il saggio di interesse associato a ciascuna tecnica, nonché le relazioni intercorrenti fra intensità capitalistica delle tecniche e distribuzione del reddito, nell'ipotesi che i soggetti adottino un comportamento massimizzante.

In quanto il capitale non è un fattore produttivo originario, deve esserne possibile, per Böhm-Bawerk, la riduzione ai fattori originari; in particolare, se per semplicità si trascura la terra, il

capitale deve poter essere ridotto al lavoro che nelle epoche precedenti al periodo considerato è stato investito nella produzione dei mezzi di produzione. L'intensità capitalistica delle tecniche di produzione può allora essere misurata in termini di periodo medio di produzione: se si considerano i periodi che intercorrono tra l'epoca in cui è stata prestata ciascuna quantità di lavoro e il momento in cui si rende disponibile il prodotto, il periodo medio di produzione sarà dato dalla media aritmetica dei singoli periodi, ponderati con le rispettive quantità di lavoro. Ciascuna tecnica di produzione ha dunque un suo periodo medio di produzione, che cresce quanto più indiretta, o 'capitalistica', è la tecnica stessa.
Si pone ora il problema dell'interesse. Posto che il capitale non è altro che il lavoro investito nelle epoche precedenti, perché mai il valore del prodotto è maggiore della somma dei valori di tutte le quantità di lavoro direttamente e indirettamente impiegate nella produzione del prodotto stesso, e comprende anche un interesse?
Per Böhm-Bawerk l'esistenza e l'ammontare dell'interesse sono spiegati da tre fattori fondamentali, i drei Gründe: le "diverse circostanze di bisogno e di approvvigionamento" nel presente e nel futuro, la "sottovalutazione del futuro" e la "superiorità tecnica dei beni presenti su quelli futuri" (i beni presenti possono essere investiti oggi e reinvestiti domani, al momento in cui si rendono disponibili; mentre i beni disponibili domani possono essere investiti soltanto domani). I primi due fattori (o 'regole fattuali', ma in realtà convenzioni) presiedono alle decisioni di consumo: la natura umana è tale che i soggetti preferiscono i beni presenti a beni futuri, così che i beni presenti possiedono un'utilità e perciò un valore superiore rispetto a quelli futuri. La speranza induce a sopravvalutare le risorse future, la mancanza di immaginazione e la debolezza della volontà determinano una sottovalutazione dei bisogni futuri: presi insieme, questi due fattori rendono l'utilità marginale dei beni presenti maggiore di quella dei beni futuri. Per provocare un'offerta di beni presenti in cambio di beni futuri occorre pagare un aggio, un interesse. L'altro fattore, il terzo, spiega l'esistenza di un prezzo di domanda dei beni presenti in termini di beni futuri, spiega cioè come mai l'utilizzatore di capitale possa pagare quell'interesse. Questo fattore è di carattere tecnico e consisterebbe nella maggior produttività dei metodi di produzione indiretti: il procedere della civiltà, nel campo tecnico, consiste precisamente nell'adozione di metodi di produzione più 'indiretti', nell''allungamento' dei processi produttivi. Tutto viene così ricondotto alla 'natura', anziché alla struttura storicamente determinata della società: è la natura dell'uomo, e della società, che spiega come l'interesse debba e possa essere pagato.

Esso deve essere pagato perché corrisponde a una rinuncia alla disponibilità di ricchezze presenti in vista di una disponibilità futura, e può essere pagato grazie alla maggiore 'produttività' dei metodi 'indiretti', 'capitalistici', di produzione.
Molte obiezioni possono essere e saranno mosse alla costruzione di Böhm-Bawerk (curiosamente analoghe, sul piano logico, a quelle con le quali egli pretese di liquidare il sistema marxiano), ma nella vulgata dei manuali essa continua a imperare.

d) Irving Fisher: impazienza e volontà

Al di fuori delle ipotesi ad hoc sulle quali si regge, la teoria austriaca del capitale (nella formulazione di Böhm-Bawerk) presenta difficoltà insormontabili per quanto riguarda la misurazione del capitale stesso e dunque la determinazione del saggio di interesse in termini di periodo medio di produzione. Irving Fisher tenta di aggirare queste difficoltà abbandonando quest'ultimo concetto, e muovendo dall'idea che "l'interesse è un indice della preferenza espressa dalla comunità per un dollaro presente rispetto a un dollaro futuro" (v. Fisher, 1930).
Per trattare correttamente del ruolo del tempo nella determinazione del saggio di interesse occorre distinguere, secondo Fisher, il concetto di fondo da quello di flusso. In particolare, "un fondo di ricchezza esistente in un dato istante di tempo è detto capitale; un flusso di servizi durante un periodo di tempo è detto reddito". Il capitale è allora qualsiasi forma di ricchezza capace di produrre un flusso di reddito. I percettori di reddito, d'altra parte, cercano di modificare le quantità di reddito disponibili per il consumo nelle diverse epoche future ricorrendo al risparmio e ai prestiti: il saggio di interesse è il prezzo che viene pagato per ottenere reddito presente in luogo di reddito futuro.
La determinazione di tale saggio dipende dall'interazione fra forze soggettive e forze oggettive. Le forze soggettive sono costituite dal "principio dell'impazienza" (o "della volontà"), che presiede alla redistribuzione dei consumi nel tempo da parte dei soggetti. Le forze oggettive, d'altra parte, consistono nel "principio delle occasioni di investimento", il quale consente di valutare i diversi progetti di investimento in termini del loro "saggio di rendimento sul costo". È questa una nozione importante, poiché strettamente connessa a quella keynesiana (anche se con essa non si identifica) di "efficienza marginale del capitale". Il saggio di rendimento sul costo è quel saggio di sconto in base al quale due o più progetti di investimento hanno lo stesso valore attuale netto. Le decisioni di investimento dipenderanno

da un confronto fra tale saggio e il saggio di interesse (o, se questo è diverso, il costo del capitale). Il saggio di interesse, a sua volta, non è determinato dall'ammontare di capitale (anche se Fisher non ne nega la produttività), bensì dall'interazione fra le preferenze temporali dei percettori di reddito e le opportunità di investimento. Nell'ipotesi di un comportamento massimizzante dei soggetti e di mercati concorrenziali, i saggi di preferenza temporale individuali e i saggi di rendimento sul costo dei diversi progetti di investimento convergeranno in un punto di equilibrio, il quale costituisce il saggio di interesse del sistema. Il saggio di interesse non dipende dunque dall'ammontare di capitale (semmai è vero il contrario, che il valore del capitale dipende dal saggio di interesse), bensì dalla preferenza temporale dei soggetti. La prospettiva tradizionale viene rovesciata e il processo produttivo praticamente scompare dall'analisi: all'analisi retrospettiva della struttura del capitale viene sostituita l'analisi prospettica delle opportunità di investimento. Per Fisher (ibid.) il principio fondamentale è che il valore del capitale in un dato momento deriva dal valore del reddito futuro che è atteso da quel capitale: "Il principio del valore presente è di fondamentale importanza nella teoria del valore e dei prezzi. Esso significa che il valore di qualsiasi articolo di ricchezza o di proprietà dipende soltanto dal futuro, non dal passato".

e) Domanda di capitale e produttività marginale

L'idea che il capitale, al pari degli altri fattori, sia produttivo, ha ovvie e importanti conseguenze sul piano della teoria della distribuzione del reddito. Per Smith il profitto era determinato dal saggio naturale del profitto, per Ricardo era un residuo, per Marx il risultato di un rapporto di sfruttamento. Per la teoria neoclassica della produttività marginale (Philip H. Wicksteed, Knut Wicksell, John B. Clark e altri) il profitto (qui, l'interesse), come qualsiasi altra quota distributiva, era univocamente determinato - date le condizioni tecniche della produzione - dalla produttività marginale del capitale. Il principio della marginalità era già presente in Ricardo, che su di esso basava la determinazione della rendita (e solo di questa). Gli economisti marginalisti generalizzano questo principio: tutti i fattori (variabili) della produzione devono essere remunerati, in equilibrio, secondo la loro produttività marginale, che è misurata dalla variazione del prodotto totale provocata dall'aggiunta o dalla sottrazione di un'unità del fattore considerato, quando sia mantenuta costante la quantità degli altri fattori.

Tale tesi ha due importanti implicazioni, una logica, l'altra normativa, riconducibili a questa

domanda: una volta che tutti i fattori della produzione siano stati remunerati secondo la loro produttività marginale, secondo il loro 'contributo' alla produzione stessa, si sarà esaurito il prodotto totale? Se così non fosse - se il prodotto totale non bastasse per una siffatta distribuzione, oppure se restasse un residuo - si tratterebbe di una teoria logicamente insoddisfacente. In effetti non tutte le funzioni di produzione godono di questa proprietà, anzi una soltanto: perché il prodotto risulti esaurito, occorre che la funzione di produzione sia di un tipo speciale, omogenea lineare (cioè con rendimenti di scala costanti). Ma ovviamente non c'è nessuna ragione per sostenere che le funzioni di produzione, nella realtà, siano necessariamente di questo tipo, quasi che si trattasse di "una sorta di misteriosa legge naturale" (Joan Robinson). Conviene osservare, inoltre, che una teoria della distribuzione basata sul principio della produttività marginale presuppone che per ciascun fattore questa possa essere calcolata indipendentemente dalla distribuzione del prodotto; occorre, in altri termini, che il valore del capitale non vari al variare della distribuzione, e questo - come oggi sappiamo, e come si vedrà più avanti - in generale non è vero.
L'implicazione normativa di questa teoria della distribuzione, d'altra parte, è che essa - quando sia soddisfatto il requisito di cui si è detto - sembra fornire un principio di giustizia distributiva: ciascun fattore della produzione deve essere remunerato secondo il suo contributo alla produzione, ed esiste un'unica configurazione distributiva di equilibrio, che è imposta dalle condizioni tecniche della produzione e non può né deve essere modificata dall'azione umana. Il profitto (l'interesse), in particolare, trova così una piena e doppia legittimazione, analitica ed 'etica'.

f) Knut Wicksell: una curiosa divergenza

Knut Wicksell condivide l'idea della scuola austriaca, secondo la quale il capitale è un insieme di mezzi di produzione prodotti, "una massa unitaria e omogenea di lavoro e terra risparmiati, accumulati negli anni" (v. Wicksell, 1901-1906). Il valore del capitale può dunque essere calcolato come somma delle remunerazioni (salari e rendite) ai servizi produttivi dei fattori originari (lavoro e terra) investiti nella produzione di beni capitali, capitalizzati in base al saggio di interesse corrente per il periodo di tempo durante il quale sono stati investiti. L'interesse (la remunerazione del capitale) è perciò la differenza fra la produttività marginale del lavoro e della terra risparmiati e la produttività marginale del lavoro e della terra correnti.

Wicksell si accorge però di una "curiosa divergenza": mentre per il lavoro e per la terra la nozione di produttività marginale è inequivoca, tale non è per il capitale nel suo complesso; dunque non si può affermare che vi sia una relazione univoca fra variazioni del capitale totale, variazioni del prodotto sociale e saggio di interesse. La spiegazione di questa curiosa divergenza (dopo Wicksell) è molto semplice: mentre il lavoro e la terra sono misurati ciascuno in termini della propria unità tecnica (ad esempio giornate o mesi lavorativi, ettari per anno), il capitale non può essere misurato in termini fisici, ma soltanto calcolato come una somma di valori di scambio, poiché ciascun singolo bene capitale è misurato in termini di un'unità estranea ad esso.
In quanto è un valore, il capitale non è una variabile indipendente (quali sono il lavoro e la terra), bensì dipende dai saggi di remunerazione. È dunque impossibile determinare il valore del capitale prima che sia stato raggiunto l'equilibrio fra produzione e consumo, per il semplice fatto che il capitale, a differenza dei fattori originari, non può esistere prima e indipendentemente dalla produzione. Se cambia il saggio di interesse, cambia il valore del capitale, e dunque voler spiegare l'interesse a partire dal valore del capitale significa ragionare in circolo.

4. Capitale e moneta

Con l'importante eccezione di Karl Marx e di altri eretici minori, incapaci di influenzare il corso del pensiero e dell'analisi economica, tutta l'economia ortodossa concepisce il capitale essenzialmente come capitale produttivo: come insieme di mezzi di produzione prodotti e non anche come capitale monetario. Questa distinzione è invece al centro delle opere, peraltro ben distinte quanto a metodi e risultati, di Joseph A. Schumpeter e di John M. Keynes. Sia pure brevemente, si deve però ricordare che spunti importanti, circa la distinzione fra capitale produttivo e capitale monetario, si trovano (oltre che in Alfred Marshall) soprattutto in Carl Menger, con la sua nozione di "capitale a disposizione". Nell'ambito delle teorie classiche e neoclassiche egemoni la nozione di capitale monetario quale potere d'acquisto capace di comandare mezzi materiali di produzione sfugge inevitabilmente, in quanto tali teorie vedono nella moneta soltanto un mezzo per agevolare le transazioni. Menger rileva invece che la moneta ha anche altre funzioni: essa può funzionare come mezzo di tesoreggiamento e di trasferimento patrimoniale nello spazio e nel tempo. In

questa prospettiva diventa possibile, e necessario, distinguere le funzioni di chi possiede potere d'acquisto e di chi investe in mezzi di produzione, e l'interesse può essere spiegato semplicemente come il prezzo che si forma nello scambio di un bene contro moneta, dove il bene consiste nel potere di comandare i mezzi di produzione: per Menger il capitale è "proprietà produttiva [considerata] come una somma di denaro impiegata produttivamente" (v. Menger, 1871).

a) Joseph A. Schumpeter: il capitale monetario nel processo capitalistico

Nella sua Teoria dello sviluppo economico Schumpeter definisce così il capitale: "Il capitale non è altro che la leva che consente all'imprenditore di sottomettere al proprio dominio i beni concreti di cui ha bisogno, nient'altro che un mezzo per disporre di certi beni per nuovi scopi o un mezzo per dettare alla produzione una nuova direzione" (v. Schumpeter, 1911). Ma cos'è questa "leva", questo mezzo di dominio?

Per l'imprenditore schumpeteriano tutti i beni di cui ha bisogno al fine di introdurre nuove combinazioni nel processo produttivo si trovano sullo stesso piano, siano essi appezzamenti di terra, prestazioni di lavoro, macchine o materie prime, o anche beni di consumo. Il modo di comportarsi dell'imprenditore nei confronti di qualsiasi bene di cui ha bisogno è sempre lo stesso: li acquista tutti in cambio di moneta, per la quale calcola (quando la possegga di già) o paga (quando la debba prendere a prestito) un interesse.

Il capitale non è l'aggregato di tutti i beni che servono agli scopi dell'imprenditore: il capitale sta di fronte al mondo dei beni, e la sua funzione consiste nel procurare all'imprenditore quei beni che devono essere impiegati produttivamente; sta dunque tra l'imprenditore e il mondo dei beni come un 'terzo agente' necessario alla produzione nell'economia di scambio. Esso costituisce il ponte fra le due cose, non prende direttamente parte alla produzione, né viene esso stesso 'trasformato': piuttosto esegue un compito che deve essere assolto prima che possa iniziare la produzione tecnica. C'è infatti un momento in cui l'imprenditore ha già il capitale necessario, ma non ancora i beni produttivi: in questo momento si può vedere con chiarezza come il capitale non si identifichi affatto con beni concreti e sia invece un 'agente autonomo'. Schumpeter cita qui, significativamente, Quesnay: "Parcourez les fermes et les ateliers, et [...] vous trouverez des bâtiments, des bestiaux, des matières premières, des meubles et des instruments de toute espèce". Una volta acquistato tutto ciò, il capitale ha adempiuto la sua funzione; l'imprenditore non ha più il capitale che gli era stato messo a disposizione, l'ha

ceduto in cambio dei mezzi di produzione, nel cui insieme consisterebbe - secondo "la concezione scientifica dominante" - il capitale stesso. In questo modo, secondo Schumpeter, si ignora però completamente la funzione del capitale di procurare beni, e si sostituisce alla visione complessiva (e 'circolare') del processo economico la supposizione "estranea alla realtà" che all'imprenditore vengano direttamente prestati quei beni di cui ha bisogno.
Il capitale è per Schumpeter un fondo di potere d'acquisto. Si tratta ora di capire in che cosa consista questo fondo di potere d'acquisto. Se ci si limitasse a dire che esso consiste in una somma di denaro, si tornerebbe alla nozione di Menger (e per un verso a quella di Fisher); ma soprattutto non è vero che per entrare nel novero degli imprenditori occorra il possesso di una siffatta somma, poiché non soltanto la moneta, bensì qualsiasi mezzo di circolazione che adempia a tale funzione è capitale. D'altra parte, se un mezzo di pagamento non serve a procurare a un imprenditore beni produttivi e a sottrarli per questo scopo a quello che era finora il loro impiego, esso non è capitale. Ciò significa che in un'economia senza sviluppo non esiste 'capitale', nel senso che il capitale non adempie alla sua funzione caratteristica di agente autonomo. Il concetto di capitale "incarna un aspetto dei processi economici che ci è suggerito soltanto dai fenomeni dello sviluppo"; si potrà dunque ridefinire il capitale come "quella somma di moneta e di altri mezzi di pagamento che è in ogni momento disponibile per essere ceduta a imprenditori".
Lla nozione secondo la quale lo sviluppo economico è sostanzialmente un diverso impiego dei servizi del lavoro e della terra già esistenti conduce Schumpeter a due 'eresie' (che lo accomunano a Marx e a Keynes); la prima è che nel processo capitalistico ha una funzione essenziale la moneta, l'altra che tale funzione spetta anche agli altri mezzi di pagamento. "I processi che avvengono nel campo dei mezzi di pagamento non sono meri riflessi dei processi che avvengono nel mondo dei beni". Di qui il ruolo centrale, nel processo capitalistico, della moneta e del credito, dunque del banchiere. L'imprenditore ha bisogno del credito nel senso di una cessione temporanea di potere d'acquisto, e ciò proprio per poter produrre, per poter introdurre le sue nuove combinazioni, per diventare un imprenditore. Questo potere d'acquisto, se non lo possiede per altra via - ma in questo caso ciò sarebbe soltanto la conseguenza di uno sviluppo precedente - se lo deve far prestare: egli può diventare imprenditore solo diventando prima debitore. Se il capitale non è altro che la leva che consente all'imprenditore di sottomettere al proprio dominio i beni concreti di cui ha bisogno, il credito è a sua volta la leva di questa sottrazione di beni.

Il credito, per Schumpeter, può essere definito come la "creazione di potere d'acquisto al fine di cederlo all'imprenditore", e non semplicemente come il trasferimento di potere d'acquisto esistente. Attraverso il credito si apre agli imprenditori l'accesso al flusso di beni della società; in un'economia caratterizzata dalla proprietà privata e dalla divisione del lavoro, la concessione del credito agisce come un'ingiunzione al sistema economico di subordinarsi agli scopi dell'imprenditore, come un ordine ai beni di cui egli ha bisogno, "come un affidamento a lui di forze produttive". Il profitto imprenditoriale, in questo quadro, assume il significato di un fenomeno autonomo (anziché conseguente al rapporto fra capitale e lavoro salariato), essenzialmente legato alla funzione dell'imprenditore in quanto leader dell'economia: "È l'espressione del valore del contributo dell'imprenditore alla produzione, esattamente nello stesso senso [anche se non nel senso della teoria della produttività marginale] in cui il salario è l'espressione di valore di ciò che 'produce' l'operaio". Il profitto imprenditoriale non è una rendita, non è un semplice residuo, non è il risultato di un rapporto di sfruttamento, e non è neppure un utile del capitale, poiché non coincide con l'interesse: l'interesse è una porzione del profitto, della quale si appropria chi fornisce all'imprenditore la liquidità che gli occorre, e ha la natura di un premio del potere d'acquisto presente su quello futuro, che si forma sul mercato della moneta e non su quello dei beni.

b) John M. Keynes: capitale, investimenti e animal spirits

Il concetto centrale, nella teoria del capitale di John M. Keynes, è quello di "efficienza marginale del capitale", cioè "quel saggio di sconto al quale il valore presente della serie di annualità, rappresentate dai ricavi attesi dal capitale durante la sua vita, eguaglia esattamente il prezzo di offerta del capitale medesimo" (v. Keynes, 1936). Questa nozione può sembrare - anche a causa di talune concessioni retoriche dello stesso Keynes - simile o addirittura identica a quella fisheriana di "saggio di rendimento sul costo", il quale, come si è già visto, è quel saggio ipotetico di interesse che, se impiegato per calcolare il valore presente di due opzioni messe a confronto, le renderebbe uguali. Di qui conviene cominciare, per mostrare come Keynes non sia affatto un epigono degli economisti neoclassici e la Teoria generale vada invece letta come critica della tradizione neoclassica (così come Il capitale di Marx va letto come critica della tradizione classica).

Al di là delle analogie formali, e a parte alcune differenze tecniche (peraltro importanti), la differenza fondamentale sta in questo: che mentre in Fisher, e in generale nella visione neoclassica, il saggio di rendimento dipende dalla produttività fisica del capitale, per Keynes

l'efficienza marginale del capitale dipende dalla scarsità del capitale stesso; scarsità che almeno per il sistema economico nel complesso è artificiale, poiché ancora oggi si è ben lontani da una situazione di saturazione dei capitali: il rendimento del capitale è una sorta di rendita improduttiva ed eliminabile. D'altra parte, mentre la teoria neoclassica determina il valore dello stock di capitale sulla base del tasso d'interesse che risulta dal confronto fra domanda e offerta dei servizi dei beni capitali, per Keynes l'efficienza marginale del capitale è il tasso di profitto atteso, ciò che l'imprenditore -mosso dai suoi animal spirits - si aspetta di ottenere, e non ciò che egli otterrà davvero; l'efficienza marginale del capitale va definita in termini dell'aspettativa di reddito e del prezzo corrente di offerta del capitale: "essa dipende dal saggio atteso di rendimento in termini di moneta, se questa venisse investita in un dato capitale di nuova produzione, non dal risultato storico di ciò che un investimento ha reso rispetto al suo costo originario se si guarda indietro a ciò che ha fruttato quando la sua vita è giunta al termine".
L'ammontare effettivo dell'investimento corrente dipenderà da un confronto fra l'efficienza marginale del capitale e il saggio di interesse corrente, poiché si realizzeranno soltanto quei progetti di investimento la cui efficienza marginale è maggiore, o almeno uguale, al saggio di interesse corrente. L'incentivo a investire dipende quindi in parte dall'efficienza marginale del capitale (dalle aspettative degli imprenditori) e in parte dal saggio di interesse; quest'ultimo tuttavia, e dunque il valore attuale di quel capitale, non può essere determinato sulla sola base della conoscenza del reddito prospettivo di un capitale o della sua efficienza marginale. Occorre dunque una teoria indipendente del saggio di interesse, che consenta poi di calcolare il valore attuale del capitale 'capitalizzando' il suo reddito prospettivo.
Per Keynes il rendimento del capitale dipende dal fatto che esso è (artificialmente) scarso. Ciò è sufficiente per consigliare di non dire che il capitale è produttivo: è assai meglio dire che esso fornisce, nel corso della sua vita, un reddito maggiore del suo costo originario. L'unica ragione per la quale un bene capitale offre la prospettiva di rendere, durante la sua vita, servizi aventi un valore complessivo superiore al suo prezzo di offerta iniziale è perché esso è scarso; e viene mantenuto scarso a causa della concorrenza del saggio di interesse: se il capitale diviene meno scarso, il suo rendimento rispetto al costo diminuirà, senza che diminuisca la sua produttività fisica. Per questa ragione Keynes si dichiara vicino alla dottrina classica: "ogni cosa è prodotta dal lavoro, coadiuvato da ciò che allora usava chiamarsi arte e che ora si chiama tecnica, dalle risorse naturali e dai risultati del lavoro

passato incorporati in attività capitali". È insomma preferibile considerare il lavoro (compresi i servizi personali dell'imprenditore e dei suoi collaboratori) come l'unico fattore di produzione, operante in un dato ambiente di tecnica, di risorse naturali, di beni capitali e di domanda effettiva.
Il problema diventa allora quello di determinare che cosa solleciti gli imprenditori a fare quel che fanno, posto che la Teoria generale si può ridurre a questa proposizione: l'occupazione è quella che i capitalisti decidono di dare, secondo le loro aspettative. Secondo lo stesso Keynes, "la teoria si può riassumere dicendo che, data la psicologia della gente, il livello della produzione e dell'occupazione complessive dipende dall'ammontare dell'investimento" (v. Keynes, 1937).

c) Keynes e il saggio di interesse: il distacco dalla tradizione
Keynes, come egli stesso scrive nel fondamentale articolo del 1937, si distacca dalla teoria tradizionale in almeno due punti.
1. La teoria ortodossa assume che noi si disponga di un tipo di conoscenza del futuro completamente diverso da quello che abbiamo in realtà. L'ipotesi di un futuro prevedibile mediante un calcolo benthamiano conduce a un'interpretazione erronea dei principî di comportamento che la necessità di agire ci costringe ad adottare, e a una sottovalutazione dei fattori nascosti di dubbio, di precarietà, di speranza, di timore. Il risultato è stato un'errata teoria del tasso di interesse: "È vero che la necessità di uguagliare i vantaggi della scelta tra il possesso di titoli o di beni capitali rende necessario che il tasso di interesse sia uguale all'efficienza marginale del capitale. Ma questo non ci dice a che livello questa uguaglianza si realizzi. La teoria ortodossa ritiene che sia l'efficienza marginale del capitale a stabilirlo. Ma l'efficienza marginale del capitale dipende dal prezzo dei beni capitali, e poiché questo prezzo determina il tasso del nuovo investimento, esso è compatibile in equilibrio con un solo dato livello del reddito monetario. Perciò l'efficienza marginale del capitale non è determinata, se non è dato il livello del reddito monetario. In un sistema nel quale il livello del reddito è soggetto a fluttuazioni, la teoria ortodossa ha un'equazione di meno di quanto serve per ottenere una soluzione. Indubbiamente la ragione per cui il sistema ortodosso non è riuscito a scoprire la lacuna sta nel fatto che esso ha sempre tacitamente assunto che il reddito sia dato, e precisamente al livello corrispondente al pieno impiego di tutte le risorse disponibili. Così, invece di essere l'efficienza marginale a determinare il tasso di interesse, è

corretto affermare che è il tasso di interesse che determina l'efficienza marginale del capitale".
2. La teoria ortodossa avrebbe scoperto il difetto di cui sopra, se non avesse ignorato la necessità di una teoria della domanda e dell'offerta del prodotto complessivo: "Non è stato osservato che la quantità dei beni di consumo che conviene produrre agli imprenditori è funzione della quantità dei beni di investimento che conviene loro produrre".
Secondo Keynes la tradizione (neo)classica del pensiero economico considera il saggio di interesse come il fattore che porta all'equilibrio la domanda di investimento con la disposizione a risparmiare. L'investimento rappresenta la domanda di risorse investibili e il risparmio rappresenta l'offerta, mentre il saggio di interesse è il 'prezzo' delle risorse investibili al quale domanda e offerta si equilibrano. Allo stesso modo in cui il prezzo di una merce si stabilisce necessariamente al punto nel quale la domanda è uguale all'offerta, così il saggio di interesse si stabilisce necessariamente, sotto il gioco delle forze del mercato, al punto nel quale l'ammontare dell'investimento a quel saggio di interesse è uguale all'ammontare di risparmio a quello stesso saggio. Keynes è disposto ad ammettere che il saggio di interesse possa forse esercitare un influsso sull'ammontare risparmiato da un reddito dato; ma il punto è proprio questo, che il reddito non è dato e che il caso veramente generale è quello in cui il livello del reddito e dell'occupazione è soggetto a fluttuazioni: "In verità la teoria classica non ha afferrato l'importanza delle variazioni del livello del reddito o della possibilità che il livello del reddito divenga effettivamente una funzione dell'ammontare di investimento per unità di tempo". L'analisi tradizionale è dunque difettosa perché non è riuscita a isolare correttamente le variabili indipendenti del sistema.
Così come l'efficienza marginale del capitale, anche il saggio di interesse per Keynes dipende in maniera decisiva dalle aspettative degli agenti economici. Ciascun individuo, una volta deciso quanta parte del suo reddito consumerà e quanta invece accantonerà in qualche forma disponibile per il consumo futuro, dovrà decidere in quale forma - liquida oppure no - conserverà le disponibilità che ha accantonato. Il saggio di interesse, in questo contesto, non può essere una ricompensa per il risparmio o l'astinenza come tali; infatti, se un uomo tesaurizza i suoi risparmi in denaro, non percepisce alcun interesse benché risparmi esattamente tanto quanto prima: il saggio di interesse è invece la ricompensa all'abbandono della liquidità per un periodo determinato. Esso misura la riluttanza di coloro che possiedono la moneta ad abbandonare il loro controllo liquido su di essa; esso non è il 'prezzo' che porta

all'equilibrio la domanda di mezzi da investire con la disposizione ad astenersi dal consumo presente: è il prezzo che equilibra il desiderio di tenere la ricchezza in forma di denaro con la quantità di denaro disponibile.
Questa preferenza per la liquidità richiede però una spiegazione. "Perché mai vi dovrebbe essere qualcuno, al di fuori delle mura di un manicomio, che desideri usare la moneta come riserva di ricchezza?". La spiegazione keynesiana è che, per motivi in parte ragionevoli, in parte istintivi, il nostro desiderio di tenere moneta come riserva di ricchezza è un barometro del nostro grado di sfiducia nelle nostre capacità di calcolo e nelle nostre convenzioni sul futuro. Sebbene questo atteggiamento verso la moneta sia esso stesso convenzionale o istintivo, esso opera, per così dire, a un livello più profondo delle nostre motivazioni. Esso subentra nei momenti in cui le più superficiali, più instabili convenzioni si sono indebolite: "Il possesso della moneta culla la nostra inquietudine, e il premio che noi pretendiamo per dividerci da essa è la misura della nostra inquietudine".

d) Keynes: capitale e (dis)occupazione

La teoria keynesiana, come si è accennato, si può riassumere così: "Data la psicologia della gente, il livello della produzione e dell'occupazione complessive dipende dall'ammontare dell'investimento". Più esaurientemente, la produzione totale dipende dalla propensione al tesoreggiamento, da come la politica monetaria influenza la quantità di moneta, dallo stato della fiducia relativamente al rendimento futuro dei beni capitali, dalla propensione alla spesa, dai fattori sociali che influenzano il livello del salario monetario. Di questi diversi fattori, però, quelli dei quali ci si può fidare di meno sono quelli che determinano il tasso dell'investimento, perché sono influenzati dalle nostre previsioni sul futuro, del quale sappiamo così poco.
In condizioni di laissez faire, d'altra parte, il livello dell'investimento normalmente non assicurerà la piena occupazione. Il nostro mondo economico oscilla, evitando i più gravi estremi delle fluttuazioni dell'occupazione e dei prezzi in ambo i sensi, intorno a una posizione intermedia sensibilmente al di sopra di quel livello minimo dell'occupazione al di sotto del quale si metterebbe in pericolo l'esistenza, ma sensibilmente al di sotto dell'occupazione piena: "La nostra sorte normale è una situazione intermedia, né disperata né soddisfacente" (v. Keynes, 1936). Di qui l'auspicio di Keynes - il quale vede nella disoccupazione il difetto principale della società capitalistica (oltre alla distribuzione

arbitraria e iniqua delle ricchezze e dei redditi) - di una socializzazione "di una certa ampiezza" dell'investimento e soprattutto dell'eutanasia del redditiero.
La ragione di una proposta politica così radicale è la radicalità dell'analisi keynesiana. La teoria keynesiana del saggio di interesse e delle decisioni di investimento, infatti, rovescia la tesi tradizionale secondo la quale un saggio di interesse moderatamente alto è giustificato dalla necessità di offrire un incentivo sufficiente al risparmio. L'ampiezza del risparmio effettivo è invece determinata dalla scala dell'investimento, e questa è favorita da un saggio di interesse basso (purché non si cerchi di stimolare in tal modo l'investimento al di là del punto corrispondente alla piena occupazione). L'eutanasia del redditiero (e dunque "l'eutanasia del potere oppressivo e cumulativo del capitalista di sfruttare il valore di scarsità del capitale") sarebbe pertanto vantaggiosa per la società, e allo stesso tempo ragionevole ed equa.
L'interesse non rappresenta il compenso di alcun sacrificio genuino, come non lo rappresenta la rendita della terra. Il possessore del capitale può ottenere l'interesse perché il capitale è scarso, proprio come il possessore della terra può ottenere la rendita perché la terra è scarsa. Ma mentre vi può essere una ragione intrinseca della scarsità della terra, non vi sono ragioni intrinseche della scarsità del capitale.
Circa questo punto - nell'ultimo capitolo: Note conclusive sulla filosofia sociale verso la quale la 'Teoria generale' potrebbe condurre - Keynes argomenta e predica così: "A lungo andare non esisterebbe una ragione intrinseca di questa scarsità, ossia non esisterebbe un sacrificio genuino, ottenibile soltanto con l'offerta del compenso dell'interesse; salvo che la propensione individuale al consumo si dimostrasse di carattere tale che il risparmio netto in condizioni di occupazione piena venisse a finire prima che il capitale fosse divenuto sufficientemente abbondante. Ma anche in tal caso, sarà ancora possibile che il risparmio collettivo per il tramite dello Stato sia mantenuto a un livello che permetta l'aumento del capitale fino al punto al quale questo non sia più scarso. Considero perciò l'aspetto del capitalismo caratterizzato dall'esistenza del redditiero come una fase di transizione, destinata a scomparire quando esso avrà compiuto la sua opera. E con la scomparsa del redditiero, molte altre cose del capitalismo subiranno un mutamento radicale. [...] Potremmo dunque mirare in pratica (poiché non vi è nulla di tutto questo che sia irraggiungibile) a un aumento del volume di capitale finché questo non sia più scarso, cosicché l'investitore senza funzioni non riceva più un premio gratuito; e a un sistema di imposizione diretta tale da permettere che l'intelligenza e la determinatezza e la capacità del finanziere, dell'imprenditore et hoc

genus omne (i quali certamente amano tanto il loro mestiere che il loro lavoro potrebbe ottenersi a molto minor prezzo che attualmente) siano imbrigliate al servizio della collettività, con un compenso a condizioni ragionevoli".

5. Controversie recenti sulla misura del capitale

Wicksell e Keynes avevano colto un problema importante, quello delle difficoltà di una misura inequivoca del capitale (e perciò della produttività marginale del capitale stesso): il capitale, a differenza degli altri fattori della produzione, non può essere misurato in termini di una sua propria unità di misura; essendo - nella sua determinazione materiale - una collezione di beni capitali eterogenei, di esso non si può dare una 'misura' aggregata altro che in termini di valore.
Della "curiosa divergenza" notata da Wicksell si è già detto. Quanto a Keynes, egli rileva - nel capitolo della sua Teoria generale sulla scelta delle unità di misura - che le unità sulla base delle quali gli economisti misurano il reddito nazionale, il livello generale dei prezzi e per l'appunto lo stock di capitale reale sono insoddisfacenti. In generale "la produzione di merci e servizi della collettività è un complesso non omogeneo, il quale a rigore non può venire misurato se non in certi casi particolari, come ad esempio quando tutti gli elementi di una certa produzione siano compresi in egual proporzione in un'altra produzione". E la difficoltà è anche maggiore quando si cerca di misurare l'incremento netto del capitale, poiché allora si deve trovare una base per un confronto quantitativo fra i nuovi beni capitali prodotti nel periodo e i vecchi beni capitali che sono stati consumati. È per questa ragione che Keynes usa, come unità di misura, unicamente moneta e lavoro e considera come unico fattore della produzione il lavoro stesso.

a) Il capitale nella funzione di produzione

Le teorie del capitale egemoni nell'accademia e nell'insegnamento (e dunque ispiratrici delle idées reçues), almeno fino ai contributi critici di Joan Robinson, di Piero Sraffa, di Pierangelo Garegnani e di Luigi Pasinetti, ignorano o aggirano il problema ricordato sopra. Il problema consiste nella possibilità di determinare una misura teorica univoca e rigorosa del capitale aggregato, la quale sia indipendente dalle variabili che sulla base di questo si vogliono determinare: saggio di salario, saggio di interesse, prezzi relativi delle merci. Il

contesto è quello della teoria marginalistica della distribuzione (v. § 3e).
In un articolo del 1953-1954 la questione viene sollevata da Joan Robinson: "Il predominio, nell'insegnamento economico di tipo neoclassico, del concetto di una funzione della produzione per mezzo della quale si dimostra che i prezzi relativi dei fattori sono funzione del rapporto in cui questi vengono impiegati, a un dato stato della conoscenza tecnica, ha avuto un effetto negativo sullo sviluppo di questi argomenti; e infatti, prendendo in esame principalmente i problemi relativi alle proporzioni tra fattori produttivi, esso ha distolto l'attenzione dai problemi, più difficili ma più interessanti, relativi alle forze che regolano le offerte dei fattori e alle cause e conseguenze dei mutamenti delle conoscenze tecniche. Per di più, la funzione della produzione è stata un potente strumento diseducativo. Allo studente di teoria economica si insegna a scrivere $Q = f(L, K)$, dove L è una quantità di lavoro, K una quantità di capitale e Q la quantità prodotta di certe merci. Gli viene poi detto di supporre che tutti i lavoratori siano uguali e di misurare L in ore-uomo di lavoro, gli viene detto qualcosa sul problema dei numeri indice, a proposito della scelta di una unità di misura per il prodotto, e poi lo si fa passare frettolosamente al problema che segue, nella speranza che egli si dimentichi di chiedere in quali unità viene misurato K. Prima di avere il tempo di porsi tale domanda egli è già diventato professore; e così abiti mentali frusti sono tramandati da una generazione all'altra".
Il problema, come ha scritto Geoffrey C. Harcourt, "è quello di trovare un'unità con cui il capitale sociale, o capitale aggregato in valore che dir si voglia, possa essere misurato come un numero; cioè un'unità che sia indipendente dai prezzi relativi e dalla distribuzione e possa quindi essere inserita in una funzione della produzione dove, insieme al lavoro, misurato anch'esso in modo opportuno, possa spiegare il livello della produzione aggregata. Inoltre, in un'economia perfettamente concorrenziale in cui le previsioni sono perfette [...], questa unità dev'essere tale da uguagliare la derivata parziale del prodotto rispetto al 'capitale' alla remunerazione del 'capitale' e la corrispondente derivata rispetto al lavoro al salario reale (in termini di prodotto). Questa unità fornirebbe dunque anche gli elementi per una teoria della distribuzione basata sulla produttività marginale. Se è possibile trovare una tale unità di misura, si possono prendere due piccioni con una fava; poiché in tal caso possiamo da un lato analizzare un sistema produttivo in cui i beni capitali - mezzi di produzione prodotti - fungono da ausiliari del lavoro (caratteristica, questa, comune a ogni società industrialmente avanzata) e, dall'altro lato, possiamo studiare la distribuzione del reddito in un sistema

capitalistico (cioè in un sistema in cui la proprietà di un capitale in valore permette ai suoi possessori di partecipare alla distribuzione del reddito nazionale sotto forma di profitti sul capitale investito, dove sia l'ammontare di tali profitti che lo stesso saggio del profitto sono in relazione alle caratteristiche tecniche del sistema produttivo)" (v. Harcourt, 1972).
Trovare quell'unità di misura, e quindi poter trattare il capitale come 'fattore' della produzione in un mondo caratterizzato dalla scarsità delle risorse e dalla razionalità massimizzante dei soggetti economici, vorrebbe inoltre dire che non ci sarebbe alcun bisogno di specificare il contesto politico e istituzionale del processo di produzione-riproduzione, e che diventerebbe superfluo o addirittura improprio il rinvio ai rapporti di forza fra le diverse classi come variabile essenziale nella determinazione della configurazione distributiva. E verrebbe meno il problema di analizzare le diverse forme che il capitale assume nelle diverse fasi del processo capitalistico, problema che Joan Robinson evoca così: "Siamo abituati a parlare del saggio del profitto realizzato da una impresa come se i profitti e il capitale fossero entrambi delle somme di denaro. Il capitale, quando consiste in attività finanziarie non ancora investite, è in effetti una somma di denaro, e così pure i ricavi netti di un'impresa sono somme di denaro. Ma le due cose non coesistono mai allo stesso tempo. Quando il capitale si presenta come una somma di moneta, i profitti non sono stati ancora realizzati. Quando i profitti (quasi-rendite) vengono realizzati, il capitale ha cessato di essere moneta ed è diventato un impianto. Ma possono succedere molte cose che fanno sì che si crei una divergenza tra il valore dell'impianto e il suo costo originario. [...] Come dobbiamo valutare il capitale rappresentato dall'impianto?" (v. Robinson, 1953-1954).
Il fatto è che in generale è impossibile concepire un valore del 'capitale in generale', che sia indipendente dalla distribuzione: saggio di salario e saggio dei profitti non sono 'prezzi' come gli altri.

b) Le 'parabole' neoclassiche

Gli "abiti mentali frusti" di cui parla Joan Robinson hanno preso la forma, come ha scritto lo stesso Harcourt, di 'parabole' sulle forze che determinano la distribuzione del reddito fra capitalisti e lavoratori salariati, sulle modalità dell'accumulazione del capitale e sulle scelte (da parte dei capitalisti) delle tecniche di produzione.
Queste parabole sono: a) a saggi dei profitti più bassi corrispondono valori del capitale pro capite più elevati; b) a saggi dei profitti più bassi corrispondono rapporti capitale-prodotto

più elevati; c) a saggi dei profitti più bassi corrispondono (tramite investimenti in metodi produttivi più 'meccanizzati') livelli di consumo pro capite più elevati (in 'stato uniforme'); e soprattutto d) in condizioni concorrenziali la distribuzione del reddito fra capitalisti e lavoratori salariati può essere spiegata sulla sola base della conoscenza delle produttività marginali e delle offerte dei fattori, in sostanza della sola conoscenza delle condizioni 'tecniche' della produzione (risorse, tecnologia e gusti dei consumatori).

c) La critica sraffiana: l'armonia neoclassica come caso particolarmente astratto

Queste quattro proposizioni non sono tutte condivise da tutti gli economisti neoclassici; esse corrispondono però alla visione ancora corrente del capitale come un 'fattore' della produzione al quale, come a tutti gli altri, spetta una remunerazione commisurata al suo 'contributo' alla produzione (alla sua 'produttività marginale'). Esse si reggono, d'altra parte, sulla possibilità di dare del capitale una misura in valore che sia indipendente dal saggio del profitto: in caso contrario il ragionamento sarebbe circolare. Una critica definitiva a una concezione siffatta del capitale è stata possibile grazie al contributo di Piero Sraffa (del quale ci si occuperà qui soltanto in questa prospettiva, e non anche in quella delle possibili riletture di Marx 'dopo Sraffa').

Una funzione della produzione descrive le varie combinazioni dei fattori fra le quali l'imprenditore potrà scegliere: nell'ipotesi che le condizioni tecniche della produzione siano date, che vi sia sostituibilità fra i fattori della produzione, che vi sia concorrenza perfetta e che l'imprenditore persegua la massimizzazione del profitto. Ciascun livello della produzione richiederà un'adeguata combinazione di fattori (un dato metodo di produzione) e la combinazione più vantaggiosa sarà quella per la quale il saggio di sostituzione dei fattori è uguale al reciproco del rapporto fra i prezzi dei fattori stessi; se i prezzi relativi dei fattori mutano, dovranno mutare anche le proporzioni in cui sono combinati i fattori. Ad esempio: se i salari aumentano e i profitti diminuiscono, in una funzione di produzione regolare ci si sposterà verso una combinazione di fattori che impieghi meno lavoro e più capitale.

Questa proposizione, tuttavia, ha un senso soltanto se vi è davvero un solo bene capitale, così che non si pone il problema di misurarlo in valore. Se invece si producono più merci - nell'ambito di un processo circolare di produzione di merci a mezzo di merci, esse stesse prodotte - quella proposizione, apparentemente sensata, diventa dubbia: vera soltanto entro ipotesi estremamente precise e restrittive. Se si potesse mostrare che in generale non è vero

che vi sia una relazione inversa fra saggio dei profitti e 'intensità di capitale', allora la nozione di capitale implicita in quella di funzione della produzione, e la conseguente teoria della distribuzione, perderebbero qualsiasi significato analitico e politico: e in effetti così è.
Si può dimostrare infatti, grazie ai lavori di Sraffa, Pasinetti, Garegnani e altri, che non vi è una relazione univoca e inversa fra saggio dei profitti e intensità di capitale, e che invece è possibile che le tecniche di produzione 'ritornino' (non nel senso storico del termine, ma in quello logico, che qui è dirimente). È cioè possibile che a seguito di un cambiamento nella distribuzione del prodotto netto fra salari e profitti, ad esempio a favore dei salari, una data tecnica di produzione, che comporta una data intensità di lavoro, sia sostituita con un'altra tecnica a più alta intensità di capitale; e che però, a un livello dei salari ancora più alto, la prima tecnica torni a essere conveniente e quindi venga a sua volta sostituita a quella che l'aveva soppiantata. Questo risultato non è un curiosum: esso mette definitivamente in crisi l'idea che le grandezze distributive siano prezzi come tutti gli altri, poiché così si dimostra che i prezzi stessi variano al variare della distribuzione del reddito, e dunque varia il valore del capitale che mediante quei prezzi deve essere calcolato.
È invece davvero curioso quanto segue, se si ricorda che la vulgata neoclassica pretende che in equilibrio il saggio dei profitti sia uguale alla produttività marginale del capitale, e che il calcolo di questa, se non si vuole ragionare in circolo, presuppone la possibilità di misurare il capitale indipendentemente dalla distribuzione. La curiosità sta in questo: il valore del 'capitale' non varia al variare del saggio dei profitti (che si vorrebbe determinare e giustificare a partire da quel valore) soltanto ed esattamente nelle stesse ipotesi in cui Böhm-Bawerk confina la validità della versione semplice della teoria marxiana del valore-lavoro, e cioè quando la composizione organica del capitale sia uniforme nelle diverse industrie; che è come dire che dal punto di vista dell'analisi economica si produce una sola merce.
Sembrerebbe una vendetta della storia; ma mentre la riduzione di Marx a caso particolare è diventata dogma, dell'astrattezza neoclassica, almeno nei manuali, non si fa cenno.
In generale non esiste una misura della quantità di capitale che possa essere usata, senza ragionare in circolo, per la determinazione dei prezzi e della distribuzione del reddito. Le conseguenze di ciò sono gravi per la teoria accettata del capitale e della distribuzione: come scrive Garegnani, "dall'aumento della proporzione tra capitale e lavoro nell'economia, quando l'interesse diminuisce, sono state infatti dedotte 'funzioni di domanda' del 'capitale' (cioè, in ultima analisi, del 'risparmio') e del lavoro e, con esse, l'idea che la distribuzione del prodotto

sociale fosse determinata dall'equilibrio fra la domanda e l'offerta di tali 'fattori di produzione'. Di qui, in particolare, la spiegazione dell'interesse (profitti) in termini di scarsità del 'capitale', e di ricompensa per l''attesa'. È difficile vedere come questa complessa struttura teorica possa essere preservata, quando la base su cui essa è eretta si rivela erronea".

d) Capitale e società: la scomparsa della categoria

Dimostrare non basta però a convincere. Neoclassici autorevoli ammettono che la teoria neoclassica aggregata è logicamente debole. Alcuni ritengono che ci si debba comunque attenere a essa, e continuare a insegnarla, per ragioni di fede, in attesa di improbabili soluzioni econometriche della questione. Altri negano che vi siano neoclassici seri che credono nelle parabole, poiché la vera teoria neoclassica sarebbe quella delle versioni moderne della teoria dell'equilibrio economico generale, che continuamente promettono, senza mai mantenere, di tener conto di tutte le forme e metamorfosi del capitale in una società capitalistica: all'interno però di un modello nel quale il capitale è ridotto a misteriose e naturali 'dotazioni iniziali'.

Nell'analisi economica contemporanea il 'capitale' scompare. Nessuno crede più che abbia senso parlare di 'leggi di movimento' della società capitalistica. Così come le 'classi' si sarebbero fatte da sanguigne realtà pallide astrazioni (secondo Schumpeter), il capitale sembra essere diventato un indistinto insieme di beni capitali eterogenei. Tutto è spostato altrove, e non ci sono più eretici influenti.

Concorrenza
di Ruggero Paladini

sommario: 1. Introduzione. 2. La concorrenza come processo: i movimenti dei capitali. 3. La concorrenza come processo: asimmetria tra i produttori: a) Cournot e il vantaggio temporaneo; b) Bertrand e l'asimmetria nel prezzo; c) Schumpeter e l'innovazione. 4. La concorrenza come processo: il meccanismo degli scambi. 5. La concorrenza come situazione caratterizzata. 6. Insoddisfazioni nella teoria della concorrenza. 7. La concorrenza potenziale e le barriere all'entrata. 8. La teoria dei giochi e i processi concorrenziali. 9. Concorrenza e

innovazione. 10. Teoria della concorrenza e politica antimonopolistica. □ Bibliografia.

1. Introduzione

La nozione di concorrenza sorge con il definirsi dell'economia politica come scienza autonoma. È strettamente connessa con l'idea di libertà, che dal punto di vista economico implica la libera iniziativa, il laissez faire, laissez passer, e con la fondamentale ipotesi che gli individui, in quanto operatori economici, si muovono secondo un comportamento razionale volto a massimizzare la propria soddisfazione.
Una delle idee più rilevanti elaborata da Adam Smith è quella per cui il libero agire di una pluralità di individui egoisti non determina il caos, ma al contrario un sistema sociale ordinato - sul quale è quindi possibile un'investigazione scientifica - e tale per cui l'operato del singolo, volto a procurarsi il maggior vantaggio possibile, torna, con apparente paradosso, a vantaggio di tutti gli altri.
Ora, per l'operare della 'mano invisibile' è essenziale la funzione della concorrenza. La concorrenza è quindi un principio 'virtuoso': il suo operare ha effetti positivi, la sua assenza, e quindi la presenza del monopolio, determina al contrario effetti negativi. Per i critici dell'economia politica e del libero mercato, invece, la concorrenza è un meccanismo che distrugge l'equilibrio sociale ed economico, peggiora la posizione dei più deboli e avvantaggia ulteriormente quella dei più forti. Tuttavia per Marx, certo non un apologeta del libero mercato, la concorrenza è il processo economico che rende possibile l'analisi scientifica del sistema capitalistico.Il termine 'concorrenza' è usato nell'analisi economica in molte accezioni: innanzitutto occorre distinguere tra concorrenza come processo economico e concorrenza come stato, cioè forma di mercato. Nella prima accezione si intende un susseguirsi logico e/o temporale di azioni e reazioni da parte degli operatori economici, che trova un termine quando si viene a determinare una situazione nella quale non vi è più per alcuno un incentivo a mutare la propria condotta: situazione detta di equilibrio.
Nella seconda accezione invece la concorrenza è individuata come un regime di mercato, caratterizzato da una serie di presupposti e di elementi tipici e tale comunque da cogliere la realtà economica della maggior parte dei mercati.Nei capitoli che seguono ci occuperemo prima delle diverse nozioni di concorrenza come processo; vedremo quindi le ragioni che hanno portato l'attenzione sulla seconda nozione di concorrenza, cioè la concorrenza come

stato, per esaminare poi il ritorno alla prima nozione che si è manifestato nell'analisi economica più recente.

2. La concorrenza come processo: i movimenti dei capitali

La prima nozione di processo concorrenziale è quella degli economisti classici, da Smith a Marx. Consiste nel processo di gravitazione dei prezzi di mercato verso i prezzi naturali, o prezzi di produzione, e nel processo di riduzione di questi ultimi, nel lungo periodo, in seguito al progresso tecnico (v. Sylos Labini, 1976). Il fenomeno economico alla base del processo concorrenziale è quello della tendenza dei capitali a muoversi verso i settori dove il saggio del profitto è più elevato. Si supponga, partendo da una situazione di equilibrio, che in un dato momento si determini un aumento della domanda in alcuni settori e una diminuzione in altri. Vi sarà una tendenza dei capitali ad allontanarsi dai settori dove i prezzi di mercato sono diminuiti e a dirigersi dove sono aumentati. Questo processo determina delle variazioni della produzione tali per cui gli squilibri tra prezzi di mercato e prezzi naturali si attenuano e, in tendenza, si annullano, ripristinando l'eguaglianza nei saggi di profitto nei vari settori.
È dunque l'ipotesi della libertà di movimento dei capitali che rende possibile la concorrenza e con essa l'analisi scientifica dei prezzi naturali, cioè una teoria del valore e della distribuzione: i prezzi naturali o di produzione sono i prezzi della libera concorrenza.I classici ritenevano che la maggior parte delle barriere poste al libero movimento dei capitali derivasse dalla pubblica autorità, cioè dalla concessione di privilegi ad alcuni operatori, o dall'imposizione di dazi e tariffe; tali interventi determinano inevitabilmente situazioni di monopolio, o comunque restrizioni alla concorrenza, con effetti negativi in termini di prezzi più elevati e produzione minore. Venivano anche riconosciute barriere naturali, dovute ad esempio ai costi di trasporto, che limitavano la zona di influenza dei produttori.
Se la libertà di movimento dei capitali è alla base del processo concorrenziale dei classici, ci si può chiedere quale ruolo svolgano la numerosità dei produttori e, connessa con questa, l'impossibilità da parte del singolo produttore di influire sul prezzo di mercato, a causa della limitatezza della sua produzione. In linea di massima si può dire che il processo concorrenziale descritto dai classici comporta la presenza di numerosi produttori in ciascun ramo; nel modello di base da essi formulato, infatti, a capitale circolante, il capitale del singolo imprenditore è limitato e non ci sono economie di scala.

3. La concorrenza come processo: asimmetria tra i produttori

a) Cournot e il vantaggio temporaneo

Il problema che si pone Augustin Cournot è il seguente: perché i duopolisti non colludono? La risposta consiste nella constatazione dell'esistenza di un vantaggio momentaneo, da parte di ciascun duopolista, nel non collaborare. Cournot assume che sul mercato si realizzi l'equilibrio tra domanda e offerta con un prezzo determinato dall'incontro delle due curve: i produttori decidono ciascuno la quantità da offrire e, data la domanda, il prezzo dipenderà dalla somma delle quantità prodotte.

Sulla base di questa impostazione Cournot stabilisce un rapporto molto preciso tra numero dei produttori e intensità della concorrenza: più alto è il numero, maggiore è la produzione e più basso è il prezzo. La tecnica analitica è quella dell'ipotesi di massimizzazione del profitto, supponendo date le quantità degli altri produttori. In questo modo si stabilisce che, dato un certo numero di imprese, l'entrata di un nuovo produttore porta a una contrazione della produzione da parte delle imprese già esistenti; ciò favorisce l'incremento della produzione da parte dell'ultimo entrato, finché, supponendo che tutte le imprese abbiano una eguale funzione di costo, la produzione non sia ripartita tra esse in parti eguali. Con l'aumento del numero delle imprese, pur essendo minore la produzione pro capite, aumenta la produzione totale, finché il prezzo non diventa eguale al costo unitario; in tale situazione di concorrenza illimitata la produzione di ciascuna impresa è infinitesima rispetto al totale, e quindi qualsiasi variazione da parte del singolo produttore non avrebbe influenza sull'equilibrio del mercato. L'intensità della concorrenza dipende quindi dal numero dei produttori: tuttavia, se l'entrata sul mercato non è libera, il numero non basta ad assicurare una vera e propria concorrenza. La posizione dei classici che davano rilievo alla libertà di entrata, e non al numero, da Cournot in poi è dunque capovolta.

b) Bertrand e l'asimmetria nel prezzo

Come si è accennato, il meccanismo di funzionamento del mercato cournotiano è il seguente: i produttori fissano le quantità e i consumatori in concorrenza tra loro determinano un unico prezzo di equilibrio al quale la produzione è venduta senza eccessi di domanda o di offerta. Questo meccanismo dovrebbe rappresentare in forma stilizzata il fenomeno del ritardo nella

produzione, e quindi dell'impossibilità da parte del singolo produttore di reagire istantaneamente a scelte effettuate dagli altri produttori. Supponiamo ora un meccanismo diverso: i produttori fissano indipendentemente il prezzo di vendita; i consumatori, noti i vari prezzi di offerta, si rivolgono a chi ha offerto il prezzo più basso; questo produttore soddisfa il mercato, mentre gli altri rimangono completamente esclusi.Il comportamento asimmetrico è qui centrato sul prezzo e il risultato che si ottiene differisce nettamente da quello cournotiano; è chiaro infatti che, a parità di costi tra i produttori, le imprese fisseranno il prezzo a livello del costo medio minimo, e ciò si verificherà purché vi siano almeno due produttori. Qui il risultato pienamente concorrenziale non dipende dal numero delle imprese.Va notato peraltro che, malgrado questa importante differenza, il meccanismo concorrenziale qui ipotizzato (che prende il nome da Joseph Bertrand, il matematico francese che avanzò questa ipotesi in una recensione critica di Cournot) è del tutto analogo a quello di Cournot: ciascun produttore fissa il prezzo in modo da avvantaggiarsi sul concorrente o evitare di essere superato da questi. Il criterio di fissare il prezzo più basso possibile presenta una duplice natura, offensiva e difensiva al tempo stesso. A parità di costi il processo asimmetrico si svolge in un tempo logico e determina immediatamente il risultato di un comportamento simmetrico: tutti i concorrenti fissano il prezzo a livello di costo (v. Bertrand, 1883).

c) Schumpeter e l'innovazione

Anche alla base del processo concorrenziale di Joseph Schumpeter vi sono il verificarsi di una situazione di asimmetria e il suo ricomporsi. Tale meccanismo tuttavia si svolge effettivamente nel tempo, e si verifica a causa del fatto che alcune imprese si trovano in una situazione diversa da quella delle altre, per una particolare capacità innovativa degli imprenditori che le dirigono. L'innovazione pone queste imprese in una posizione di vantaggio e di profitti più alti, cioè in una situazione di temporaneo monopolio; ma questa situazione è appunto solo temporanea, in quanto si verifica una serie di atti di imitazione e di emulazione da parte delle altre imprese, che non vogliono permanere in una situazione di inferiorità. In tal modo avviene la diffusione dell'innovazione e quindi la scomparsa del potere monopolistico. Se questo è l'effettivo funzionamento del meccanismo concorrenziale, è chiaro che monopolio e concorrenza non sono due ipotesi inconciliabili, ma semplicemente due momenti diversi ed entrambi necessari di un comune processo di sviluppo. Schumpeter

considera anche l'ipotesi che il meccanismo non operi nella seconda fase e quindi che l'innovazione non possa generalizzarsi; in tal caso la posizione di vantaggio dell'impresa diviene permanente e il profitto si trasforma in una rendita di monopolio. Tale ipotesi, peraltro, appare atipica nel sistema di Schumpeter, mentre il caso normale è rappresentato dal processo concorrenziale sopra delineato.

Va detto che sin dalla sua prima opera (Teoria dello sviluppo economico) Schumpeter sottolineava come uno dei casi di innovazione da lui considerati, cioè il mutamento nell'organizzazione della produzione, comportasse l'aumento delle dimensioni aziendali attraverso l'introduzione della produzione su larga scala.In un'opera successiva (Cicli economici) vengono distinte due diverse fasi storiche, quella del capitalismo concorrenziale, in cui le dimensioni delle imprese sono piccole rispetto al mercato e le innovazioni comportano normalmente la costituzione di nuove imprese, e quella del capitalismo dei grandi trust, dove invece si assiste all'affermarsi delle grandi unità produttive e il processo innovativo si verifica prevalentemente all'interno di tali aziende, senza quindi la nascita di nuovi complessi.

Tuttavia nella sua ultima opera (Capitalismo, socialismo, democrazia) Schumpeter esprime una valutazione positiva delle pratiche monopolistiche poste in essere da queste grandi imprese, e critica la politica antimonopolistica degli Stati Uniti. Se infatti la concorrenza non avviene tra le piccole imprese che producono tutte le stesse merci con gli stessi metodi, ma tra gli imprenditori innovatori e gli imprenditori imitatori, allora va considerato che la grande impresa svolge un ruolo determinante proprio nel campo dell'innovazione. Quindi iniziative come i contratti vincolanti a lungo termine, la fissazione di prezzi rigidi, e simili, sono certamente pratiche monopolistiche che determinano la sospensione del funzionamento della concorrenza per un certo periodo di tempo e il conseguimento di sovraprofitti da parte dei grandi gruppi; ma sono al tempo stesso forme di assicurazione contro i rischi cui questi ultimi vanno incontro e che essi non affronterebbero altrimenti, così come non li affrontano le piccole imprese concorrenziali dei libri di testo. Inoltre in tali modi si ottengono i mezzi necessari per affrontare programmi di ricerche a vasto raggio; queste pratiche monopolistiche proteggono quindi lo sviluppo più che danneggiarlo: è un'affermazione non più paradossale di quella per cui le automobili possono correre più veloci proprio perché sono dotate di freni.

4. La concorrenza come processo: il meccanismo degli scambi

Una terza nozione di processo concorrenziale è quella elaborata nell'ambito dell'impostazione neoclassica da Francis Edgeworth. La scuola neoclassica di Stanley Jevons, Léon Walras e Alfred Marshall fonda la propria teoria del valore sulla nozione di utilità marginale e sul processo di scambio. La concorrenza è quindi intimamente legata al funzionamento del mercato; ora, se avviene uno scambio tra due individui è perché ciascuno apprezza il bene dell'altro più del proprio. Tuttavia in un atto isolato di baratto i termini di scambio, o prezzi relativi tra i due beni, non possono essere fissati con esattezza, ma si possono stabilire solo dei prezzi minimi e massimi, che ciascuno è disposto a ricevere o a pagare.L'idea fondamentale è che questa indeterminatezza cede il posto a un'estrema determinatezza nel caso di un mercato perfettamente funzionante, e cioè in un mercato in cui vi sono numerosi scambisti, che agiscono indipendentemente e sono ben informati sulla situazione del mercato. In questo caso non esiste più spazio per l'abilità nella trattativa tra le parti, perché per ogni bene in un dato mercato si stabilisce un solo prezzo di equilibrio tra le quantità offerte e quelle domandate.

Il problema è di chiarire in qual modo vengono annullate le tendenze collusive tra alcuni scambisti ai danni di altri. Jevons afferma esplicitamente che si ha concorrenza se non vi è collusione per limitare l'offerta e determinare rapporti innaturali di scambio; del resto già Smith aveva parlato di spontanee tendenze collusive tra i produttori. L'analisi di Edgeworth si propone quindi il compito di spiegare il processo che porta dall'indeterminatezza del baratto alla determinatezza del prezzo relativo in un perfetto terreno concorrenziale. Tale processo analizzato astrattamente richiede due condizioni: la libertà di ricontrattazione da parte di ciascun soggetto con qualunque altro, in modo che lo scambio avvenga effettivamente solo dopo che tutti gli operatori avranno verificato che non esiste un'occasione migliore; la perfetta divisibilità dei beni oggetto dello scambio, che si traduce nella possibilità da parte di ciascun individuo di scambiare dosi infinitesime e quindi di disporre, ultimati gli scambi, di un insieme continuo di possibili combinazioni di beni.

Va sottolineato che l'ipotesi di ricontrattazione impedisce la formazione di tendenze collusive, e questo fenomeno si verifica non solo quando vi sono numerosi, al limite infiniti, scambisti, ma già quando abbiamo un paio di venditori e compratori. La condizione di numerosità è peraltro necessaria per eliminare l'indeterminatezza del prezzo di equilibrio. L'ipotesi di ricontrattazione svolge una funzione essenziale, soprattutto perché elimina la

possibilità e la convenienza degli operatori a stabilire accordi collusivi. Ora questa ipotesi, essenziale nell'analisi astratta, deve trovare un sostituto, per quanto imperfetto, per i normali mercati competitivi; l'ipotesi di Edgeworth è quella della libera informazione tra gli scambisti. Egli immagina scambisti presenti fisicamente in uno stesso luogo o collegati per telefono. Questo requisito di una perfetta conoscenza della situazione di mercato diverrà, come vedremo, uno dei requisiti della concorrenza perfetta come forma di mercato.

5. La concorrenza come situazione caratterizzata

Nella maggior parte degli autori tra Ottocento e Novecento la concorrenza non viene analizzata quale processo determinato da precisi meccanismi, ma come una situazione caratterizzata da alcuni risultati e legata a presupposti che vengono ipotizzati come esistenti. La concorrenza si presenta come un fenomeno definibile sotto vari aspetti che, a seconda degli interessi dei vari autori, vengono diversamente sottolineati. Così, ad esempio, la concorrenza viene definita come il meccanismo di incontro tra domanda e offerta legato al grado di perfezione del mercato (Walras), oppure come eguaglianza dei ricavi marginali in tutti i settori (Clark), e infine come quella situazione in cui la curva di domanda per il singolo produttore è infinitamente elastica. Questa definizione di concorrenza è implicitamente presente negli autori citati in precedenza, in Walras come in Marshall. Proprio nell'appendice matematica dei Principî di economia, ad esempio, Marshall definisce correttamente il ricavo marginale come la somma algebrica delle due componenti: variazione della quantità moltiplicata per il prezzo, e variazione del prezzo moltiplicata per la quantità. Ora la variazione del prezzo operata da una singola impresa risulta estremamente limitata - o addirittura trascurabile - in un mercato in cui le imprese sono molto numerose. Come si vede il ragionamento è del tutto analogo a quello di Cournot.D'altra parte Marshall parla anche di un mercato particolare di ciascun produttore, immerso, per così dire, in quello più ampio dell'intero settore industriale, e di una curva di domanda relativamente rigida, in riferimento alla singola impresa. Questa non è presentata come una situazione inconciliabile con l'ipotesi di concorrenza; Marshall ne parla come di una situazione da esaminare in un'ulteriore approssimazione alla realtà e, in special modo, quando si fa riferimento al breve periodo e in condizioni di recessione. Tuttavia, volendo definire analiticamente la concorrenza perfetta, Arthur C. Pigou e gli altri autori della scuola marshalliana ipotizzano l'eguaglianza tra ricavo

marginale e prezzo, e in questo non si può dire che tradiscano esplicitamente Marshall.
Su questa definizione si incentra una delle critiche fondamentali di Piero Sraffa (v., 1926). Se il prezzo, infatti, è dato, la singola impresa può trovare un limite alla produzione solo se i costi marginali, e quindi anche i medi, sono crescenti. La curva dei costi dell'impresa deve quindi presentare la tipica forma a U.Nel caso della produzione agricola e mineraria questa definizione della concorrenza perfetta può essere facilmente conciliabile con il reale funzionamento del mercato. In questi settori la produzione avviene a costi crescenti e il singolo produttore è realmente price taker; non a caso, peraltro, in questi mercati il compito di determinare l'equilibrio tra domanda e offerta è svolto da un limitato numero di intermediari, la cui essenziale funzione è già stata sottolineata da Walras.
Nei settori industriali, invece, questo modello di concorrenza va incontro a due formidabili critiche: innanzitutto le condizioni di produzione non possono essere descritte da una curva a U quanto piuttosto da una curva a L, e in effetti i produttori, al livello di produzione di equilibrio, si trovano ancora in una situazione di costi decrescenti. Inoltre il singolo produttore industriale non è normalmente un price taker, non fronteggia un'unica domanda comune agli altri produttori, sia proveniente dalla massa dei consumatori sia da intermediari specializzati. Il problema tipico dell'impresa industriale deriva invece dal fatto che una maggior produzione sarebbe perfettamente possibile a prezzi costanti, ma la possibilità di espanderla è limitata alla domanda. Il ricavo marginale non coincide più col prezzo; la seconda componente di Marshall è rilevante e sistematicamente presente.
Un mercato composto da numerosi produttori non è quindi necessariamente un mercato che approssima, per quanto imperfettamente, il modello puro della concorrenza perfetta; le imprese che vi operano possono essere del tutto indipendenti le une dalle altre, e tuttavia dotate ciascuna di una propria curva di domanda inclinata negativamente. Sraffa invita quindi a dirigere l'analisi non verso il modello di concorrenza perfetta ma verso l'opposto, il modello di monopolio. Le cause che determinano la rottura del mercato in una serie di micromercati collegati tra loro sono estremamente varie; il punto da sottolineare è che esse non sono elementi di frizione, ma fattori sistematici e permanenti.
Le ulteriori argomentazioni di Sraffa indicano una soluzione analoga alla massimizzazione congiunta dei profitti, limitata però dalla possibilità di entrata di qualche concorrente potenziale; l'entrata è peraltro ostacolata proprio dalle cause che determinano un mercato preferenziale per le imprese esistenti, sicché questi ostacoli potrebbero essere superati solo in

caso di profitti particolarmente superiori al normale.È ben noto come l'articolo di Sraffa abbia svolto un ruolo fondamentale nella critica dell'apparato marshalliano e abbia ispirato il lavoro di Joan Robinson sulla concorrenza imperfetta. Peraltro la Robinson non approfondisce le indicazioni di Sraffa sul tema dell'entrata limitata: si limita a svolgere l'analisi dell'equilibrio del monopolista con l'uso della nozione di ricavo marginale e dell'equilibrio di tangenza tra costo medio e prezzo.
L'ordine di considerazioni che porta Edward Chamberlin a sviluppare la sua teoria della concorrenza monopolistica come teoria più generale, in grado di comprendere la concorrenza perfetta come caso particolare, è diverso nell'impostazione e nelle finalità da quello di Sraffa, anche se per molti aspetti le analisi presentano delle somiglianze, come ad esempio per quanto riguarda le cause che determinano le curve di domanda individuali. Si considerino i seguenti elementi caratterizzanti un mercato: 1) numerosità delle imprese; 2) libertà di entrata; 3) effetto trascurabile dell'iniziativa del singolo produttore sugli altri. Se aggiungiamo l'ipotesi di un prodotto omogeneo abbiamo la concorrenza perfetta e curve di domanda orizzontali per i singoli produttori, mentre se aggiungiamo un prodotto eterogeneo abbiamo curve di domanda inclinate. Dunque la concorrenza perfetta è la situazione in cui l'elemento della differenziazione del prodotto tende a scomparire. Il risultato è ben noto: tangenza della curva dei costi, decrescente, con la curva di domanda, anch'essa decrescente; l'elasticità della domanda nel punto di equilibrio è maggiore dell'unità.
Questo equilibrio concorrenziale del 'largo gruppo' di Chamberlin si differenzia dalla concorrenza perfetta, quindi, solo per l'elasticità finita della domanda; come conseguenza le dimensioni delle imprese sono subottimali, ma questa caratteristica non va considerata come un aspetto negativo, bensì come la necessaria conseguenza degli elementi che soddisfano le preferenze dei consumatori. Si tenga presente che Chamberlin ipotizza anche uno stato di conoscenza perfetta, e pertanto non sono elementi di ignoranza a determinare l'inclinazione delle curve di domanda. L'impostazione di Chamberlin si propone quindi di salvare la sostanza dell'impostazione marshalliana eliminandone l'aspetto più debole.Il problema di questa impostazione consiste nella compatibilità tra l'ipotesi di differenziazione del prodotto e le due ipotesi di libertà di entrata e di non rivalità, cioè di mancanza di effetti percepibili dell'iniziativa del singolo.
Quest'ultima ipotesi implica che le conseguenze della manovra su prezzo e quantità da parte della singola impresa si diffondano uniformemente su tutti gli altri concorrenti e siano quindi

trascurabili. Viceversa l'azione dell'insieme degli altri produttori è rilevante e determina uno spostamento della curva di domanda del singolo.
Tuttavia la natura stessa degli elementi che determinano la differenziazione - caratteristiche oggettive del prodotto o soggettive del produttore, localizzazione geografica - fa sì che, definito un gruppo di prodotti simili ma differenziati, esisterà sempre al loro interno una graduazione nella differenziazione, con la conseguenza che l'iniziativa di un singolo produttore avrà delle ripercussioni differenziate sugli altri. Connesso con questo problema si presenta quello della libertà d'entrata; essa sarà più o meno difficile, nel senso che comporterà costi più o meno elevati, in riferimento a diversi prodotti e produttori del largo gruppo. Queste osservazioni portano a concludere che la presenza simultanea delle quattro caratteristiche della concorrenza monopolistica di largo gruppo non sia logicamente sostenibile, e che quindi una teoria della concorrenza fondata su questa base presenti inconvenienti analoghi a quelli dell'impostazione marshalliana.
Al posto di un largo gruppo avremmo allora una catena di piccoli gruppi i cui prodotti sono differenziati; ogni produttore si trova a far parte di più gruppi che parzialmente si sovrappongono. In ciascuno di essi esiste una rivalità potenziale e le variazioni dei prezzi di ciascun produttore troverebbero una reazione da parte degli altri. Questa situazione, analizzata da Chamberlin come oligopolio più differenziazione, è più simile a quella che aveva delineato Sraffa ed esplorato la Robinson, e non è un caso che anche Chamberlin arrivi a conclusioni in parte simili. In questa situazione infatti è probabile che i concorrenti non usino la manovra del prezzo come strumento concorrenziale, ma piuttosto le spese di vendita; inoltre, nella misura in cui esiste la libertà di entrata, si determina un eccesso di capacità produttiva, mentre se la libertà di entrata è limitata persistono sovraprofitti. La differenziazione del prodotto non è più corrispondente a quella richiesta dai gusti e dalle necessità dei consumatori, ma diviene eccessiva. La concorrenza non implica più l'efficienza nell'allocazione delle risorse.
Questa conclusione critica cui si perviene tramite la concorrenza monopolistica era stata prevista da almeno due autori, Vladimir Karpovič Dmitriev e Harold Hotelling. Il primo, in un saggio critico su Cournot, pone in luce il fatto che il processo di concorrenza comporta inevitabilmente degli sprechi in termini di eccessi di produzione invenduti o di scorte, in modo analogo agli armamenti che le nazioni devono produrre al fine di garantire il mantenimento della pace; le armi vengono prodotte e devono essere prodotte anche se si è

consci che non saranno mai usate (v. Dmitriev, 1898-1902).
Il secondo autore, esaminando in particolare il fenomeno della differenziazione spaziale, cioè della diversa localizzazione delle imprese, mostra come il meccanismo alla Bertrand non necessariamente fa sì che l'impresa che fissa il prezzo più alto perda tutta la clientela; vi è infatti un fattore di vicinanza che può rendere ancora conveniente al consumatore rivolgersi a una certa impresa (più vicina) anche se il prezzo è più alto. Esaminando il comportamento delle imprese volte a scegliere la migliore localizzazione, Hotelling mostra che la tendenza delle imprese è di concentrarsi geograficamente in una stessa zona, laddove dal punto di vista sociale l'ottimizzazione della localizzazione dovrebbe seguire un criterio del tutto diverso. Questa conclusione viene poi generalizzata affermando che il processo concorrenziale caratterizzato da qualche forma di differenziazione non comporta il raggiungimento di una situazione di efficienza (v. Hotelling, 1929).

6. Insoddisfazioni nella teoria della concorrenza

La posizione della nozione di concorrenza agli inizi degli anni cinquanta appare per alcuni versi paradossale: mentre da un lato, infatti, essa assume un ruolo fondamentale per una serie di risultati ottenuti dall'analisi dell'equilibrio economico generale (al punto che John Hicks dichiara che mettere in dubbio l'ipotesi di concorrenza pura equivale a minare buona parte dei risultati della teoria economica), da un altro lato, quello dell'analisi degli equilibri parziali, i risultati sono negativi.In effetti le ben note proprietà ottimali del sistema concorrenziale, enunciate da Walras e ulteriormente definite da Vilfredo Pareto, richiedono l'ipotesi di mercati di libera concorrenza e quindi un comportamento da price takers degli operatori. In tali condizioni sarebbe assicurata la massima soddisfazione possibile per tutti gli operatori, soggetta alle famose condizioni di un unico prezzo d'equilibrio, cioè un prezzo tale da eguagliare domanda e offerta e da essere eguale ai costi dei servizi necessari per produrlo.Walras sostanzialmente si limita a contare il numero delle equazioni e delle incognite, cosa questa che non assicura né la stabilità né l'unicità, e proprio sul tema dell'unicità è stato criticato da Knut Wicksell. Da questo punto di vista Gerard Debreu e Kenneth J. Arrow portano nei primi anni cinquanta un contributo analitico fondamentale sul tema dell'esistenza dell'equilibrio. Tuttavia questa analisi tralascia il problema del meccanismo di funzionamento del mercato di concorrenza; il doppio mercato d'asta di

Walras, ad esempio, non trova in realtà riscontro nel funzionamento della maggior parte dei mercati, in particolare di quelli industriali. La teoria assiomatica dell'equilibrio economico generale deve quindi assumere esplicitamente come postulato il comportamento da price takers degli operatori, senza essere in grado di spiegare come e se il funzionamento di un mercato e degli operatori conduca a questo risultato.
D'altra parte, anche sulla base dell'analisi degli equilibri parziali i risultati sono sostanzialmente negativi. La concorrenza pura è riconoscibile solo nei mercati agricoli e minerari; in quelli industriali il modello che cerca di adattarsi meglio è quello della concorrenza monopolistica di largo gruppo di Chamberlin, che - come si è visto - ha grosse difficoltà di coerenza interna e inoltre presenta risultati non corrispondenti alle caratteristiche dei mercati industriali. È infatti assente il dato più saliente della concentrazione industriale, e quindi delle economie di scala e delle discontinuità tecnologiche. Non a caso gli esempi più frequenti di tale modello si rifanno al mercato delle vendite al dettaglio: in tale settore John Stuart Mill e Knut Wicksell avevano osservato che l'eccessivo numero dei dettaglianti non è a vantaggio ma a svantaggio dei consumatori, capovolgendo con questa affermazione quanto era stato sostenuto proprio a questo proposito da Smith.
Né l'analisi dell'oligopolio sembra fornire qualche utile indicazione; questa forma di mercato viene esaminata in genere con modelli di mercato chiuso, con pochi produttori in grado di influenzarsi a vicenda. Ora in tale situazione non è utilizzabile con profitto il metodo di partire da alcuni dati, o variabili esogene, e, attraverso l'ipotesi di comportamento massimizzante, trovare la soluzione di equilibrio, mentre lo è con i modelli della concorrenza perfetta e del monopolio. Malgrado numerosi sforzi l'impressione generale è che sia confermata l'idea che la gamma di casi intermedi non sia riducibile a una coerente analisi economica; l'ipotesi più probabile resta quella di una massimizzazione congiunta dei profitti (v. Fellner, 1949), ma si tratta pur sempre di un'ipotesi tra le tante possibili.

7. La concorrenza potenziale e le barriere all'entrata

La concorrenza proveniente da nuove imprese è talvolta indicata da vari autori come la causa che limita la facoltà delle imprese di elevare i prezzi, anche in situazioni di monopolio.Il problema dei costi d'ingresso trova un ampio spazio nei primi tentativi di elaborare una teoria che non sia basata sull'ipotesi di massimizzazione di breve periodo, ma su un pur vago

concetto di profitto normale di lungo periodo (v. Andrews, 1949; v. Fellner, 1949). È tuttavia con Joe Bain e Paolo Sylos Labini che il problema della concorrenza potenziale e dei fattori limitanti viene posto al centro di una teoria della determinazione dei prezzi nei mercati industriali.

Già in un primo lavoro Bain (v., 1949) aveva utilizzato l'idea dell'entrata potenziale come modo per limitare la fissazione di un prezzo di tipo monopolistico, ma è con il lavoro sulle barriere alla nuova concorrenza che l'autore sviluppa l'analisi delle cause da cui dipende il prezzo limite, inteso come un livello di prezzo che non stimola l'entrata di nuove imprese. Il concetto di entrata è definito come l'aggiunta di nuova capacità produttiva a quella esistente nel settore; per quanto restrittiva, tale definizione ha dei vantaggi in un'analisi di tipo parziale, e si ricollega direttamente alla nozione dei classici. Le cause che determinano le barriere sono: differenziazione del prodotto, vantaggi di costo in termini assoluti, elevati fabbisogni iniziali di capitale, economie di scala. Si ricorderà che uno dei punti deboli del modello di largo gruppo di Chamberlin è la contemporanea presenza delle condizioni di differenziazione del prodotto e di libertà di entrata. Bain sottolinea appunto che la differenziazione comporta inevitabilmente un vantaggio del produttore già esistente rispetto a quello nuovo, vantaggio che si riflette nei costi aggiuntivi cui l'impresa nuova deve andare incontro. Nella sua ampia indagine su venti settori produttivi Bain ha indicato proprio la differenziazione come il fattore che più contribuisce alle barriere all'entrata, non solo tra i beni di consumo ma anche tra quelli strumentali.

Tralasciamo le altre due forme di barriera e veniamo alle economie di scala, fenomeno essenzialmente tecnologico. Il fatto che per la produzione di determinate merci siano necessari impianti di notevoli dimensioni e quindi ingenti capitali iniziali è stato sottolineato da molti autori. I due aspetti rilevanti sono: 1) esiste una quantità di prodotto minima al di sotto della quale i costi fissi hanno una tale incidenza da rendere il prodotto assolutamente non competitivo; 2) esiste una tendenza del costo medio, al di là del livello minimo, a mantenersi costante fino alla piena utilizzazione dell'impianto, quando i costi medi divengono fortemente crescenti (v. Bain, 1956).Sylos Labini (v., 1956) analizza solo le barriere all'entrata che derivano da economie di scala. L'ipotesi sull'andamento dei costi è in parte simile a quella di Bain, ma con qualche diversità significativa: il costo medio è composto da un costo fisso, unitario, che diminuisce al crescere della produzione, e un costo variabile costante: quindi più che un livello minimo esiste una fascia di produzione dove il

costo medio è fortemente decrescente. Ma più importante è il fatto che Sylos Labini sottolinea l'importanza delle discontinuità di scala; esistono diverse tecnologie con coppie di costi fissi e variabili inversamente correlati, mentre la massima capacità produttiva per ciascuna tecnologia è funzione diretta del costo fisso.
Sylos Labini suppone che ciascuna impresa operi con un solo stabilimento e che ogni stabilimento, compreso quello del potenziale entrante, produca al massimo della capacità; mentre la prima ipotesi è solo semplificativa, la seconda ha maggiore importanza e si può giustificare proprio ricordando che il costo medio minimo si trova appunto a quel livello. Implicitamente, quindi, vi è l'idea che a ciascuna singola impresa, indipendentemente dalla dimensione, convenga ampliare il più possibile la produzione, anche se ciò comporta un prezzo più basso, perché questo fatto è compensato da un costo medio più basso.Le caratteristiche della domanda, nonché le caratteristiche della tecnologia, determinano la strategia delle grandi imprese; queste hanno l'obiettivo di non far entrare altre grandi imprese e a questo scopo fissano la produzione a un livello tale che, nel caso di entrata di un'altra grande impresa, la quantità totale determinerebbe un prezzo più basso del costo minimo. In questo modo viene fornita una base teorica alle ipotesi di variazioni del prezzo secondo la formula del costo pieno (mark up). La teoria della concorrenza potenziale, salutata come una svolta risolutiva per spiegare il comportamento concorrenziale delle moderne strutture produttive, ha ricevuto una serie di critiche che vanno dalla negazione della nozione stessa a diverse ipotesi e complicazioni del modello di base, soprattutto in contesti dinamici.La critica più radicale è stata condotta alla nozione stessa di barriera all'entrata; si è argomentato che per l'esistenza di un ostacolo all'entrata di un concorrente potenziale deve sussistere un costo addizionale, netto, rispetto a quelli sopportati dalle imprese operanti. Come si è visto, per Bain questo è un caso di barriera, e la situazione più tipica è costituita da un brevetto; per i critici invece solo in questa situazione si può parlare di barriera. È stata quindi sviluppata un'analisi delle implicazioni in termini di benessere che ciò può comportare; in questa sede si può solo accennare al legame tra questa critica e quella che si è sviluppata all'interno della teoria dei giochi, su cui ci soffermeremo più avanti.
Una seconda critica riguarda invece il prezzo di esclusione, che concorrerebbe a determinare il prezzo effettivamente praticato dalle grandi imprese. Infatti se tutte le imprese, operanti o potenziali, sono a conoscenza delle funzioni di domanda e di costo, allora appare decisiva l'ipotesi formulata dall'impresa potenziale circa il comportamento delle imprese operanti nel

caso del suo ingresso. Quindi le imprese operanti possono anche fissare un livello di prezzo più elevato di quello d'esclusione; si tratta di vedere se esse hanno effettivamente una capacità produttiva tale da rendere credibile l'ipotesi eventualmente formulata dall'impresa potenziale circa un loro atteggiamento non collaborativo o addirittura aggressivo.La capacità produttiva inutilizzata sarebbe la vera arma strategica che rende credibile la barriera all'entrata. L'implicazione di questa critica è che il ruolo della concorrenza potenziale e gli ostacoli a essa sarebbero essenziali per comprendere il comportamento delle imprese, ma piuttosto dal lato dell'eccesso di capacità che da quello dei prezzi d'equilibrio.Agli inizi degli anni ottanta una nuova teoria generale delle strutture dei mercati industriali, quella dei mercati accessibili (contestable markets), ha di nuovo posto il tema della concorrenza potenziale al centro di un'analisi dei meccanismi di competizione tra le imprese. L'ipotesi di base della teoria dei mercati accessibili è quella della concorrenza sui prezzi, alla Bertrand, e ciò distingue questa impostazione da quella di Bain e Sylos Labini che ipotizzano un meccanismo alla Cournot.
La conseguenza è che il risultato concorrenziale si raggiunge immediatamente anche con due sole imprese, e anzi una delle due imprese può essere anche un'impresa potenziale, cioè pronta a entrare qualora ritenga di poter praticare un prezzo tale da assicurarsi il mercato. Per questo motivo l'impresa formalmente monopolista non può mantenere un livello di prezzo maggiore del costo, pena la perdita del mercato a favore dell'impresa non ancora operante (v. Baumol e altri, 1982).Una delle differenze della teoria dei mercati accessibili rispetto a quella di Bain-Sylos Labini è che, mentre in quest'ultima un ruolo rilevante è svolto dai costi fissi, nella prima il costo fisso pone una barriera all'entrata solo se si tratta di un costo irrecuperabile (sunk). Un settore caratterizzato, ad esempio, da alti costi fissi ma non irrecuperabili è quello del trasporto, in particolare quello aereo: la teoria dei mercati accessibili ha svolto un ruolo importante nel campo delle premesse teoriche delle politiche di deregulation.

8. La teoria dei giochi e i processi concorrenziali

L'ultima analisi della concorrenza come processo era stata quella di Schumpeter del 1912; per circa quarant'anni non vi erano più stati contributi sulla concorrenza come meccanismo da spiegare e non da presupporre. Uno dei motivi era la grossa difficoltà di analizzare le

situazioni in cui gli operatori sono pochi; uno dei presupposti di base all'analisi della concorrenza come forma di mercato era, in fondo, che ciascuna impresa si limitava a considerare il prezzo, che per essa era un dato, e la funzione del costo, senza predisporre alcuna particolare strategia nei confronti dei rivali.

La teoria del comportamento economico di John von Neumann e Oskar Morgenstern è stata formulata esplicitamente sulla base della teoria dei giochi, secondo cui il risultato del gioco dipende dall'interazione tra la propria condotta e quella degli altri partecipanti; il criterio di razionalità, comune a tutti i partecipanti, non può essere quello proprio dell'impostazione tradizionale. Le metodologie sviluppatesi nel filone della teoria dei giochi hanno riportato l'attenzione degli studiosi sul concetto di concorrenza come processo; sono stati individuati due diversi approcci, definiti l'uno approccio cooperativo e l'altro approccio non cooperativo. Il primo fa riferimento all'impostazione di Edgeworth, il secondo a quella di Cournot. In entrambe le impostazioni gli operatori economici hanno un atteggiamento attivo nelle scelte riguardanti prezzi o quantità; il risultato rilevante è che al crescere del numero degli agenti economici le soluzioni di equilibrio convergono verso l'equilibrio concorrenziale, cioè verso una situazione in cui gli operatori agiscono come price takers e in cui valgono le condizioni walrasiane (v. Shubik, Strategy..., 1959; v. Shapley e Shubik, 1967; v. Mas-Colell, 1982). Entrambi gli approcci hanno rappresentato un sensibile passo in avanti dal punto di vista della teoria dell'equilibrio economico generale, perché hanno permesso di ottenere i risultati walrasiani senza dover assumere come postulato iniziale quello degli operatori come price takers . L'equilibrio competitivo viene raggiunto in modelli chiusi attraverso la sola estensione del numero degli operatori; la numerosità degli agenti economici appare quindi come la caratteristica fondamentale della concorrenza basata sulla teoria dei giochi. Sia l'approccio cooperativo che quello non cooperativo, tuttavia, presentano dei problemi e degli aspetti non risolti in modo soddisfacente. Vi è innanzitutto l'ipotesi della conoscenza perfetta da parte degli operatori economici; si è mostrato che in situazioni in cui l'informazione è limitata e costosa il processo di convergenza al risultato competitivo incontra dei problemi (v. Rubinstein, 1982). Ma il problema più serio è posto dalla difficoltà di trovare una convincente spiegazione alla ragione per cui in un modello chiuso - in cui quindi i produttori non temono l'entrata di nuovi rivali - non si determinano fenomeni di coalizioni tra imprese. La questione, è stato osservato (v. Arrow, 1974), è anche connessa con l'asimmetria nella distribuzione delle informazioni; è più facile infatti che gli accordi si formino tra gli operatori

dello stesso settore piuttosto che tra alcuni venditori e alcuni compratori. Per affrontare questi problemi sono stati compiuti dei tentativi di introdurre limitazioni e vincoli alle ipotesi di comportamento razionale e massimizzante degli operatori, in modo analogo, quindi, alla teoria della razionalità limitata di Herbert Simon. L'estensione dell'equilibrio non cooperativo alla teoria della concorrenza potenziale e delle barriere all'entrata è un altro tema sul quale si è sviluppata dalla fine degli anni settanta una crescente letteratura. Dal punto di vista formale, infatti, non occorrono particolari complicazioni per costruire un gioco in cui vi siano un'impresa operante e un'impresa potenziale entrante: si tratta di esaminare due momenti della strategia delle due imprese, il momento della scelta se entrare o no, e quello della scelta dei prezzi o della quantità di produzione.

Un primo risultato di rilievo fu ottenuto dimostrando che la teoria del prezzo di esclusione alla Bain-Sylos Labini non trova conferma nell'ambito dei giochi non cooperativi. La ragione è sostanzialmente analoga a quella adottata da coloro che ritengono infondata la nozione stessa di barriera all'entrata, a parte il caso del vantaggio assoluto nei costi. Infatti l'impresa potenziale entrante non può credere all'ipotesi di una produzione invariata da parte delle imprese esistenti, in quanto sa che tale comportamento non sarebbe nell'interesse delle imprese già operanti, una volta che l'entrata sia avvenuta. In questo caso, infatti, a quelle imprese conviene l'accordo piuttosto che un atteggiamento non cooperativo (v. Friedman, 1979). Un discorso analogo vale per la minaccia da parte dell'impresa operante dell'uso della sua capacità produttiva come deterrente contro la potenziale entrante: in realtà anche questa minaccia è poco credibile agli occhi della potenziale entrante.

Sulla base di una distinzione effettuata da Thomas Schelling si è cercato quindi di individuare i casi in cui la minaccia può essere credibile: sono i casi in cui l'impresa operante sul mercato pone in essere un comportamento che la obbliga a un atteggiamento non cooperativo nel caso di entrata effettiva dell'altra impresa. Un esempio tipico è quello dell'impresa che stipula accordi con i clienti, garantendo un prezzo almeno pari a quello che potrebbe essere offerto da un'altra impresa. Se la sanzione, in caso di mancato rispetto dell'accordo, è sufficientemente elevata, questo comportamento pone in atto una minaccia credibile e viene quindi a costituire una vera barriera all'entrata (v. Schelling, 1960; v. Stiglitz, 1985).

Anche l'esistenza di una capacità produttiva inutilizzata può costituire una minaccia credibile, non in se stessa, ma in quanto collegata, come del resto è plausibile che sia, a costi fissi

perduti di tale ampiezza da rendere comunque non conveniente la produzione a un potenziale entrante che voglia entrare aggiungendo i suoi costi fissi a quelli già esistenti. Sono stati comunque condotti dei tentativi di conciliare la teoria dell'entrata potenziale e del prezzo d'esclusione con l'equilibrio non cooperativo; in sostanza si sono introdotte limitazioni nei gradi di conoscenza da parte dei produttori: ad esempio, i potenziali entranti potrebbero non conoscere la funzione del costo del produttore operante sul mercato, e potrebbero assumere il prezzo che questi pratica come indicatore dei costi. Ma forse il vero problema risiede nelle limitazioni delle assunzioni di base dei giochi non cooperativi. Ad esempio lo schema impresa operante-impresa potenziale entrante assume implicitamente che la prima impresa sappia quanti (e quali) siano i concorrenti potenziali. Vale la pena di sottolineare che ciò che importa non è che i potenziali entranti siano in numero di uno, ma che siano in numero definito e in una ben definita sequenza dal punto di vista temporale.
Diversamente si pone invece il problema nel caso in cui l'impresa non sappia quanti e quali concorrenti potenziali può trovarsi a fronteggiare; in questo caso un atteggiamento non cooperativo volto a contrastare l'entrata del primo potenziale entrante può servire all'impresa per costruirsi una 'reputazione' tale da scoraggiare l'entrata di altre imprese e risultare redditizia a lungo andare (v. Kreps e Wilson, 1982). Situazioni di questo tipo si possono paragonare ai risultati raggiunti nell'applicazione della teoria dei giochi all'etologia (v. Smith, 1982): ad esempio il comportamento 'borghese', non cooperativo sul proprio terreno e cooperativo sul terreno altrui, potrebbe suggerire un modello valido anche per il comportamento dell'impresa. Altri sviluppi interessanti si possono cogliere nei tentativi di applicare la nozione di strategie miste al comportamento delle imprese (v. Dasgupta e Maskin, 1986). In ogni caso un accordo tra teoria dell'entrata potenziale e del prezzo di esclusione e teoria dei giochi sembra implicare il superamento di alcuni aspetti della nozione di equilibrio propria dei giochi non cooperativi.

9. Concorrenza e innovazione

Per lungo tempo i rapporti tra i processi innovativi e i regimi di mercato sono stati oggetto più di ipotesi convenzionali che di vere e proprie analisi. Se si prescinde da Schumpeter, occorre giungere a Josef Steindl e a Sylos Labini per un approfondimento degli effetti dell'innovazione nei diversi regimi di mercato; in sostanza l'ipotesi era che in presenza della

concorrenza si determina un processo di diffusione dei frutti del progresso tecnico in tutto il sistema attraverso la riduzione dei prezzi; mentre invece posizioni monopolistiche determinano il permanere dei guadagni all'interno dei settori dove si è verificata l'innovazione. Questa differenza avrebbe una serie di conseguenze sul livello degli investimenti e sulla distribuzione del reddito. In base a questa ipotesi si ritiene che il regime di mercato concorrenziale produca una spinta all'innovazione più forte di quello monopolistico; la concorrenza agirebbe come una 'frusta' sugli imprenditori, indipendentemente dalle loro inclinazioni naturali. In questa visione tuttavia la concorrenza viene concepita implicitamente come presenza di un alto grado di rivalità, effettiva o potenziale, tra le imprese, mentre la situazione monopolistica sarebbe caratterizzata proprio dall'assenza di tale rivalità. La critica di Schumpeter a questa tesi, e sulla sua scia quella di Galbraith, consiste, a ben vedere, in due distinte proposizioni: si nega innanzitutto che la rivalità sia propria del regime concorrenziale e la sua assenza di quello monopolistico, così come sono definiti tradizionalmente; si ritiene poi che la capacità finanziaria di predisporre la somma necessaria alle spese di ricerca sia maggiore nelle imprese dotate di potere di mercato, dato che le soglie minime necessarie per avviare programmi di ricerca sono piuttosto elevate ed esistono forse economie di scala nella produzione della ricerca.
Il dibattito sviluppatosi negli anni sessanta ha fornito alcune indicazioni interessanti ma non conclusive; il saggio di Arrow (v., 1962) può essere considerato come la prima dimostrazione analitica della tesi convenzionale. Si può notare tuttavia che la maggiore propensione all'innovazione del regime concorrenziale rispetto a quello monopolistico viene a dipendere dalle stesse ragioni che stabiliscono la superiorità in termini di efficienza allocativa della concorrenza rispetto al monopolio; infatti se l'equilibrio di mercato nella concorrenza è definito dall'eguaglianza tra costo marginale e prezzo, mentre nel monopolio è definito dall'eguaglianza tra il ricavo marginale e il costo marginale, allora il monopolio si caratterizza per una minore quantità prodotta e un più alto prezzo, e quindi esso comporta una 'perdita secca' per la collettività. In tale situazione un'innovazione che riduca il costo di produzione determina un vantaggio maggiore nell'ipotesi di concorrenza che in quella di monopolio, esattamente per le stesse ragioni per le quali si determina una perdita secca.
Il problema è tuttavia se sia corretto caratterizzare nei termini anzidetti il meccanismo di determinazione del prezzo nei due regimi di mercato. In effetti l'innovazione determina tipicamente una situazione di monopolio, almeno temporaneamente, dell'impresa che ha

avuto successo; il problema si presenta quindi come una questione di scelta tra una condizione di efficienza statica e una di efficienza dinamica.Le indagini empiriche hanno mostrato che esiste un ammontare minimo necessario per dare il via a progetti di ricerca, ma non esistono economie di scala; che le spese di ricerca e sviluppo crescono con la dimensione dell'impresa, ma in proporzione oltre un certo limite il rapporto decresce (con qualche eccezione, per esempio l'industria chimica). Sembra quindi che situazioni di 'eccesso di concorrenza', intesa come grande numerosità di imprese, o di 'troppo limitata concorrenza' nel caso opposto non giovino all'innovazione; sembra quindi necessario un certo grado di potere di mercato che non deve però essere eccessivo (v. Kamien e Schwartz, 1982).
L'analisi teorica dell'innovazione e delle spese per la ricerca e lo sviluppo è stata condotta da molti studiosi attraverso la teoria dei giochi; la metodologia è evidentemente quella dei giochi non cooperativi e vi hanno trovato un'applicazione le due ipotesi di concorrenza, quella alla Cournot e quella alla Bertrand. Nella prima i frutti della ricerca dipendono essenzialmente dalle spese effettuate da ciascuna impresa, indipendentemente da ciò che fanno le altre, mentre nella seconda ipotesi il frutto della ricerca, in senso economico, si ha solo quando l'impresa sia la prima a giungere a un determinato risultato; esso dipende quindi dalla brevettabilità o meno dell'innovazione.I risultati che si ottengono riflettono questi due diversi approcci. Nel primo caso, ad esempio, si mostra come il volume della ricerca dipenda dal numero dei concorrenti, ma la caratteristica nuova messa in luce è che le spese di ricerca hanno un effetto sulla struttura del mercato, ponendo un limite al numero delle imprese che vi possono operare e costituendo così una barriera all'entrata, con evidenti riflessi sulla determinazione dei prezzi. Nel secondo caso invece è dubbio che esistano soluzioni di equilibrio; in sostanza si può dire che in questo caso il problema delle imprese è decidere se le probabilità di vittoria nella ricerca sono buone o meno. Solo nel caso che siano buone conviene investire. Ciò può portare sia a eccessi di spesa per la ricerca, sia a carenze rispetto a ciò che sarebbe socialmente adeguato.
Un altro aspetto interessante che emerge quando la ricerca e lo sviluppo sono caratterizzati da meccanismi di gara, cioè nel caso della piena brevettabilità dell'innovazione da parte dell'impresa che ottiene per prima il risultato, è che l'analisi del comportamento strategico in situazioni asimmetriche indica la possibilità per l'impresa dominante di determinare delle barriere all'entrata attraverso una politica di ricerca volta alla 'saturazione' delle innovazioni. In questo caso avremmo una produzione di innovazioni che non vengono utilizzate se non

quando gli investimenti effettuati in precedenza siano stati completamente ammortizzati (v. Dasgupta e Maskin, 1986).
La letteratura sul tema non è ancora giunta a una soddisfacente sistemazione della tematica; i problemi connessi con la formalizzazione di situazioni complesse sono infatti numerosi. I casi reali difficilmente rientrano nei casi estremi analizzati: la brevettabilità dell'innovazione non è mai né piena né nulla, i rapporti tra azionisti, managers, addetti ai laboratori sono complessi e i risultati dipendono da ipotesi specifiche sulle motivazioni delle singole parti (v. Stiglitz, 1985). Allo stato attuale si può dire che esiste una generale convinzione che tra innovazione e concorrenza ci siano una notevole interdipendenza e nessi di causalità reciproci.

10. Teoria della concorrenza e politica antimonopolistica

Nei paesi anglosassoni, in particolare negli Stati Uniti, la politica antimonopolistica trova le sue radici nella common law, e ciò spiega perché l'antitrust sia una elaborazione dei tribunali. È chiaro tuttavia che vi sono sempre state strette relazioni tra teoria economica e antitrust e da questo punto di vista può sembrare sorprendente constatare che la maggioranza degli economisti ha un atteggiamento critico verso la legislazione antitrust, considerata in fondo inutile, se non controproducente dal punto di vista dello sviluppo dei processi concorrenziali. Questo atteggiamento critico è basato sull'evoluzione della teoria economica che, come si è cercato di mostrare nelle pagine che precedono, si è andata via via allontanando da una visione della concorrenza che la considera come caratteristica di un sistema atomistico, in cui ciascuna impresa è priva di potere nella fissazione del prezzo, e dove vige quindi la regola fondamentale dell'eguaglianza tra prezzo e costo marginale. In effetti, quando gli economisti ritenevano fondamentalmente valida questa concezione della concorrenza, la loro valutazione antitrust era certamente più positiva; nel noto pamphlet di Henry Simons (v., 1948) l'eliminazione del monopolio privato in tutte le sue forme era indicata come il primo obiettivo di un programma positivo per il laissez faire, e la Federal Trade Commission veniva considerata come la più importante delle agenzie governative.
Con la crisi della teoria tradizionale della concorrenza si pone in dubbio che il mercato atomistico abbia le caratteristiche di efficienza ipotizzate e si indicano invece nella libertà d'entrata e nell'innovazione le vere radici dei processi concorrenziali, e quindi la base di

un'efficienza intesa in senso dinamico.Come conseguenza di ciò alcune categorie dell'antitrust vengono poste in discussione: si prenda ad esempio il caso del predatory pricing. Sulla base della teoria tradizionale l'esistenza di una politica sleale di concorrenza sul prezzo sarebbe confermata da un prezzo inferiore al costo marginale, ma le teorie della concorrenza monopolistica e dell'oligopolio hanno dimostrato che quella condizione non è né necessaria né sufficiente perché l'atteggiamento dell'impresa che lo persegue possa essere qualificato come predatory pricing (v. Scherer, 1976).

Le analisi economiche che hanno colpito più in profondità la legislazione antitrust sono quelle che hanno messo in discussione il principio secondo cui gli accordi tra le imprese ricadono nella giurisdizione dei tribunali sulla base di una presunzione iuris et de iure, cioè di quella che nella tradizione statunitense è definita la per se rule. La distinzione, in effetti, tra tendenze alla collusione tra imprese diverse e tendenze alla monopolizzazione era stata stabilita dallo stesso Sherman act, dando luogo a due sezioni separate della legge; veniva considerata in modo nettamente più severo la prima sezione rispetto alla seconda. Le teorie dei costi di transazione e dei diritti di proprietà di Oliver Williamson e di Harold Demsetz hanno portato a una riconsiderazione degli accordi verticali tra le imprese; tali accordi possono essere spiegati da considerazioni di efficienza produttiva in presenza di costi di transazione, piuttosto che dalla smithiana conspiracy against the public (v. Williamson, 1975; v. Demsetz, 1982).

Il colpo più duro all'antitrust è stato inferto dall'analisi di Lester Telser, che ha sottoposto a critica anche la condanna dei cartelli orizzontali tra le imprese, condanna che si poteva considerare come l'unico punto sicuro di incontro tra giuristi dell'antitrust ed economisti. Basandosi sulla teoria dei giochi, in particolare sull'approccio cooperativo, Telser ha sostenuto che in condizioni di eccesso di capacità e/o di costi decrescenti può mancare un equilibrio stabile nel settore. In queste circostanze la fissazione di quote di produzione può essere un modo per raggiungere una soluzione di equilibrio che altrimenti mancherebbe (v. Telser, 1987). Il dibattito su queste tesi è in corso e al momento non è possibile prevederne gli sviluppi; si può tuttavia ricordare che nelle situazioni di assenza di equilibrio, quali quelle descritte da Telser e che egli stesso ritiene più diffuse nei mercati industriali di quanto gli studiosi dell'equilibrio economico generale non abbiano ammesso finora, i prezzi perdono la loro fondamentale funzione di allocazione delle risorse; il problema è allora se la realizzazione dell'equilibrio debba essere lasciata all'iniziativa delle parti o non richieda

piuttosto qualche forma di intervento pubblico.

Ricchezza
di Roberto Scazzieri

sommario: 1. Introduzione. 2. Ricostruzione storico-analitica del concetto. 3. Lavoro, scarsità e ricchezza. 4. Ricchezza, capitale e reddito: l'analisi classica dei fondi di ricchezza e flussi di reddito. 5. Reddito, consumo e ricchezza: ciclo vitale e capitale intergenerazionale. 6. Ricchezza, reddito e dinamica economica. □ Bibliografia.

1. Introduzione

La ricchezza di un sistema economico, oppure di un singolo individuo o gruppo sociale, può essere identificata con l'estensione del suo 'potere di disposizione' su beni e servizi (sia materiali sia immateriali). Questa definizione, piuttosto generale, può comprendere sia la disponibilità di beni e servizi a un dato istante o periodo di tempo, sia la loro disponibilità sotto il profilo intertemporale (quindi attraverso un'intera sequenza di periodi).
Il concetto di ricchezza è più 'antico' e, per così dire, fondamentale rispetto al concetto di reddito: il potere di disposizione collegato alla persistenza di un fondo o patrimonio (ed eventualmente alla sua trasferibilità da una generazione all'altra) è tradizionalmente apparso come una base più solida della disponibilità di un potere d'acquisto (reddito) ricorrente ma non necessariamente collegato a diritti certi riguardo al possesso e all'utilizzazione di beni e servizi.Secondo alcuni autori l'esistenza di ricchezza (individuale o sociale) è condizione necessaria per la formazione o il calcolo del reddito (v. Marshall, 1890; v. Edgeworth, 1925-1926; v. Fisher, 1906). Secondo altri autori, invece, la stessa ricchezza (intesa come potere di disposizione su beni e servizi) si caratterizza essenzialmente come effetto della formazione di reddito e della sua destinazione al risparmio piuttosto che al consumo (v. Smith, 1776; v. Storch, 1815; v. Keynes, 1936, cap. 16; v. Modigliani e Brumberg, 1954; v. Modigliani, 1986).
La diversità di approcci nell'analisi della ricchezza si sovrappone ma non coincide

interamente con la distinzione fra le concezioni della ricchezza individuale e quelle della ricchezza sociale. Nel primo caso si mette in evidenza la capacità individuale di produrre reddito (potere d'acquisto) e si identifica la ricchezza con gli elementi del patrimonio materiale, finanziario o intellettuale (capitale umano) a cui può essere ricondotta la formazione del potere individuale di acquisto. Nel secondo caso, invece, si mette in evidenza la capacità collettiva di produrre un flusso di beni e servizi che possano essere destinati alla formazione di nuovo reddito (prodotto netto), anziché semplicemente al mantenimento della capacità produttiva esistente.

2. Ricostruzione storico-analitica del concetto

Nella lingua italiana l'uso antico del termine ricchezza (al singolare) fa riferimento alla condizione generale di chi ha a disposizione quantità rilevanti di beni materiali o di denaro (secoli XIII-XIV). Le caratteristiche astratte di tale condizione sono individuate con lucidità già da Ferdinando Galiani (v., 1963, p. 130) nel suo saggio Della moneta (1750): "Ricchezza è il possesso d'alcuna cosa che sia più desiderata da altri che dal possessore".
L'impiego del termine ricchezze (al plurale) si distingue dal corrispondente singolare per la maggiore attenzione riservata ai beni fisici posseduti, o comunque tenuti a disposizione, da un soggetto (o comunità) a cui va riferita la condizione di ricchezza (cfr., ad esempio, C. Beccaria, Elementi di economia pubblica, Milano 1804). Si deve a Gerolamo Boccardo il richiamo alla distinzione, già operata da Bastiat (v., 1850), fra il significato relativo del termine 'ricchezza' (considerato caratteristico del linguaggio ordinario) e il suo significato assoluto (considerato caratteristico della categoria economica). Infatti, nel linguaggio ordinario si dice "ricco chi non è povero; e ricchezza significa il contrario di povertà", mentre nel linguaggio dell'economia politica il termine ricchezza "ha un significato assoluto, ed esprime in generale il complesso delle cose utili e godibili" (v. Boccardo, 1877², p. 910). Boccardo aggiunge: "Quanto più le utilità gratuite si moltiplicano [grazie ai progressi dell'industria], tanto è maggiore la [quantità] di ricchezze effettive possedute dalla nazione e dal genere umano. Ma un individuo, una famiglia, una limitata agglomerazione di persone sono tanto più relativamente ricchi, quanto è maggiore la somma di valori che possiedono". Infatti "la somma di ricchezze che possono ottenere per via di scambio, sulla massa delle ricchezze esistenti, è proporzionale alla somma dei valori medesimi" (ibid., p. 914).

Nella lingua inglese il termine wealth fa riferimento, in alcuni degli usi più antichi, alla "condizione di essere prospero e felice" (Oxford English dictionary, s. v.) e denota quindi una condizione generale di benessere. Presto tuttavia (impieghi del termine in questa accezione si trovano già a partire dal XIII-XIV secolo) il termine passa a indicare l'abbondanza degli oggetti posseduti da un individuo oppure da una comunità.
L'espressione wealth of nations (ricchezza delle nazioni) è usata già in epoca anteriore alla Inquiry di Adam Smith, in riferimento alle ricchezze collettive di un sistema economico (cfr., ad esempio, W. Petty, Two essays on political arithmetick concerning the people, housing, hospitals, etc. of London and Paris, 1687). Accanto al termine wealth è usato già attorno ai secoli XIII-XIV anche riches, che si caratterizza per una maggiore attenzione rivolta agli aspetti individuali e materiali della ricchezza. In ambito strettamente economico, si deve a Smith un'implicita distinzione fra riches e wealth, secondo cui "ciascun uomo è ricco o povero a seconda del grado in cui può permettersi di godere delle necessità, comodi e piaceri della vita umana" (v. Smith, 1776, ed. 1976, p. 47). In una società caratterizzata da una divisione del lavoro sufficientemente estesa, tuttavia, difficilmente il singolo individuo può soddisfare direttamente, con le risorse a sua disposizione e con il suo lavoro, una quota significativa delle proprie esigenze e desideri. In queste circostanze, i beni a disposizione di ciascun individuo dipendono dalla sua capacità (potere) di utilizzare a proprio vantaggio un particolare schema di divisione del lavoro. Il termine wealth (esplicitamente collegato - nell'opinione di Thomas Hobbes - all'esercizio di una forma di potere) richiama direttamente l'esistenza di un sistema complesso di rapporti sociali in cui la capacità di disporre del lavoro altrui passa in primo piano rispetto all'immediata disponibilità di certi beni (riches).
La dualità di significati che consente di distinguere wealth e riches ritorna negli economisti successivi a Smith, benché la terminologia usata non sia sempre la stessa. In particolare, si trova in Robert Torrens un impiego del termine wealth che mette in evidenza (nell'espressione man of wealth) "una quantità e significa abbondanza delle comodità e dei lussi della vita" (v. Torrens, 1821, p. 3). Viene in questo modo messa in evidenza una concezione di ricchezza come 'fondo', in cui è centrale la disponibilità di certi beni (o servizi) in un dato momento piuttosto che il corrispondente processo di acquisizione.Jeremy Bentham si distingue invece per la concezione della ricchezza come 'flusso', approfondita attraverso l'analisi del corrispondente processo di formazione: "Tutta la ricchezza è alternativamente la produzione spontanea della terra, oppure il risultato del lavoro impiegato nella coltivazione

della terra o sui materiali che la terra fornisce" (v. Bentham, 1825, p. 237). Le connessioni tra ricchezza fondo e ricchezza flusso sono studiate da John Stuart Mill, che individua anche in modo preciso la distinzione fra concezione individuale e concezione sociale della ricchezza. Francis Ysidro Edgeworth riprende alcuni aspetti di questa analisi sottolineandone le idee di scambiabilità e utilità e mettendo in risalto la possibilità di escludere materialità, trasferibilità e limitazione di quantità (ma non l'utilità) dai requisiti necessari della ricchezza (v. Edgeworth, 1925-1926, pp. 660-661).
Nella lingua francese il termine richesse (al singolare) fa inizialmente riferimento all"abbondanza' di qualunque specie di valore reale o monetario, per acquisire in seguito il significato più puntuale di potere d'acquisto connesso a capacità produttive reali (cfr. É. Bonnot de Condillac, Le commerce et le gouvernement considérés relativement l'un à l'autre, Paris 1776). Nell'impiego al plurale, il termine richesses indica soprattutto beni concreti considerati nel loro insieme. In ambito strettamente economico si osserva, già a partire dal XVIII secolo, una connessione diretta fra ricchezza, capacità produttiva e flusso di beni e servizi acquisiti per mezzo di essi (cfr. R. Cantillon, Essai sur la nature du commerce en général, London 1755).Si deve a Frédéric Bastiat un ampio esame della distinzione fra ricchezza effettiva e ricchezza relativa: "Si attribuiscono a buon diritto due significati al termine ricchezza: la ricchezza effettiva, vera, [...] ovvero la somma delle utilità che il lavoro umano, col concorso della natura, mette a disposizione delle società, e la ricchezza relativa, cioè la quota-parte proporzionale di ciascuno nella ricchezza generale, quota-parte che si determina in base al valore" (v. Bastiat, 1850, p. 211). La distinzione fra ricchezza effettiva e ricchezza relativa introduce un interessante collegamento fra analisi della ricchezza e teoria del valore, richiamando l'attenzione sul ruolo del valore (e della scambiabilità dei beni) nel determinare la ricchezza relativa di individui o gruppi sociali, ma non nel determinare la corrispondente ricchezza effettiva.

3. Lavoro, scarsità e ricchezza

Negli scritti degli economisti fisiocratici e classici il concetto di ricchezza è spesso usato in riferimento alla formazione di un flusso di beni (e servizi) prodotti e annualmente riproducibili all'interno di un sistema economico. In questo modo la teoria classica della ricchezza sposta la sua attenzione dalla ricchezza come grandezza fondo alla ricchezza come

grandezza flusso. Come ha osservato Luigi Pasinetti "'dovizia di beni' può significare una dotazione o fondo di beni esistenti, cioè ricchezza fondo (ricchezza da patrimonio), oppure può significare un cospicuo flusso periodico di beni e servizi, cioè ricchezza flusso o ricchezza da reddito [...] Anche se i due tipi di ricchezza non sono naturalmente disgiunti, la relazione tra i due non è affatto semplice, né chiara, né invariante rispetto all'evoluzione dei sistemi economici" (v. Pasinetti, 1975, p. 4).
Alcune anticipazioni dell'approccio fisiocratico e classico si trovano nel già citato Essai sur la nature du commerce en général di Richard Cantillon (v., 1755, p. 1), in cui si legge che "la terra è l'origine o materia da cui tutta la ricchezza è prodotta. Il lavoro dell'uomo è la forma che la produce: e la ricchezza di per sé non è altro che le necessità, comodi e agi della vita". Il collegamento fra ricchezza e produzione, già esplicito in Cantillon, viene precisato da François Quesnay (Tableau économique, 1758) e dagli altri autori fisiocratici attraverso la considerazione del prodotto netto del sistema economico. Alcuni anni dopo A.R.J. Turgot pubblica le Réflexions sur la formation et la distribution des richesses (1766), in cui l'attenzione per la formazione (produzione) e distribuzione di ricchezza implica "il riferimento immediato a una nozione di ricchezza flusso" (v. Pasinetti, 1975, p. 5). Il cambiamento di prospettiva rispetto alla concezione di ricchezza fondo riguarda non soltanto le coordinate analitiche del concetto di ricchezza, ma lo stesso processo che è alla base della formazione di ricchezza. Infatti, quando la ricchezza di un sistema economico è costituita da beni prodotti e riprodotti, essa "viene continuamente consumat[a] e rimpiazzat[a] dallo stesso processo di produzione; non è più qualcosa di dato, ma qualcosa che viene esso stesso prodotto, rinnovato e accresciuto" (ibid.).
La distinzione tra ricchezza fondo e ricchezza flusso assume carattere centrale nella teoria economica classica che, sotto questo profilo, prelude a risultati successivi dell'analisi economica per quanto riguarda l'imperfetta sovrapposizione di ricchezza individuale e ricchezza sociale (si mette in evidenza come non necessariamente processi individuali di accrescimento della ricchezza conducano all'aumento della ricchezza sociale). Questo punto di vista è espresso chiaramente da John Stuart Mill: "Per un singolo individuo può essere ricchezza qualunque cosa che, per quanto inutile in se stessa, gli possa consentire di richiedere ad altri una parte del loro fondo [stock] di cose utili e piacevoli. Si consideri ad esempio un mutuo di mille sterline acceso sopra un fondo di terra. Esso costituisce ricchezza per la persona a cui assicura un reddito [...] Ma non costituisce ricchezza per l'intero paese;

se l'impegno contrattuale fosse annullato, il paese non sarebbe né più povero né più ricco" (v. Mill, 1848, ed. 1886, vol. I, p. 9).
La distinzione milliana è ripresa nella General theory di John Maynard Keynes, in cui è esplicito il riconoscimento del fatto che il risparmio individuale presuppone un trasferimento di ricchezza da un soggetto all'altro, ma non implica necessariamente la formazione di nuova ricchezza: "Qualsiasi atto di risparmio comporta inevitabilmente un trasferimento 'forzato' di ricchezza verso colui che risparmia, benché questi possa a sua volta subire un danno a causa del risparmio di altri. Questi trasferimenti di ricchezza non richiedono la creazione di nuova ricchezza" (v. Keynes, 1936, p. 212).
Luigi Pasinetti ha di recente ripreso la distinzione classica fra ricchezza individuale e ricchezza sociale mettendo in evidenza che "il risparmio del sistema economico nel suo insieme, determinato in termini fisici dall'ammontare totale degli investimenti, non deve essere confuso con le ampie possibilità di risparmio personale che sono aperte ai singoli individui" (v. Pasinetti, 1981; tr. it., p. 176). In particolare, nel caso di un'economia di puro lavoro (in cui i beni sono prodotti mediante l'esclusivo impiego di lavoro e sono assenti i beni capitali) non è possibile alcuna formazione aggregata di risparmio, dal momento che non è richiesta alcuna formazione di capitale al fine di espandere la capacità produttiva. Il reddito consiste esclusivamente di beni di consumo, e un'eventuale eccedenza del reddito prodotto rispetto al reddito consumato non potrebbe avere alcun effetto sulla capacità produttiva futura. Questa situazione, tuttavia, "non comporta alcuna impossibilità, per i singoli individui, di posporre o anticipare il consumo nel tempo, ossia non comporta alcuna impossibilità di risparmio personale, positivo o negativo. Ciascun individuo può sempre astenersi dal consumare parte del suo reddito corrente e posporre il consumo nel tempo, a condizione che egli trovi qualche altro individuo o altri individui che siano disposti a fare l'opposto" (ibid., p. 177).
Esiste, a prima vista, una distanza profonda fra la concezione classica della ricchezza sociale (nazionale) come flusso di beni prodotti misurabile in termini di 'comando' sul lavoro altrui (Smith) e la concezione successiva (postclassica) in cui la ricchezza di un individuo oppure di un sistema economico viene collegata principalmente all'esistenza di un fondo di risorse utili, misurabili in base a un indicatore di scarsità (v. Senior, 1836; v. Carver, 1904). Alla base del diverso ruolo assegnato alla producibilità e alla scarsità nella teoria della ricchezza esiste una diversa rappresentazione analitica del sistema economico. Gli economisti classici

partono dalla considerazione della capacità di produrre beni (e servizi) all'interno di un sistema di interdipendenze caratterizzato dalla divisione del lavoro fra individui e gruppi sociali. In questa prospettiva, la misurazione della ricchezza in termini di lavoro comandato (Smith, Malthus) riflette un elemento centrale della divisione del lavoro (la possibilità di disporre del lavoro altrui sulla base di un sistema sufficientemente affidabile di regole contrattuali o norme sociali) e consente di trascurare la considerazione immediata della scarsità. Gli economisti postclassici, invece, fanno riferimento diretto alla relazione fra dotazione di risorse da un lato e necessità o desideri degli individui e dei gruppi dall'altro. In questo caso il power of purchasing della ricchezza (Smith) risulta collegato alla scarsità piuttosto che alla producibilità, e il sistema rilevante di interdipendenze sociali fa riferimento alle relazioni di scambio piuttosto che all'organizzazione della produzione. La possibilità che un bene sia acquisito attraverso lo scambio oppure attraverso la produzione è collegata alla "scarsità ovvero all'insufficiente offerta dell'oggetto in questione" (v. Carver, 1921, p. 43) e la stessa divisione del lavoro fra una pluralità di individui può apparire come un aspetto accessorio della 'riproduzione economica' (inclusiva sia dello scambio sia della produzione). In quest'ultima prospettiva si può concludere che un aspetto caratterizzante della ricchezza è il costo positivo di acquisizione (riproduzione) dei beni sia attraverso la produzione in senso proprio sia attraverso lo scambio (v. anche Ferrara, 18891891, vol. I., pp. 472-492). L'applicazione del concetto di ricchezza agli schemi di contabilità nazionale suggerisce alcuni approfondimenti della prospettiva analitica appena delineata, e individua possibili utilizzazioni (non necessariamente in alternativa fra loro) delle concezioni classica e postclassica della ricchezza. Ad esempio, la misurazione della ricchezza in termini di lavoro comandato (approccio classico) oppure di scarsità relativa e costo di riproduzione (approccio postclassico) può mettere in evidenza aspetti diversi di un processo di dinamica strutturale. Le misurazioni di tipo classico risultano di immediata efficacia quando si tratta di individuare gli effetti di un cambiamento delle tecniche di produzione sulla capacità complessiva del sistema economico di produrre beni e servizi. In particolare, l'attenzione di Smith per la misurazione della ricchezza in termini di lavoro comandato mette in evidenza come la ricchezza di un sistema economico possa essere considerata come un 'fondo' atto a produrre beni e servizi ed eventualmente ad accrescere la sua stessa capacità produttiva nel corso del tempo. Secondo Smith "il lavoro svolto in un anno in ciascuna nazione è il fondo originario che fornisce tutti i beni necessari e utili che essa consuma in un anno" (v. Smith,

1776, ed. 1976, p. 10); tuttavia il flusso di beni prodotti diviene esso stesso un fondo (stock, nella terminologia smithiana) di capacità produttiva che può costituire la base per un ulteriore incremento di ricchezza. In questa prospettiva la misurazione della ricchezza sociale in termini di lavoro comandato mette in evidenza la capacità di accrescimento della ricchezza attraverso l'espansione della forza lavoro occupata. Secondo Smith il 'valore reale' del prodotto netto di un sistema economico "esprimeva [...] il limite all'incremento nel numero dei lavoratori produttivi ottenibile in quella nazione, e costituiva perciò il miglior indice del suo potere di accumulare e accrescere così la propria prosperità" (v. Garegnani, 1960, p. 192).

Le misurazioni della ricchezza di tipo postclassico richiamano l'attenzione su altri aspetti della ricchezza individuale e sociale. In particolare, i riferimenti alla scarsità relativa (Senior) e al costo di riproduzione (Ferrara) mettono in evidenza il collegamento fra ricchezza e struttura delle preferenze e consentono di misurare gli effetti del passaggio di un bene dalla categoria dei beni liberi a quella dei beni con costo di riproduzione positivo, oppure gli effetti di un cambiamento nei gusti individuali e collettivi. Nel primo caso, nuovi beni (come acqua e aria pura) potrebbero entrare a far parte del fondo di ricchezza di un sistema economico; nel secondo caso, beni precedentemente collegati al funzionamento di un "sistema industriale efficiente" (v. Carver, 1921, p. 45) potrebbero essere esclusi dagli elementi costitutivi della ricchezza, ad esempio perché espressione di una struttura dei gusti che induce gli esseri umani a desiderare "oggetti che procurano un danno" e che sono quindi all'origine di "deterioramento piuttosto che [di] benessere nazionale" (ibid.).

La misurazione della ricchezza in termini di lavoro comandato richiama l'attenzione sulle caratteristiche intrinsecamente dinamiche della formazione di ricchezza in un'economia di produzione (in cui sono rilevanti la capacità di produrre ricchezza e la caratteristica della ricchezza come fondo di prodotti). D'altra parte, la misurazione della ricchezza in termini di scarsità relativa e costo di riproduzione richiama l'attenzione sulle caratteristiche che rinviano al grado di riproducibilità dei beni e alla struttura delle preferenze. Diversi contesti storici e operativi rinviano all'uno o all'altro dei due criteri di misurazione. In particolare, i confronti della ricchezza prodotta in sistemi economici sufficientemente simili si possono più agevolmente stabilire in termini di lavoro comandato, mentre i confronti fra sistemi caratterizzati da differenze profonde nel grado di producibilità dei beni o nella struttura delle preferenze rinviano alla considerazione degli elementi soggettivi appena esaminati. Infatti

sistemi economici analoghi sotto il profilo delle categorie di beni producibili e non producibili, e sotto il profilo dei beni a cui si collegano 'utilità positiva' oppure 'utilità negativa', possono essere confrontati in termini di una misura oggettiva della ricchezza, quale quella espressa dal lavoro comandato di Smith e Malthus (tale unità di misura permette di valutare le possibilità di espansione di un sistema economico considerando il 'potere di disposizione' collegato a un particolare fondo di prodotti). Il lavoro comandato, tuttavia, difficilmente può consentire un confronto significativo in termini di ricchezza fra sistemi economici molto distanti fra loro per quanto riguarda i beni producibili e non producibili, oppure i beni forniti di utilità positiva o negativa. Infatti in questo caso la considerazione del lavoro comandato elimina differenze nella struttura produttiva e nella struttura dei gusti che può essere invece interessante prendere in considerazione: il lavoro comandato di un sistema economico in cui il criterio della producibilità è esteso a comprendere alcune risorse naturali (completamente o parzialmente rinnovabili) ha un significato profondamente diverso dal lavoro comandato di un sistema economico in cui le risorse naturali appartengono esclusivamente alla categoria dei beni non producibili (cosicché non vi possono essere quantità di lavoro destinate ai processi di mantenimento delle risorse).

4. Ricchezza, capitale e reddito: l'analisi classica dei fondi di ricchezza e flussi di reddito

I contributi di Pietro Verri e Adam Smith richiamano l'attenzione sulle caratteristiche centrali della concezione classica relativa alle caratteristiche intertemporali del reddito e della ricchezza. Nelle Meditazioni sull'economia politica Pietro Verri puntualizza il concetto di ricchezza flusso e la relazione tra flussi di produzione e consumo nella dinamica della ricchezza: "due oggetti principalmente bisogna osservare, e sono annua riproduzione, e consumazione annua. In ogni Stato si riproduce per mezzo della vegetazione e delle manifatture, e in ogni Stato si consuma. [...] Quando il valore totale della riproduzione equivale al valore dell'annua consumazione, quella nazione persevera nello stato in cui è quando tutte le circostanze sieno eguali. Deperisce quella nazione in cui l'annua consumazione eccede la riproduzione annua, e migliora invece quello Stato in cui l'annua riproduzione sopravanza il consumo" (v. Verri, 1771, pp. 18-19).
Nella Wealth of nations di Smith passa in primo piano la relazione fra capital (capitale) e revenue (reddito) in cui, a differenza del nesso esaminato da Verri, acquista importanza

centrale il collegamento tra una grandezza flusso (il revenue) e una grandezza fondo (lo stock di beni prodotti e non destinati al consumo immediato). La relazione tra fondi e flussi nella dinamica della ricchezza è inizialmente descritta da Smith facendo riferimento alla distinzione fra capitale e reddito individuali: "Quando il fondo [stock] posseduto da un individuo è appena sufficiente a mantenerlo per pochi giorni o poche settimane, raramente egli pensa di derivare un qualsiasi reddito da esso. Egli lo consuma nel modo più parsimonioso possibile, e cerca con il suo lavoro di acquisire qualcosa che possa prendere il suo posto prima che sia interamente consumato. In questo caso il suo reddito è derivato esclusivamente dal suo lavoro" (v. Smith, 1776, ed. 1976, p. 279). La formazione di reddito è in questo caso un risultato esclusivo dell'attività lavorativa direttamente rivolta alla formazione di beni, e non può fare riferimento all'impiego di un fondo di ricchezza allo scopo di accrescere la capacità di formazione del reddito.In un passo immediatamente successivo, tuttavia, Smith riconosce l'importanza della formazione di fondi nel processo di creazione della ricchezza flusso: "Quando [un individuo] possiede un fondo sufficiente a mantenerlo per mesi o per anni, egli naturalmente si sforza di ottenere un reddito dalla maggior parte di esso e di riservare per il suo consumo immediato soltanto quella parte che lo possa mantenere fino a quando questo reddito inizi a essere percepito" (ibid.).
In particolare, la disponibilità di fondi di ricchezza (nella forma di fondi di beni prodotti) è considerata la condizione necessaria per l'introduzione della divisione del lavoro e quindi per l'avvio di un processo di crescita della produttività del lavoro basato su forme significative di apprendimento: "Un tessitore non può dedicarsi interamente alla sua particolare attività a meno che non vi sia accumulato da qualche parte, in suo possesso oppure in possesso di qualche altra persona, un fondo sufficiente per mantenerlo e per rifornirlo di tutte le materie prime e attrezzi del suo lavoro, finché egli non abbia non solo completato ma venduto la sua tela. Questa forma di accumulazione deve evidentemente precedere l'applicazione del suo lavoro a una tale particolare attività per un tempo così prolungato" (pp. 276-277). La formazione (accumulazione) di un fondo di ricchezza flusso (o capitale) è quindi per Smith la condizione necessaria perché un sistema economico possa avviare un circolo virtuoso basato sull'estensione della divisione del lavoro, gli incrementi di produttività e l'allargamento dei mercati: "Poiché l'accumulazione di un fondo deve, nella natura delle cose, precedere la divisione del lavoro, il lavoro può essere sempre più suddiviso a condizione che quel fondo sia sempre più accumulato" (p. 277).

La connessione tra ricchezza flusso e formazione di fondi attraverso l'accumulazione di uno stock di prodotti è centrale nella teoria smithiana della dinamica della ricchezza: "La proporzione fra capitale e reddito, quindi, sembra ovunque regolare la proporzione fra industriosità e pigrizia. Ogni qual volta predomina il capitale, prevale l'industriosità. Ogni qual volta predomina il reddito, prevale la pigrizia. Qualunque aumento o diminuzione di capitale, quindi, tende in modo naturale ad aumentare oppure a diminuire il volume reale di attività, il numero di lavoratori produttivi, e di conseguenza il valore di scambio del prodotto annuo della terra e del lavoro di un paese, quindi la ricchezza reale e il reddito dei suoi abitanti" (p. 337).
L'analisi smithiana delle relazioni intertemporali tra grandezze fondo e grandezze flusso mette in evidenza come il passaggio alla concezione di ricchezza flusso basata sulla produzione annua di ricchezza e sulla formazione di prodotto netto (Verri, Smith) richieda un'identificazione dei fondi coerente con la centralità dei processi produttivi. I fondi, infatti, sono considerati da Smith come stocks di prodotti piuttosto che come dotazioni di risorse. In questo modo la teoria classica della ricchezza flusso viene ad assumere la caratteristica di uno schema logico basato sulla considerazione di asimmetrie temporali nel coordinamento tra i processi di formazione dei prodotti e i processi di formazione degli stocks (in primo luogo di beni di consumo) necessari per consentire l'introduzione della divisione del lavoro e la produzione di attrezzi. Le caratteristiche strutturali dei processi produttivi determinano, secondo Smith, una specifica articolazione temporale dei flussi e fondi di ricchezza, e quindi il ruolo rispettivo di capital e revenue nella formazione di potere d'acquisto all'interno del sistema economico. Questa prospettiva suggerisce anche un particolare punto di vista per quanto riguarda le relazioni tra accumulazione individuale di ricchezza (risparmio individuale), formazione di fondi di ricchezza (stocks di prodotti) a livello dell'intero sistema economico (risparmio aggregato) e caratteristiche dinamiche del sistema economico complessivo. Infatti, il ruolo del risparmio individuale (accumulazione individuale) risulta strutturalmente distinto dalla spesa di revenue benché, dal punto di vista aggregato, sia il risparmio sia la spesa possano essere considerati come particolari forme di consumo della ricchezza flusso prodotta. Tuttavia si tratta di un consumo effettuato da soggetti (o gruppi di soggetti) collocati in modo molto diverso nel processo di formazione della ricchezza. A questo proposito Smith osserva: "Ciò che è annualmente risparmiato è consumato altrettanto regolarmente [del reddito] che è annualmente speso, e inoltre quasi nello stesso periodo di

tempo; ma è consumato da un insieme diverso di persone. La porzione di reddito che un uomo ricco spende annualmente è nella maggior parte dei casi consumata da ospiti oziosi e domestici, che non lasciano alcunché dietro di sé come risultato del loro consumo. Quella quota [di reddito] che egli risparmia annualmente, dal momento che è immediatamente impiegata come capitale allo scopo di percepire un profitto, finisce per essere consumata anch'essa, e quasi nello stesso tempo, ma da una diversa categoria di persone, da lavoratori, operai e artigiani che riproducono con un profitto il valore del loro consumo annuale. [...] Il consumo è lo stesso ma i consumatori sono diversi" (pp. 337-338).

5. Reddito, consumo e ricchezza: ciclo vitale e capitale intergenerazionale

Come si è visto, il nesso fra ricchezza e consumo è centrale nella concezione preclassica della ricchezza, in cui questa è considerata in primo luogo come un fondo di beni e servizi almeno potenzialmente consumabili (v. cap. 2). Gli economisti classici richiamano l'attenzione sui collegamenti tra formazione della ricchezza e produzione di beni, e al tempo stesso sul ruolo del consumo produttivo (consumo destinato al mantenimento di lavoratori produttivi) nei processi di utilizzazione dei fondi di ricchezza (v. cap. 4). Nella riflessione economica postclassica l'attenzione per i flussi di reddito e per il collegamento tra reddito e mantenimento del capitale (o della ricchezza) conduce alla considerazione esplicita delle relazioni fra dinamica della ricchezza individuale e scelte intertemporali di consumo e risparmio. Irving Fisher (v., 1906 e 1930) mette in evidenza che la distinzione tra flussi di reddito e fondi di ricchezza riflette spesso la valutazione dei singoli individui riguardo all'attualizzazione dei flussi attesi di potere d'acquisto. Ad esempio, la ricchezza di uno stesso individuo potrà essere considerata maggiore o minore a seconda che lo stesso flusso atteso di entrate sia attualizzato facendo riferimento, rispettivamente, a un saggio di sconto minore oppure maggiore. Lo stesso Fisher illustra il ruolo delle scelte intertemporali nella valutazione di reddito e ricchezza (capitale).
Il collegamento fra reddito e ricchezza si presenta, nella prospettiva fisheriana, soprattutto dal punto di vista ex ante delle valutazioni dei singoli. La ripresa di interesse per la concezione classica della ricchezza, che può essere collegata al contributo di John Maynard Keynes, suggerisce il superamento della concezione fisheriana, e quindi dell'ipotesi che "il risparmio non consista soltanto nell'astensione dal consumo presente ma nel collocare

simultaneamente un ordine specifico per un consumo futuro" (v. Keynes, 1936, p. 210). Il contributo keynesiano è caratterizzato dalla concezione della formazione di ricchezza come processo ex post, il cui risultato è una diminuzione netta della domanda per beni di consumo anziché "una sostituzione della domanda di consumo futuro al posto della domanda di consumo presente" (ibid.). Uno sviluppo interessante della teoria keynesiana della ricchezza si ha con la teoria del ciclo vitale dei consumi e risparmi, che mantiene la concezione dell'accumulazione di ricchezza come processo ex post, ma ne individua i collegamenti con le propensioni e le scelte degli individui (e dei gruppi sociali) per quanto riguarda la distribuzione intertemporale di risparmio e consumo. Alla base della teoria del ciclo vitale è l'idea che "il risparmio positivo in certe fasi della vita [sia] compensato dal risparmio negativo in altre: ciò che Harrod ha suggestivamente denominato 'gobba del risparmio' (hump saving)" (v. Modigliani, 1975, ed. 1992, p. 178). Secondo questa prospettiva, la distribuzione di consumo e risparmio all'interno del ciclo vitale è sufficiente a determinare un tasso positivo di formazione della ricchezza anche in assenza di un saggio positivo di espansione del sistema economico nel suo complesso: "In un'economia stazionaria - con tasso di crescita della popolazione e della produttività pari a zero - la quota di risparmio deve essere zero, indipendentemente dal ciclo vitale della quota del risparmio delle singole famiglie, poiché il risparmio positivo di coloro che presentemente accumulano in vista di un successivo risparmio negativo sarebbe esattamente compensato da coloro che si trovano in questa seconda fase. Nel contempo [...], anche assumendo che la sola fonte basilare della 'gobba del risparmio' [sia] il ritiro dall'attività, ci si [può] attendere che una società stazionaria [tenga] un sostanziale volume di attività patrimoniali in rapporto al reddito" (p. 185). La teoria del ciclo vitale è stata successivamente estesa a considerare il peso relativo della formazione di ricchezza come risultato di trasferimenti intergenerazionali (eredità) rispetto alla formazione di ricchezza attraverso il risparmio all'interno della stessa generazione (pp. 222-224). A questo proposito, si è giunti alla conclusione che è possibile "avere una forte ricchezza anche senza passaggio ereditario" (v. Modigliani, 1996, p. 22) e che, anche ammettendo i trasferimenti ereditari, "il risparmio, così come inteso nel modello del ciclo vitale, cioè nella sua funzione di livellare il consumo, sembrerebbe ancora essere la fonte più importante di ricchezza del nostro sistema" (ibid.). Malgrado questo risultato di carattere macroeconomico (basato su stime della quota di ricchezza ereditaria e di quella dovuta al ciclo vitale), è stata riconosciuta l'importanza del movente dell'eredità nella spiegazione di comportamenti

individuali diversi da quelli inizialmente considerati nel modello del ciclo vitale, come il fatto che "l'attività di decumulo (della ricchezza), se mai esiste, sia minima nell'età avanzata" (v. Modigliani, 1975, ed. 1992, pp. 234-235).

La teoria del ciclo vitale suggerisce un interessante collegamento fra i risultati collettivi (aggregati) delle scelte individuali e la critica keynesiana della teoria di Fisher sulla sostituzione di consumo futuro a consumo presente. I contributi sul ciclo vitale di risparmio e consumo, infatti, mettono in evidenza come tale sostituzione possa avere le caratteristiche di un risultato ex post connesso alle dinamiche demografiche (struttura della popolazione per classi di età) e indipendentemente dall'atteggiamento dei singoli individui nei confronti della sostituzione intertemporale. In altre parole, la struttura demografica può determinare la quota di reddito consumata e la quota destinata alla formazione di nuova ricchezza senza fare diretto riferimento alla preferenza intertemporale dei singoli: la propensione al risparmio positivo oppure a quello negativo può essere considerata un dato strutturale collegato all'età e sostanzialmente indipendente da un'esplicita massimizzazione degli indici di utilità.

In questo modo la teoria del ciclo vitale, benché esplicitamente collegata (in alcune sue formulazioni) alla teoria della sostituibilità intertemporale fra consumo presente e futuro (v., ad esempio, Modigliani, 1988) mostra anche evidenti connessioni con la concezione secondo cui i comportamenti di risparmio riflettono propensioni psicologiche e assetti istituzionali o demografici di fatto indipendenti dalla logica di una teoria della scelta razionale (v., a questo riguardo, Del Vecchio, 1915 e 1956; v. Quadrio Curzio, 1965).

Il problema del collegamento fra reddito e ricchezza secondo una prospettiva intertemporale (che è al centro dei contributi appartenenti alla tradizione analitica del ciclo vitale) rimane essenziale nello studio dei trasferimenti di potere d'acquisto (reddito e ricchezza) da una generazione all'altra. Il trasferimento intergenerazionale può riguardare sia flussi di reddito sia fondi di ricchezza. Nel primo caso si può trattare (ad esempio) del trasferimento di titoli a cui sia collegato un rendimento di tipo ricorrente; nel secondo caso si può trattare di immobili o di altri 'fondi' privi di un diretto 'comando' su beni e servizi e tali da richiedere un processo di liquidazione. Nei trasferimenti intergenerazionali di potere d'acquisto è centrale il ruolo di assetti istituzionali e demografici (quali regimi di proprietà e di trasferimento per eredità, saggio di crescita della popolazione, mutamenti nella composizione della popolazione per coorti di età) che, in modo più accentuato che nel caso dei trasferimenti di medio periodo all'interno di una stessa generazione (ciclo vitale), tendono a prevalere rispetto

agli elementi di sostituzione intertemporale (consumo futuro al posto di consumo presente), più evidenti nel breve periodo.

6. Ricchezza, reddito e dinamica economica

Le osservazioni precedenti consentono di individuare una precisa linea di evoluzione che conduce il concetto di ricchezza a costituire (forse più del concetto di reddito) un indicatore delle 'possibilità di acquisizione' godute da un particolare individuo, gruppo sociale o sistema economico in una prospettiva di medio e lungo periodo. Alla base di questa essenziale caratteristica del concetto di ricchezza è il suo riferimento a situazioni consolidate piuttosto che a un flusso corrente di beni e servizi. Per questa ragione, i cambiamenti nella concezione prevalente di ricchezza possono essere considerati come un indicatore importante del passaggio da una fase all'altra nelle relazioni di lungo periodo fra aspetti economici fondamentali quali la dinamica delle popolazioni da un lato e la dinamica delle risorse a loro disposizione dall'altro.Sistemi economici stazionari o quasi stazionari tendono a privilegiare una concezione della ricchezza basata sulla considerazione dei 'fondi', cioè della quantità di potere d'acquisto accumulato in oggetti o titoli di possesso certo e valore reale indipendente (almeno in parte) dalla struttura degli scambi. D'altra parte i sistemi economici caratterizzati da rilevanti trasformazioni della quantità e della composizione delle risorse disponibili tendono a mettere in evidenza caratteristiche di 'flusso' della ricchezza, e quindi l'esistenza, le dimensioni e il saggio di crescita del potere d'acquisto corrente (reddito), e la sua accumulazione in uno stock di cui è possibile studiare i processi (talora alterni) di formazione oppure progressivo esaurimento (un esempio è costituito dai processi di accumulazione e consumo studiati nella teoria del ciclo vitale).Non si può escludere che un rallentamento di lungo periodo nei processi di crescita del reddito finirebbe per mettere nuovamente in risalto la dimensione di 'fondo' della ricchezza, e quindi la sua capacità di assegnare a individui e gruppi sociali una sorta di 'protezione reale' nei confronti di condizioni di volatilità dei flussi di reddito (a questo proposito v. Buchanan, 1993, p. 35). D'altra parte, situazioni di accentuata dinamica sociale (favorita da particolari dinamiche demografiche e da regimi di trasferimento della ricchezza da una generazione all'altra) possono mettere in evidenza flussi di ricchezza (da un individuo all'altro, da una generazione all'altra) che non siano ancora assestati in fondi sufficientemente consolidati rispetto alle dinamiche di mercato, oppure

rispetto ad altri processi di accumulazione del trasferimento del potere d'acquisto.(Sono lieto di riconoscere il mio debito intellettuale nei confronti del gruppo di studiosi dell'International research network 'The wealth of nations in economic theory', ferma restando la mia responsabilità esclusiva per il contenuto specifico di questo scritto).

Distribuzione della ricchezza e del reddito

di Alberto Quadrio Curzio

sommario: 1. Introduzione. 2. La teoria sociale della distribuzione del reddito: a) principî generali e richiami storici; b) lo schema teorico di base: salari, profitti, rendite; c) prezzi e distribuzione; d) antagonismi sociali e distributivi e tecniche nella statica; e) rendita e risorse naturali scarse; f) la dinamica delle grandezze distributive: classici e post-keynesiani. 3. Teoria funzionale della distribuzione: a) principî generali e richiami storici; b) schema teorico aggregato di base e teorema di Eulero; c) formulazione microeconomica, domanda e remunerazione dei fattori; d) equilibrio generale e distribuzione. 4. Teoria delle forme di mercato e distribuzione: a) una possibile 'intersezione' delle precedenti teorie; b) il grado di monopolio; c) il grado di oligopolio. 5. La distribuzione personale del reddito: a) le curve di Lorenz; b) alcuni dati statistici. 6. La ricchezza e la sua distribuzione: a) le difficoltà delle analisi della ricchezza; b) teorie sulla distribuzione e sull'accumulazione della ricchezza: cenni; c) analisi empiriche della ricchezza in Italia. 7. Conclusione. □ Bibliografia.

1. Introduzione

I problemi della distribuzione del reddito e della ricchezza sono da sempre rilevanti nella scienza economica in termini di analisi teorica, di indagine empirica e storica, di politica economica. Alcuni importanti studiosi hanno addirittura fondato tutti i problemi della economia politica sulle teorie della distribuzione del reddito. Queste centralità richiedono dunque notevole selezione per non disperdere l'analisi. La nostra scelta di temi sarà la seguente: a) teorie 'sociali' della distribuzione del reddito e teorie 'funzionali', con riferimento a un singolo paese o sistema economico; b) analisi teorico-storico-empiriche, che denominiamo teoria 'mista', sulla dinamica della distribuzione del reddito in economie

singole e comparate; c) analisi teorico-empiriche sulla distribuzione 'personale' del reddito; d) la distribuzione della ricchezza: cenni.
Quanto al metodo svolgeremo analisi storico-analitiche e analitico-empiriche. Tracceremo cioè un'evoluzione selettiva della teoria in base a stilizzazioni analitiche (e non esegeticamente) dei contributi di vari economisti ed esamineremo dati statistici stilizzati mediante l'analisi economica. In entrambi i casi punteremo alla chiarezza piuttosto che alla completezza della rassegna e della critica, rinviando a tali fini ad altri contributi e ben sapendo che la teoria della distribuzione ha varie componenti - che evidenzieremo - talvolta complementari, talvolta alternative, talvolta antagonistiche. L'uso che faremo di alcune relazioni formali ha scopo di sintesi e confronto tra teorie, pur non attribuendo agli stessi simboli sempre lo stesso significato.

2. La teoria sociale della distribuzione del reddito

a) Principî generali e richiami storici

I punti su cui la teoria sociale è stata costruita in circa due secoli sono i seguenti.
1. La presenza di categorie, ceti o classi sociali che entrano in contatto con il processo di produzione esercitando su di esso una funzione o un diritto o un potere o una combinazione di questi. Le categorie sociali 'stilizzate' sono quelle dei 'lavoratori', dei 'capitalisti', dei 'proprietari terrieri' (secondo la denominazione ottocentesca che ha mantenuto vitalità pur cambiando via via di contenuto); esse avanzano richieste sulla ripartizione del prodotto netto (cioè la differenza tra produzione complessiva e mezzi di produzione impiegati), o reddito, del processo di produzione in un certo periodo (di solito l'anno).
2. L'esistenza di categorie di redditi in cui si stabilisce una corrispondenza tra categorie sociali, processo di produzione, prodotto netto (reddito). Sono le rendite dei 'proprietari terrieri', o più in generale di chi apporta alla produzione risorse scarse; i profitti dei 'capitalisti', o più in generale di chi apporta alla produzione imprenditorialità e capitale monetario; i salari dei 'lavoratori' dipendenti, o più in generale di chi apporta al processo produttivo servizi lavorativi.
Quanto detto spiega perché si può parlare di teoria 'sociale' della distribuzione. Al centro dell'analisi non è infatti il problema della remunerazione di un singolo fattore della produzione, in base al solo nesso prestazione-remunerazione o in base alle sole relazioni tra

domanda e offerta del fattore, ma è piuttosto quello della distribuzione del prodotto netto nella società e tra i suoi ceti.
Questa teoria ha avuto grande impulso con i classici (per una sintesi, v. Cannan, 1893; v. Pivetti, 1987) - a cominciare da Adam Smith (1776) - e in particolare con David Ricardo (1815) (v. Sraffa, 1951). Egli si concentrò (v. De Vivo, 1987) dapprima sul profitto, con l'Essay on profits del 1815, poi progettò il trattato The principles of rent, profit and wages ritenendo che: "questi principî sono così interrelati e così connessi con ogni elemento appartenente alla scienza dell'economia politica che io considero della massima importanza una appropriata visione degli stessi" (v. Ricardo, 1815; ed. 1952, vol. VI, pp. 315316). Dopo Ricardo la teoria sociale ebbe molte varianti tra cui una ideologizzata con Karl Marx (v., 1867-1894). Dal 1870 circa bisogna passare agli anni trenta di questo secolo per una ripresa-riformulazione-estensione della teoria sociale secondo due filoni complementari: uno macrodinamico, che fa capo a Keynes (v., 1936) e a Kalecki (v., 1939), l'altro multisettoriale, che fa capo a Sraffa (v., 1960).

b) Lo schema teorico di base: salari, profitti, rendite
Lo schema di base della teoria sociale è quello di Ricardo, così stilizzabile (v. Pasinetti, 1960):

$Y' = W + P + R$, (1)
$W = wL$ (2)
$Y = Y' - W = P + R$. (3)

Per la (1) la produzione (Y´), che per ora si suppone data da una sola merce ('grano'), si ripartisce tra salari (W), profitti (P), rendite (R). Per la (2) i salari totali sono dati dal salario unitario (w) per il numero dei lavoratori occupati (L).
Due tipi di salari vanno distinti: a) il salario naturale unitario (o prezzo naturale del lavoro) wn, che è determinato dalle necessità naturali e sociali di sussistenza del lavoro in un periodo storico e perciò non è costante nel tempo; b) il salario di mercato unitario (o prezzo di mercato del lavoro) wm, che dipende dalle condizioni di domanda e offerta del lavoro. Può accadere che wm superi wn: ne segue un aumento della popolazione e dell'offerta di lavoro con una discesa di wm verso wn. Per converso, se wm scende sotto wn, calano la popolazione

e l'offerta di lavoro e wm risale verso wn. Per la (3) i salari naturali sono tutto capitale circolante, cioè mezzi di produzione, e quindi non entrano nel prodotto netto, che dunque si distribuisce tra rendite e profitti.
La rendita è data da

$$R = Y - P = Y' - yuL \cong (y1 - yu) + (y2 - yu) + ...+ (yu - yu), \quad (4)$$

cioè dalla produzione meno quella che si avrebbe se tutti i lavoratori producessero come l'ultimo occupato, cioè quello occupato sulla terra meno fertile (yu). La rendita è generata dalla diminuzione di produttività del lavoro, dovuta alla decrescente fertilità e alla scarsità delle terre, e dalla concorrenza tra capitalisti-imprenditori, che rende il saggio di profitto pari a quello del capitale impiegato sulla terra meno fertile.Il profitto totale è infine una grandezza residuale, o calcolata per differenza, data dalla (5), sulla cui base si determina il saggio di profitto (π), dato dalla (6):

$$P = Y - R = Y' - W - R = (yu - w)L; \quad (5)$$
$$\pi = P/W = (yu/w) - 1. \quad (6)$$

Esso potrebbe anche intendersi come una forma di 'rendimento' del capitale circolante (W) e come una 'remunerazione' per il ruolo che i capitalisti-imprenditori svolgono nella produzione. Dalla (6) risulta che tra salario unitario e saggio di profitto esiste una relazione statica inversa per cui quando cresce w cala π.
In sintesi grafica il modello è esprimibile tramite il diagramma della fig. 1, dove è tracciata la curva di produttività ('marginale') del lavoro: $y=f_1 (L)$.
Questa trattazione riguarda un'economia dove si produce una sola merce (il 'grano'): pur nella sua estrema schematizzazione, essa evidenzia la struttura della distribuzione sociale in un contesto uniperiodale (e poi dinamico). Vari problemi rimangono aperti, in particolare quello delle relazioni tra prezzi e distribuzione in una economia a pluralità di merci e quello dell'accumulazione e della dinamica.

c) Prezzi e distribuzione
Introduciamo il problema delle relazioni tra prezzi e distribuzione per una economia a più

merci riferendoci al calcolo di π. Essendo $\pi = P/K$, dove K è il capitale totale, bisogna allora conoscere i prezzi delle merci da cui dipende il valore sia dei profitti totali sia del capitale; ma per calcolare i prezzi bisogna conoscere π. Vi è dunque una circolarità.
Ricardo tentò una soluzione con la teoria del valore-lavoro, assumendo che il prezzo delle merci fosse dato dalla quantità di lavoro, scelto come numerario, incorporata direttamente e indirettamente nelle merci stesse; non risolse comunque il problema e non eliminò i disturbi dei prezzi. Anche Marx affrontò il problema, allargando la teoria del valore-lavoro a una teoria generale dello sfruttamento del lavoro e impostando (ma non risolvendo) il cosiddetto problema della trasformazione dei valori nei prezzi (v. Garegnani, 1960; v. Blaugh, 1962).
Alla soluzione giunse nel 1960 Piero Sraffa (v., 1960; v. Pasinetti, 1975), mediante una teoria multisettoriale espressa dal sistema a m equazioni prezzi-distribuzione

$$A' P(1 + \pi) + Lw = P, \quad (7)$$

ove A' è una matrice di coefficienti tecnici di tipo leontieviano, P il vettore incognito dei prezzi, L il vettore dei coefficienti di lavoro. Sraffa individua un numerario (merce tipo o merce standard) con il quale la relazione tra π e w può essere analizzata senza i disturbi dei prezzi.
Con una doppia normalizzazione (sul valore del prodotto netto del sistema tipo e sul lavoro totale occupato) Sraffa giunge alla relazione

$$\pi = s(1 - \alpha), \text{ con } 0<\alpha<1, \quad (8)$$

ove $\alpha = (wL)/Y \quad (9)$

da cui $\quad w = Y\alpha/L \qquad (9')$

Nella (8) s è il saggio uniforme di prodotto netto e α è la quota dei salari sul valore del prodotto netto del sistema economico.
Per concludere sulla teoria sociale si consideri l'esogenità di π oppure di w (ovvero di α). La teoria studia infatti le relazioni antagonistiche (inverse) tra w e π, non la loro simultanea determinazione. Una delle due deve essere determinata esogenamente essendo questa una teoria 'aperta' che deve essere 'chiusa' con altre. Perciò è teoria flessibile e capace di interpretare varie situazioni tramite la variabile esogena.

d) Antagonismi sociali e distributivi e tecniche nella statica

Per Ricardo tra salario e profitto c'era un antagonismo che nel lungo periodo operava sempre contro il salario spingendolo alla sussistenza; esisteva anche un antagonismo tecnico-naturalistico tra rendita e profitto che operava contro il profitto, ma che poteva essere superato con il progresso tecnico che, a parità di lavoro occupato, determina un aumento dei profitti e una riduzione delle rendite (v. fig. 1: spostamento dalla curva f_1 alla curva f_2). In tale impostazione è difficile leggere un conflitto necessario tra classi sociali per quel che riguarda la distribuzione. Marx, partendo da Ricardo, elaborò invece una teoria del conflitto tra classi a proposito della distribuzione, che emerge già dalla seguente relazione per la determinazione di π basata su flussi annui:

$$\pi = P/(K + W) = P/(K + wL) = (P/W)/[(K/W) +1] = ss/(co + 1) \quad (10)$$

K è il capitale costante (ammortamento annuo del capitale fisso più le materie prime); W è il capitale variabile (o salari annui); P è il plusvalore; Y´=P+K+W è il prodotto nazionale lordo e Y=P+W quello netto. Le grandezze K e W sono definite in termini di lavoro necessario a produrle e P è pluslavoro. Su tale base si definiscono il saggio di plusvalore (ss) - che è una misura dello sfruttamento - e la composizione organica del capitale (co). La teoria della distribuzione di Marx andrebbe letta unitamente alla teoria del valore-lavoro. La crescita di π dovuta a quella di ss comporta un aumento dello sfruttamento dei lavoratori da parte dei capitalisti. Se ss cala per l'aumento del salario unitario sopra le sussistenze, cioè se i lavoratori riescono a far diminuire lo sfruttamento, diminuirà π e ciò spingerà i capitalisti a cambiare le tecniche riducendo l'impiego di lavoro e aumentando quello delle macchine, cioè aumentando co. Ciò determinerà disoccupazione riducendo di nuovo il salario. Secondo questa impostazione, che contiene parecchi errori, esiste un conflitto sociale insanabile riguardo alla distribuzione.

Sraffa (v., 1960) si stacca sia da Ricardo che da Marx nell'analisi dell'antagonismo tra w e π, non ricercando 'leggi naturali' alla Ricardo né 'leggi social-capitalistiche' alla Marx. Egli studia la relazione statica (a tecnologia data) inversa (antagonista) tra salari e profitti. La sua teoria consente di vedere subito come il progresso tecnico possa eliminare o attenuare l'antagonismo. Infatti, in determinate condizioni e sotto specifiche ipotesi circa le quote

distributive, un aumento di s da s_1 a s_2 (v. la relazione 8 e la fig. 2) e un aumento di Y/L (v. la relazione 9) permettono un aumento simultaneo di w e π a parità di α.

e) Rendita e risorse naturali scarse

Anche la rendita, ch'ebbe ampio spazio in Ricardo, va riconsiderata. Le risorse naturali scarse, le materie prime e la rendita, che spesso sono state tralasciate dalla teoria generale della distribuzione, ora hanno ripreso rilevanza con l'emergere di nuove scarsità e con la generalizzazione del modello a tutti i mezzi di produzione non prodotti o non riprodotti e scarsi che entrano nei processi produttivi circolari (v. Quadrio Curzio, 1987).

Una teoria della rendita, elaborata da Sraffa (v., 1960) e generalizzata da Quadrio Curzio (v., 1967, 1977 e 1990), si basa su k equazioni del tipo

$$(a_{1h}p_1 + a_{2h}p_2 + \ldots + a_{nh}p_n)(1 + \pi) + l_{1h}w + t_{1h}r_{1h} = p_1 \quad (11)$$

con h=1,..., k e a, l, t coefficienti tecnici leontieviani. Ogni equazione riguarda una stessa merce ('grano') prodotta con una diversa risorsa naturale ('terra') scarsa, su cui si genera una rendita (r_1) che sarà tanto maggiore quanto più il processo che utilizza la risorsa t_1 è efficiente rispetto al meno efficiente in uso, sul quale non ci sarà rendita in quanto la stessa ha natura differenziale. Le k equazioni (11) vanno affiancate alle m equazioni (7) per risolvere il sistema prezzi-distribuzione. È intuitivo come le variazioni della grandezza distributiva esogena (w o π) influenzino le rendite, ma anche come la rendita influenzi la grandezza distributiva endogena. Dunque la rendita 'complica' molto le relazioni tra w e π.

f) La dinamica delle grandezze distributive: classici e post-keynesiani

La teoria sociale ha rivolto grande attenzione alla dinamica.

Per Ricardo la dinamica nasce dalla accumulazione del profitto e cambia tutte le grandezze distributive (v. fig. 1): salari, rendite e profitti. Più precisamente l'andamento di tali grandezze sarebbe il seguente.

Salari. Se ci si trova in una situazione di eguaglianza tra domanda e offerta di lavoro in corrispondenza di wn, l'accumulazione porta a un aumento della domanda di lavoro, a un wm maggiore di wn e quindi a un aumento della popolazione e di L. Ne segue un calo di wm, che ritorna verso wn, anche se questo richiede tempo. Nel lungo periodo crescono sia W che L in

proporzioni tali che tendenzialmente wm converge al livello costante di wn. Nella dinamica naturale cresce dunque il monte salari W=Lwn, fermo restando wn.
Rendite. L'accumulazione porta a utilizzare terre via via meno fertili e/o ad aumentare l'impiego di lavoro su una stessa terra. Ne segue una continua diminuzione della produttività dell'ultimo lavoratore che - come si vede dalla (4) - fa crescere la rendita, la quale diventa massima nello stato stazionario (posizione L_2).
Profitti. L'accumulazione porta a una riduzione continua di π, come risulta dalla (6), in quanto la crescita del sistema economico genera una diminuzione di yu e di π. Poiché il profitto è la fonte dell'accumulazione e della dinamica, ne segue che, quando yu=w, π diventa zero e si giunge allo stato stazionario dove cessa la crescita (posizione L_2). L'ovvia pecca di questa teoria è che essa sottovaluta il progresso tecnico, il quale, aumentando la produttività (spostamento dalla curva f_1 alla curva f_2 nella fig. 1 e passaggio da uno stato stazionario L_2 a uno L_3) può condurre a risultati diametralmente opposti a quelli previsti da Ricardo.
Anche Marx studiò la dinamica, proponendo, tra l'altro, una 'legge' sulla caduta tendenziale del saggio di profitto. In base alla (10), poiché l'accumulazione e la scelta delle tecniche accrescono K/W, cioè co, se si suppone che ss cresca meno (Marx ne suppone poi addirittura la costanza), ne segue il calo di π nel tempo. Questa 'legge' è insostenibile per molti motivi, tra cui il seguente: supporre ss costante vuol dire non tenere conto che con co cresce la produttività del lavoro.
La rilevanza dei contributi classici e di Marx, pur criticabili sotto vari aspetti, dipende dal fatto che in essi è riconosciuta la centralità della relazione tra distribuzione del reddito, saggio di profitto, accumulazione del capitale, scelta delle tecniche, progresso tecnico.
Imperniati sulla stessa centralità sono i due grandi contributi contemporanei: quello post-keynesiano, che si connette a quello neoricardiano, e quello kaleckiano che assume una sua autonomia collocandosi come teoria 'mista'.
Nella teoria post-keynesiana π e la distribuzione si legano al saggio di crescita dell'economia tramite le propensioni al risparmio. La teoria è stata formulata da Kaldor (v., 1956) e generalizzata da Pasinetti (v., 1962 e 1974), che qui seguiremo. La distribuzione del reddito netto è

$$Y = W + P, \quad (12)$$

dove

$P = Pc + Pw$ (13)

è la somma dei profitti che vanno agli imprenditori-capitalisti (Pc) e dei profitti (interessi) che vanno ai lavoratori (Pw) in quanto il loro risparmio si traduce in capitale (Kw) che genera interessi (iKw).
Il risparmio (S) è

$S = Sc + Sw$, (14)

dove

$Sw = Sw(W + PW)$ e $Sc = ScPc$, (15)

essendo Sc il risparmio dei capitalisti-imprenditori, Sw quello dei lavoratori, sw e sc le rispettive propensioni al risparmio dei lavoratori e dei capitalisti.
L'equilibrio dinamico di piena occupazione del lavoro e del capitale è dato, alla Harrod-Domar, da

$s = vGn$, (16)

dove

$s = S/Y$, $v = K/Y$; (17)

Gn è il saggio naturale di crescita, che esprime la dinamica della popolazione e del progresso tecnico. Supponendo che gli investimenti (I) necessari per la crescita stabilita dalla (16) siano effettuati in prospettiva dell'ampliamento dei mercati, la condizione di equilibrio dinamico

$I = S$, (18)

è da cui si ottiene

$I = sw(W + Pw) + scPc = swY + (sc - sw)Pc$, (19)

$(P/Y) = [1/(Sc - sw)](I/Y) - [sw/(sc-sw)]$. (20)

Questa ultima è la quota del reddito che va ai capitalisti-imprenditori e che dà la distribuzione tra categorie sociali, essendo quella per i lavoratori 1-(Pc/Y). Ma essa non esprime più la quota dei profitti sul reddito, che è invece data dalla (21), da cui risulta come, nelle democrazie sviluppate di mercato, non ci sia più coincidenza tra categorie sociali e di reddito. A ciò segue che il rapporto tra i profitti dei capitalisti-imprenditori e il capitale (Pc/K) non si identifica più con il saggio di profitto del sistema economico:

$(P/Y) = (P/Y) + (Pw/Y)$, (21)

$\pi = (P/K) = (Pc + Pw)/K$. (22)

Se il saggio di interesse è eguale al saggio di profitto, ipotesi plausibile in equilibrio di lungo periodo, dalle (21) e (22) si ha:

$(P/Y) = (1/sc)(I/Y)$ (23)

$(P/K) = (1/sc)(I/K) = (1/sc)Gn$ (24)

Queste relazioni, e in particolare la (24) detta 'equazione di Cambridge', hanno grande importanza per quanto segue.

1. Nella dinamica di lungo periodo la propensione al risparmio dei lavoratori influenza la distribuzione del reddito tra le categorie sociali ma non la distribuzione tra profitti e salari e neppure il saggio di profitto.
2. Per ciascuna categoria sociale i profitti nel lungo periodo sono proporzionali ai suoi risparmi, cioè al capitale che si è formato con i risparmi. I capitalisti-imprenditori, dalle cui decisioni dipende il processo di produzione e di investimento, hanno, con le loro scelte di risparmio, un ruolo strategico per tutta la dinamica.
3. Se sc, v, Gn sono costanti nel tempo, il mantenimento dell'equilibrio di piena occupazione di lungo periodo implica la costanza del saggio di profitto e quella delle quote distributive tra salari e profitti. A ciò s'associa che il salario unitario deve aumentare allo stesso saggio di

crescita della produttività, il che significa che i benefici del progresso tecnico vanno ai lavoratori (v. Pasinetti, 1974) senza escludere un aumento dei profitti.
4. π è determinato esogenamente dal saggio naturale di crescita e dalla propensione al risparmio dei capitalisti-imprenditori. Sono quindi i salari la grandezza distributiva endogena, 'residuale', pur in crescita continua nel tempo.
5. Gli investimenti determinano i profitti e non viceversa. Gli investimenti per la dinamica di piena occupazione determinano il livello dei risparmi necessari e questi sono soddisfatti mediante una adeguata distribuzione del reddito tra salari e profitti e tra categorie sociali.
Ulteriori osservazioni su questa teoria riguarderebbero i dibattiti critici (v. Pasinetti, 1974; v. Baranzini, 1987, per una sintesi) e le estensioni, specie quelle multisettoriali (v. Pasinetti, 1981; v. Quadrio Curzio, 1987).

3. Teoria funzionale della distribuzione

a) Principî generali e richiami storici

La teoria funzionale della distribuzione matura con la grande svolta marginalistica o neoclassica (denominazioni che useremo come sinonimi) dopo il 1871. Il numero degli economisti e delle scuole che l'hanno costruita, pur con differenze, rende impossibile persino una sommaria rassegna sia in un contesto di equilibrio parziale che in uno di equilibrio generale (per rassegne: v. Stigler, 1941; v. Ferguson, 1969; v. Bliss, 1987).
Il principio marginalista venne elaborato in varie applicazioni (domanda, utilità, impresa, benessere, ecc.) che si sono via via integrate perfezionando un principio generale. Secondo tale principio i prezzi dei beni e i prezzi dei fattori di produzione sono determinati da regole comuni e in particolare dalle forze di offerta e di domanda esercitate da soggetti che massimizzano i propri obiettivi (v. Bliss, 1987).
Consideriamo innanzitutto i contributi di equilibrio parziale alla teoria della distribuzione, dove spiccano i nomi di fondatori come P.H. Wicksteed (v., 1894), J.B. Clark (v., 1899), A. Marshall (v., 1890), K. Wicksell (v., 1893). Questi autori, pur partendo dai problemi dei classici per spiegare la distribuzione tra grandi categorie sociali (lavoratori, capitalisti, proprietari terrieri), categorie di redditi (salari, profitti, rendite), categorie di fattori di produzione (lavoro, capitale, terra), se ne staccano poi nettamente. Secondo Wicksteed, che Sraffa definisce "il purista della teoria marginale" (v. Sraffa, 1960, p. VI), i classici, con la

distinzione di ogni categoria sociale-fattore-reddito, rendono "impossibile coordinare le leggi della distribuzione" (v. Wicksteed, 1894; tr. it. p. 126). Bisogna "portare le diverse leggi della distribuzione in una più stretta relazione reciproca. La base comune a queste leggi non deve essere cercata nella natura speciale dei servizi resi dai vari fattori, ma nel fatto, a tutti comune, di prestare un servizio. Se è possibile ottenere una misura oggettiva del servizio reso da ciascun fattore nel suo impiego marginale, allora vi sarà la possibilità di coordinare la distribuzione del prodotto su questa base" (v. Wicksteed, 1894; tr. it. p. 127).
La teoria marginalistica passa dal problema classico della remunerazione di grandi categorie (sociali, di reddito, di fattori) a una teoria generale della produzione e del valore, che riassorbe la spiegazione della distribuzione nella spiegazione del pagamento dei servizi di ogni tipo di fattore produttivo, senza che il 'nome' del fattore abbia alcuna rilevanza. Così da una teoria (classica) sociale della distribuzione si passa a una teoria (neoclassica) funzionale, dove vi è una stretta connessione tra la prestazione di ogni fattore, la sua domanda e disponibilità, la sua remunerazione. Da ciò sono state ricavate persino proposizioni di tipo 'etico'. Clark ha valutato la teoria neoclassica come un principio di giustizia distributiva in quanto il comportamento dei soggetti economici non può scostare le remunerazioni dalla loro produttività marginale (v. Blaugh, 1962).

b) Schema teorico aggregato di base e teorema di Eulero
Iniziamo con una versione neoclassica aggregata - che noi denomineremo 'Wicksteed-Wicksell' -, la quale consente di ben vedere prima il raccordo con i classici e poi il distacco dagli stessi. Essendo Y il prodotto netto, T la terra, L il lavoro, K il capitale, si consideri la funzione di produzione (25) continua, omogenea di primo grado, derivabile, con derivate prime parziali positive e seconde negative, e l'identità distributiva (26):

$Y = F(L, T, K)$, (25)
$Y = wL + rT + \pi K$. (26)

In regime di concorrenza perfetta la remunerazione dei fattori sarà data dalle rispettive produttività marginali. Si avrà quindi il cosiddetto 'esaurimento del prodotto', secondo il teorema di Eulero:

$$Y = (\partial Y/\partial L)L + (\partial Y/\partial T)T + (\partial Y/\partial K)K. \quad (27)$$

Una variante più particolare e di grande successo della (25) è la funzione:

$$Y = cL1\text{-}\alpha K\alpha \quad (28)$$

dove c e α sono costanti. Questa funzione di produzione, detta 'funzione Cobb-Douglas', ha molte proprietà logico-intuitive. Tra queste le produttività marginali del lavoro e del capitale

$$\partial Y/\partial L = c(1\text{-}\alpha)(K/L)\alpha = w \quad (29)$$

e

$$\partial Y/\partial L = c\alpha(L/K)1\text{-}\alpha = \pi \quad (30)$$

dalle quali risulta che il saggio di salario e quello di profitto dipendono anche dall'impiego relativo di L e K. A loro volta le quote relative di salari e profitti sul reddito nazionale sono espresse, rispettivamente, dalla (31) e dalla (32),

$$(\partial Y/\partial L/Y) = (1\text{-}\ \alpha)/L \quad (31)$$

e

$$(\partial Y/\partial L/Y)/Y = \alpha/K, \quad (32)$$

da cui discende la famosa proposizione sulla costanza delle quote distributive, dati i parametri della tecnologia.
Molti sono i problemi a proposito di queste relazioni. Ne segnaliamo tre: quelli che riguardano, rispettivamente, l'esaurimento del prodotto, i coefficienti e la misurazione del capitale.
Circa l'esaurimento del prodotto il problema non si poneva per i classici, in quanto una grandezza distributiva era residuale. Ma qui tutte sono determinate simultaneamente, sicché nulla garantisce l'esatta distribuzione. Il lungo dibattito svoltosi ha riguardato in gran parte il

teorema di Eulero (v. Stigler, 1941; v. Quadrio Curzio, 1972) e le funzioni omogenee di primo grado. Altri problemi sono sorti per la fissità assoluta o relativa dei coefficienti di produzione e per l'indivisibilità di alcuni servizi produttivi che, laddove si verificasse, renderebbe impossibile applicare la teoria marginale (v. Stigler, 1941).
Circa il capitale, si può dire che la precedente impostazione, mentre è relativamente accettabile quando si ha a che fare con i fattori L e T, ciascuno dei quali può considerarsi espresso in entità fisiche internamente omogenee, diventa inutilizzabile quando si usa il 'fattore capitale'. Questo è infatti un insieme di beni eterogenei, unificabile solo in termini di valore, e a tal fine è necessario conoscere il saggio di profitto, che però, a sua volta, può essere determinato solo in base al valore del capitale. Vi è dunque una 'circolarità' per la soluzione della quale si sono impegnati molti economisti - da Wicksell (v., 1893), fino a Samuelson (v., 1962) e altri -, ma i problemi sono rimasti (v. Robinson, 1954; v. Garegnani, 1960; v. Pasinetti, 1966; v. Harcourt, 1972; v. Bliss, 1975). La conclusione è che, in assenza di una misurazione delle grandezze aggregate, la produttività marginale del capitale è indefinita.
Ciò detto, la miglior formulazione della teoria marginalistica della distribuzione si ha passando a un livello microeconomico e a uno di equilibrio generale.

c) Formulazione microeconomica, domanda e remunerazione dei fattori
La maggior forza interpretativo-intuitiva della teoria marginalistica della distribuzione si ha in riferimento all'impresa.
La formulazione elementare (v. Dorfman, 1987) si fonda sulle seguenti ipotesi: a) le imprese massimizzano il profitto; la domanda dei fattori di produzione e il loro impiego sono basati sul confronto tra prezzi dei fattori, produttività, prezzi dei prodotti; b) le funzioni o tecniche di produzione sono in proporzioni variabili, per cui eguali successivi aumenti di un fattore, fermi restando gli impieghi degli altri, generano incrementi successivi decrescenti della produzione fisica; c) i mercati dei fattori e quelli dei prodotti sono in concorrenza perfetta.In termini formali si ha:

$P = R^* - C;$ (33)
$g(q1, q2, ..., qm; f1, f2, ..., fn) = 0.$ (34)

L'impresa massimizza il profitto (33), dove R* sono i ricavi e C i costi, subordinatamente alla funzione di produzione implicita, differenziabile e senza argomenti negativi, (34); da scegliere sono le qi, le quantità dei prodotti, e le fi, le quantità dei fattori. Su base annua i ricavi e i costi sono, rispettivamente,

$R = \sum piqi$ (35)

e

$C = \sum cifi$, (36)

ove pi sono i prezzi dei prodotti e ci i prezzi dei fattori. La scelta delle qi e delle fi che massimizzano P è data dalle seguenti condizioni 'marginalistiche'.
1. Il saggio marginale di sostituzione tra due fattori di produzione, cioè - nella relazione (37) specificata qui di seguito - la variazione di fj che sostituisce una variazione di fi, ferma restando la produzione, deve soddisfare la condizione (38), secondo la quale devono essere eguali i costi di sostituzione dei fattori che, variando in direzione opposta mantengono costante la produzione:

$dfj/dfi = -(\partial g/\partial fi)/(\partial g/\partial fj)$; (37)
$(dfj/dfi) = ci/cj$; $dfj \bullet cj = dfi \bullet ci$. (38)

2. Il saggio marginale di 'trasformazione' di un prodotto in un altro, dato dalla (39), restando fermi gli impieghi dei fattori, deve soddisfare la condizione (40), per cui le variazioni di produzione di beni in direzioni opposte devono essere eguali in termini di valore:

$(dyi/dyk) = -(\partial g/\partial yk)/(\partial g/\partial yi)$; (39)
$(dqi/dqk) = pk/pi$; $dqi \bullet pi = dqk \bullet pk$. (40)

3. La produttività marginale di ogni fattore, data dalla (41), deve soddisfare la (42), per cui si ha l'eguaglianza tra il valore di un prodotto incrementale e il valore del fattore incrementale necessario per ottenerlo:

$(dqi/dfj) = -(\partial g/\partial fj)/(\partial g/\partial qi)$; (41)

$(dqi/dfj) = cj/pi$; $dqi \bullet pi = dfj \bullet cj$. (42)

Questa è la sintesi della teoria marginalistica della distribuzione in condizioni di equilibrio parziale. Essa è, in fondo, una teoria della domanda dei fattori da parte di imprese operanti in condizioni di concorrenza perfetta e in presenza di date disponibilità di fattori di produzione. Una critica importante a questa impostazione è quella che prende di mira l'irrealismo dell'ipotesi di concorrenza perfetta; tale critica ha dato luogo a un lungo dibattito sulle forme di mercato non concorrenziali (v. Chamberlin, 1933; v. Robinson, 1933).

d) Equilibrio generale e distribuzione

Nella teoria dell'equilibrio economico generale il sistema economico viene concepito come un grandioso blocco unitario nel quale confluiscono le decisioni di tutti gli agenti relativamente a tutti i beni, tra i quali vi sono i vari tipi di servizi di lavoro, le terre di diverse qualità, le miriadi di beni capitali fisicamente individuabili. Nei modelli, tutti di carattere uniperiodale, predisposti da Walras (v., 1874-1877) e perfezionati da Pareto (v., 1896-1897 e 1906) i prezzi di tutti i beni, compresi i fattori produttivi, originari e non, sono determinati simultaneamente, con l'intervento del banditore, in modo da pareggiare la domanda e l'offerta di ciascun bene. Non si dà quindi alcun problema della distribuzione distinto dal problema generale della simultanea determinazione dei prezzi di tutti i beni (v. Equilibrio economico).

4. Teoria delle forme di mercato e distribuzione

a) Una possibile 'intersezione' delle precedenti teorie

Entrambe le precedenti teorie contengono elementi per comprendere la distribuzione, malgrado i loro contrasti spesso aspri. Una linea 'intermedia' o 'mista' sembra essere quella che parte da Kalecki ed è stata sviluppata da altri, tra cui Sylos Labini. È una teoria centrata sulle relazioni prezzi-costi-distribuzione con riferimento alle forme di mercato prevalenti nel settore manifatturiero delle economie industrializzate e che rivolge molta attenzione ai fenomeni storico-stilizzati della distribuzione.

b) Il grado di monopolio

La teoria di Kalecki (v., 1939 e 1954) ha rilevanza per l'analisi della dinamica della quota salari sul reddito nazionale in rapporto alla dinamica della relazione costi-prezzi e al 'grado di monopolio'. In base alla relazione (43) (v. anche Laski, 1987), dove P sono i profitti lordi, B gli stipendi, k il rapporto di mark up, W i salari, M il costo delle materie prime (o più in generale dei materiali di base), la quota relativa (α) dei salari sul reddito nazionale (Y) (in breve la 'quota salari') è data dalla (44), ovvero dalla

$P + B = k(W + M)$; (43)

$\alpha = W/[W + k(W + M)] = [W/(W + P + B)] = W/Y$; (44)

$\alpha = 1/[1 + k(1 + j)]$. (45)

Le grandezze cruciali per spiegare α sono dunque k, cioè il rapporto di mark up rispetto ai costi del lavoro e ai costi dei materiali di base, e j, cioè il rapporto tra costi dei materiali e costi del lavoro:

$k = (P + B)/(W + M)$; (46)

$j = M/W$. (47)

Queste grandezze dipendono dal meccanismo di formazione dei prezzi. I prezzi delle materie prime sono determinati dalla domanda, mentre quelli dei beni finiti sono determinati dai costi, in quanto la loro offerta è elastica in virtù di riserve di capacità produttiva. Nella struttura oligopolistica (specie nel manifatturiero) i prezzi dei prodotti finiti sono stabiliti dai produttori che applicano un mark up sui costi primi medi per pagare gli stipendi e ottenere dei profitti. Il rapporto di mark up dipende a sua volta dal 'grado di monopolio', che è funzione del grado di competizione sul prezzo. Da queste premesse possono essere tratte almeno quattro conclusioni.

1. Un aumento sia del grado di monopolio, e quindi di k, sia dei prezzi delle materie prime rispetto al costo-salario, e cioè di j, causa un decremento della quota salari. Storicamente la stabilità osservata è dovuta invece al movimento contrapposto di k e di j.
2. In Gran Bretagna la quota salari non è cambiata molto tra il 1880 (41,4%) e il 1913 (39,4%). Ciò è dovuto a una relativa costanza statistica di j e di k. Dal 1913 al 1935, invece, j

è calato notevolmente, in quanto sono calati i prezzi delle materie prime rispetto ai costi-salario. Essendo rimasta pressoché invariata la quota dei salari (tra il 43% e il 41,8%), Kalecki ne deduce che è notevolmente aumentato il grado di monopolio, e cioè k. Negli Stati Uniti la quota salari è rimasta abbastanza stabile tra il 1909 (37,9%) e il 1925 (40,2%), in presenza di un notevole aumento del grado di monopolio (e quindi di k), che è stato controbilanciato dalla caduta di j (v. Kalecki, 1939).

3. Il rapporto di mark up (k) ha andamento anticiclico: aumenta nelle depressioni, per le decisioni imprenditoriali tese a ricostituire i profitti che calano con la produzione, e diminuisce nei boom, dove si rafforza il potere dei sindacati. A loro volta i prezzi delle materie prime determinati dalla domanda (e di conseguenza j) sono ciclici: crescono con la domanda nei boom e calano nelle depressioni. Perciò i movimenti di k e di j almeno parzialmente si compensano determinando una quota salari costante nel breve periodo.

4. La concentrazione dell'industria tende nel lungo periodo ad aumentare il 'grado di monopolio' e vari settori industriali diventano 'oligopolistici'. Ciò può essere controbilanciato dal venire meno di imperfezioni del mercato, per la caduta dei costi di trasporto, per la standardizzazione dei beni, per l'organizzazione degli scambi, ecc. "I risultati a cui si perviene in questo saggio hanno un aspetto più generale. Un mondo in cui il grado di monopolio determina la distribuzione del reddito nazionale è un mondo che si allontana molto dal modello della libera concorrenza [forma di mercato che] come descrizione dello stato normale dell'economia capitalistica è puramente un mito" (v. Kalecki, 1939; tr. it. p. 234).

c) Il grado di oligopolio

Consideriamo adesso il contributo di Sylos Labini, concentrandoci su un suo lavoro (v. Sylos Labini, 1979), raccordo di sue varie analisi che si fondano sulla sua teoria dell'oligopolio (v. Sylos Labini, 1956), mentre Kalecki postula piuttosto un grado di monopolio.

Sylos Labini rileva che una risposta generale al problema della relazione costi-prezzi non può essere data, perché, malgrado le molte stime econometriche, varie equazioni includono elementi di domanda che oscurano tale relazione. Questo problema viene superato utilizzando il modello teorico di oligopolio (v. Sylos Labini, 1956 e 1984) per l'industria manifatturiera, dove si presentano tre tipi di oligopolio: concentrato, differenziato e misto.

Si parte dalle seguenti relazioni, che incorporano il principio del costo pieno nel modo più

semplice:

$p = av = a(w/z) + am$; (48)
$v = (w/z) + m$. (49)

In queste relazioni p è il prezzo, a è il margine proporzionale (mark up sui costi diretti), v è il costo diretto o variabile, dato dalla somma del costo del lavoro e delle materie prime per unità di prodotto, w/z è il rapporto tra salari monetari orari e produttività oraria in termini reali, e cioè il costo del lavoro per unità di prodotto, m è il costo delle materie prime per unità di prodotto.
Sylos Labini stima econometricamente per vari paesi la relazione che lega le variazioni costi-prezzi nell'industria manifatturiera, concludendo che, mentre la traslazione dei costi delle materie prime sui prezzi è abbastanza completa e simmetrica, la traslazione dei costi del lavoro tende a essere parziale e asimmetrica a seconda del tipo di interdipendenze oligopolistiche e a causa della concorrenza internazionale.
Passando alla relazione prezzi-costi-distribuzione nell'industria manifatturiera, Sylos Labini, dopo aver stimato la (48) per l'Italia e gli Stati Uniti, passa a considerare la quota salari sul valore aggiunto data dalla (50), dove, essendo X il volume della produzione, $Yi=piX-mX$ è il valore aggiunto industriale, $W=(w/z)X$ è il totale salari, $Df=p-m$ è il deflattore del valore aggiunto. La quota salari varia dunque inversamente ad a e a m.

$W/Yi = (1/a) - (m/Df)a^*$; $a^* = (a - 1)/a$. (50)

Il margine proporzionale dà il potere di mercato dei produttori industriali, che, a settore isolato, può esprimersi come capacità sia di aumentare i prezzi sia di ridurre i salari.
La relazione inversa con m indica un antagonismo tra lavoratori industriali e produttori di materie prime (e/o tra paesi industrializzati importatori e paesi produttori di materie prime). Un antagonismo esiste dunque tra il margine proporzionale e la quota salari, ma anche tra questa e il prezzo delle materie prime. Se quest'ultimo aumenta, la quota salari cala, a meno che il margine proporzionale non cali in una misura che compensa la crescita del prezzo delle materie prime. A sua volta la quota profitti sul valore aggiunto industriale è data dalla (51), dove $P=Yi-W-Co$ rappresenta i profitti totali e Co il valore complessivo dei costi indiretti. Se

a si riduce, lo segue la quota profitti, che cala anche al crescere dei costi generali.

$(P/Yi) = a*(pi/Df) - (Co/Yi)$. (51)

Passando dal piano teorico a quello empirico, si può precisare quanto segue.
1. In Italia e in Gran Bretagna la quota profitti ha mostrato una tendenza a calare, sia pure con fluttuazioni, almeno negli ultimi anni settanta, e così è stato, sia pure meno rapidamente, negli Stati Uniti e in altre economie di mercato avanzate.
2. Le cause di questo declino sono: la competizione internazionale, che agisce come freno sui prezzi; l'aumento del potere contrattuale dei sindacati; l'aumento del potere di mercato dei paesi in via di sviluppo produttori di materie prime e di fonti di energia; l'aumento dei costi generali unitari nel lungo periodo.
3. L'aumento dei costi indiretti unitari è dovuto più all'aumento relativo degli stipendi che a quello degli ammortamenti e di altri oneri in conto capitale. A causa di ciò la quota profitti può rimanere costante solo se cresce il margine proporzionale sui costi diretti, e questo può accadere nel lungo periodo solo se a periodi di aumento dei costi seguono periodi di cadute dei costi di sufficiente intensità e durata. Tuttavia la quota dei costi generali unitari, mentre nel breve periodo cresce a un saggio stabile, largamente indipendente dal saggio di crescita della produzione, cala al crescere del reddito in fasi di rapida ripresa economica e di prosperità. In questi periodi la quota profitti cresce anche con un margine proporzionale lordo stabile o calante.
4. Per quel che riguarda le materie prime importate, l'aumento dei loro prezzi acutizza l'antagonismo tra lavoratori e imprenditori industriali. Le difficoltà delle economie industrializzate per l'aumento di tali prezzi (specie di quello del petrolio) derivano non tanto dalla riduzione del margine proporzionale quanto da quella della domanda, che aumenta i costi unitari generali e comprime la quota profitti.
Il contrasto tra paesi produttori di materie prime e paesi produttori di manufatti può essere visto tramite le variazioni di m/Df nella (50), che è anche un modo per vedere i termini di scambio tra materie prime e prodotti industriali.La conclusione è che esiste un antagonismo trilaterale tra lavoratori, industriali, produttori di materie prime.
Su tale base sarebbe di grande interesse esaminare le molte stime econometriche sulla dinamica delle quote relative. Citiamo solo quelle fatte da Sylos Labini sulla distribuzione del

valore aggiunto manifatturiero in Italia e Stati Uniti. In Italia dal 1951 al 1978 la quota salari e stipendi è cresciuta di 17,3 punti, la quota salari è cresciuta di 5,7 punti, la quota costi indiretti e tasse è cresciuta di 13,7 punti e la quota profitti è calata di 19,4 punti.
Negli Stati Uniti la situazione è più complessa, perché si sono ridotte sia la quota profitti che la quota salari, mentre sono aumentate la quota salari più stipendi e la quota costi indiretti più tasse.

5. La distribuzione personale del reddito

a) Le curve di Lorenz

L'analisi della distribuzione personale, o familiare, del reddito fornisce la misura del grado di diseguaglianza dei redditi individuali, o familiari, in un dato paese, la loro evoluzione, il confronto fra le distribuzioni del reddito di paesi diversi. Si tratta di una linea d'analisi avviata da Pareto (v., 1896 e 18961897) e proseguita da Lorenz (v., 1905) e da Gini (v., 1910) nonché da una serie di statistici e di economisti alla ricerca della miglior misura della concentrazione del reddito (v. Dagum, 1987).
Lo strumento più diffuso per rappresentare e analizzare la distribuzione personale, o familiare, del reddito è la 'curva di Lorenz', che stabilisce una corrispondenza tra quote di soggetti (individui o famiglie) di una popolazione e quote di reddito nazionale che vanno a questi soggetti. Si consideri il diagramma riportato nella fig. 3: sulle ascisse sono indicate le percentuali della popolazione e sulle ordinate quelle del reddito nazionale. La curva di Lorenz ha un andamento lineare in caso di equidistribuzione, cioè quando ogni percentuale di popolazione riceve la stessa percentuale di reddito. Le due curve sotto la retta 0B esprimono invece diversi gradi di diseguaglianza nella distribuzione: la curva I corrisponde a un paese in cui la distribuzione è più egualitaria che nel paese cui corrisponde la curva II.
Se le curve di Lorenz di due distribuzioni di reddito non si intersecano, allora presenta maggior concentrazione la distribuzione la cui curva sta al di sotto dell'altra.
L'ordinamento delle distribuzioni di reddito in base alle curve di Lorenz è solo parziale nel caso di intersezione delle curve; ma è possibile ricorrere a indici che consentono un ordinamento completo in termini di maggiore o minore concentrazione. Nel caso di curve di Lorenz che si intersecano è tuttavia possibile che diversi indici diano luogo a diversi ordinamenti.

Una linea di indagine che ha molto attirato l'attenzione degli economisti (v. Atkinson, 1975; v. Atkinson e Harrison, 1978; v. Sen, 1973; v. Kakwani, 1987) è quella relativa all'utilizzo delle curve di Lorenz per lo studio delle connessioni tra ordinamenti nella distribuzione del reddito (e della ricchezza) e benessere sociale e per le misure della diseguaglianza e della povertà.

b) Alcuni dati statistici

Consideriamo brevemente l'analisi empirica che ha fornito interessanti elementi sulla distribuzione personale come risulta dai dati della Banca Mondiale che sono tra i più diffusi.Sulla base di tali dati (v. World Bank, 1992), facciamo alcuni confronti, avvertendo tuttavia che tali confronti non sempre sono attendibili, per la diversità degli anni di riferimento e per altre particolarità di stima.

Ogni popolazione esaminata è stata suddivisa in cinque parti uguali, dette, appunto, 'quintili' (in altre parole un quintile corrisponde a un 20% della popolazione): il I quintile corrisponde al 20% più povero della popolazione, il V al più ricco.

La Nuova Zelanda ha una distribuzione più diseguale della Spagna e un reddito medio pro capite superiore, anche se, a tale riguardo, la Spagna ha guadagnato terreno dal 1985 al 1990.Il Belgio ha una distribuzione meno diseguale della Spagna e della Nuova Zelanda e un reddito medio pro capite superiore a quello spagnolo, anche se relativamente peggiorato. Il Regno Unito ha una distribuzione più diseguale del Belgio con un reddito medio pro capite superiore e in relativo miglioramento. Il confronto tra Italia e Regno Unito dà una distribuzione più equa per l'Italia per i primi due quintili e per il V, ma più diseguale per il III e il IV. Ciò significa che sono i ceti medio e medio-alto quelli relativamente meno coperti rispetto al Regno Unito, mentre per quelli basso, medio-basso e alto la distribuzione è più equa. Il reddito medio pro capite italiano ha notevolmente guadagnato terreno dal 1985 al 1990. Il confronto fra Italia e Olanda rivela per quest'ultima una distribuzione meno diseguale a parità, circa, di reddito pro capite. Il confronto tra Giappone, Germania e Stati Uniti rivela come i primi due abbiano una distribuzione meno diseguale degli ultimi. Passando ad alcuni paesi in via di sviluppo, si rileva che il Bangladesh ha una discreta equidistribuzione, purtroppo della miseria. La Colombia ha una fortissima concentrazione del reddito nel 20% più ricco della popolazione e quindi ha una forte diseguaglianza. Botswana e Brasile presentano una situazione ancora più diseguale. In Brasile il 20% più

ricco della popolazione riceve il 62,6% del reddito nazionale. Un'analisi adeguata della distribuzione nei paesi in via di sviluppo, specie in quelli di grandi dimensioni come il Brasile, dovrebbe considerare anche gli squilibri territoriali che si associano a concentrazioni di classi povere e ricche.Limitatamente ai paesi sviluppati, facendo dei confronti entro ogni quintile, si rileva quanto segue.
I quintile: il massimo si ha per il Giappone (8,7) e il minimo per l'Austria (4,4); il Giappone dà quindi ai suoi ceti più modesti una quota di reddito maggiore dell'Austria.
II quintile: il massimo si ha per il Belgio (13,7) e il minimo per la Nuova Zelanda (10,8).
III quintile: il massimo si ha per la Norvegia (18,9) e il minimo per la Nuova Zelanda (16,2).
IV quintile: il massimo si ha per la Danimarca (25,6) e il minimo per la Svizzera (22,1).
V quintile: il massimo si ha per la Svizzera (44,6) e il minimo per il Belgio (36,0).L'analisi di questi dati richiede sempre molta prudenza. Il fatto che la Svizzera abbia la massima concentrazione di reddito nel quintile di popolazione più alto non significa che i suoi ceti meno abbienti godano di un benessere inferiore rispetto ai ceti più ricchi di altri paesi, dal momento che il reddito pro capite svizzero è di gran lunga superiore a quello di tutti gli altri paesi.
Questa riflessione apre la strada ai problemi relativi alla distribuzione del benessere e della ricchezza.

6. La ricchezza e la sua distribuzione

a) Le difficoltà delle analisi della ricchezza

Si pensa comunemente che il problema della ricchezza sia il problema centrale, almeno storicamente, dell'economia politica: ne è dimostrazione immediata il titolo dell'opera che i più considerano la pietra angolare della scienza economica contemporanea, e cioè La ricchezza delle nazioni (1776) di Adam Smith. Ma l'analisi della distribuzione della ricchezza non sembra aver raggiunto una elaborazione paragonabile alla teoria della distribuzione del reddito, principalmente perché il concetto di ricchezza è stato oggetto di molti equivoci o, quantomeno, di molte definizioni che hanno dato luogo ad altrettante linee di ricerca. Per lungo tempo la ricchezza è stata considerata come stock di beni materiali e più spesso come stock di risorse naturali e, al più, di metalli preziosi. Con la rivoluzione industriale la ricchezza è stata intesa come flusso di produzione e capacità di produzione.

Molti economisti hanno parlato di ricchezza trattando invece del reddito. Lo stesso Smith considerava ricchezza la capacità di produrre flussi di reddito nel tempo piuttosto che gli stocks di beni materiali in un dato momento (v. Blaugh, 1962; v. Pasinetti, 1975). Questa scelta, implicita o esplicita, si è spesso generalizzata, tant'è che oggi per misurare la 'ricchezza' nei diversi paesi si guarda innanzitutto al livello del reddito (totale e pro capite). Altri hanno identificato la ricchezza con l'utilità derivante da servizi e quindi, in definitiva, con il benessere (v. Heilbroner, 1987).
Ciò premesso, noi procederemo come segue: da un lato daremo elementi di rassegna sulle teorie della distribuzione della ricchezza, nell'impossibilità, data la loro natura non omogenea, di presentarne una esposizione organica, come per quelle del reddito; dall'altro accenneremo ad alcune rilevazioni empiriche sulla ricchezza in Italia.

b) Teorie sulla distribuzione e sull'accumulazione della ricchezza: cenni
Come detto, il processo di distribuzione e di accumulazione della ricchezza ha occupato gli economisti sin dal tempo dei classici, nella maggior parte dei casi limitatamente però al ruolo che lo stesso svolge nel processo di formazione e di distribuzione del reddito. Più recentemente i marginalisti hanno riconsiderato l'intera problematica, soprattutto per quanto concerne la distinzione tra ricchezza del ciclo vitale e intergenerazionale (v. Kessler e Masson, 1988). L'approccio dei marginalisti ha tuttavia un taglio fondamentalmente microeconomico, in quanto il processo di accumulazione e di distribuzione non è che la semplice aggregazione, pur tenendo conto della piramide delle età e di altri fattori, dei parametri individuali.
Un'analisi esaustiva del processo di distribuzione e accumulazione della ricchezza dovrebbe invece partire da un quadro macroeconomico ben preciso, sia pure specificato attraverso certe regole per integrarvi i fattori microeconomici come il ciclo vitale, la struttura demografica della popolazione, ecc. Questo approccio è stato proposto dai post-classici e dai post-keynesiani, fra cui Modigliani, il fondatore della teoria del ciclo vitale (v. Modigliani, 1949; v. Modigliani e Brumberg, 1954), secondo la quale la propensione al risparmio e l'accumulazione della ricchezza dipendono dal rapporto tra anni di lavoro e anni di post-lavoro. I soggetti tendono infatti a livellare i consumi nel corso della vita, bilanciando gli stessi attraverso l'accumulazione, prima, e poi, nella vecchiaia, la decumulazione del risparmio.

L'approccio proposto dai post-classici e dai post-keynesiani, oltre che sulle proprietà viste a proposito della distribuzione del reddito, dovrebbe gettar luce sui meccanismi di natura storica che stanno alla base della formazione e della conservazione di classi differenti, anche per una diversa dotazione di capitale intergenerazionale: questo elemento è evidentemente importante per analizzare il processo di crescita del sistema e il suo grado di capitalizzazione. Diversi sforzi in questa direzione sono stati fatti in questi ultimi anni. Atkinson (v., 1975) ha sviluppato un semplice modello di distribuzione funzionale dei risparmi del ciclo vitale, basato su un modello marginalista. Vaughan (v. Kessler e Masson, 1988) ha considerato alcuni problemi relativi alla distribuzione della ricchezza in un modello microeconomico con la presenza di due classi di risparmiatori: i lavoratori e i capitalisti. Pasinetti (v., 1974) ha espresso il capitale di equilibrio delle classi dei lavoratori e dei capitalisti all'interno del modello distributivo di Kaldor e Pasinetti. Baranzini (v., 1991) ha in particolare mostrato che le classi tradizionali del modello ricardiano e post-keynesiano in generale possono coesistere nel sistema, entrambe con due forme di capitale (quello del ciclo vitale e quello intergenerazionale), pur notando che le condizioni di sopravvivenza a lungo termine per le classi che traggono reddito principalmente dal capitale o dalla rendita sono legate a una propensione al risparmio superiore alla media e, soprattutto, a una propensione alla trasmissione del capitale intergenerazionale (a favore cioè dei propri figli) più forte di quella delle altre classi socioeconomiche. Ancora Baranzini mostra che la presenza di rischio e di incertezza riguardo al rendimento dei risparmi tende a polarizzare la società in classi socioeconomiche differenti, il che sembrerebbe rafforzare la validità di un'analisi dei processi distributivi (sia del reddito che della ricchezza) basata, tra l'altro, sulla presenza di gruppi sociali differenti.

La ricerca svolta in quest'ambito negli ultimi due o tre decenni conferma, in sostanza, la validità di un approccio globale all'ampio problema della distribuzione, in base al quale si parte da un modello macroeconomico ben preciso (sia esso di natura squisitamente classica oppure neoricardiana o post-keynesiana) e successivamente lo si integra con sottosistemi di natura microeconomica che spieghino i comportamenti degli individui, delle dinastie o delle classi. In quest'ultimo caso la teoria del ciclo vitale e i meccanismi di trasmissione intergenerazionale del capitale, sia pur basati su una funzione di utilità molto generale, possono risultare preziosi nel comporre il quadro generale dell'analisi.

c) Analisi empiriche della ricchezza in Italia

Le analisi empiriche della ricchezza e della sua distribuzione sono moltissime (cfr., per esempio, "Review of studies in income and wealth"), sicché è impossibile darne conto in termini sintetici come s'è fatto per il reddito. Gli strumenti di analisi impiegati vanno dalle curve di Lorenz alle indagini campionarie relative a varie classificazioni economico-giuridiche di ricchezza.

Limitiamoci perciò a citare l'indagine campionaria svolta periodicamente (la prima volta nel 1986) dalla Banca d'Italia, concernente i bilanci delle famiglie italiane con riferimento sia al reddito sia alla ricchezza. Dall'indagine del 1989 (v. Banca d'Italia, 1991) risulta che la ricchezza reale familiare è in media di 124,1 milioni di lire, circa 3,6 volte il reddito medio. Tale ricchezza viene investita per l'86,2% in immobili, per il 10,4% in partecipazioni in aziende, per il 3,5% in oggetti di valore. Le famiglie che possiedono almeno un immobile sono pari al 67,4%.La ricchezza reale netta è in media più elevata al centro (152,6 milioni) che al nord (123,4 milioni) e al sud (116,1 milioni), ed è più elevata per i lavoratori autonomi (251,2 milioni) che per i lavoratori dipendenti (105,5 milioni) e i pensionati (95 milioni). Dal punto di vista della diffusione delle attività finanziarie risulta che l'88,0% delle famiglie possiede almeno uno strumento finanziario. In particolare l'80,9% delle famiglie possiede depositi bancari, il 12,2% depositi postali, il 22,5% titoli di Stato e il 7,4% altri strumenti finanziari.Questi brevi richiami aprono la strada a quella che dovrebbe essere una analisi economico-istituzionale che non potremo fare qui.

7. Conclusione

La precedente analisi si è concentrata sulle principali teorie economiche della distribuzione del reddito e della ricchezza. La teoria sociale si è molto modificata in quasi due secoli. La sua originaria attenzione alle classi sociali stilizzate (lavoratori, capitalisti, proprietari terrieri) si è spostata, col tempo, verso categorie sociali molto più articolate e cioè percettori di salari, di profitti, di rendite e anche di redditi misti. Costante è però rimasta la sua attenzione per i nessi con la produzione, il valore, la crescita, pur nella autonomia relativa di queste componenti nell'ambito di una teoria generale della dinamica. Così si è passati dalla impostazione classica originaria di Ricardo alle moderne versioni multisettoriali di Sraffa, a quelle macroeconomiche di Kaldor e di Pasinetti, a varie altre che si ricollegano al filone

keynesiano.Talvolta la teoria sociale è stata criticata per aspetti ideologici.
La teoria funzionale ha avuto uno sviluppo molto minore rispetto alle origini, collegate in particolare a Wicksteed e a Wicksell. Essa è rimasta infatti, nella formulazione microeconomica, la teoria di massimizzazione dei profitti dell'impresa che domanda fattori di produzione sulla base del confronto tra l'andamento del loro prezzo, della loro produttività, del prezzo dei prodotti, il tutto in regime di concorrenza perfetta. Le critiche d'irrealismo avanzate nei confronti di questa teoria non ne hanno diminuito la diffusione nell'ambito della manualistica; questa popolarità è dovuta sia all'eleganza formale sia all'apparente intuibilità della teoria in questione.
La teoria delle forme di mercato (grado di monopolio e oligopolio), da Kalecki (v., 1939) a Sylos Labini (v., 1956 e 1979) è stata da noi presentata come teoria 'mista', intermedia tra le due precedenti (sociale e funzionale). Le relazioni tra imprese e percettori di salari dipendono dai prezzi delle materie prime, dei prodotti, dai costi di produzione, dal livello di attività economica. Le 'forze' di mercato che determinano la distribuzione sono dunque almeno tre: imprese, sindacati, produttori di materie prime. In tal modo ci si avvicina a una spiegazione più completa delle effettive dinamiche storiche della distribuzione.
La teoria personale della distribuzione (v. Lorenz, 1905) è stata da noi presentata nella sua forma analitica più statistico-quantitativa che economico-qualitativa. Gli economisti hanno però lavorato molto sulle cause della diseguaglianza rispetto alle quali la teoria personale rappresenta una sorta di sintesi nella quale confluisce una grande quantità di fattori sociali, economici, istituzionali. Per capire quali elementi confluiscono in questa teoria bisognerebbe però ampliare l'analisi ad aspetti socioeconomici.
Questo passaggio è anche opportuno per avere una analisi completa della distribuzione della ricchezza. Noi abbiamo trattato questo argomento nella parte conclusiva per due ragioni: perché dal punto di vista della teoria economica esso è ancora molto magmatico; perché abbiamo circoscritto l'analisi alle teorie sociali, funzionali e personali della distribuzione del reddito, tenendo anche conto della teoria del ciclo vitale del risparmio.Il passo successivo di questa trattazione, che è essenzialmente di analisi economica, dovrebbe essere verso la teoria sociale delle classi, dei ceti, della stratificazione, della grande gamma di tipologie di redditi, per arricchire la quale appaiono necessarie anche strumentazioni di altre discipline e molti riferimenti istituzionali a determinati periodi. Ciò dovrebbe portare a un'analisi della crescita delle classi medie nelle loro diverse connotazioni (urbana, impiegatizia, ecc.) e al

superamento della polarizzazione imprenditori-lavoratori, anche in relazione all'evoluzione della struttura produttiva e alla distribuzione territoriale dello sviluppo. Una analisi qualitativa e quantitativa di questa natura deve tenere conto di elementi socioeconomici, anche sotto il profilo storico e istituzionale, ed essere svolta anche comparativamente tra sistemi economici (si veda, per esempio, Sylos Labini, 1986).
È dunque di estrema importanza stabilire un legame tra analisi economica e analisi sociale, così da avere una riflessione socioeconomica che interpreti anche la realtà dei fenomeni della distribuzione periodo per periodo. Una riflessione di questo tipo è particolarmente confacente alla tradizione delle scienze sociali in Italia pur nella distinzione tra discipline e metodi.

ECONOMIA PUBBLICA

di Ruggero Paladini

Economia pubblica

sommario: 1. Premessa. 2. Uno sguardo generale ai dati. 3. Il mercato e l'intervento pubblico. a) I fallimenti del mercato. b) I beni meritori. 4. La redistribuzione del reddito. 5. La stabilizzazione del reddito. 6. Ascesa e crisi dell'intervento pubblico. a) Le cause della crescita: le caratteristiche strutturali. b) Altre cause: scelte pubbliche, partiti, burocrati. 7. Struttura ed evoluzione dei bilanci pubblici. a) Aspetti generali. b) Alcuni approfondimenti. 8. La politica di bilancio degli anni ottanta. 9. Alcune considerazioni generali. 10. La riforma dell'intervento pubblico. a) La fornitura dei servizi. b) L'intervento assistenziale. c) L'intervento previdenziale. 11. Conclusioni. □ Bibliografia.

1. Premessa

Un'analisi dell'economia del settore pubblico presuppone lo sviluppo delle moderne nazioni europee e quindi di strutture statali in cui le spese e le entrate della 'corona' vengono a separarsi chiaramente dal patrimonio privato delle case regnanti; il grande sviluppo dell'economia pubblica è poi un fenomeno sostanzialmente proprio del XX secolo: nei secoli precedenti il livello della spesa pubblica era estremamente ridotto, salvo i momenti in cui le

guerre comportavano una intensa mobilitazione di risorse. Del resto anche durante i periodi di pace la spesa militare e di ordine pubblico assorbiva il grosso delle risorse statali.
Nell'Inghilterra di due secoli fa la quota della spesa pubblica sul reddito nazionale era vicina al 10%; per tutto il lungo periodo dell'industrializzazione questa quota cresce solo lievemente, tanto che alla vigilia della prima guerra mondiale essa è prossima al 13%. La prima guerra mondiale porta ovviamente a un forte aumento della percentuale, che negli anni venti scende, senza tornare però al livello prebellico; la lunga depressione degli anni trenta fa di nuovo crescere la quota, che si impenna con la seconda guerra mondiale. Nel secondo dopoguerra la spesa pubblica crescerà ancora costantemente di quasi venti punti percentuali, arrivando alla metà del reddito nazionale.
Un discorso analogo si può ripetere per gli altri paesi a capitalismo sviluppato, e si può notare che i paesi di seconda industrializzazione presentano, sin dall'inizio, quote di spesa pubblica più elevate: ad esempio l'Italia parte da una percentuale iniziale del 14%, che cresce fino al 18% negli anni precedenti la prima guerra mondiale. Il secondo dopoguerra è caratterizzato da una percentuale leggermente inferiore al 30%, che giunge a superare il 50% negli anni ottanta.
Ciò che appare sorprendente è che il periodo successivo alla seconda guerra mondiale è un periodo di elevata crescita economica, e tuttavia la crescita della spesa pubblica è stata per tutti i paesi superiore a quella del reddito; per l'insieme dei paesi dell'Europa occidentale la quota si colloca a cavallo del 50% del PIL (prodotto interno lordo), mentre negli Stati Uniti e nel Giappone si mantiene a livelli inferiori.
Proprio il caso degli Stati Uniti è emblematico di questa tendenza, in quanto la società americana si è formata sulla base di valori che esaltano l'iniziativa individuale e privata e che guardano con sospetto alle interferenze governative; ebbene, la spesa pubblica, che agli inizi del Novecento si aggirava sul 78% del PIL, e nel fatidico 1929 era ancora sul 10,5%, superava il 20% nel secondo dopoguerra, per giungere addirittura al 34% nel 1989 (v. Sylos Labini, 1991).
Va considerato inoltre che queste percentuali non forniscono un'indicazione completa del peso del settore pubblico nei paesi a capitalismo avanzato: le imprese pubbliche, nazionali o locali, non vengono contabilizzate tra le spese della pubblica amministrazione, anche se vanno ovviamente tenute presenti per una valutazione dell'influenza del ruolo dello Stato.
Così va anche tenuto presente il peso della normativa economica - dal settore monetario e

creditizio al mercato del lavoro - cui ci si riferisce col termine di regulation.
Nei capitoli che seguono si cercherà di esaminare alcuni aspetti statistici su spese ed entrate, i motivi che spiegano l'intervento pubblico e le cause dell'aumento della spesa pubblica e del debito, le politiche di bilancio effettuate dai diversi paesi, le alternative nella gestione della spesa pubblica. L'attenzione si concentrerà sui quattro grandi paesi europei, sul Giappone e sugli Stati Uniti: oltre a costituire la parte più rilevante delle economie occidentali sviluppate, sono anche i paesi dove è più agevole ottenere i dati statistici necessari per il confronto. Si può inoltre notare che gli altri paesi europei presentano dati che non si discostano, come livelli e come tendenze, da quelli esaminati, e la stessa cosa vale per paesi extraeuropei come il Canada o l'Australia.

2. Uno sguardo generale ai dati

Un primo esame dei dati statistici si ricava dalla tab. I, dove vengono riportate le percentuali rispetto al PIL delle spese e delle entrate delle amministrazioni pubbliche: governi centrali, enti locali ed enti della sicurezza sociale. I dati dell'OECD (1985) pongono a confronto il 1960 con il 1982; nel dopoguerra e negli anni cinquanta si è infatti sviluppata una rilevante attività legislativa che ha determinato in tutti i paesi aumenti estremamente rilevanti dell'incidenza sul PIL; in molti paesi, ma come vedremo non in Italia, i primi anni ottanta sono quelli in cui le percentuali arrivano a toccare i livelli più elevati. Si può notare come l'Italia, partendo da un livello inferiore a quello degli altri grandi paesi europei, e di poco maggiore di quello statunitense, arrivi nel 1982 a livelli nettamente superiori; peraltro, le percentuali italiane sono quelle calcolate sulla base della vecchia serie di contabilità nazionale, mentre sulla base della nuova serie esse risultano meno elevate: per esempio, la quota del 53,7% del 1982 risulta, con la nuova contabilità, del 50,1. Il divario tra spese ed entrate è aumentato per tutti i paesi, ma di nuovo in Italia il fenomeno è nettamente più rilevante.
Nella tab. II viene presentata la quota dei consumi pubblici (in cui la principale componente di spesa è costituita dagli stipendi dei dipendenti pubblici) sui consumi finali; nei ventitré anni l'aumento delle percentuali è rilevante, ma inferiore a quello delle spese rispetto al PIL; la posizione italiana non si distingue particolarmente rispetto a quella degli altri paesi. Per quanto riguarda invece la quota degli investimenti pubblici rispetto al totale, la tab. III mostra

come nei paesi europei la situazione non sia mutata significativamente, mentre sono rilevanti la caduta della percentuale statunitense e l'aumento di quella giapponese.
Il peso dei bilanci pubblici nella formazione del reddito familiare viene evidenziato nella tab. IV, che riporta i conferimenti (stipendi e trasferimenti di sicurezza sociale) e i prelievi (imposte dirette e contributi sociali); le variazioni nel ventennio sono di nuovo rilevanti, e la situazione italiana non si differenzia da quella degli altri paesi europei.
Gli ultimi due aspetti generali riguardano il peso dell'occupazione pubblica e delle imprese pubbliche; la tab. V mostra l'aumento delle quote dell'impiego pubblico sul totale, con la netta differenziazione tra il Regno Unito e il Giappone. Si può notare come Regno Unito e Stati Uniti fossero nel 1960 molto vicini come quote, ma, mentre gli Stati Uniti hanno avuto un leggero incremento, questo è stato molto sostenuto per il Regno Unito.
Per quanto riguarda le imprese pubbliche, va chiarito che i criteri di classificazione non sono omogenei come lo sono, ad esempio, le voci dei bilanci pubblici: vi rientrano, in generale, quelle imprese che hanno una personalità giuridica autonoma, anche nella forma di società per azioni (ad esempio le partecipazioni statali italiane); viceversa, alcune imprese vengono ricomprese nell'ambito dell'organizzazione statale, come articolazioni di un ministero, e quindi spese ed entrate rientrano nel bilancio statale; questo è il caso, in Italia, delle poste e, fino al 1989, delle ferrovie. Le tabb. VI e VII presentano dei dati che coprono gli anni settanta per quanto riguarda la quota degli investimenti e dell'occupazione; occorre però una certa cautela nell'effettuare i confronti tra i vari paesi, dato che i criteri usati per identificare le imprese pubbliche sono differenti. Si può comunque notare il peso maggiore della quota degli investimenti rispetto a quella dell'occupazione, e come, ad esempio, la percentuale dell'Italia sia più elevata di quella della Germania per quanto riguarda gli investimenti, mentre il contrario accade per l'occupazione.
Per quanto riguarda il prelievo fiscale, la tab. VIII fornisce la ripartizione percentuale tra il 1969 e il 1971: come si può notare, le differenziazioni nella struttura fiscale sono notevoli, con Francia e Italia da un lato e Giappone e Stati Uniti dall'altro; vedremo come, dopo una quindicina di anni, le differenze si siano notevolmente ridotte. Come si può notare nella tab. I, i deficit pubblici hanno avuto nel ventennio una tendenza al peggioramento; la tab. IX mostra più dettagliatamente questo fenomeno scomponendo in tre sottoperiodi i rapporti medi del deficit rispetto al PIL; come si può notare, la posizione italiana è particolarmente anomala, in quanto parte da una posizione di deficit nel primo sottoperiodo e subisce un

marcato peggioramento. Per quanto riguarda la quota del debito pubblico (cioè della somma dei deficit dei vari periodi) sul PIL, il discorso è più articolato, in quanto il rapporto risente della quota del deficit, ma anche della crescita nominale del PIL stesso. La tendenza generale - come si vedrà meglio più avanti (v. cap. 8) - è stata caratterizzata da un aumento del rapporto debito/PIL. In conclusione, da questi dati emerge una chiara tendenza all'aumento del peso del settore pubblico nell'economia e la posizione dell'Italia non si discosta, in linea di massima, da quella degli altri paesi europei, se non per quanto riguarda il deficit e il debito pubblico. Passiamo ora a esaminare come gli economisti abbiano affrontato il problema della presenza del settore pubblico nell'economia; torneremo a esaminare i dati più in dettaglio per approfondire le tendenze degli anni ottanta e verificarne, su quella base, gli aspetti teorici.

3. Il mercato e l'intervento pubblico

a) I fallimenti del mercato

L'atteggiamento degli economisti classici (intesi in senso lato) era in generale ostile all'intervento pubblico; in parte ciò dipendeva dal fatto che le varie forme di privilegi monopolistici e di regolamentazioni corporative erano frutto di elargizioni (non gratuite) dei sovrani. Anche la spesa pubblica era vista come una sottrazione di risorse allo sviluppo dell'economia; tipica in questo senso è l'analisi di David Ricardo. Occorre tuttavia ricordare che il principe degli economisti classici, Adam Smith, alla fine del libro IV della Ricchezza delle nazioni, aveva aggiunto ai compiti del sovrano, oltre a quelli tradizionali della difesa e dell'ordine pubblico, anche quello di occuparsi di "certain public works and certain public insitutions, which it can never be for the interest of any individual, or small number of individuals, to erect and maintain; because the profit would never repay the expense to any individual or small number of individuals, though it may frequently do much more than repay it to a great society"; questo brano (e quello analogo del libro V, cap. 1, parte III) mostra come il teorico della "mano invisibile" riconosca l'esistenza di casi in cui vi è divergenza tra interesse individuale e interesse collettivo.

L'analisi delle ragioni che giustificano l'intervento pubblico ha occupato uno spazio crescente nel lavoro degli economisti; molti dei temi e dei motivi tradizionalmente da questi indicati possono essere ricondotti a due grandi filoni: il primo li inquadra dal lato della tecnologia produttiva e si connette quindi alle caratteristiche di un'economia industrializzata, il secondo

li riconduce alla dinamica dei rapporti sociali e politici. Nel primo filone si possono far rientrare i tradizionali 'fallimenti del mercato' costituiti dai beni pubblici, dalle esternalità, dai monopoli naturali. Nel caso di beni pubblici ci troviamo di fronte a un tipo di beni o servizi che il mercato non è in grado di erogare, in quanto non risulta possibile, o il costo è troppo oneroso, stabilire il principio di esclusione, cioè realizzare l'identificazione dei consumatori, il che permette la possibilità di applicare un prezzo per il consumo del bene o servizio. Un tipico esempio è costituito dalle strade urbane, con relativa pulizia e illuminazione; un altro tipico esempio sono i fari lungo una costa. Il fruitore del bene non può essere individuato e isolato, e dunque un produttore privato non potrebbe richiedere la remunerazione della prestazione svolta. È necessario l'intervento dell'operatore pubblico, e si discute se tale intervento possa portare al rispetto di un criterio di efficienza tale da assicurare una situazione di 'ottimo paretiano' (v. Benessere, Stato del).

Nel caso di beni e servizi con caratteristiche di esternalità, il fenomeno del bene pubblico puro si presenta in forma attenuata: il bene è escludibile, e quindi viene prodotto dal mercato, ma la condizione di efficienza realizzabile in concorrenza perfetta - prezzo = costo marginale - non si verifica poiché i costi e i benefici privati non coincidono con quelli sociali, in quanto nella produzione o nel consumo vi sono degli effetti, positivi o negativi, che investono soggetti terzi (v. Pigou, 1928). Un caso classico di esternalità negativa è rappresentato dalle diverse forme di inquinamento, uno di esternalità positiva dall'istruzione. Anche in questo caso è necessario un intervento pubblico che riconduca all'equilibrio tra calcolo privato e calcolo sociale; i modi di intervento possono essere diversi, andando dall'assunzione diretta del servizio, all'applicazione di tasse o sussidi, da parte dello Stato, alla sola attribuzione di specifici diritti agli operatori economici, lasciandoli poi alla libera interazione di mercato (v. Coase, 1960). Su ciò si tornerà più avanti; per ora possiamo notare che comunque è necessario un intervento dell'operatore pubblico.

Nel caso dei monopoli naturali la ragione dell'intervento deriva dall'impossibilità di funzionamento del meccanismo concorrenziale; classici casi sono quelli dei beni non riproducibili, come la fonte di acqua minerale di Augustin Cournot. In generale si tratta di quei beni e servizi caratterizzati da tecnologie che richiedono capitali molto elevati; l'impossibilità della concorrenza può essere esaminata sotto due profili: da un lato, il fatto che i costi fissi sono in buona misura irrecuperabili, per cui un produttore eviterà di duplicare gli impianti là dove si è già installato un altro produttore; dall'altro, il fatto che i costi medi

sono decrescenti rispetto alla produzione, il che dà un vantaggio incolmabile al primo arrivato. In tale modo viene spiegata l'esistenza di imprese pubbliche, anche se non è questo il solo strumento di intervento dello Stato.
Accanto a questi tradizionali fallimenti del mercato, negli ultimi anni sono stati suggeriti altri motivi che possono giustificare l'intervento pubblico e che si rifanno anch'essi a una impossibilità per il meccanismo concorrenziale di determinare una situazione efficiente; nei casi in questione l'attenzione è posta sulla mancanza di un adeguato livello di informazione di cui possano venire in possesso gli operatori. Il modello di concorrenza perfetta prescrive infatti uno stato di piena conoscenza da parte degli acquirenti delle caratteristiche del prodotto; anche se ciò costituisce una forzatura, si ritiene che per una larga gamma di beni e servizi il grado di conoscenza dei prodotti che il consumatore può acquisire sia a un livello sufficiente da non creare un problema di allontanamento dal modello concorrenziale.
Esistono però certi beni e servizi dove ciò non si realizza, e il campo dei servizi educativi e sanitari è uno di questi; in questi settori esistono seri vincoli di irreversibilità nelle scelte, tali per cui l'eventuale conoscenza acquisita ex post risulta inutilizzabile per il consumatore.
Più tecnicamente, il problema che si pone è quello di un'asimmetria informativa tra venditore e acquirente; un altro settore in cui il problema può portare a inefficienze allocative è quello delle assicurazioni. Nel caso delle assicurazioni sulla vita, ad esempio, i premi dovrebbero riflettere la speranza media di vita, nella misura in cui non sia possibile discriminare tra gli assicurati, accertando ad esempio dei difetti cardiaci, e richiedere quindi premi più elevati. Ciò può portare al risultato che coloro che ritengono di essere meno esposti a rischi non trovino conveniente sottoscrivere l'assicurazione, portando così a una diminuzione della speranza media di vita degli assicurati, a un aumento dei premi e quindi all'allontanamento di una parte dei potenziali assicurati; si realizza così un fenomeno di 'selezione avversa', per superare il quale si richiede un sistema pubblico di sicurezza sociale, o comunque un intervento pubblico che permetta una protezione per quegli operatori più esposti a rischi che il mercato privato non copre. Questo può essere il caso di alcuni settori economici, come il credito o l'agricoltura, o di alcuni rischi sociali come quelli di disoccupazione o di invalidità; si parla in questi casi di un intervento dovuto ad una incompletezza dei mercati (v. Stiglitz, 1988).

b) I beni meritori

Un'altra ragione che è stata addotta per giustificare l'intervento dello Stato è l'esistenza di beni dei quali si ritiene opportuno aumentare la produzione, o dei quali al contrario si vorrebbe scoraggiare il consumo. In questi casi non si tratta di problemi di inefficienza allocativa; i criteri di ottimalità possono essere rispettati, ma l'ammontare del bene che il mercato determina spontaneamente viene ritenuto o insufficiente o, al contrario, eccessivo: latte e alcolici possono essere i due esempi opposti.Richard Musgrave (v., 1959), cui si deve la prima elaborazione di questa impostazione, parla di un intervento coattivo sulle scelte individuali; ora questo può essere inteso come un caso di paternalismo volontariamente accettato dalla collettività, oppure come il desiderio di una maggioranza di intervenire sulle scelte della minoranza. Ad esempio, nel primo caso i membri della collettività, pur consapevoli della dannosità del fumo, ritengono di non saper resistere al medesimo e richiedono interventi che li obblighino a limitarsi; nel secondo caso invece una maggioranza di non fumatori impone a una minoranza di fumatori limiti al fumo (anche nella forma di prezzi più elevati), limiti di cui la minoranza farebbe volentieri a meno.Si può ritenere che coloro che sostengono l'intervento dello Stato per beni ritenuti meritevoli di tutela abbiano anche in mente un obiettivo redistributivo, secondo un criterio che è stato definito (v. Tobin, 1970) di egualitarismo specifico. Si ritiene cioè che vi siano beni che presentano caratteristiche che li distinguono dagli altri, tali per cui debbano essere allocati secondo criteri diversi dal mercato; essi dovrebbero essere forniti a tutti i membri della collettività senza poter essere oggetto di compravendita: esempi tipici sono certe prestazioni sanitarie, l'istruzione di base, gli alloggi popolari. Da questo punto di vista i beni meritori verrebbero a costituire una delle articolazioni del Welfare State integrando le misure di sicurezza sociale proprie di questo.

4. La redistribuzione del reddito

Alcune delle motivazioni addotte per spiegare l'intervento pubblico puntano sulla necessità di operare una redistribuzione del reddito: misure rivolte contro i rischi di disoccupazione, di invalidità, o di fluttuazione dei prezzi agricoli hanno evidenti contenuti redistributivi, tanto che, come si vedrà, vi sono autori che sostengono che la vera ragione che spiega l'intervento dello Stato sia quella redistributiva. Tuttavia in questi casi, come si è visto, le ragioni addotte sono altre, e si riassumono nell'idea che il mercato privato porterebbe a risultati inefficienti, o

che certi beni presentano una natura tale che si ritiene opportuno escluderli dal normale gioco della domanda e dell'offerta.
Qui invece l'attenzione va posta sulla redistribuzione del reddito quale obiettivo primario dell'intervento pubblico, di cui si possono considerare quattro distinte motivazioni. La prima è quella che si fonda sulla distinzione tra bisogni necessari e bisogni di lusso, e sulla necessità di assicurare, a coloro che non ne hanno i mezzi, il soddisfacimento dei bisogni necessari; questa tesi la ritroviamo tra i pensatori religiosi e in quelli socialisti. La seconda è propria degli utilitaristi inglesi e ha alla base due ipotesi: eguaglianza delle funzioni di utilità di tutti gli individui e decrescenza dell'utilità marginale del reddito; se l'obiettivo dell'intervento pubblico - la funzione di benessere sociale, si sarebbe detto successivamente - è quello di massimizzare il benessere della collettività, allora prelevare una somma a chi è più ricco e darla a chi è più povero fa aumentare l'utilità collettiva.
Gli utilitaristi classici, da Bentham a Pigou, ipotizzavano la confrontabilità delle funzioni di utilità, prestando così il fianco alla critica paretiana dell'inconfrontabilità; più recentemente, John Rawls (v., 1971) ha riproposto l'obiettivo redistributivo come derivante da una scelta contrattuale tra individui che ignorano quale sarà il posto che occuperanno nella società, e che, essendo avversi al rischio, preferiscono che il livello più basso di reddito sia il più elevato possibile. L'ineguaglianza nella distribuzione può essere giustificata solo in quanto porti a un livello minimo di reddito più elevato di quanto si avrebbe con una distribuzione più egualitaria.
Secondo una terza impostazione, che evita anch'essa la critica paretiana all'utilitarismo classico, gli individui traggono soddisfazione sia dal proprio reddito, sia dal fatto che coloro che ne sono privi, in tutto o in parte, ottengano un miglioramento. La povertà è penosa per chi la soffre, ma la visione della povertà arreca disturbo al non povero, per cui una redistribuzione di reddito può presentare caratteristiche di ottimalità paretiana (v. Hochman e Rodgers, 1969). La ragione per cui sarebbe necessario un intervento pubblico è simile a quella addotta per i beni pubblici puri per i quali non vale il principio di esclusione (v. § 3a): se l'intervento fosse lasciato alla beneficenza privata, vi sarebbe una tendenza degli individui ad aspettare che il trasferimento venisse effettuato da qualcun altro (cosiddetto comportamento del free rider), col risultato che l'ammontare di redistribuzione sarebbe inferiore a quello ottimale; da questo punto di vista l'intervento pubblico si spiegherebbe con motivi simili ai fallimenti di mercato, ed è interessante che esso venga accettato anche da un

liberista come Milton Friedman (v., 1962).
Le motivazioni redistributive cui sinora si è fatto cenno partono dal presupposto dell'esistenza, in un dato periodo di tempo, di una quota della popolazione che si trova in condizioni di povertà, si tratti di giovani o di anziani, di disoccupati o di membri di famiglie numerose. In tutti i casi l'intervento mira a redistribuire una quota del reddito prodotto in quel periodo. Un obiettivo redistributivo particolare è quello invece che riguarda le diverse generazioni: ad esempio, quelle la cui vita attiva si è svolta in periodi di bassi tassi di crescita, come nel periodo tra le due guerre mondiali, avrebbero livelli di risparmio nettamente più esigui delle generazioni del secondo dopoguerra, quando i tassi di crescita erano nettamente più elevati. Ciò pone un problema di equità intergenerazionale, che può essere mitigato dai trasferimenti familiari, ma non risolto. I sistemi pensionistici a ripartizione, che sono necessariamente sistemi pubblici, costituiscono il modo per affrontare questo particolare aspetto della redistribuzione, che non è rivolta tanto alle situazioni singole quanto a quelle che sono comuni a intere generazioni (v. Samuelson, 1958).

5. La stabilizzazione del reddito

Sinora i motivi individuati per giustificare l'intervento dello Stato possono essere ricondotti o alla tematica dei fallimenti del mercato o a quella della redistribuzione del reddito; adesso passeremo a illustrare quello che può essere definito come il tema del fallimento macroeconomico del mercato, tema che scaturisce principalmente dall'analisi keynesiana. Per quanto infatti vi siano studiosi che hanno cercato di ottenere risultati keynesiani partendo dall'ipotesi di un mercato a concorrenza imperfetta, la maggior parte degli autori che possono essere qualificati come keynesiani ritiene che il sistema capitalistico di laisser faire difficilmente possa riuscire a conseguire e mantenere un livello di piena occupazione delle risorse, anche se può essere sufficientemente efficiente dal punto di vista dell'allocazione delle risorse impiegate.
La difficoltà di realizzare una piena occupazione delle risorse non dipende dal meccanismo microeconomico di funzionamento del mercato, ma da un paradosso macroeconomico in cui gli operatori del mercato possono imbattersi; l'esempio storico tipico è quello della 'grande depressione' degli anni trenta. Fotografando la situazione economica di quegli anni negli Stati Uniti o nell'Europa sviluppata, si sarebbero potuti notare alti livelli di disoccupazione e di

capacità inutilizzata delle imprese, scorte elevate di materie prime e di altri prodotti, e spesso un sistema bancario con sufficiente liquidità. I disoccupati avrebbero accettato volentieri di lavorare al salario corrente, e le imprese li avrebbero volentieri assunti per produrre se avessero avuto soddisfacenti prospettive di domanda, nel qual caso si sarebbero rivolte al sistema bancario per quei finanziamenti che le banche avrebbero volentieri concesso: ogni operatore attendeva quindi che l'iniziativa partisse da un altro. Secondo la 'teoria dei giochi' diremmo che il sistema economico durante la grande depressione era in preda al 'dilemma del prigioniero'. La tesi centrale di Keynes è infatti che, qualora si trovi in una situazione di sottoccupazione, esso non riesca a uscirne con le sue sole forze, cioè basandosi sui meccanismi del mercato, o che comunque non riesca a farlo in tempi sufficientemente rapidi da non causare alla collettività tensioni politiche e sociali laceranti (questo è il senso del famoso detto "nel lungo periodo siamo tutti morti", spesso citato a prova di un presunto disinteresse di Keynes per i problemi di lungo periodo). La fuoruscita da questo stato di stagnazione può essere prodotta solo da una domanda esterna al mercato; e, prescindendo dal caso in cui un paese approfitti della domanda estera, questa domanda autonoma che funziona da stampella per il sistema economico non può che essere la spesa pubblica.Il processo moltiplicativo che la spesa pubblica innesca permette quindi una ripresa economica, e il problema diventa allora quello di impedire una nuova caduta, o una ripresa troppo accelerata; in sostanza di smorzare le fluttuazioni cicliche. La politica di stabilizzazione deve operare sia attraverso la politica monetaria che attraverso quella di bilancio; la prima deve essere espansiva, in modo da stimolare con bassi tassi d'interesse gli investimenti, mentre la seconda deve essere più restrittiva, in modo da tenere sotto controllo la domanda aggregata (v. Samuelson, 1955). La regola di massima è quella del pareggio di bilancio non su base annuale ma pluriennale, in modo da alternare dei deficit nelle fasi basse del ciclo a dei surplus nelle fasi alte; un altro modo di formulare la politica di bilancio consiste nel prevedere un avanzo del bilancio calcolato a livello di piena occupazione (full employment surplus) (v. Blinder e Solow, 1974).

Un bilancio tendenzialmente in pareggio non è peraltro un bilancio neutro, poiché gli effetti moltiplicativi di una spesa diretta dello Stato sono maggiori di quelli delle imposte; la dimostrazione che anche mantenendo un bilancio in pareggio si può avere un aumento del reddito nazionale (v. Haavelmo, 1945), che nel modello keynesiano più semplice è pari all'ammontare dell'incremento di spesa pubblica, ebbe un ruolo importante nella critica alla

tradizionale regola del pareggio di bilancio. Livelli elevati di spesa e di entrate hanno poi dei positivi effetti automatici di stabilizzazione; nelle fasi di discesa del ciclo la spesa non scende, anzi sale per i maggiori sussidi alla disoccupazione, mentre le imposte scendono (tanto più in quanto le imposte dirette sono progressive), mentre nelle fasi di espansione le imposte crescono più delle spese, realizzando quello che Walter Heller (v., 1966) ha definito il fiscal drag (termine che in Italia ha assunto invece il significato di aumento dell'incidenza dell'IRPEF a causa dell'inflazione). Una politica di stabilizzazione dei cicli più brevi dell'economia deve quindi basarsi su sistemi di stabilizzazione automatica, sia dal lato del bilancio, che dal lato monetario; una politica volta a mantenere un adeguato livello di occupazione dei fattori produttivi, e cioè in sostanza un adeguato livello di crescita, richiede delle politiche discrezionali di intervento, nonché un coordinamento delle politiche economiche tra i vari Stati. Il sistema dei cambi fissi di Bretton Woods e le politiche degli anni cinquanta e sessanta costituiscono un riferimento storico in cui le ricette keynesiane hanno trovato larga applicazione, seppure con modalità e intensità diverse, nei paesi a capitalismo sviluppato.
Occorre tuttavia ricordare che l'impostazione keynesiana della crescita, nella formulazione in particolare di Roy Harrod (v., 1948), permette di individuare un possibile limite critico al tasso 'garantito' di crescita determinato dal bilancio pubblico; il tasso garantito può essere definito come il tasso potenziale massimo di crescita del sistema economico, dato un certo rapporto tra capitale e prodotto e una certa propensione al risparmio. Il bilancio pubblico, anche se è in pareggio e in maggior misura se è in disavanzo, viene a ridurre il tasso garantito. Di conseguenza, mentre un aumento della spesa pubblica determina una spinta positiva alla crescita effettiva del reddito in un dato periodo, allo stesso tempo abbassa il livello massimo di crescita potenziale; a un livello basso del rapporto spesa pubblica-PIL ciò non costituisce un problema, ma a livelli elevati, invece, gli effetti destabilizzanti degli aumenti del peso del bilancio si possono far sentire (v. Paladini, 1984).

6. Ascesa e crisi dell'intervento pubblico

a) Le cause della crescita: le caratteristiche strutturali

Come si è visto nel cap. 2, sino agli inizi degli anni ottanta la spesa pubblica ha continuato ad aumentare il suo peso sul reddito nazionale nei vari paesi sviluppati, e la stessa cosa è

accaduta alle imposte, anche se la crescita di queste ultime non ha impedito un notevole aumento del debito pubblico, che, come vedremo, solo in alcuni paesi è stato stabilizzato nel corso degli anni ottanta; l'intervento pubblico è quindi salito sul banco degli accusati, e ai fallimenti del mercato e ad altre ragioni di intervento si sono contrapposti il fallimento dello Stato e la necessità di un drastico ridimensionamento del suo ruolo, sia come organo di spesa, sia come imprenditore (privatizzazioni) sia come regolatore (deregulation). In questo capitolo vedremo quali siano le ragioni che possono spiegare la crescita continua del peso del settore pubblico nelle economie sviluppate e i motivi del 'fallimento' dello Stato.Una prima ragione è stata individuata in un elemento strutturale di modifica delle economie dei paesi a capitalismo sviluppato: l'industrializzazione e l'urbanizzazione hanno determinato una crescita dei beni pubblici, delle esternalità, dei beni di merito. In questo senso la crescita del ruolo dello Stato sarebbe connessa allo sviluppo economico. Questa era in sostanza la tesi avanzata da Adolf Wagner (v., 1877-1901) e da Francesco Saverio Nitti (v., 1903), ripresa in parte da autori come William Baumol (v., 1952) e Richard Musgrave (v., 1969). Una seconda ragione, anch'essa di natura strutturale, è nota come 'il morbo di Baumol' (v. Baumol, 1967). Il fenomeno individuato riguarda in realtà una larga parte del settore dei servizi; in questo settore dell'economia la produttività cresce a un ritmo molto minore di quanto non avvenga nell'industria, e ciò per una ragione di fondo, e cioè che il lavoro costituisce allo stesso tempo l'input e l'output del processo. Secondo l'esempio dell'autore, un quartetto di Brahms non può che essere suonato oggi da quattro musicisti nello stesso tempo in cui quattro musicisti lo suonavano cento anni fa.

Si supponga ora che la collettività voglia consumare beni prodotti dal settore ad alta produttività e servizi prodotti dal settore a bassa produttività secondo un rapporto fisso, e che d'altra parte i lavoratori del settore a bassa produttività vogliano avere incrementi salariali eguali a quelli di cui beneficiano i lavoratori dell'altro settore. L'inevitabile conseguenza è che le risorse si dovrebbero spostare dal settore ad alta produttività verso quello a bassa produttività; se i servizi di quest'ultimo sono privati, il loro prezzo aumenterebbe tanto che non ne sarebbe possibile un consumo proporzionato agli altri, per cui tenderebbero a sparire; questo può essere il caso di certi artigiani, come i sarti. Se il settore è pubblico, dovremmo assistere a una crescita della spesa per consumi collettivi, e in particolare per retribuzioni, dovuta a un aumento degli occupati pubblici rispetto a quelli privati.

b) Altre cause: scelte pubbliche, partiti, burocrati

Altre analisi centrano invece la loro attenzione sui meccanismi politici di determinazione delle scelte pubbliche in regimi di democrazia e di suffragio universale. Una prima formulazione può essere considerata quella di Allan Peacock e Jack Wiseman (v., 1961) che stabilisce una dialettica tra i partiti al governo e gli elettori: i primi sono propensi ad aumentare la spesa pubblica, perché ne traggono popolarità, mentre ai secondi non piace essere sottoposti per questo motivo a un aumento di tassazione. In condizioni normali ciò fa sì che la quota di spesa e di entrata rispetto al reddito si mantenga costante, ma in circostanze eccezionali (tipico esempio la guerra) la soglia di resistenza dei cittadini può essere superata, con la conseguenza che terminata l'emergenza la quota della spesa e delle entrate non torna al livello iniziale, ma si consolida a un livello più elevato; i contribuenti hanno avuto un fenomeno di assuefazione.

Questa tesi ha ricevuto varie critiche (per l'Italia, v. Brosio e Marchesi, 1986), basate sul fatto che essa non può spiegare il continuo trend di aumento delle quote di spesa e di entrata in situazioni, come quelle del secondo dopoguerra, che non possono considerarsi eccezionali; essa è stata recentemente ripresa da Charles Schultze (v., 1992), il quale, considerando a parte la spesa pensionistica, ha notato un fenomeno di stabilità della pressione fiscale e della tendenza alla trasformazione della spesa militare in spesa civile; rifacendosi agli studi di Daniel Kahneman e Amos Tversky (v., 1986), Schultze parla di un meccanismo asimmetrico degli elettori, analogo a quel framing effect, per cui la disutilità della perdita di un dato ammontare viene percepita in modo molto più netto dell'utilità di un guadagno dello stesso ammontare. Gli elettori si opporrebbero a un aumento di imposte per finanziare spese civili, ma non alla sostituzione di spese militari con spese civili, preferendo questa alternativa a una diminuzione della pressione fiscale.

Si potrebbe riprendere la tesi di Peacock e Wiseman inserendola nel modello macroeconomico keynesiano; si tratterebbe di aggiungere l'ipotesi (abbastanza plausibile) di una propensione alla spesa pubblica, da parte dei governi, sostanzialmente pari all'unità (v. Steindl, 1966): in altre parole i governi spendono tutto ciò che entra nelle casse dell'erario. Data una certa crescita delle componenti autonome della domanda (investimenti ed esportazioni), ne deriva un certo aumento del reddito, e data una certa flessibilità automatica delle imposte, un aumento delle entrate; se questo aumento viene speso, ne deriva un ulteriore aumento del reddito, quindi delle entrate e poi delle spese. Il risultato può portare a

un tasso di aumento della spesa pubblica anche maggiore di quello del reddito, e contribuisce a spiegare perché, anche in periodi di forte aumento del reddito, come gli anni cinquanta e sessanta, si è assistito a un aumento della spesa pubblica. Per i periodi successivi, invece, in cui il tasso di crescita del reddito è calato, l'ipotesi non funziona, per cui si tratta di passare a spiegazioni basate sulla relativa rigidità della spesa pubblica rispetto alle altre componenti della domanda globale.

Secondo altre analisi, il problema fondamentale deriva dalla caratteristica essenziale che distingue la sfera delle decisioni pubbliche da quella delle decisioni private, che avvengono sul mercato; infatti, in quest'ultimo vale il principio 'una lira, un voto', mentre nelle scelte pubbliche vale il principio 'una testa, un voto' (v. Cosciani, 1977). Da questo deriverebbe una tendenza a utilizzare la spesa pubblica a scopo redistributivo. Se infatti la maggioranza dei cittadini ha un reddito inferiore a quello medio (il reddito mediano è inferiore al reddito medio), una spesa eguale per tutti - sotto forma di consumi pubblici o di trasferimenti e finanziata da un'imposta strettamente proporzionale - porta a un vantaggio per una maggioranza e a uno svantaggio per una minoranza. Si tratta dell'utilizzazione del teorema dell'elettore mediano formulato da Howard Bowen (v., 1948), la cui radice si può rinvenire nell'osservazione di Alexis de Tocqueville per cui, confrontando un regime democratico e uno assolutista, si dovrebbe trovare nel primo una spesa pubblica più elevata. Sulla base di questa tesi dovremmo avere una spesa per servizi pubblici che vengono utilizzati effettivamente dalla generalità dei cittadini piuttosto che da gruppi più limitati, che versino in circostanze di particolare bisogno; un analogo discorso vale per la spesa di trasferimento.

Per altri autori il meccanismo dell'eccessiva espansione della spesa pubblica risulta dalla spinta di gruppi relativamente ristretti interessati a particolari provvedimenti di spesa, il cui onere viene ripartito su un numero di contribuenti molto più vasto. I rappresentanti politici (prevalentemente i partiti) operano la mediazione necessaria perché si formi la maggioranza sui provvedimenti a cui i singoli gruppi sono interessati, e cioè, in un certo senso, per unire i singoli gruppi. Anche se di volta in volta ci può essere chi ci guadagna e chi ci perde, in effetti il processo implica un certo grado di 'illusione finanziaria', in cui ciascun gruppo crede di scaricare sugli altri l'onere della spesa cui è particolarmente interessato. Poiché il risultato è un eccesso di spese e di imposte, è possibile ipotizzare un rovesciamento dello schema, che può vedere vincenti i contribuenti interessati a ridurre le imposte (v. Mueller, 1989).

In via generale le analisi di questo filone di studio si basano sull'estensione ai rappresentanti

politici - che compongono le assemblee legislative - e ai funzionari pubblici - che gestiscono i bilanci - delle ipotesi di comportamento tipico del consumatore, che massimizza l'utilità, o del produttore, che massimizza il profitto. Un accenno al tema della somiglianza tra gli imprenditori che massimizzano i profitti e i politici che massimizzano i voti era già stato formulato da Joseph Schumpeter (v., 1942), ma fu sviluppato compiutamente da Anthony Dawns (v., 1957). Secondo questo autore, "i partiti elaborano politiche per vincere le elezioni e non vincono le elezioni per formulare politiche"; esiste quindi una competizione in politica che presenta analogie con quella del mercato, e infatti Dawns si ispira al modello di concorrenza spaziale di Harold Hotelling (v. Concorrenza), descrivendo un processo di convergenza verso il centro dello schieramento politico di due partiti, uno progressista e uno conservatore. Analogamente William Niskanen (v., 1971) ipotizza che i funzionari pubblici non abbiano un comportamento linearmente ispirato a quelli che sono i fini generali dell'organizzazione di cui fanno parte, secondo la teoria weberiana, ma siano dei massimizzatori della propria funzione di utilità, nella quale entrano elementi più strettamente egoistici, quali, oltre ovviamente alla remunerazione, il numero dei sottoposti e l'ammontare delle risorse da gestire, elementi che condizionano potere e prestigio personali. La conseguenza è una tendenza dei funzionari a spingere l'ammontare dei servizi erogati, e quindi delle risorse utilizzate, al di là di quanto sarebbe desiderabile secondo un criterio di efficienza.

Non tutti gli autori che si ispirano a queste analisi economiche della politica e della burocrazia giungono al risultato di un eccesso di spesa pubblica, o più in genere di intervento pubblico; si può ricordare come lo stesso Dawns (v., 1960) ritenesse, al contrario, che il meccanismo della concorrenza elettorale porti a una dimensione troppo ridotta dei bilanci pubblici, giungendo quindi a una conclusione eguale a quella sostenuta poco tempo prima da John Galbraith (v., 1958). È stata in particolare la 'scuola delle scelte pubbliche' di James Buchanan a sottolineare che l'insieme dei decisori pubblici viene a costituire una sorta di moderno Leviatano, dotato di un potere monopolistico e mirante a massimizzare la quantità di risorse che vengono prelevate coattivamente ai cittadini: di qui l'indicazione di limitare con regole costituzionali il potere pubblico, attraverso vincoli di bilancio in pareggio e maggioranze qualificate.

Secondo questa scuola tale potere assolutistico sarebbe tanto maggiore quanto più accentrate sono le spese e le entrate pubbliche, ovvero quanto minore è l'attinenza tra le forme di

erogazione della spesa e quelle di prelievo delle entrate. Di conseguenza, sistemi ampiamente decentrati in materia di bilanci pubblici e un ricorso a tariffe, tasse e imposte di scopo piuttosto che a imposte di natura erariale dovrebbero mettere un freno al potere discrezionale degli organismi pubblici (v. Brennan e Buchanan, 1980).

7. Struttura ed evoluzione dei bilanci pubblici

a) Aspetti generali

Un'analisi della struttura della spesa pubblica e della sua evoluzione negli anni ottanta può consentire di introdurre qualche elemento statistico utile per orientarci rispetto alle diverse impostazioni analizzate nei capitoli precedenti. Nella tab. X viene presentata la ripartizione per funzioni della spesa dei sei paesi considerati, alla fine degli anni settanta e ottanta; vengono individuate quattro grandi categorie: beni pubblici, Welfare State, servizi economici e servizio del debito. I beni pubblici vengono divisi in servizi generali e spese per la difesa (o meno eufemisticamente spese militari), il Welfare State in tre beni di merito (istruzione, sanità, abitazione) e in spese di sicurezza sociale, distinte in trasferimenti alle famiglie (pensioni, indennità, sussidi) e altre spese di amministrazione e di trasferimento a istituzioni di beneficenza non aventi scopo di lucro.

Come si nota, la quota più consistente è costituita dalle spese per il Welfare State, divise in una parte leggermente più piccola di beni meritori e una più grande di sicurezza sociale. I sei paesi sono divisi in gruppi di tre: in testa Francia e Germania, poi Regno Unito e Italia, infine Giappone e Stati Uniti. Nel caso dei beni di merito la posizione statunitense si differenzia da quella degli altri paesi, che hanno quote piuttosto vicine tra loro, e ciò per la scarsa presenza pubblica nel settore della sanità; più disperse sono le quote di spesa per la sicurezza sociale. Si può anche notare come esista una relazione inversa tra spese per la difesa e servizi economici.Risulta abbastanza chiaro come il tradizionale motivo dei beni pubblici abbia un potere esplicativo piuttosto limitato, e come beni di merito e obiettivi redistributivi presentino una maggiore rilevanza; tuttavia, anche nel campo delle spese per il Welfare State non si può escludere che considerazioni di efficienza - mercati incompleti e asimmetrie informative - abbiano svolto un loro ruolo (v. Barr, 1992). Dato che, in generale, la spesa per la sicurezza sociale rappresenta una delle quote più rilevanti della spesa pubblica, è opportuno dare uno sguardo più particolareggiato alla sua composizione percentuale; ciò viene fatto nella tab. XI,

dove appaiono le quote della spesa per pensioni, malattia e invalidità, sostegno delle famiglie, disoccupazione e per altre forme di interventi assistenziali, tra cui quelli a favore della maternità.Il paese che si presenta come il più anomalo per composizione percentuale è l'Italia; la spiegazione è relativamente semplice, in quanto nella spesa per malattia e invalidità sono comprese le pensioni di invalidità, che da sole costituiscono il 26% della spesa per la sicurezza sociale. Ora, come è noto (v. Franco e Morcaldo, 1990), questa spesa ha una natura estremamente ibrida, in quanto, accanto alla funzione che formalmente dovrebbe assolvere, è servita anche per erogare pensioni di vecchiaia ad anziani che non avevano i requisiti previdenziali necessari, nonché sussidi nelle zone del Sud dove minori sono le possibilità di lavoro. Questo spiega perché la quota delle pensioni vere e proprie risulti più bassa relativamente ad altri paesi, così come i sussidi di disoccupazione. Italia a parte, si può segnalare il basso livello delle forme assistenziali negli Stati Uniti, e la quota relativamente elevata di sussidi a sostegno della famiglia dell'Inghilterra e del Giappone. Per quanto riguarda la struttura del prelievo fiscale, la tab. XII presenta la ripartizione percentuale alla metà degli anni ottanta: si può notare come, rispetto al quindicennio precedente (v. tab. VIII), la struttura del prelievo nei vari paesi sia diventata più simile; ad esempio, se consideriamo l'imposizione diretta, l'eccezione più rilevante è costituita dalla Francia, mentre l'Italia, che partiva da percentuali simili, si è portata a un livello lievemente superiore alla Germania. Si noti come esista un rapporto inverso tra contributi sociali e imposizione diretta.
Dall'insieme dei dati esaminati si possono tirare alcune conclusioni di massima. Il fenomeno principale che ha mosso l'intervento pubblico è stato la costruzione del Welfare State, secondo le linee essenziali tracciate dall'UK Social insurance and allied services del 1942, più noto come il Beveridge report; esso prevede la fornitura pubblica dell'istruzione e della sanità e un intervento nel campo dell'abitazione popolare, nonché un sistema pubblico di trasferimento delle risorse dalle generazioni attive a quelle a riposo, oltre ad altre forme di assistenza (disoccupazione, invalidità, famiglia). In questo ambito vi sono sostanziali somiglianze tra i paesi europei, mentre a un minor livello nel campo dell'assistenza si colloca il Giappone (dove in buona parte funzioni di questo genere vengono svolte a favore dei dipendenti dalle imprese); a un livello ancora più basso si collocano gli Stati Uniti in quanto, oltre alla spesa assistenziale, è nettamente più contenuta la spesa sanitaria. In effetti, la maggior quota di spesa pubblica statunitense rispetto al Giappone dipende dalla forte differenza nelle spese militari; al netto di queste, la quota di spesa pubblica giapponese sul

PIL eccede quella americana di circa tre punti.
Dall'altra parte, i due paesi che hanno una minore spesa per la sicurezza sociale, e quindi un minor effetto di redistribuzione del reddito, sono anche i due paesi che hanno un'imposizione diretta più elevata; si può pertanto dire che, in una certa misura, ottengono dal lato del prelievo quell'effetto redistributivo che non viene realizzato con la spesa.

b) Alcuni approfondimenti

Per approfondire i vari elementi che, nel processo di costruzione del Welfare State, hanno portato a un aumento della spesa pubblica sul PIL possiamo prendere in considerazione la spesa per l'istruzione; posto l'obiettivo di fornire i servizi di istruzione alla popolazione, i fattori che determinano la crescita di questa spesa sono dati dalla quota della popolazione giovanile sul totale, dalla progressiva scolarizzazione di tale settore della popolazione, dai maggiori inputs di lavoro (e secondariamente di strumenti) messi a disposizione, nonché dalla crescita più sostenuta delle remunerazioni degli insegnanti (e altri addetti) e dei prezzi degli strumenti rispetto al deflattore del PIL. La tab. XIII mostra la scomposizione percentuale dei quattro fattori che hanno determinato i tassi di crescita negli anni sessanta e settanta; come si può vedere, il fattore principale è stato l'aumento reale degli inputs. Il 'morbo di Baumol' (il quarto fattore) ha quindi avuto un ruolo relativamente minore.Un recente lavoro di Daniele Franco (v., 1992) permette di approfondire questa analisi per l'Italia; la tab. XIV mostra la variazione sul PIL della quota di spesa del Ministero della Pubblica Istruzione, scomposta nei vari rapporti riportati, da cui si ricava che i fattori che hanno favorito la crescita della percentuale di spesa per l'istruzione sono l'aumento della scolarizzazione e del rapporto tra personale scolastico e statali (impressionante la crescita del personale non docente). Il 'morbo di Baumol' risulta praticamente assente, anzi il rapporto tra retribuzione unitaria del personale del Ministero della Pubblica Istruzione e reddito nazionale pro capite, da solo, avrebbe fatto diminuire la quota della spesa per l'istruzione sul PIL.
La principale voce di spesa del settore della sicurezza sociale è quella pensionistica; trattandosi di una spesa per trasferimento, i fenomeni dati dall'interazione tra inputs reali e loro prezzo, tipica della spesa per servizi, non si verificano. I fattori che determinano variazioni della quota di spesa pensionistica sul PIL sono: il fattore demografico, ovvero la quota di anziani sulla popolazione; il grado di copertura, cioè il rapporto tra pensionati e anziani; il rapporto tra pensione media reale e reddito medio pro capite reale, nonché il

rapporto relativo tra prezzi al consumo e deflattore del PIL. La tab. XV riporta il peso di questi quattro fattori nel determinare la crescita della quota della spesa pensionistica sul PIL. Come si può notare, le differenze tra paesi e tra i due sottoperiodi sono rilevanti; in generale si può dire che il fattore demografico è stato importante in particolare in Germania e nel Regno Unito, mentre dei notevoli mutamenti nelle politiche pensionistiche sono avvenuti in Giappone e negli Stati Uniti per quanto riguarda il rapporto reale tra pensione e reddito pro capite. Nel caso del Giappone ciò deriva anche da un rallentamento dei tassi di crescita del reddito, ma ancor di più dalla scelta di rivalutare i trattamenti pensionistici. Per l'Italia si può segnalare la caduta del grado di copertura (pensioni su popolazione di età superiore a 65 anni) nel secondo sottoperiodo, determinato in particolare dalle pensioni di assistenza e di invalidità; per le pensioni previdenziali di vecchiaia il grado di copertura (che riguarda la popolazione di età superiore a 60 anni) è leggermente aumentato (v. Franco, 1992).

8. La politica di bilancio degli anni ottanta

Passiamo ora a esaminare le linee generali della politica di bilancio perseguita dai sei paesi negli anni ottanta; come si può notare nella tab. XVI, ogni paese ha seguito differenti linee di condotta: la Germania, il Regno Unito e il Giappone hanno diminuito, in diversa misura, la spesa pubblica al netto degli interessi sul debito; il Regno Unito è riuscito anche a ridurre il livello del debito in modo tale da far diminuire la spesa per interessi (grazie anche alla politica di privatizzazioni effettuata), mentre il Giappone ha effettuato un notevole incremento della pressione fiscale. La Francia e l'Italia hanno aumentato invece la spesa pubblica al netto degli interessi e allo stesso tempo hanno aumentato la pressione fiscale, ma mentre in Francia ciò ha permesso di contenere la spesa per interessi, dato il basso livello di partenza del debito pubblico, la stessa cosa non si è verificata per l'Italia, che ha visto un notevole aumento della spesa per interessi.
Come conseguenza, il rapporto tra debito pubblico e PIL è aumentato in tutti i paesi, con l'eccezione del Regno Unito, ma il fenomeno presenta particolare gravità per l'Italia. Conviene esaminare più in dettaglio il problema posto dal debito pubblico, in un periodo in cui i tassi d'interesse reali sono stati portati da una serie di fattori, tra cui le politiche monetarie degli Stati Uniti prima e della Germania poi, a livelli estremamente elevati, che non hanno praticamente confronto con il passato. Nelle tabb. XVIIa, XVIIb, XVIIc e XVIId

sono riportati l'indebitamento al netto della spesa per interessi, i tassi reali d'interesse a breve e lunga, nonché il tasso reale di crescita dell'economia. Come si può notare, l'indebitamento è sceso gradatamente fino ad azzerarsi, sicché il deficit è venuto a identificarsi sempre di più con la spesa per interessi; il fatto che i tassi d'interesse superino quelli di crescita segnala una situazione di grave difficoltà a contenere la crescita del debito. Né si può affermare con sicurezza che gli alti livelli dei tassi d'interesse dipendano dal livello del debito, dato che, se è vero che i tassi italiani sono stati, guardando agli anni più recenti, maggiori di quelli statunitensi e giapponesi, sono però del tutto paragonabili a quelli degli altri paesi europei, paesi dove il debito pubblico è a livelli molto più bassi. Paradossalmente il fenomeno che si presenta è il seguente: l'Italia, dopo aver seguito nei decenni precedenti una politica di bilancio fondamentalmente espansiva per ragioni cui si accennerà più avanti, è giunta in una situazione nella quale, se si prescinde dal passato, e quindi dal debito pregresso, si ha un equilibrio tra spese ed entrate. È tuttavia il macigno del debito che, unito a un livello particolarmente elevato dei tassi d'interesse, determina uno squilibrio dei conti pubblici e alimenta la propria crescita, non solo in assoluto, ma, cosa più grave, rispetto al PIL. La crisi del cambio italiano del settembre 1992 trova in questa situazione una delle sue motivazioni. Se le politiche che hanno portato a questi tassi reali così elevati non cambieranno, le prospettive per l'Italia vanno da una politica di bilancio estremamente restrittiva per alcuni anni (con eventuale ricorso a misure straordinarie) a una crisi del debito con processi di cancellazione attraverso misure quali il consolidamento o la gestione forzosa.

Se osserviamo la tab. XVIII possiamo notare che l'aumento della spesa pubblica al netto degli interessi è stato di 3,6 punti percentuali dal 1979 al 1991; se questo aumento non vi fosse stato, e se le entrate totali avessero avuto invece l'andamento storico, il debito pubblico sarebbe incominciato a calare dal 1986, e nel 1991 sarebbe risultato inferiore di circa 15 punti percentuali. In verità questa valutazione presuppone che il reddito sarebbe cresciuto allo stesso tasso, il che costituisce un'ipotesi criticabile, dato che la minore spesa avrebbe determinato una riduzione della domanda aggregata. Tuttavia si può pensare che i tassi d'interesse avrebbero potuto essere in qualche misura inferiori, e questo avrebbe avuto invece degli effetti positivi sulla crescita del reddito e sulla spesa per interessi. Per quanto valutazioni del genere siano difficili e largamente opinabili, è probabile che il costo sociale di una politica più restrittiva del bilancio sarebbe risultato minore di quanto non saranno le misure che attendono in futuro i cittadini italiani.

Un'altra osservazione è legata all'individuazione delle componenti che hanno determinato la crescita della quota della spesa al netto degli interessi: la componente che è cresciuta di più è quella delle prestazioni sociali (2,6 punti), seguita dalle retribuzioni (0,8) e dai consumi intermedi (0,4); le altre voci (investimenti, trasferimenti correnti alle imprese) sono rimaste nell'insieme invariate. In effetti, però, l'aumento della quota della spesa per retribuzioni è dovuto integralmente alla crescita dei contributi sociali; al netto di questa componente le retribuzioni sono rimaste sostanzialmente invariate intorno al 9%; d'altra parte, pur essendo aumentata la retribuzione degli addetti alle amministrazioni pubbliche rispetto ai dipendenti che producono beni e servizi destinati alla vendita (il rapporto si è incrementato di un 16,7% nei tredici anni considerati, recuperando in parte il terreno perduto negli anni precedenti), tale retribuzione è cresciuta in misura leggermente inferiore al PIL, per cui l'invarianza della quota complessiva è dipesa dalla crescita del numero dei dipendenti pubblici (a un tasso medio dell'1,8%, cifra del tutto simile a quella registrata negli anni ottanta dagli altri paesi). In sostanza, a parte la questione della spesa per interessi, il problema si pone soprattutto sul lato della spesa per la sicurezza sociale, e in particolare per la spesa pensionistica (e ancora più specificamente per le pensioni di vecchiaia e superstiti). Come si è potuto notare, la crescita in Italia è stata spinta in particolare dal forte aumento del rapporto tra trattamento pensionistico medio e retribuzione media, anche se in futuro il fattore demografico inciderà più sensibilmente.

9. Alcune considerazioni generali

La crescita del ruolo pubblico nell'economia dei paesi a capitalismo sviluppato è derivata essenzialmente dal perseguimento di alcuni obiettivi che possono sintetizzarsi in due aspetti: erogazione di alcuni servizi che si ritiene debbano essere forniti a tutti i cittadini, almeno in una data misura, e creazione di una struttura di trasferimenti volta alla riduzione dei rischi sociali nonché all'attuazione di una redistribuzione del reddito sia intergenerazionale che intragenerazionale. Questo obiettivo è stato perseguito con modalità diverse e in grado diverso dai vari paesi, ma comunque in modo rilevante: certamente sono stati ottenuti dei risultati significativi, ma allo stesso tempo ci si è trovati di fronte alla forte caduta dei tassi di crescita del reddito negli anni settanta e ottanta, nonché alla altrettanto forte crescita dei tassi d'interesse negli anni ottanta.

Questo è stato il fattore principale che ha causato l'aumento della pressione fiscale, ma allo stesso tempo l'aumento del debito pubblico. In queste condizioni i paesi hanno dato luogo con diversa intensità a politiche di bilancio restrittive, con una rilevante eccezione: l'Italia. Nei primi anni settanta il notevole aumento del debito è stato causato da un aumento della spesa cui non ha fatto seguito un analogo aumento delle entrate, mentre in seguito le maggiori entrate sono servite a finanziare una serie di interventi a sostegno dell'economia, tra cui la fiscalizzazione degli oneri sociali; l'aumento del debito è stato inoltre causato da finanziamenti concessi alle imprese pubbliche. Negli anni ottanta, come si è visto, l'aumento deciso della pressione fiscale ha più che compensato l'aumento delle spese per la sicurezza sociale, ma non è riuscito a compensare il forte aumento della spesa per interessi; si tenga presente che i finanziamenti alle imprese pubbliche sono continuati per buona parte degli anni ottanta, e che nel complesso essi rappresentano circa il 20% del debito pubblico.
Il problema della pressione fiscale è reso più difficile dal permanere di un elevato tasso d'evasione nel settore delle piccole imprese e del lavoro autonomo; il fenomeno è comune anche agli altri paesi, ma risulta più serio in Italia, sia per una maggiore importanza dei soggetti in questione (circa doppia rispetto agli altri paesi europei), sia per una minore efficienza dell'amministrazione fiscale. Vi è inoltre da aggiungere un elemento di cui è difficile valutare l'importanza, ma che è sicuramente presente: un aspetto della tax revolt deriva da una valutazione negativa che viene data sia del livello dei servizi pubblici, sia dei flussi della spesa per trasferimento. Un esempio tipico è costituito dalle pensioni di invalidità, che, come si è visto, costituiscono una caratteristica tipicamente italiana, di cui vengono lamentati la natura discrezionale e l'uso clientelare nelle regioni del Sud.
Il problema della qualità dei servizi pubblici e della spesa assistenziale non viene sollevato solo in Italia; le due critiche fondamentali sono da un lato (servizi) l'assenza di un meccanismo di concorrenza che assicuri l'efficienza delle prestazioni, dall'altro (trasferimenti) gli effetti negativi sull'offerta di lavoro e di risparmio. Prima di passare a qualche cenno sulle proposte di modifica delle modalità di intervento dello Stato è opportuno riflettere sul fatto che il problema dell'efficienza e dell'efficacia dei servizi pubblici è più complesso di quanto non appaia dalla semplice formula dell'assenza di concorrenza; è un luogo comune notare la differenza tra i servizi italiani e quelli di altri paesi, ma quello cui si presta minor attenzione è il fatto che essi sono notevolmente differenziati all'interno dell'Italia stessa. Se si considerano i servizi sanitari, si può dire che essi migliorano, anche

sensibilmente, passando dal Sud verso il Nord e dalle grandi città verso le piccole; per l'istruzione, invece, il rapporto tra città grandi e piccole si pone in modo diverso. Un limite di molte analisi sull'argomento è che esse pretendono di individuare un unico meccanismo di causazione per spiegare l'inefficienza pubblica, quando il problema è piuttosto quello di spiegare l'enorme differenza di efficienza tra una scuola e un'altra, tra una struttura sanitaria e un'altra. Osservazioni analoghe possono essere formulate verso la scuola delle scelte pubbliche: lo Stato Leviatano ha articolato le sue strutture in modo notevolmente diverso da paese a paese, e ha reagito in modo diverso ai grandi shocks che hanno investito l'economia mondiale. Il meccanismo della democrazia e dei partiti politici ha operato in modo diverso; da alcuni studi (v. Wadhwani e Kong, 1992) risulta ad esempio che la scelta tra sistemi di rappresentanza proporzionale e sistemi maggioritari ha scarso riflesso sulla performance economica dei paesi, fatta eccezione per il livello del debito pubblico. Ciò potrebbe interpretarsi nel senso di una maggiore difficoltà per i paesi a rappresentanza proporzionale ad adottare decisioni sufficientemente restrittive delle politiche di bilancio. D'altra parte, da un altro studio (v. Grilli e altri, 1991) risulta una relazione tra grado di indipendenza riconosciuta istituzionalmente alle banche centrali e successo nel contenimento del tasso d'inflazione; non vi sarebbero invece relazioni né col sistema di rappresentanza politica né con la performance reale dell'economia.
Ancora, il ruolo della burocrazia pubblica come causa di inefficienza potrebbe essere visto come un problema di scarsa autonomia dei funzionari pubblici dai partiti: sarebbe quindi la mancanza di una sufficiente indipendenza dei funzionari rispetto ai governi la causa del problema, piuttosto che il contrario. Un confronto tra la situazione italiana e quella francese, ad esempio, farebbe propendere verso questo tipo di interpretazione.

10. La riforma dell'intervento pubblico

a) La fornitura dei servizi

Sull'opportunità di ridurre il peso dell'intervento pubblico e di mutare le modalità dell'intervento stesso esiste da tempo una notevole attenzione delle forze politiche e sociali, nonché di studiosi appartenenti a diverse discipline sociali. In questo capitolo intendiamo indicare nelle linee generali le proposte di riforma, rinviando al capitolo conclusivo alcune considerazioni finali.

Possiamo iniziare con le proposte di riforma dell'intervento per quanto riguarda l'erogazione dei servizi pubblici. Non verranno prese in considerazione le posizioni estreme, che negano l'opportunità stessa di un interesse dello Stato per quanto riguarda alcuni servizi quali l'istruzione o la sanità, ritenendo che le scelte in materia debbano essere lasciate alla sfera decisionale degli individui, senza interferenza pubblica (v. Friedman e Friedman, 1980). La prima proposta che consideriamo, che si pone agli antipodi della gestione pubblica e centralizzata del servizio, è quella dei buoni a destinazione vincolata: lo Stato, cioè, si limita a erogare a ciascun cittadino un buono di cui egli potrà servirsi sul libero mercato per servizi educativi, sanitari o di altra natura. In tale modo si otterrebbe il risultato di garantire un certo livello di copertura dei bisogni meritevoli di tutela pubblica, lasciando però al mercato tutte le scelte per quanto riguarda le caratteristiche delle prestazioni, in modo che il processo concorrenziale possa esplicare in pieno i propri effetti in termini di efficienza (v. IEA, 1967). Le proposte in materia si articolano in modo diverso a seconda del tipo di servizio specifico; in molti casi i buoni sono personalizzati, e quindi non vendibili sul mercato, mentre in altri casi possono non avere questa caratteristica; nel caso siano personalizzati, si tratta di decidere se siano o meno integrabili monetariamente, in quanto ciò può avere una notevole importanza sulla domanda e quindi sul tipo di prestazioni offerte (v. Atkinson, 1987). In tutti i casi il criterio è quello di utilizzare al massimo il mercato privato intervenendo attraverso un sistema di sussidi, e ciò al duplice scopo di garantire un ammontare di consumi più elevato di quanto accadrebbe in assenza del sussidio, e in particolare un adeguato livello individuale.In una diversa ipotesi, la produzione del servizio viene effettuata dal settore privato, ma una serie di caratteristiche circa il tipo di servizio da offrire vengono decise dall'operatore pubblico, il quale sceglie il produttore privato sulla base di una gara d'asta. La prima formulazione di questo criterio si deve a Edwin Chadwick e risale a oltre un secolo fa (v. Chadwick, 1859); il tema è stato trattato anche da autori italiani come G. Montemartini (v., 1902), e più di recente da un gruppo di giovani economisti della "Rivista trimestrale" (v. Boitani e altri, 1980). Il consumo del servizio verrebbe deciso collettivamente tramite scelte pubbliche, ma erogato privatamente; l'efficienza sarebbe ottenuta da una concorrenza per il settore, non da una concorrenza nel settore, cosa questa che permetterebbe di sfruttare adeguatamente le economie di scala.In un'ultima ipotesi, l'erogazione del servizio verrebbe ricondotta al settore pubblico, ma con un'articolazione e un'autonomia maggiori di quanto non avvenga con sistemi centralizzati che escludono quasi totalmente la scelta dei cittadini (v.

Artoni, 1992). Gli erogatori dei servizi dovrebbero quindi avere maggiore flessibilità e autonomia di quanto non accada nel caso di organi di una struttura statale, il che implicherebbe anche un'autonomia finanziaria, con ampi margini di fissazione di prezzi e tariffe. Si pone quindi un problema di sussidi agli utenti, in analogia ai criteri ispiratori dell'ipotesi dei buoni.

b) L'intervento assistenziale

Come si è potuto notare nel cap. 6, il peso della spesa assistenziale varia notevolmente tra i diversi paesi, e all'interno di essa variano i destinatari e il sistema dei trasferimenti. Questa componente di spesa ha sempre suscitato critiche e riserve, si può dire fin dai tempi delle Poor laws. Le misure assistenziali sono considerate inutili e controproducenti: inutili perché, come diceva una canzoncina popolare inglese dell'Ottocento, se si invertissero le parti tra un povero e un ricco, entro poco tempo essi tornerebbero al proprio posto (dentro un comodo letto e sotto uno scomodo ponte); controproducenti in quanto l'assistenza fiacca la spinta a lavorare e a intraprendere, generando un continuo aumento di persone che finiscono col vivere di assistenzialismo (con le ovvie conseguenze sul bilancio pubblico).

L'ipotesi opposta, espressa con la massima chiarezza da William Beveridge, prevede, al contrario, che adeguati livelli di assistenza, cui partecipino i soggetti stessi che ne potrebbero beneficiare, possano determinare quella freedom from want che rappresenta un beneficio effettivo per la società: se il rischio viene ridotto, gli individui sono disposti a rischiare di più (v. anche Meade, 1989). Una prima ipotesi di riforma dei sistemi assistenziali propone tuttavia di eliminare le diverse forme di assistenza (agli invalidi, ai disoccupati, alle famiglie numerose) a favore di un sistema unico e generale. Una proposta in questo senso venne avanzata, nello stesso anno in cui uscì il Beveridge report, da parte di lady Juliet Rhys Williams (v., 1943) e prevedeva l'erogazione generalizzata di un sussidio, chiamato 'dividendo sociale', a tutti coloro che ne facessero domanda, indipendentemente dal reddito e dalla condizione lavorativa. L'unica condizione per ottenere il dividendo sociale era - da parte dei disoccupati - la disponibilità ad accettare un lavoro che venisse offerto da un apposito ufficio di collocamento; la clausola rivelava un tipo di visione comune anche a Beveridge, secondo cui l'individuo doveva meritarsi ciò che la società faceva per lui. La spesa per il dividendo sociale, che avrebbe assorbito qualunque altra forma assistenziale, implicava uno sforzo finanziario non piccolo, e richiedeva una imposta generale su tutti i redditi (a

eccezione del dividendo sociale) piuttosto elevata (la Rhys Williams la poneva al 40%). In effetti quello che si viene a realizzare è un ampio prelievo e un'altrettanto ampia distribuzione del reddito; la posizione di ciascun membro della collettività dipende dal rapporto fra il suo reddito e il reddito medio. Se il suo reddito coincide con quest'ultimo, il suo reddito disponibile non muta: l'imposta sul reddito gli viene restituita sotto forma di dividendo sociale. Tanto minore è il reddito personale rispetto a quello medio, tanto maggiore è il vantaggio, e viceversa.

In base a una seconda ipotesi il trasferimento dovrebbe essere limitato a integrare il reddito di coloro che si trovano nelle fasce più basse delle classi di reddito; in questo caso il trasferimento diviene condizionato e lo scopo è quello di portare (o comunque avvicinare) a un certo livello di reddito coloro che si trovano al di sotto. Nella forma più pura tutta la differenza di reddito viene integrata dal sussidio; poiché ci si rende conto che in tal modo vi sarebbe un forte disincentivo al lavoro, l'integrazione della parte mancante del reddito viene limitata a una quota percentuale. Nella versione di Milton Friedman (v., 1962) questo sistema di trasferimenti si baserebbe sulla stessa income tax, che, oltre a prevedere un'imposta per i redditi superiori a una certa cifra, prevedrebbe un sussidio per chi si trovasse al di sotto (imposta negativa sul reddito), ma sono possibili numerose varianti di questo schema (v. Atkinson, 1987).

Una posizione originale sulla politica assistenziale, che anticipava i merit wants di Musgrave e l'egualitarismo specifico di Tobin, è quella di Ernesto Rossi; enunciata nell'immediato dopoguerra (v. Rossi, 1946), è stata ripresa dallo stesso autore (v. Rossi, 1956) e da altri (v. Fuà, 1957). Secondo questa impostazione la redistribuzione dovrebbe avvenire attraverso l'erogazione di beni - "l'alloggio, il vestiario, il vitto, l'istruzione e gli altri beni che si ritengono necessari per assicurare un minimo di vita civile" - attraverso una produzione standardizzata. La distribuzione in natura avrebbe il vantaggio di evitare che i bisognosi sperperino il denaro dei sussidi in consumi voluttuari, e assicurerebbe più facilmente che la distribuzione delle risorse vada effettivamente a coloro che ne hanno bisogno: "Le persone facoltose, infatti, difficilmente si adattano ad abitare in una casa popolare, anche se possono ottenerla gratuitamente, a prendere i loro pasti in una mensa a menu fisso, [...] a indossare vestiti tipici". Come si vede, si tratta di una posizione che è agli antipodi di quella espressa da Friedman, e che è stata suffragata da nuovi argomenti a supporto negli sviluppi moderni dell'economia dell'informazione asimmetrica (v. Petretto, 1988).

c) L'intervento previdenziale

Come è noto, le teorie del ciclo vitale individuano una natura previdenziale nella formazione del risparmio familiare, che costituisce la parte principale del risparmio aggregato (v. Castellino e Fornero, 1990). In tutti i paesi lo Stato è intervenuto attraverso una serie di norme che fissano degli obblighi circa il versamento di una parte del reddito ottenuto durante la vita attiva degli individui, e dei diritti a ottenere delle pensioni durante il periodo di ritiro. Le normative in materia sono notevolmente complesse, ma si possono individuare due tipi di schemi previdenziali: quello del sistema a ripartizione e quello del sistema a capitalizzazione. Nel sistema a ripartizione la generazione attiva subisce un prelievo dal proprio reddito che viene distribuito alla generazione a riposo; ciò implica quello che è stato definito (v. Samuelson, 1958) un contratto tra le generazioni, in cui ciascuna paga le pensioni alla precedente e si aspetta che lo stesso avvenga al momento in cui si ritirerà dal lavoro; la prima generazione che è andata a riposo, senza aver versato contributi, riceve quindi un beneficio netto da parte della prima generazione attiva che versa i contributi. Nel sistema a capitalizzazione, invece, ogni generazione accumula delle quote del proprio reddito che vengono investite, per ottenere un flusso di reddito, nel periodo del ritiro, che dipenderà dal rendimento degli investimenti effettuati. Inizialmente, quindi, si viene a costituire un fondo che crescerà rapidamente in una prima fase, fino al momento in cui i flussi in entrata da parte delle generazioni in attività verranno bilanciati da quelli in uscita delle generazioni che vanno a riposo.

Osservando i due sistemi a regime, si può osservare che se l'economia di un paese procede a un tasso di crescita costante, e se il tasso di interesse (reale) è eguale al tasso di crescita, i sistemi risultano equivalenti (v. Aaron, 1966); se il tasso di crescita è maggiore (minore) del tasso d'interesse, il sistema a ripartizione (capitalizzazione) è preferibile. Tuttavia questo non è che uno degli aspetti preliminari del problema, che difficilmente può risolvere la scelta tra i due criteri, visto che i rapporti tra i due tassi sono variati (negli anni sessanta era maggiore il tasso di crescita, negli anni ottanta quello d'interesse) e possono variare in futuro.

A favore dei regimi a ripartizione si può infatti osservare che essi rendono esplicito il fatto che le pensioni si pagano sempre col reddito corrente del paese, anche se formalmente il sistema è a capitalizzazione (v. Barr, 1987); essi permettono una notevole riduzione del rischio cui vanno incontro i regimi a capitalizzazione, dato che la variabilità dei tassi effettivi

di rendimento degli investimenti finanziari è certamente maggiore della variabilità del tasso di crescita (v. Munnel, 1987). A favore dei regimi a capitalizzazione, invece, si argomenta che essi evitano la riduzione della propensione al risparmio propria del sistema a ripartizione (v. Feldstein, 1978) e rappresentano un forte stimolo allo sviluppo dei mercati dei capitali, creando fondi che rendono possibile effettuare investimenti a lungo termine (v. Friedman, 1980).
La discussione che si svolge su questi temi è molto articolata; il regime a ripartizione entra in difficoltà quando i fattori demografici fanno crescere il rapporto tra generazioni a riposo e generazioni attive, e quando fenomeni economici fanno diminuire la quota del reddito da lavoro dipendente. Sarebbe necessario quindi che il sistema presentasse una certa flessibilità per potersi adeguare a tali mutamenti; ma a ciò è di ostacolo la certezza dei diritti che accompagna tradizionalmente il sistema e che ne costituisce uno dei vantaggi. D'altra parte non è chiaro se il sistema a ripartizione determini una caduta della propensione al risparmio (v. Modigliani, 1986), mentre il sistema a capitalizzazione presenta costi di gestione più elevati; tuttavia l'inefficienza dell'operatore pubblico può far venire meno quelle economie di gestione che in teoria sarebbero proprie del sistema a ripartizione.

11. Conclusioni

Oltre cento anni fa l'economista francese H. Leroy Beaulieu (cit. in Brosio, 1986) sosteneva che il limite all'aumento della quota della spesa pubblica sul reddito nazionale era il 15%; cinquant'anni più tardi Colin Clark (v., 1945) poneva questo limite al 25%. Più recentemente Milton Friedman (v., 1976) poneva il limite al 60%. Mentre l'economista francese temeva una fuga dei capitali e degli stessi cittadini e quello australiano l'inflazione, l'economista americano vede in questo limite un problema di difesa della libertà individuale; tuttavia il limite in cento anni si è quadruplicato e probabilmente Friedman sottoscriverebbe senza esitazioni una quota che a Clark sembrava foriera di disastri economici.
In base agli stessi dati che abbiamo esaminato, tuttavia, si può ritenere che il processo di crescita della spesa pubblica nei grandi paesi a capitalismo sviluppato sia arrivato a livelli difficilmente superabili, date le attuali tendenze della crescita economica. Non si può escludere che una ripresa del tasso di crescita delle economie sviluppate possa portare a qualche aumento della spesa pubblica e in genere dell'intervento statale, ma questa ipotesi si

basa su scenari che al momento sembrano lontani e incerti.
È dunque probabile che nel prossimo futuro assisteremo a politiche di bilancio restrittive, e ciò tanto più in quanto i tassi d'interesse si manterranno ai livelli elevati che hanno raggiunto nell'ultimo decennio. La fase che attraverseremo sarà quindi di consolidamento o di ripiegamento nei livelli raggiunti già dalla spesa pubblica, e in questo quadro dovremo collocare i necessari tentativi di aumento dell'efficienza e dell'efficacia (v. Buglione e France, 1990), e il perseguimento di obiettivi redistributivi ed equitativi.
Tentare di migliorare qualitativamente la spesa pubblica in una fase di stretta quantitativa può sembrare impresa impossibile; in realtà nel campo dell'erogazione dei servizi i due obiettivi sono potenzialmente compatibili; del resto, la rapida espansione dei servizi pubblici è stata considerata come causa del loro peggioramento qualitativo (v. Hirschman, 1980), ed è probabile che l'esperienza italiana sia una buona prova di questa ipotesi.
Sembra che l'erogazione dei servizi pubblici debba prevedere una diversità di formule e comunque una flessibilità nella gestione maggiore che in precedenza. Difficilmente, tuttavia, il metodo dei buoni potrà avere ampio spazio e dare dei risultati apprezzabili; in settori come l'istruzione e la sanità i possibili fallimenti del mercato sembrano consistenti e tali da sconsigliare modifiche così radicali.
Per quanto riguarda le forme assistenziali, è difficile immaginare che le linee di riforma vadano nella direzione di un trasferimento generalizzato quale il dividendo sociale; anche se l'applicazione dello schema dell'elettore mediano sembrerebbe suggerire una potenzialità del genere, è probabile che le diverse forme di assistenza debbano essere più differenziate per poter essere efficaci (v. Vianello, 1989). Tuttavia il problema delle forme assistenziali è che esse possono divenire oggetto di forte ostilità nel momento in cui da trasferimenti che si rivolgono a bisognosi non identificabili a priori, si passa a trasferimenti a bisognosi identificati (immigrati, residenti in una particolare area, ecc.).
Per quanto riguarda la previdenza, è probabile che nei paesi, come l'Italia, dove il peso della previdenza pubblica a ripartizione è elevato si debba far posto a un maggior spazio per la previdenza privata a capitalizzazione; tuttavia sembra plausibile ritenere che fino a un certo livello di reddito vi debba essere una garanzia del sistema pubblico, che elimini gli elementi di rischio propri dei sistemi privatistici; quale debba essere questo livello (il reddito medio, ad esempio) è cosa che potrà dipendere da circostanze diverse nei diversi paesi (v. Pizzuti, 1990).

Per quanto riguarda l'Italia, in particolare, le linee di riforma dovrebbero consistere nella eliminazione di alcune anomalie su cui ci siamo già soffermati e che determinano delle distorsioni nel nostro sistema di sicurezza sociale: certi aspetti delle pensioni di invalidità, della cassa integrazione, dei trattamenti pensionistici privilegiati. In generale il sistema previdenziale dovrebbe acquistare una maggiore flessibilità, anche se ciò comporta difficili scelte rispetto al grado di garanzia dei diritti che un sistema a ripartizione deve necessariamente presentare. Uno sviluppo dei fondi pensione dovrà necessariamente investire il sistema di trattamento di fine rapporto. Si pone inoltre il problema di una sostituzione dei contributi che gravano sulla massa salariata con imposte che gravino più in generale sul valore aggiunto del paese.

Nel campo dell'imposizione gli sforzi profusi da molti economisti accademici, per uno spostamento dell'imposizione diretta dal reddito al consumo, non sembrano aver dato risultati di rilievo; la tendenza attuale è semmai verso un allargamento della base imponibile, con l'eliminazione o la riduzione di una serie di esenzioni e deduzioni, e una riduzione delle aliquote marginali. Peraltro difficilmente i contributi obbligatori potranno essere inseriti nel reddito imponibile, mentre problemi più difficili si pongono per le agevolazioni fiscali di cui godono i sistemi pensionistici privati.

In Italia due sembrano le direzioni in cui il sistema del prelievo tenderà a muoversi (v. Ceriani e altri, 1992): la riduzione dell'area dell'evasione ed elusione fiscale e un maggior grado di autonomia impositiva degli enti subalterni. Per quanto riguarda il problema dell'evasione, ciò comporta l'allontanamento dal concetto di reddito effettivo proprio della riforma tributaria del 1973, in riferimento all'area del reddito d'impresa minore e dei lavoratori autonomi; per quanto concerne la maggiore autonomia, essa riguarderà sicuramente la proprietà immobiliare, e si estenderà probabilmente anche alle imposte dirette, piuttosto che alle indirette.

Passando invece al problema delle imprese pubbliche, l'attuale tendenza alla privatizzazione probabilmente continuerà, ma con un'intensità ridotta, in particolare nel Regno Unito, dove il processo è stato particolarmente intenso. Esso investirà certamente anche l'Italia, in quanto, accanto a considerazioni di efficienza produttiva, vengono posti obiettivi di riduzione del debito pubblico, nonché dell'influenza dei partiti politici. Questi due obiettivi, in realtà, dovrebbero avere un ruolo del tutto marginale sui problemi delle privatizzazioni: la vendita di imprese pubbliche non può essere la soluzione del problema del debito pubblico italiano, così

come non lo è stata per il Regno Unito, che ha fatto ricorso, a questo scopo, a politiche di bilancio fortemente restrittive. Se considerazioni di efficienza produttiva consigliano il ritorno a mani private nella gestione di un'impresa, la privatizzazione va perseguita per questo specifico obiettivo, indipendentemente dal fatto che il Tesoro ne ricaverà un beneficio. Quanto all'idea di diminuire l'eccesso di potere dei partiti, riducendo il numero delle imprese dirette da imprenditori di nomina pubblica, essa costituisce una illusione, dato che le possibilità di influenza del settore pubblico sulle imprese private sono molto articolate. Anche in questo caso, quello che conta è la creazione di regole del gioco trasparenti, che diano l'autonomia necessaria ai managers e che li chiamino a rispondere del loro operato. Questa considerazione ci porta a un ultimo problema, che in Italia costituisce uno dei temi più rilevanti. La funzionalità del settore pubblico non potrà migliorare se non quando i funzionari, i professori, i medici avranno la percezione che lavoro, remunerazione e in generale carriera dipendono dall'impegno e dalle capacità personali, e non da altri fattori. Questo implica certamente la necessità di tagliare quei legami impropri che si sono costituiti nel tempo con i partiti politici e i sindacati, e che si sono sovrapposti e sostituiti a regole costituzionali molto chiare, a cui sarà necessario tornare. Sarà opportuno considerare anche una serie di problemi che riguardano i rapporti tra la macchina amministrativa pubblica, il governo e il parlamento, e rivedere anche il sistema dei controlli, ma il problema principale è quello di ricostituire una dirigenza della grande macchina statale che possa confrontarsi con quelle degli altri paesi europei: se ciò non avverrà, il settore pubblico continuerà a costituire un freno allo sviluppo economico e civile del paese.

Economia e Politica del lavoro
di Carlo Dell'Aringa

sommario: 1. Introduzione. 2. Il contributo degli economisti classici. 3. I marginalisti e la teoria della domanda di lavoro. 4. Gli istituzionalisti e il ruolo del sindacato. 5. I primi tentativi di sintesi neoclassica. 6. Il dibattito antimarginalista nel secondo dopoguerra. 7. La moderna economia del lavoro. 8. Una nuova sintesi? □ Bibliografia.

1. Introduzione

L'economia del lavoro è uno dei campi specialistici in cui si divide l'economia politica. Due filoni di pensiero si contendono, da sempre, il predominio in questo campo: quello di origine neoclassica, che oggi è decisamente dominante, e quello di origine istituzionale, che era dominante alcuni decenni or sono e che ancora oggi comunque riscuote, da parte degli specialisti, notevoli simpatie. La controversia tra le due scuole e la differenza fra i due approcci non sta tanto o solo nei contenuti e nell'oggetto di studio, quanto nella metodologia utilizzata. Ricordare come questa diversità di approccio è nata e si è evoluta nel tempo è fondamentale per capire come vengono affrontati oggi i problemi del lavoro e quali sono le politiche del lavoro che vengono suggerite da economisti che appartengono a diverse scuole di pensiero.

2. Il contributo degli economisti classici

Gli economisti classici studiavano soprattutto la crescita del reddito nazionale e la sua distribuzione fra le diverse classi sociali: i proprietari di terra, i capitalisti e i lavoratori. L'approccio ai problemi economici era di tipo aggregato. Vi fu un intenso dibattito sulle determinanti del livello generale dei salari e sulla loro tendenza a fissarsi a livello di sussistenza, ma poca attenzione veniva rivolta ad aspetti microeconomici e strutturali quali il processo di determinazione dei salari a livello aziendale e le ragioni delle differenze salariali fra i lavoratori.
Adam Smith rappresentò un'eccezione rispetto a questo approccio. Egli fu certamente, come è ben noto, un economista classico e come tale utilizzò essenzialmente un approccio di tipo aggregato ai problemi del lavoro, ma dedicò anche parte del suo interesse scientifico a problemi di tipo strutturale e microeconomico. Uno dei suoi contributi, molto citato e ripreso anche oggi nella letteratura sull'economia del lavoro, riguarda i differenziali salariali. Innanzitutto egli anticipò la critica - poi sviluppata dagli economisti del lavoro - al concetto, tipico della scuola classica, di un unico tasso di salario. Egli osservava che il prezzo del lavoro non può essere accertato in modo sicuro, dal momento che prezzi diversi sono spesso pagati nello stesso posto e per lo stesso tipo di lavoro. Ciò non dipende solo dalla diversa abilità e dalle diverse capacità dei lavoratori, ma anche dalle disponibilità dei datori di lavoro.

Così quando si parla di tasso di salario o di salari in genere - egli sosteneva - di fatto si sta semplificando ciò che nella realtà è una complessa e diversificata struttura dei salari. Smith non era nemmeno convinto che vi fosse una tendenza all'uguaglianza dei diversi salari. La tendenza all'uguaglianza riguardava - a suo parere - non i salari, bensì i "vantaggi netti" legati al posto di lavoro. Il complesso dei vantaggi e degli svantaggi legati a ogni particolare tipo di occupazione - egli sosteneva - o è perfettamente uguale o tende continuamente all'eguaglianza. Se esistessero occupazioni molto più vantaggiose di altre, vi sarebbero molti lavoratori che verrebbero attirati dalle prime, mentre le seconde andrebbero deserte. Di conseguenza vi sarebbe una tendenza a ridurre la differenza tra vantaggi e svantaggi di entrambe (v. Smith, 1937, p. 9).Il concetto di "vantaggio netto" (insieme a quello corrispondente di "differenze salariali compensatrici") è rimasto un concetto fondamentale della moderna economia del lavoro.

La trattazione che Smith ha offerto dei differenziali salariali rappresenta una delle migliori analisi che mai siano state condotte sull'argomento. Egli osservava che queste differenze pecuniarie nascono proprio da vantaggi e svantaggi dei posti di lavoro. Essi riguardano innanzitutto la piacevolezza o meno del lavoro svolto; poi la relativa facilità e il costo richiesto per impararlo; inoltre la stabilità o meno del posto; ancora la fiducia che si deve avere in coloro che sono chiamati a svolgerlo e, infine, la probabilità di avere più o meno successo in quel particolare tipo di occupazione. Queste cinque caratteristiche - Smith sottolineava - possono causare differenze tra i salari delle diverse occupazioni, differenze che "compensano" appunto vantaggi e svantaggi di altra natura. L'esistenza quindi di differenze salariali di tipo "compensativo" fa sì che i vantaggi netti dei vari posti di lavoro si uguaglino (ibid., pp. 99, 100, 114, 116). Smith non mancava di evidenziare le condizioni affinché ciò si verificasse: prima fra tutte la libertà di movimento dei lavoratori fra i vari tipi di posti di lavoro. In difetto di questa, si verificherebbero disuguaglianze nei vantaggi netti. Come si sa, Smith criticava le misure dirette a ridurre la libertà d'iniziativa e nel caso specifico criticava le misure dirette a limitare l'accesso ad alcune professioni a un numero ristretto di lavoratori oppure a limitare la mobilità geografica dei lavoratori.

Per quanto riguarda il problema della determinazione del livello generale dei salari e del suo andamento nel tempo, il contributo di Smith fu altrettanto importante. Il miglioramento delle capacità produttive del lavoro e delle condizioni di vita, dipendeva, secondo Smith, dalla specializzazione e dalla divisione del lavoro. Il suo contributo fondamentale fu di non

considerare la divisione del lavoro come un fenomeno connesso con e reso possibile dalle diverse qualità e capacità dei lavoratori, i quali, sulla base di queste diversità si distribuivano i compiti. Secondo Smith la divisione del lavoro accresce la destrezza dei lavoratori, fa risparmiare tempo nei processi produttivi e favorisce l'invenzione di nuove macchine: tutto ciò dà luogo a rendimenti crescenti. I costi decrescenti, cioè la produttività crescente, dipendevano solo dal fatto che ciascun lavoratore si concentrava in una piccola operazione del processo produttivo. La divisione del lavoro e i rendimenti crescenti che essa comportava dipendevano solo dall'organizzazione e dalle macchine esistenti e non dalla diversa abilità dei singoli lavoratori.

Lo sviluppo di questa idea della possibilità di far crescere la produttività del lavoro influenzò la stessa teoria di Smith sulla determinazione del salario e sul suo andamento nel tempo. Egli rifiutò l'idea degli studiosi che lo avevano preceduto secondo cui il salario tende a fissarsi al livello di sussistenza. L'enfasi con cui egli descriveva le caratteristiche dinamiche del sistema economico e la sua possibilità di crescita, e quindi di miglioramento delle condizioni di vita, lo conduceva a dissentire da quella visione pessimista: se mai, il suo pessimismo riguardava le conseguenze della crescente monotonia del lavoro provocata dalla progressiva suddivisione delle mansioni (ibid., libro V, cap. I, parte III, art. II). L'opera di Smith, come quella di tutti i grandi pensatori, è ricordata per una molteplicità di contributi di contenuto e anche di indirizzo ideologico diversi e suscettibili di sviluppi differenti. Se il suo contributo sui differenziali salariali è rimasto fondamentale per lo sviluppo successivo della teoria del lavoro basata sull'azione delle forze competitive del mercato, quello sulle regole generali che guidano, nel tempo, la dinamica del livello generale dei salari e della produttività si è fatto apprezzare per una ricchezza di contenuti che derivava non solo dalle qualità dello studioso di problemi economici, ma anche e soprattutto da quelle dello storico e del filosofo sociale.

Anche John Stuart Mill dedicò particolare attenzione al problema dei differenziali salariali; egli non solo sviluppò le idee originali di Smith, ma seppe dare ulteriori contributi. In particolare Mill accentuò, ancor più di Smith, l'importanza dei fattori sociali. Secondo Mill la gerarchia, in ordine di importanza, dei vari tipi di lavoro, dipende dal ceto sociale dei lavoratori e dalla 'casta' cui essi appartengono. I salari di cui ciascuna casta gode dipendono a loro volta dal ritmo di crescita non tanto della popolazione nel suo complesso (come certi economisti classici ritenevano), quanto della popolazione della specifica casta di appartenenza. Questo concetto anticipa quello poi elaborato da John E. Cairnes sui "gruppi

non competitivi". Mill sosteneva inoltre che le retribuzioni di quei mestieri e di quelle professioni che richiedevano anni d'istruzione e di addestramento, avevano la caratteristica di "salari di monopolio", data l'impossibilità, per la gran massa dei lavoratori, di conseguire il livello d'istruzione necessario a svolgerli. Egli notava ancora come le retribuzioni delle donne fossero ben inferiori a quelle degli uomini, a parità di lavoro svolto. Una possibile ragione era da ricercarsi - secondo Mill - nel fatto che la donna non rappresentava il lavoratore principale all'interno della famiglia. Una ragione ancora più importante, argomentava poi, era il numero limitato delle occupazioni aperte alle donne: questo fatto causava una concentrazione delle donne nelle poche occupazioni loro aperte, con conseguente riduzione dei salari ivi pagati. È qui che ha origine l'ipotesi dell"affollamento' delle donne in particolari mestieri, che tanta attenzione riceve nel dibattito attuale sul problema della discriminazione.
Mill manifestava quindi una fiducia ben minore di quella manifestata da Smith nella capacità delle forze di mercato di uguagliare i salari o i vantaggi netti; riteneva che vi fossero importanti fattori socioistituzionali che impedivano il pieno dispiegarsi della concorrenza nell'attribuzione dei diversi posti di lavoro. Questa linea di pensiero venne ripresa e sviluppata da uno degli ultimi economisti della scuola classica: J. E. Cairnes. Cairnes riconosceva che la concorrenza fra i lavoratori non è mai completa e che il concetto secondo cui ogni lavoratore abbandona il proprio lavoro per andare a cercare un lavoro meglio retribuito è una pura finzione, frutto della mente dell'economista, che non ha alcun fondamento nel mondo reale. Il mondo reale che Cairnes ha presente è quindi molto diverso da quello cui pensava Smith. Nel variegato mondo degli economisti classici si trovavano quindi analisi di specifici e importanti temi di economia del lavoro, che erano fra loro agli estremi opposti. Queste divergenze anticipavano i contrasti che si sarebbero sviluppati nei decenni successivi fra coloro che avrebbero privilegiato il ruolo delle forze di mercato e coloro che invece avrebbero considerato i fattori socioistituzionali come quelli rilevanti per interpretare i fatti del mondo del lavoro.
Anche Cairnes comunque riconosceva un ruolo alla concorrenza, sia pure limitato. Per avere quella che egli definiva una "effettiva concorrenza industriale" (v. Cairnes, 1874, pp. 61-67) non era di per sé necessario che ciascun lavoratore fosse in grado di entrare in una data occupazione, occorreva semplicemente che fosse disponibile un certo numero di lavoratori. Per ogni tipo di occupazione, sosteneva Cairnes, esisteva un certo numero di lavoratori che, per capacità, educazione e ceto sociale, era in grado di trovare un posto di lavoro

corrispondente. Il contributo originale di Cairnes fu appunto questo: mentre riteneva quasi inesistente la concorrenza fra diversi gruppi di lavoratori appartenenti a occupazioni diverse, sosteneva peraltro che i lavoratori di ciascuna occupazione costituivano un vero e proprio mercato competitivo. È questo il famoso concetto di "non competing groups", cioè di gruppi fra loro non concorrenti, che verrà ampiamente utilizzato in tutta la letteratura successiva di economia del lavoro, soprattutto da quella componente della disciplina che privilegerà gli aspetti socioistituzionali del mercato del lavoro.

3. I marginalisti e la teoria della domanda di lavoro

La novità dell'analisi marginalista del lavoro, basata sul concetto di produttività marginale, non stava tanto nel concetto di produttività, quanto nella prospettiva particolare in cui si poneva per analizzare il problema della domanda di lavoro. La teoria della produttività marginale è una teoria microeconomica; si interroga non sui meccanismi di formazione della massa salariale, ma piuttosto su quelli di formazione del salario in una data azienda, o meglio in una azienda che voglia massimizzare il profitto. L'analisi della produttività marginale insegna che l'impresa domanda lavoro fino al punto in cui la produttività dell'ultimo lavoratore assunto è uguale al salario a lui pagato, qualunque esso sia. Dal punto di vista teorico nulla vieta che l'impresa assuma lavoratori fino al punto in cui la produttività si annulla, se non fosse che un salario pari a zero non è plausibile dal punto di vista pratico. Il punto di forza dell'analisi marginalista è la prospettiva da cui si pone. Essa non si chiede come si distribuisce il prodotto nazionale fra i lavoratori e gli altri percettori di reddito, ma si chiede piuttosto quanto lavoro verrà richiesto da una impresa che voglia rendere massimo il suo profitto. Al centro dell'attenzione è qui la decisione dell'impresa sul numero di lavoratori da assumere. Ciò porta a sua volta al concetto di domanda di lavoro in funzione del salario: un concetto fondamentale che influenzerà lo sviluppo di tutta l'economia del lavoro successiva.La prima e chiara enunciazione della teoria della produttività marginale è contenuta nell'opera di John Bates Clark (v., 1902). Secondo Clark esiste una legge naturale che regola la distribuzione del reddito nella società ed essa fa sì che la quota di reddito di ogni soggetto economico corrisponda al contributo di questo soggetto all'attività produttiva. Egli si oppose alla distinzione che gli economisti classici facevano fra produzione e distribuzione del reddito e insistette sull'unità e simultaneità di questi due processi. La teoria

di Clark della determinazione del salario era costruita sulla teoria della rendita di Ricardo e sul principio della teoria marginale del valore. Secondo Clark due sono i fondamentali fattori produttivi, il lavoro e il capitale, ed essi possono essere combinati fra loro in modi diversi. La teoria classica dei rendimenti decrescenti era stata costruita sull'assunto della disponibilità limitata della terra. Clark generalizzò questo caso, estendendolo anche al fattore produttivo capitale e si chiese quale sarebbe il risultato se, con capitale fisso, si aumentasse il numero dei lavoratori. L'aggiunta di nuovi lavoratori costringerebbe quelli già esistenti a lavorare con minori dotazioni di capitale e quindi l'aumento di prodotto netto complessivo che ulteriori lavoratori potrebbero garantire si ridurrebbe progressivamente. Ed è l'ultimo lavoratore impiegato, secondo Clark, che è cruciale dal punto di vista della determinazione dei salari. Infatti il salario si fissa al livello dell'incremento di prodotto ottenuto con l'ultimo lavoratore, al livello cioè della produttività marginale del lavoro quando tutto il lavoro disponibile è impiegato.

Fra i fondatori della scuola marginalista l'economista che più di tutti contribuì allo sviluppo dell'economia del lavoro fu certamente Alfred Marshall. Egli seguì un approccio non esclusivamente di tipo aggregato e in questo seguì l'insegnamento di Smith. Prestò inoltre attenzione ai fattori socioistituzionali e in questo continuò la tradizione inglese di Mill e Cairnes. La base del suo pensiero era certamente vicina a quella di Clark. Ogni imprenditore - sosteneva al pari di Clark - cerca di fare un uso delle risorse il più vantaggioso possibile e per ottenere questo risultato deve applicare il principio di sostituzione fra i fattori produttivi, regolando il livello d'impiego di lavoro fino al punto in cui il suo costo aggiuntivo è pari al suo contributo alla produzione (v. Marshall, 1953, p. 825). Ma Marshall si differenziò da Clark, negando che la produttività marginale potesse essere considerata una teoria dei salari. Poteva essere una teoria della domanda di lavoro, ma nulla diceva dell'offerta. E l'offerta di lavoro, come già i classici avevano sostenuto, è una componente importante della determinazione dei salari. Secondo Marshall una trattazione veramente scientifica della determinazione dei prezzi deve sempre basarsi sull'azione congiunta delle forze della domanda e dell'offerta, anche se - egli aggiungeva - i cambiamenti della domanda sono molto più frequenti e veloci di quelli dell'offerta. Così, in linea generale, secondo l'autore, più breve è il periodo di tempo che noi consideriamo, maggiore è l'attenzione che noi dobbiamo dedicare alle condizioni della domanda e all'influenza da essa esercitata sul valore; più lungo è il periodo di tempo e maggiore è l'influenza che il costo di produzione (le condizioni di

offerta) ha sul valore dei beni e dei fattori di produzione. La sua distinzione fra breve e lungo periodo fu fondamentale per lo sviluppo della teoria economica. Di fatto la novità del concetto di produttività marginale indusse Marshall a privilegiare lo studio della domanda di lavoro. Il suo contributo in questo campo fu estremamente importante. A tutt'oggi in ogni libro di testo di economia del lavoro si fa riferimento ai quattro fattori 'marshalliani' che influenzano l'elasticità della domanda di lavoro. La domanda di lavoro è tanto più elastica, quanto maggiori sono: 1) la facilità con cui il lavoro può essere sostituito con altri fattori produttivi; 2) l'elasticità della domanda del prodotto finale rispetto al suo prezzo; 3) l'importanza relativa del costo del lavoro nel costo complessivo; 4) l'elasticità di offerta degli altri fattori produttivi.Al pari degli economisti tedeschi della scuola storica e degli economisti del lavoro della scuola istituzionalista (di cui si parlerà in seguito), Marshall riconobbe il carattere mutevole delle istituzioni economiche e insistette sul fatto che i principi teorici da lui proposti non potevano essere considerati validi per tutti i tempi e tutti i luoghi. Egli faceva notare, fra l'altro, il carattere ineguale del potere contrattuale dei capitalisti e dei lavoratori e considerava la crescente influenza del movimento sindacale come uno dei fenomeni più importanti della recente storia inglese (ibid., p. 750). Osservava inoltre che il termine 'concorrenza' racchiudeva in sé un significato un po' amaro in quanto tendeva a implicare egoismo e indifferenza, e in ogni caso non era del tutto appropriato per descrivere le caratteristiche più importanti della società in cui viveva (ibid., p. 6).
Eppure queste importanti qualificazioni rimasero solo tali: qualificazioni di un'analisi delle vicende economiche di tipo essenzialmente deterministico, basata su un tentativo di sintesi della teoria classica dell'offerta e della teoria marginalista della domanda. Così nel sistema marshalliano, a dispetto del riconoscimento della complessità della realtà e del ruolo delle mutevoli istituzioni, lo studio del lavoro veniva essenzialmente 'incapsulato' in un ragionamento basato sull'operare delle forze del mercato concorrenziale.

4. Gli istituzionalisti e il ruolo del sindacato

Il termine institutional economics è ormai da lungo tempo ampiamente utilizzato nel linguaggio e nella letteratura delle scienze sociali; esso rimanda a quel gruppo di studiosi americani che, come John R. Commons, Thorstein Veblen, Robert F. Oxie e altri, tentarono, nel periodo a cavallo tra gli anni venti e trenta, di riformulare la teoria economica inserendovi

una maggiore dose di realismo rispetto al pensiero tradizionale, impostato in termini di puro funzionamento dei mercati concorrenziali. Per questi studiosi, oltre alle forze del mercato concorrenziale, le istituzioni del capitalismo, nonché i fatti della storia economica, dovevano costituire gli ingredienti dell'analisi e dello studio, se si voleva avviare veramente un avanzamento della scienza economica. Lo sviluppo dell'economia del lavoro come campo della ricerca applicata fu influenzato enormemente da questa scuola di pensiero. Una delle caratteristiche comuni a questi primi studiosi di problemi del lavoro era l'insoddisfazione nei confronti della teoria economica dominante. Richard T. Ely, uno dei precursori più famosi di questa scuola, che diede importanti contributi verso la fine del secolo scorso, sosteneva chiaramente che il sistema economico non consiste semplicemente in una somma di individui. Questa nozione, per Ely, rappresentava un assunto fittizio, sostenuto dagli esponenti della scuola individualista classica, la quale immaginava che gli uomini vivessero in condizioni di "isolamento barbaro" (v. Ely, 1884, pp. 49-50). Allo stesso modo Commons, che fu allievo di Ely e che fu forse il più importante esponente della scuola istituzionalista americana, tentò di riformulare la teoria economica basandosi sul concetto di azione collettiva, azione che doveva controllare, liberare ed espandere l'azione individuale (v. Commons, 1884, p. 22). Precursori della scuola istituzionalista si trovano certamente anche in Inghilterra. Gli scritti e le attività di Sidney e Beatrice Webb riflettevano non solo il loro desiderio di attuare riforme sociali ma anche il loro rifiuto della teoria economica e il loro desiderio di rifondare su nuove basi la scienza sociale. Essi scrissero il loro Industrial democracy nel 1897. Fu la prima seria analisi economica del ruolo svolto dal sindacato nella società industriale. Eppure una ricerca di questo tipo era quasi inconcepibile a quel tempo; essa non era considerata come parte integrante della teoria economica e il sindacato non era considerato come oggetto di studio da parte degli economisti. Quanto lontano fosse il lavoro dei Webb dalla scienza economica del tempo, furono gli stessi autori a riconoscerlo. Essi ritenevano di aver studiato le caratteristiche economiche del sindacato non nell'ambito astratto del mercato concorrenziale, bensì in quello del mondo corrente degli affari, così come essi lo conoscevano (v. Webb e Webb, 1897, p. VIII).

Ma è certamente negli Stati Uniti che la corrente istituzionalista si afferma maggiormente. Si può anzi dire che la nascita dell'economia del lavoro in parte coincide con gli sforzi per rifondare la scienza economica operati alla fine del secolo scorso da questi economisti americani, che per lo più avevano studiato in Germania. Ed è così che nacque la famosa

scuola del Wisconsin, i cui componenti furono sempre caratterizzati da un intenso zelo rivolto verso la politica e le riforme sociali, nonché da un atteggiamento critico sia verso la teoria economica tradizionale che verso i confini convenzionali tracciati fra le varie scienze sociali. Questi studiosi erano favorevoli a un approccio che oggi chiameremmo interdisciplinare, anzi a un approccio unico, ancor prima che interdisciplinare. Certamente Ely, Commons e Perlman si consideravano innanzitutto economisti, ma la loro produzione spaziava dalla storia al diritto, dalla sociologia alla psicologia e anche all'etica. Il volume di Selig Perlman, A theory of the labor movement, veniva generalmente adottato nei corsi di economia del lavoro, ma aveva ben poco a che fare col metodo tradizionale dell'analisi economica; si inseriva piuttosto nella tradizione ispirata al famoso detto di Commons, secondo cui gli studiosi del lavoro dovevano "uscire e scoprire cosa c'è nella testa dei lavoratori". E Perlman scoprì che nella testa dei lavoratori c'era un pensiero ossessivo, che consisteva nella scarsità di posti e occasioni di lavoro: problema trascurato dalla teoria tradizionale fiduciosa - al contrario - nella capacità del sistema di libera concorrenza di garantire la piena occupazione (v. Perlman, 1928, pp. 242-243).
Ciò che questi profondi conoscitori del mondo del lavoro non riuscirono invece a realizzare fu la costruzione di una coerente teoria economica; in proposito la valutazione degli studiosi è unanime. Lo sforzo di Commons e dei suoi colleghi fu interamente assorbito dallo studio di specifiche istituzioni e dalla preparazione di specifici piani di riforma sociale. Certamente tutto quel complesso di esperienza, diretta osservazione, coinvolgimento personale e studio di casi che li contraddistinse, avrebbe potuto costituire l'ingrediente necessario per costruire un corpo di concetti di applicabilità generale, cioè un corpo dottrinario. Ma a questo non si arrivò. Non vi riuscirono i primi istituzionalisti e non vi riuscirono i loro allievi. Anzi, questi ultimi ereditarono soprattutto l'intenso interesse per i problemi del lavoro e i problemi sociali in genere, unito a un profondo scetticismo circa l'utilità della teoria economica dominante; non ereditarono l'interesse a rifondare la scienza economica. L'interesse scientifico si diresse in altre direzioni e contribuì a dare contenuto ad altre discipline, come la sociologia del lavoro o, ancor più, lo studio delle relazioni industriali. È nelle grandi scuole di relazioni industriali del mondo anglosassone che si conserva il prezioso contributo di quegli illuminati studiosi del mondo del lavoro.

5. I primi tentativi di sintesi neoclassica

I tentativi di sintesi dei due filoni, istituzionalista ed economico, che verranno operati nel secondo dopoguerra, affondano le loro radici in alcuni contributi teorici che apparvero fra le due guerre mondiali, per merito soprattutto di alcuni economisti non particolarmente specializzati nel campo dei problemi del lavoro: Pigou, Hicks e Douglas.Arthur C. Pigou fu il successore di Marshall alla cattedra di Cambridge e il suo nome è legato, fra l'altro, al tentativo di integrare i principi economici con l'analisi dei problemi del lavoro. Il suo famoso libro, The economics of welfare, fu descritto da Schumpeter come la più grande trattazione di economia del lavoro effettuata da un economista teorico (v. Schumpeter, 1954, p. 948). L'approccio di Pigou era sia microeconomico che macroeconomico; l'analisi spaziava dalla questione relativa ai fattori che determinano il volume del reddito nazionale ai problemi di efficienza allocativa. La fama di questo libro è certamente dovuta al contributo fondamentale che portò alla moderna economia del benessere, nonché alla teoria della finanza pubblica. Ma anche la parte dedicata ai temi di economia del lavoro è consistente. Dal momento che il suo interesse era rivolto all'individuazione dei fattori che massimizzano il prodotto nazionale, i problemi del lavoro venivano visti soprattutto in quest'ottica. Si interessò pertanto dei limiti che venivano fissati all'orario di lavoro e degli strumenti per legare la remunerazione alla produttività (ad esempio attraverso il cottimo). La preoccupazione di Pigou era che l'eccessiva regolamentazione dell'attività lavorativa potesse comportare una riduzione dell'efficienza con cui il lavoro veniva impiegato nel processo produttivo. Parte del dibattito sulla cosiddetta 'flessibilità' del lavoro, che è particolarmente vivace proprio ai giorni nostri, trova un precedente importante in questa trattazione di Pigou.
Un secondo importante contributo diretto a conciliare l'economia del lavoro con la teoria economica si deve a John R. Hicks e al suo volume The theory of wages, apparso nel 1932. Hicks riconosceva che la nascita del sindacato e il diffondersi della legislazione sociale tesa a proteggere i lavoratori costituivano l'avvenimento più importante del secolo dal punto di vista dei problemi del lavoro e in particolare della determinazione dei salari. Ma egli sosteneva anche che la teoria della determinazione dei salari non è altro che un caso speciale della più generale teoria del valore e che gli interventi dello Stato e del sindacato nel mondo del lavoro non sono di per sé sufficienti per cambiare la struttura fondamentale di questa teoria (v. Hicks, 1932, pp. I-VI).Come Pigou, anche Hicks aveva un forte interesse per le relazioni industriali e il suo The theory of wages includeva molte qualificazioni, eccezioni e

riferimenti storici e istituzionali importanti, tali da controbilanciare in una certa misura l'approccio ceteris paribus dell'analisi teorica. Nel complesso, tuttavia, il contributo di Hicks si colloca sostanzialmente nell'alveo della tradizione neoclassica e il suo sforzo di adattare il metodo neoclassico all'analisi del conflitto industriale - in cui la quantità rilevante non è la quantità di lavoro domandata e offerta, bensì la lunghezza degli scioperi - ha rappresentato un contributo importante per i successivi tentativi d'integrare gli studi sul lavoro con la teoria economica.

Un terzo contributo in questa direzione venne da un economista americano, Paul H. Douglas, il quale può essere ricordato come uno dei precursori dell'integrazione dell'economia del lavoro non solo con l'analisi teorica, ma anche con le raffinate tecniche quantitative. Il suo The theory of wages del 1934 può essere considerato un'opera complementare rispetto a quella di Hicks. Hicks aveva incominciato la sua analisi con la teoria della produttività marginale; il suo approccio seguiva la tradizione neoclassica di tipo deduttivo ed era caratterizzato dallo scarsissimo uso di dati statistici. Il lavoro di Douglas, al contrario, rappresentò un notevole sforzo di natura statistica al fine di verificare la teoria della produttività marginale. Douglas si chiedeva in particolare se i rendimenti decrescenti fossero una semplice astrazione senza alcun legame con la realtà, oppure se essi rispecchiassero i fatti concreti. A questo proposito riteneva doveroso investigare a fondo le funzioni di offerta dei fattori produttivi e verificare se il processo di distribuzione del reddito che si verificava nella realtà fornisse un supporto ai risultati dell'analisi teorica di tipo deduttivo (v. Douglas, 1934, p. XII). Fu su queste basi di partenza che Douglas cominciò a misurare la crescita del capitale e della quantità di lavoro impiegati nell'economia statunitense dal 1899 al 1922 e poi a misurare il contributo relativo di entrambi i fattori al flusso di produzione di quel lungo periodo storico. Con l'aiuto di un matematico, Charles W. Cobb, sviluppò l'analisi di una funzione di produzione che era stata proposta originariamente da Wicksell (v., 1901-1906) e che fu applicata ai dati statistici. La funzione di produzione, che da allora fu denominata 'Cobb-Douglas', è la più utilizzata nelle analisi empiriche e nelle esemplificazioni teoriche di tipo neoclassico: vi sono oggi pochi strumenti di analisi teorica ed empirica che possano essere considerati altrettanto famosi.

6. Il dibattito antimarginalista nel secondo dopoguerra

Gli economisti che nel secondo dopoguerra si dedicarono ai problemi del lavoro, con un approccio più analitico di quello seguito dai primi economisti del lavoro di scuola istituzionalista, furono certamente influenzati dai lavori di Pigou, Hicks e Douglas. Di fatto le opere di questi grandi economisti rappresentarono i libri di testo su cui si formarono gli economisti del lavoro della successiva generazione. A dispetto di tutte le vicende storiche, la teoria tradizionale dei prezzi passò quasi indenne attraverso la crisi, prima, e la guerra mondiale, poi. Negli anni cinquanta questa teoria veniva insegnata nei corsi introduttivi delle università di tutto il mondo più o meno allo stesso modo di come Marshall l'aveva esposta nei suoi testi, mezzo secolo prima.

Furono soprattutto gli economisti del lavoro americani che si opposero a un uso acritico e a un'estensione semplice e automatica della teoria microeconomica tradizionale alla spiegazione dei principali fatti sociali ed economici del tempo. Uno di questi fatti interessava particolarmente quegli economisti: negli anni quaranta il livello e la struttura delle retribuzioni erano pesantemente influenzati nelle principali industrie - come quelle dei mezzi di trasporto, della gomma, dell'acciaio e dei prodotti tessili - dalla contrattazione collettiva (soprattutto a livello aziendale). Il salario venne sempre più visto non tanto come un prezzo di mercato, quanto come un prezzo amministrato e come il risultato, in larga misura, delle decisioni di poche persone sedute al tavolo delle trattative sindacali. Come tale, quel particolare prezzo condivideva alcune delle caratteristiche dei prezzi dei mercati imperfetti e l'interesse a studiare questi mercati era certamente stimolato dagli studi innovativi di Edward H. Chamberlin e Joan Robinson sulla concorrenza imperfetta.

Come è ben noto agli studiosi di economia politica, quella generale insoddisfazione nei confronti della teoria tradizionale sfociò in un interessante dibattito che si concretizzò in una serie di articoli apparsi sull'"American economic review". Un esponente importante della critica antimarginalista fu Richard A. Lester, che si può dire fosse essenzialmente un economista del lavoro. Lester (v., 1946, p. 71) attaccò la funzione di domanda di lavoro così come l'aveva formulata Marshall. Egli condusse un'indagine presso un gruppo di imprese e scoprì che queste, nel domandare lavoro, erano più attente a ciò che succedeva nel mercato del loro prodotto che non al salario pagato ai lavoratori. Lester criticò il concetto di produttività marginale, attribuendogli scarso significato operativo a livello di azienda. Non tutti gli economisti del lavoro del tempo erano sulle posizioni estremiste di Lester, ma certamente i dubbi, se non sui fondamenti, almeno sull'utilità della teoria marginalista, erano

molto diffusi in questo ambiente. L'economia del lavoro che emergeva era una disciplina che affrontava tutti i problemi più importanti del mondo del lavoro: tutti quelli che avevano a che fare con le condizioni di lavoro e di vita dei lavoratori. I problemi della povertà, della sicurezza sociale, dell'assistenza, della legislazione a favore dei lavoratori e altri problemi, di portata sociale oltre che strettamente economica, formarono i contenuti sia della ricerca che della didattica della disciplina. Questa concezione allargata dell'oggetto di studio contribuì ancora una volta, come era successo nella prima parte del secolo, a dare un carattere multidisciplinare all'economia del lavoro. Certamente c'era ora un nocciolo economico - e questo, nonostante le critiche, i dubbi e le riserve della gran parte degli studiosi, era sostanzialmente microeconomico tradizionale - ma l'economia del lavoro, così come era intesa, si allargava a temi di sociologia, psicologia, diritto e politica. Dal punto di vista metodologico la ricerca empirica usava strumenti semplici. Non si parlava, come si fa invece oggi, di econometria, bensì di statistica economica e anziché l'analisi di regressione si usavano semplici tabelle di dati.

Anche negli anni cinquanta e sessanta, come nei decenni precedenti, la letteratura di economia del lavoro era dominata dagli studi sui sindacati e sulla contrattazione collettiva, anche se si notava uno sforzo maggiore per affrontare gli argomenti in modo più analitico. In proposito rimane famoso il dibattito che si sviluppò, sempre negli Stati Uniti, sulla natura del sindacato e sugli obiettivi che un'istituzione di questo tipo intendeva perseguire. Alcuni economisti del lavoro, come John T. Dunlop, consideravano il sindacato essenzialmente come un'istituzione economica alla pari dell'impresa e, come tale, interessata al raggiungimento di un obiettivo di carattere economico. Per Dunlop lo scopo del sindacato era quello di massimizzare la massa salariale; per raggiungere questo obiettivo doveva tener conto della funzione di domanda di lavoro da parte delle imprese, cioè della relazione esistente fra salario e occupazione (v. Dunlop, 1950², p. 33). Per altri invece, come Arthur M. Ross, il sindacato doveva considerarsi come un'istituzione prettamente politica e non come un'organizzazione governata secondo i calcoli economici di convenienza tipici dell'impresa industriale. Come organizzazione politica, e non solo economica, era soggetta a comportamenti basati anche sull'imitazione fra gruppi e sulla rivalità fra diverse organizzazioni. I concetti di pattern bargaining (modello di contrattazione) e di orbits of coercive comparisons (ambiti entro cui si effettuano i confronti) sono termini coniati da Ross, che sono entrati stabilmente negli studi di economia del lavoro e di relazioni industriali (v.

Ross, 1948, p. 51). Il tema degli effetti dell'azione sindacale è un argomento 'classico' nel dibattito fra economisti tradizionali e studiosi del lavoro più legati al ruolo svolto dalle istituzioni. Questo confronto fra opzioni teoriche e ideologiche diverse è sempre stato ed è tuttora presente nel campo dell'economia del lavoro.

La scarsa considerazione che gli economisti neoclassici nutrono nei confronti dei sindacati riflette soprattutto le loro convinzioni circa il ruolo e gli obiettivi dei sindacati stessi. Per quanto a essi venga in genere riconosciuto un importante ruolo sociale, si ritiene tuttavia che il loro principale compito sia di alzare i salari oltre il loro livello di mercato. I sindacati agiscono cioè come monopolisti, limitando l'offerta di lavoro attraverso il conflitto, o limitando l'accesso di lavoratori a un dato mestiere. Al pari dei monopolisti nei mercati dei prodotti, i sindacati sono visti come una 'imperfezione' che conduce artificialmente ad alti salari e a una conseguente distorsione nell'uso delle risorse. Oltre l'inefficienza cosiddetta 'allocativa' prodotta da questo effetto sui salari, vi è spesso - secondo questa opinione - anche un effetto dannoso sulla produttività, provocato dalle varie forme di conflitto e di pratiche restrittive (restrittive dell'iniziativa imprenditoriale all'interno dell'impresa).

Gli economisti del lavoro del primo dopoguerra avevano un'opinione diversa dei sindacati ed essa si basava sul giudizio relativo al carattere concorrenziale del mercato del lavoro. Lester, il più critico fra questi economisti, affermò che il mercato del lavoro era per sua natura il più imperfetto dei mercati (v. Lester, 1951, p. 31). Una delle imperfezioni, secondo questo autore, era dovuta alla forte capacità di controllo esercitata dalle imprese sui mercati locali del lavoro. A questo proposito, Lloyd Reynolds e Cinthia Taft (v., 1956, p. 369) affermavano che il potere di monopsonio dei datori di lavoro non era affatto trascurabile dove i sindacati erano assenti. Questo potere di monopsonio non derivava, secondo questi due autori, dal fatto che esistevano solo grandi imprese, bensì dalla tacita collusione esistente fra tutte le imprese, piccole o grandi che fossero, le quali evitavano di farsi fra loro concorrenza nel mercato del lavoro.

Caratteristica fondamentale di questa corrente di pensiero era la scarsa fiducia nei meccanismi di mercato al fine di garantire salari cosiddetti di 'equilibrio'. Fra le cause più importanti venne indicata la scarsa mobilità dei lavoratori. Di fatto da numerose ricerche empiriche risultò che, per lo stesso tipo di prestazioni di lavoro e nello stesso mercato locale del lavoro, venivano pagati salari alquanto diversi fra loro. Il fatto che esistessero gruppi diversi di lavoratori non in concorrenza fra di loro - fatto che, come si è visto, era già stato

messo in luce nel secolo scorso da Stuart Mill e Cairnes - risultava abbastanza chiaramente da queste ricerche empiriche. Clark Kerr (v., 1954) un altro noto esponente di questo gruppo, coniò un nuovo termine per descrivere questa segmentazione fra gruppi di lavoratori: egli parlò di "balcanizzazione" del mercato del lavoro.
Quegli stessi economisti del lavoro, inoltre, avevano un'idea diversa degli obiettivi e delle funzioni del sindacato. Essi ammettevano che uno dei suoi obiettivi consisteva nel far alzare i salari, anche oltre il livello del mercato. Riconoscevano inoltre che certe pratiche sindacali potevano nuocere all'efficienza e alla produttività delle imprese. Ma essi sottolineavano, molto di più degli economisti neoclassici, il ruolo positivo del sindacato. Ancora Lester, per esempio, argomentava che le ambizioni più importanti dei lavoratori erano: 1) la sicurezza economica; 2) l'opportunità di migliorare le proprie capacità professionali; 3) un trattamento equo; infine, 4) il senso di appartenenza alla comunità di lavoro. Ora non era affatto detto che il libero mercato potesse soddisfare nel modo migliore queste giuste esigenze. Anzi almeno per una, la quarta, Lester era convinto che non potesse essere soddisfatta affatto in un mercato concorrenziale (v. Lester, 1951). Proprio a causa dei limiti del meccanismo di mercato e della contrattazione individuale, i lavoratori decidevano di organizzarsi, formare sindacati e contrattare collettivamente il loro salario.

7. La moderna economia del lavoro

Alla fine degli anni cinquanta l'interesse degli studiosi di problemi del lavoro si spostò verso temi più legati alla macroeconomia, come ad esempio la disoccupazione, l'inflazione, il ruolo dei sindacati nell'accentuare le tendenze inflazionistiche, le politiche dei redditi, ecc. Di questi interessi si parlerà più avanti.
L'economia del lavoro di tipo microeconomico riprese vigore verso la fine degli anni sessanta. Si sviluppò soprattutto negli Stati Uniti, su basi sostanzialmente diverse da quelle che avevano caratterizzato la disciplina nei decenni precedenti: in questa fase le basi teoriche di riferimento sono decisamente neoclassiche. I contenuti dei libri di testo americani di economia del lavoro, che cominciano ad apparire proprio in quegli anni, mostrano chiaramente il cambiamento avvenuto rispetto al periodo precedente. I vecchi libri di testo erano essenzialmente di tipo descrittivo e privilegiavano i contenuti storici e istituzionali. I testi moderni danno spazio essenzialmente a temi di contenuto microeconomico tradizionale;

la loro struttura è ormai quella nota, che viene seguita anche nei testi più recenti, e cioè basata sulla suddivisione tra offerta di lavoro, domanda di lavoro e determinazione dei salari. È la struttura della tradizionale teoria dei prezzi di impronta neoclassica. I testi moderni, inoltre, fanno continuo riferimento ai risultati della letteratura empirica, la quale utilizza essenzialmente i metodi dell'econometria. È un mondo sostanzialmente diverso da quello dei manuali degli anni quaranta e cinquanta.

Come si giustifica questo profondo mutamento? Da un lato la trasformazione affonda le proprie radici nelle nuove tendenze della teoria economica, dominata sempre più dalla tradizione neoclassica; dall'altro lato essa è dovuta alla disponibilità di fonti statistiche sempre più ricche e allo sviluppo delle nuove tecniche econometriche, sostenute a loro volta dalle capacità di calcolo degli elaboratori elettronici. I cambiamenti in questi campi erano stati particolarmente intensi negli ultimi decenni ed era quindi logico attendersi che anche l'economia del lavoro, al pari di altre discipline di contenuto economico, ne sarebbe stata influenzata.Si può senz'altro affermare che vi è stato più progresso nella microeconomia del lavoro in questi ultimi vent'anni che non nei sessant'anni precedenti. Le nuove direzioni della ricerca economica si possono così riassumere.

1. La teoria del capitale umano elaborata da Jacob Mincer (v., 1974), Theodore W. Schultz (v., 1961) e Gary S. Becker (v., 1975). Questi autori hanno sviluppato le intuizioni felici di A. Smith e Pigou. La teoria considera l'istruzione e l'addestramento come fenomeni analoghi all'investimento in capitale fisico, e quindi come oggetto di decisioni che devono confrontare, come in un qualsiasi calcolo di convenienza, i costi immediati con il flusso di maggior reddito che l'investimento rende possibile lungo tutta la vita lavorativa. Questa teoria ha stimolato una letteratura ormai vastissima, che mette in relazione l'abilità e le doti innate, da un lato, e l'istruzione e l'addestramento dall'altro, coi loro rispettivi e distinti effetti sulle capacità di guadagno dei singoli lavoratori. È quindi un'importante teoria anche dei differenziali salariali.

2. La teoria della discriminazione. Anche questa, e la letteratura cui ha dato luogo, è tipicamente americana ed è facile comprenderne la ragione. Essa è stata stimolata da eventi importanti quali la nascita del movimento per i diritti civili e di quelli a difesa delle minoranze etniche, nonché dei movimenti a favore delle donne. La presa di coscienza da parte dell'opinione pubblica dell'importanza di questi fenomeni ha favorito lo sviluppo di questo filone di ricerca, il quale, sia a livello teorico (microeconomico) che empirico, ha

esplorato le ragioni delle differenze retributive e di opportunità di lavoro fra lavoratori di razza e sesso diversi e ha cercato di individuare quanto di queste differenze può essere spiegato da fattori economici e quanto da pratiche di discriminazione che niente hanno a che fare coi calcoli di convenienza operati dalle imprese.

3. L'approccio del ciclo di vita applicato alle decisioni riguardanti l'offerta di lavoro. Franco Modigliani aveva sostenuto, già negli anni cinquanta, che le decisioni relative al consumo, al risparmio e all'investimento vengono prese con riferimento all'arco temporale che abbraccia la vita intera dei soggetti economici coinvolti. Partendo da queste considerazioni, bastava solo un piccolo passo in avanti per argomentare che anche le decisioni relative all'offerta di lavoro - vale a dire quanto lavorare, che tipo di lavoro fare, quanto investire in capitale umano e così via - vengono prese considerando non solo le condizioni attuali, ma anche quelle future, di lungo periodo, del soggetto e della sua famiglia. Questo approccio teorico è stato molto fertile dal punto di vista delle molteplici ipotesi che ha permesso di formulare e di verificare empiricamente.

4. L'analisi dell'offerta di lavoro è stata molto approfondita ed è stata arricchita da un nuovo filone di pensiero che va sotto il nome di 'economia della famiglia'. Questa si basa sulla banale ma importante considerazione che i lavoratori non sono persone che, nella generalità dei casi, vivono da sole. Essi in genere vivono in una famiglia e ciò comporta una inevitabile interrelazione delle decisioni dei vari componenti della famiglia riguardanti la loro attuale e futura attività lavorativa. L'offerta di lavoro di ciascuno dei componenti è legata a quella degli altri e le stesse offerte di lavoro sono solo una parte di quel complesso di decisioni che riguardano importanti aspetti della vita familiare: dove vivere, quanto spendere e quanto risparmiare, cosa comprare, quanto investire nella formazione dei figli, quale attività scegliere per le persone anziane, ecc. L'offerta di lavoro viene considerata, in questo quadro teorico, come una parte del più ampio problema dell'uso del tempo (la vera risorsa scarsa) da parte di tutti i componenti della famiglia.

5. Il riconoscimento dell'esistenza di mercati del lavoro interni all'azienda, soprattutto se questa è di grandi dimensioni, nonché la diffusione dei rapporti di lavoro di lunga durata hanno comportato notevoli progressi nell'analisi della domanda di lavoro. In particolare hanno incoraggiato i tentativi di teorizzare, utilizzando il solito assunto neoclassico di un comportamento razionale teso a massimizzare l'utilità e il profitto, le decisioni delle imprese relative alla gestione del personale, all'investimento nell'attività di reclutamento e di

selezione, all'addestramento sul posto di lavoro, alle strategie retributive e dei profili di carriera, all'aggiustamento delle ore di lavoro e dell'occupazione nelle diverse fasi cicliche, e così via. Si è presa piena coscienza del fatto che il rapporto di lavoro è più simile a un 'matrimonio' che alla compravendita di un qualche bene di consumo. Il rapporto deve quindi essere regolato da un contratto adeguato. La natura del contratto è cruciale affinché il lavoratore riceva ciò cui aspira, non solo il salario ma un trattamento equo, una certa stabilità del reddito e, dall'altra parte, il datore di lavoro ottenga una prestazione effettuata con impegno, sforzo, capacità professionale. Non è possibile ottenere tutto questo con continui controlli reciproci o con continue rinegoziazioni del rapporto di lavoro. Sarebbe troppo costoso in termini di tempo e di conflitto. Meglio affidarsi a contratti di lunga durata, i quali prevedano al contempo gli incentivi giusti affinché le parti si comportino secondo le modalità previste e volute. Lo studio dei contratti di lavoro ha aperto interessanti filoni di ricerca teorica ed empirica.

6. Gli sviluppi della teoria dell'informazione e della 'ricerca', che all'inizio veniva utilizzata nell'analisi dei mercati dei prodotti, sono stati molto utili per analizzare il funzionamento del mercato del lavoro. Il problema del matching (cioè il mutuo incontro) di domanda e di offerta di lavoro viene affrontato con strumenti di analisi alquanto sofisticati, che sono ormai patrimonio di tutti gli economisti che si interessano di questo problema. Diverse sono le implicazioni interessanti di questo approccio. È stato rigorosamente dimostrato un risultato messo in luce numerose volte dagli economisti del lavoro del passato, e cioè che con un'informazione imperfetta e costosa un mercato del lavoro concorrenziale non può eliminare le differenze retributive fra posti di lavoro anche simili ma dislocati in punti diversi del mercato. È emerso inoltre, come sviluppo della tradizionale spiegazione neoclassica della disoccupazione, che ciò che appare tempo perso durante i periodi di disoccupazione, è di fatto un investimento di tempo nella ricerca di un posto di lavoro migliore di quello immediatamente disponibile. Anche la disoccupazione, quindi, in questo quadro teorico, fa parte di quel complesso di decisioni, ottime e razionali, che riguardano l'uso del tempo disponibile.

7. La natura della disoccupazione. Continua a essere dibattuto in ogni caso il tradizionale tema relativo alle cause e alla natura della disoccupazione. Come è noto, la distinzione fra disoccupazione volontaria e involontaria rimane il cuore della teoria keynesiana. Secondo i neoclassici tutta la disoccupazione, a parte quella di natura frizionale, è di carattere

volontario, cioè è causata dalla non disponibilità dei lavoratori ad accettare salari reali più bassi di quelli esistenti. Keynes sosteneva invece che la disoccupazione involontaria poteva esistere nel caso in cui, di fronte a un piccolo aumento nel livello dei prezzi relativamente al livello del salario monetario, sia l'offerta aggregata di lavoro disposta a lavorare al salario monetario corrente, sia la domanda di lavoro corrispondente a quel salario, fossero entrambe maggiori del volume esistente di occupazione (v. Keynes, 1936, p. 15). Questa definizione implica che si è in presenza di disoccupazione involontaria quando esiste un eccesso di offerta di lavoro a un dato valore del salario monetario; in questa situazione la disoccupazione può essere eliminata con un'espansione della domanda aggregata e con una riduzione del salario reale.

La disoccupazione keynesiana viene talvolta identificata con la disoccupazione ciclica, cioè con la disoccupazione che si manifesta durante la fase recessiva del ciclo economico, quando l'investimento e il consumo si contraggono e il sistema economico non è più in grado di generare lo stesso volume di posti di lavoro della fase precedente di espansione.

Accanto alla disoccupazione ciclica, che deriva da una deficienza di breve periodo della domanda aggregata che non riesce a sfruttare la capacità produttiva del sistema, può esistere anche una disoccupazione che persiste più a lungo, indipendentemente dalle alterne vicende congiunturali. Anche nei punti di massimo del ciclo la produzione può essere inferiore al livello necessario per garantire il pieno impiego. In certe condizioni la cosiddetta 'disoccupazione tecnologica' può essere un esempio di questo secondo tipo di disoccupazione, legata alle caratteristiche del processo di sviluppo più che a quelle del ciclo economico. La disoccupazione di natura tecnologica deriva dalla sostituzione di capitale a lavoro indotta dal progresso scientifico e tecnologico. I lavoratori che sono resi esuberanti dall'introduzione, nelle imprese in cui sono occupati, di nuove tecnologie, talvolta riescono a trovare un lavoro alternativo nella stessa impresa o in altre imprese; ma se le loro capacità professionali e le loro qualificazioni non sono più richieste dal mercato del lavoro, a causa di un diffuso processo di ammodernamento delle tecniche di produzione e di cambiamento dei contenuti dei posti di lavoro, può succedere che la loro riqualificazione richieda parecchio tempo. Occorre anche considerare che i lavoratori anziani sono generalmente poco disponibili a sottoporsi ai processi di apprendimento necessari per la riqualificazione. Il fatto di essere in una fase espansiva caratterizzata da un'elevata domanda aggregata può non essere sufficiente per eliminare, nel breve periodo, questo tipo di disoccupazione.

Su ciascuno degli argomenti elencati esistono ottime rassegne della letteratura, come ad esempio nel volume di saggi raccolti da Orley Ashenfelter e Richard Layard (v., 1986).In questo elenco si potrebbero inserire anche il sindacato e la contrattazione collettiva. Vi sono infatti diversi economisti del lavoro che si interessano degli stessi problemi di cui si interessavano Dunlop e Ross, e cioè degli obiettivi del sindacato, mentre altri economisti studiano le relazioni fra il sistema di contrattazione collettiva e la performance economica. Nel complesso però questi argomenti occupano molto meno spazio nella letteratura di economia del lavoro di quanto non ne occupassero nel periodo dell'immediato dopoguerra. Il lavoro di ricerca sugli aspetti storici, giuridici e istituzionali del sindacato e della contrattazione collettiva è stato progressivamente trascurato dagli economisti del lavoro e fatto proprio dagli studiosi di altre discipline. Di fatto la maggioranza delle ricerche in questo campo viene svolta da studiosi di relazioni industriali. Questa disciplina si è progressivamente staccata dall'economia del lavoro e ha raggiunto da tempo la sua completa autonomia.

Un altro argomento presente nella vecchia letteratura e che continua a ricevere grande attenzione è il ruolo che il sindacato e i salari svolgono, a livello macroeconomico, nel processo inflazionistico e nella determinazione della disoccupazione. Nella ricchissima letteratura che fa riferimento alla 'curva di Phillips' (dal nome dell'economista che per primo verificò empiricamente una relazione inversa fra disoccupazione e dinamica salariale) i sindacati, come altri possibili fattori istituzionali, possono svolgere un ruolo fondamentale. È merito di grandi economisti, come Hicks ad esempio, aver affrontato l'argomento e aver richiamato l'attenzione sull'importanza che questi elementi istituzionali svolgono nel quadro macroeconomico. Vent'anni dopo aver scritto The theory of wages Hicks (v., 1955) riconosceva che i salari sono determinati dall'interagire di fattori non solo economici ma anche sociali. Questa peraltro era la convinzione del principale economista del lavoro inglese di questo dopoguerra: Henry Phelps Brown (v., 1972). Come per Phelps Brown, anche per Hicks è fondamentale il concetto di fairness (equità) delle strutture retributive. Le spinte salariali verso l'alto, con conseguenti tensioni inflazionistiche, sono tanto maggiori quanto meno accettate dal punto di vista sociale sono le differenze retributive esistenti. In questo modo si mettono in moto rincorse salariali fra vari gruppi di lavoratori, ed è sulla base di queste importanti considerazioni - che avvicinano molto il pensiero di questo autore a quello degli economisti dell'immediato dopoguerra - che Hicks sviluppa i suoi successivi e

importanti contributi di analisi macroeconomica in tema di equilibrio, con prezzi e salari rigidi (v. Hicks, 1974). Gli economisti d'impronta più decisamente neoclassica danno minore importanza, anche nell'analisi macroeconomica, al ruolo dei sindacati e dei fattori istituzionali. Essi concentrano l'attenzione sui meccanismi di funzionamento del mercato del lavoro nelle alterne fasi del ciclo economico e sui fattori che impediscono un aggiustamento istantaneo verso condizioni di equilibrio, nonché sull'esistenza e sulla misurazione di quello che viene definito il tasso di disoccupazione 'naturale', vale a dire il tasso di disoccupazione di equilibrio (di tipo essenzialmente frizionale) verso cui il sistema spontaneamente tende. Questi argomenti rientrano in quel filone di pensiero che Edmund Phelps (v., 1970) ha definito "i fondamenti microeconomici della macroeconomia". Con lo sviluppo di quest'area di ricerca i confini fra la macroeconomia dell'occupazione e dei salari e l'economia del lavoro in senso stretto si distinguono sempre meno.

8. Una nuova sintesi?

Jacob Viner definì una volta l'economia come ciò che gli economisti studiano. L'economia del lavoro dovrebbe essere definita allo stesso modo: è ciò che viene studiato dagli economisti del lavoro. La saggezza contenuta nel paradosso di Viner sta nel riconoscere implicitamente che l'economia non è definita una volta per tutte in un dato momento storico, ma che ciò che gli economisti fanno, sia in termini di problemi che affrontano, sia in termini degli strumenti di analisi che utilizzano, cambia col passare del tempo. L'economia, come tutte le scienze, è in continuo sviluppo. Ogni tentativo di riassumere o di sintetizzare lo 'stato dell'arte' di un dato campo disciplinare deve essere fatto riconoscendo che esso è storicamente condizionato.

Forse la considerazione più importante da fare sullo stato della moderna economia del lavoro è che essa, oggi più che mai, è strettamente allineata e inserita nella teoria economica intesa in senso lato. Rispetto a quella dei decenni precedenti, la moderna economia del lavoro è più preoccupata di attenersi al metodo scientifico. È meno preoccupata di arrivare a conclusioni utili per indirizzare gli interventi di politica economica; se risultati di questo tipo vengono raggiunti, essi sono spesso solo il prodotto laterale - non sempre voluto e cercato - dell'analisi economica. La ricerca è indirizzata soprattutto verso argomenti che si prestano a essere investigati in termini quantitativi con l'ausilio delle tecniche econometriche. La chiave di

lettura fondamentale dei problemi del lavoro è la teoria del mercato concorrenziale; i risultati empirici tendono a confermare che il mercato funziona abbastanza bene, e le ipotesi basate sull'assunto di soggetti che massimizzano il loro obiettivo economico vengono in genere confermate dall'evidenza dei fatti osservati. Nella misura in cui in questo approccio può scorgersi una tendenza ideologica, questa è di natura conservatrice e ciò è in linea con le tendenze della cultura economica che si è imposta in questi ultimi due decenni. Per di più, nel caso specifico, può essere di qualche rilevanza il fatto che molto di questo lavoro di ricerca proviene da economisti come Becker, Schultz, Mincer dell'Università di Chicago, che notoriamente è sempre stata, per tradizione, a favore del libero funzionamento del mercato, critica nei confronti del movimento sindacale, perplessa circa l'utilità dell'intervento dello Stato nel campo economico e sociale.

Non bisogna certo dimenticare o trascurare l'importanza dell'area di dissenso esistente nei confronti della teoria dominante. Al contrario, occorre ricordare che un significativo gruppo di economisti americani, fra i più attivi ricercatori nel campo del lavoro, ha espresso a più riprese dissenso nei confronti dell'approccio neoclassico e lo ha giudicato inadeguato per l'analisi dei problemi del lavoro. Questo gruppo, che in una recente rassegna di Glen G. Cain (v., 1976) è stato denominato come quello dei teorici del mercato del lavoro "segmentato", comprende i nomi di S. Bowles, Peter B. Doeringer, Michael J. Piore, Lester C. Thurow. Forse il contributo più importante è stato, all'inizio degli anni settanta, quello di Doeringer e Piore sul funzionamento del mercato del lavoro interno alle imprese (v. Doeringer e Piore, 1971). Il contributo di questi due autori è importante perché raccoglie l'eredità degli economisti americani del filone istituzionalista e costituisce al contempo il punto di riferimento per la ricca letteratura sulla segmentazione del mercato del lavoro che è stata prodotta negli ultimi due decenni. A questa letteratura hanno dato contributi fondamentali anche economisti del lavoro europei (v. Wilkinson, 1975; v. Maurice e altri, 1982; v. Marsden, 1986). La maggiore importanza riconosciuta al ruolo che i sindacati e la legislazione sociale svolgono in Europa, nonché la tradizione dell'approccio storico-istituzionale che ha caratterizzato la ricerca economica in questo continente sono certamente fra i motivi che spiegano come la critica all'approccio neoclassico sia particolarmente forte 'al di qua' dell'Atlantico. Lungo queste linee di ricerca si collocano anche i contributi di diversi economisti del lavoro italiani, e fra questi il nome da ricordare per primo è certamente quello di Ezio Tarantelli (v., 1986).

Gran parte delle critiche alla moderna economia del lavoro sono concentrate sull'assunto neoclassico della concorrenza perfetta. Sono in fondo le stesse critiche che, a cominciare dalla fine del secolo scorso, furono portate da economisti del lavoro appartenenti al filone istituzionalista, come i Webb, Ely, Commons. La sostanza è la stessa e consiste, in sintesi, nella mancanza di realismo storico e istituzionale. I critici, numerosi, degli studi sul capitale umano, sostengono ad esempio che il ruolo dell'istruzione non può essere adeguatamente individuato entro gli stretti confini dell'analisi neoclassica dell'investimento; deve invece essere analizzato in termini più ampi, riconoscendo la funzione sociale dell'istruzione. Allo stesso modo i temi della povertà, della distribuzione del reddito, della discriminazione, e altri ancora, non possono essere adeguatamente studiati senza tener conto dei loro risvolti sociali. Non è escluso che queste critiche alla fine portino qualche frutto e che nei prossimi anni (o decenni) queste considerazioni di carattere socioistituzionale trovino uno spazio maggiore anche nell'ambito della teoria tradizionale. Albert Rees, che di quest'ultima è uno dei padri fondatori, partendo dalla considerazione che "la ricerca accademica svoltasi nell'ambito dell'economia del lavoro è andata forse troppo avanti, per quanto riguarda il rigore e la quantificazione dei fenomeni, a discapito di un sano radicamento nella storia e nelle istituzioni", prevede che "i migliori economisti del lavoro della prossima generazione saranno esperti sia di istituzioni che di econometria e avranno la capacità di 'miscelare queste componenti' con accortezza e bravura" (v. Rees, 1977, p. 4). I prossimi anni, o i prossimi decenni, ci diranno se questa rappresenti una previsione accurata o un desiderio senza possibilità di realizzazione. Che i due approcci, se non proprio alternativi, siano molto difficili da conciliare ce lo ricorda Clark Kerr in questo suggestivo passaggio, che sintetizza molto bene anche l'esperienza di coloro che non hanno mai rinunciato a considerare l'economia del lavoro come qualcosa di più di una semplice branca della teoria economica tradizionale: "In ogni caso, noi, gli economisti del lavoro revisionisti, abbiamo finito con l'essere come piccole volpi che conoscono molte cose, e non come ricci che sono votati a una sola teoria che tutto racchiude. Noi siamo andati in cerca di piccole verità, non dell'unica irresistibile verità; o almeno di quelle piccole verità che correggono la grande verità. Noi siamo diventati volpi pluraliste, non ricci monisti. 'Le volpi conoscono molte cose', dicevano i greci antichi, e a ragione si può aggiungere, perché le cose da conoscere sono molte [...]. Ma molti nel mondo della cultura vogliono semplicità e certezza - l'unica grande cosa che il riccio conosce - per guidare i loro pensieri; una visione centrale che dia coerenza a tutte le

loro percezioni; un'unica spiegazione che dia senso a tutti gli eventi. C'è una eterna tensione fra coloro che seguono il cammino pluralista e coloro che seguono quello monista [...]. Alcune volpi cercano di aiutare i ricci, anche se i ricci raramente apprezzano l'offerta di aiuto. E ciò è triste per il mondo della cultura" (v. Kerr, 1988, p. 9).

Economia e politica industriale
di Giovanni Zanetti

Economia e politica industriale

sommario: 1. Nuovo campo di studi o evoluzione della teoria tradizionale? 2. Alla ricerca di un nuovo paradigma. 3. La presa di coscienza dell'imperfezione dei mercati. 4. Contributi dell'analisi empirica: nuove linee di studio. 5. Le imprese come istituzioni. 6. Lettura evoluzionista del mutamento economico. 7. Nuove forme di concorrenza nei mercati accessibili. 8. Ruolo dell'innovazione nel processo di concorrenza dinamica. 9. Dall'economia alla politica industriale. □ Bibliografia.

1. Nuovo campo di studi o evoluzione della teoria tradizionale?

A partire dai primi anni sessanta è comparso e si è gradualmente diffuso negli statuti di alcune facoltà universitarie italiane (economia e commercio, scienze politiche, ingegneria) un nuovo insegnamento individuato con nomi diversi (economia e politica industriale, economia industriale, economia dei settori industriali, ecc.). Del pari, è gradualmente emersa un'ampia letteratura sull'argomento, dai manuali ai volumi riguardanti aspetti particolari, ai saggi specialistici; non sono mancate inoltre rilevanti ricerche empiriche mentre sono nate riviste scientifiche specificamente dedicate a questo settore.Malgrado tale rapida affermazione, non s'è avuto un comune modo di intendere i contenuti oggetto della materia: le interpretazioni sono infatti oscillate da letture del sistema economico, senza soluzione di continuità con le proposte della teoria neoclassica, ad altre di taglio più aziendale o centrate su momenti particolari dell'attività produttiva (innovazione, tipologia della concorrenza, ecc.).
In realtà questo insieme di studi traspone quanto nella ricerca economica anglosassone è stato

raccolto sotto il titolo industrial organization; esso ha preso crescente consistenza con l'aumentare di complessità dell'apparato produttivo e con la crescita delle imperfezioni dei mercati, in rapporto alle quali l'esistente impianto teorico e conoscitivo è apparso affetto da discrepanze e lacune.In particolare, l'economia industriale ha trovato spazi importanti lasciati aperti dalle insufficienze interpretative dell'approccio neoclassico la cui indubbia concezione sistemica e globale del funzionamento dell'apparato economico si era andata sempre più dimostrando condizionata da ipotesi di base fortemente limitative.
S'è di conseguenza avvertita la necessità di una linea di approfondimento che non perdesse di vista la realtà, le scelte, i risultati delle imprese come unità costitutive dell'apparato produttivo e considerasse al tempo stesso le interrelazioni tra le stesse e il sistema che ne risulta. S'è precisato in questo senso un campo di studio che partecipa a un tempo degli aspetti microeconomici e macroeconomici della realtà, largamente centrato sull'attività di produzione di beni e servizi, sulle relazioni intercorrenti tra operatori e massa degli utenti-consumatori, sugli aspetti organizzativo-istituzionali da ciò sollecitati.L'acquisizione di adeguati livelli di conoscenza della struttura e del funzionamento dell'apparato genera le condizioni per passare a momenti propositivi nella fattispecie di interventi intesi a sopperire a carenze o a promuovere miglioramenti là dove si rivelino necessari.
L'impianto teorico dell'economia industriale crea, in tal senso, le condizioni per precisare le linee di fondo di una politica industriale, concepibile in parte come alternativa, in parte come complemento alle politiche macroeconomiche, non di rado praticate in condizioni d'urgenza e nel perdurare delle insufficienze nascenti dall'inesistenza o dall'inadeguatezza di interventi capaci di incidere sugli aspetti strutturali dell'apparato produttivo e distributivo.

2. Alla ricerca di un nuovo paradigma

Il processo attraverso il quale il campo di studio s'è venuto precisando ha avuto carattere graduale, e si è manifestato mano a mano che andava emergendo l'insufficienza interpretativa dell'approccio neoclassico. Esso ha interessato un apprezzabile periodo di tempo (40-45 anni) nel corso del quale sono emersi numerosi contributi, diversi tra di loro per obiettivi, metodo di lavoro, lessico e terminologia adottati. L'economia industriale, quale oggi è conosciuta, è la risultante di questo lavoro di analisi e di questi approfondimenti.Il punto di partenza va identificato nella fragilità della teoria tradizionale nel momento in cui si è posta di fronte al

tema della produzione e dell'analisi dell'offerta. Essa fondava in effetti il proprio impianto analitico su ipotetiche unità operative dagli attributi tecnologici estremamente semplici (si tratta normalmente di imprese monoprodotto, operanti con processi produttivi descritti tramite funzioni di produzione continue e implicanti di norma non più di due fattori produttivi). Inoltre, quelle stesse unità produttive risultano oggi, a un esame critico, quali soggetti preprogrammati sull'obiettivo della massimizzazione del profitto e come tali concepiti funzionalmente rispetto all'intento finale di identificare le condizioni descrittive dell'equilibrio generale.

È comprensibile come, da concezioni siffatte, derivasse una sostanziale insoddisfazione in merito alla capacità di interpretare il comportamento economico dell'impresa nel concreto, rispetto al quale l'analisi appariva eccessivamente astratta e, al limite, addirittura fuorviante. Altrettanto poco condivisibile era, di conseguenza, la possibilità d'interpretare le linee evolutive di fondo del sistema delle imprese colto nei suoi aspetti strutturali e nella sua dinamica.Si è delineata una precisa esigenza di aprirsi alla considerazione attenta dei risultati che si venivano acquisendo in altri campi di studio: dall'analisi della tecnologia e dei comportamenti individuali o collettivi, alla teoria dell'organizzazione, all'approfondimento statistico ed econometrico. Dalle insufficienze neoclassiche ha preso le mosse dunque un iter di ricerca fondato essenzialmente su due diversi filoni: il primo, caratterizzato dall'ulteriore elaborazione e adattamento delle proposizioni teoriche di partenza nella presa di coscienza progressiva dell'imperfezione crescente e mutevole dei mercati (v. Tirole, 1988); il secondo, qualificato dai contributi della ricerca empirica, fondata sia sul pur frammentario metodo dei casi, sia sugli studi sistematici di insiemi sufficientemente ampi di informazioni (v. Schmalensee e Willig, 1989).Il duplice approccio di studio poteva in realtà generare le premesse per una nuova e poco costruttiva contrapposizione tra teoria e pratica, stante la difficoltà indubbia di saldare una linea interpretativa - tutto sommato ancorata alle esigenze di integrazione e di quadratura tra equilibri parziali ed equilibri generali - con la varietà dei risultati di analisi introspettive di specifiche problematiche o di particolari situazioni, per loro natura non gravate dalla preoccupazione di trovare posto in una sintesi coerente.

Una lettura definitiva e unitaria ha stentato e tuttora stenta a emergere. Il lavoro di ricerca ha nondimeno posto in evidenza acquisizioni precise sul piano della conoscenza: dal cambiamento della struttura e del grado di concentrazione nell'apparato produttivo, al delinearsi della separazione tra proprietà e direzione, all'emergere di scelte e linee strategiche

quale risultato della contrapposizione e interazione di comportamenti diversi. Su questi frammenti di realtà, faticosamente derivati con caratteri di sufficiente stabilità, sono maturate le prime proposte di sintesi interpretative tuttora in fase evolutiva.

3. La presa di coscienza dell'imperfezione dei mercati

Il fatto cruciale sul quale matura la consapevolezza dell'imperfezione dei mercati è la non omogeneità dell'offerta, originata dalle politiche del prodotto intese a differenziarlo. Negli spazi di mercato, prima considerati perfettamente uniformi, si generano aree di influenza, più o meno accentuata, di ciascuna impresa, ciò che consente di abbandonare l'ipotesi della totale impossibilità d'influire sul livello del prezzo esistente sul mercato. L'abbandono dell'immagine di una funzione di domanda parallela all'asse delle ascisse, con ovvia costanza dell'incasso marginale, e la sua sostituzione con l'immagine nuova di una funzione caratterizzata da una certa pendenza negativa, coincidono con l'ammissibilità di un certo campo di scelte aperto alla singola unità produttiva. La concezione dell'impresa quale particella preprogrammata sul massimo profitto in funzione della determinabilità dell'equilibrio generale, esce gradualmente smorzata per lasciare spazio all'esercizio di un minimo di individualità, almeno a livello di coppie volumi/prezzo. I contributi di P. Sraffa (v., 1925 e 1926), J. Robinson (v., 1933) ed E. H. Chamberlin (v., 1933) implicano questo grado di realismo, sintetizzabile nel lasciare emergere la possibilità di una certa manovra del prezzo aperta anche all'impresa di dimensione minore.
Questo avvicinamento alla realtà torna tuttavia a stemperarsi allorché la proposta interpretativa della concorrenza monopolistica riprende come obiettivo la determinazione di un volume complessivo di offerta e di un livello di prezzo valido per tutti gli operatori. Chamberlin formula a tal proposito la famosa ipotesi 'eroica' tesa a conciliare l'inconciliabile, ossia a riportare all'omogeneità quanto s'è dimostrato essere differenziato. L'ipotesi passa attraverso la definizione di una domanda di settore, identificata da livelli di prezzo e determinazione di quote di mercato, quale sintesi tra le aspettative di domanda delle singole imprese e la ridefinizione delle stesse per effetto dell'estendersi uniforme delle singole scelte. Allo stesso modo viene considerato un andamento dei costi unitari medi e marginali caratterizzato dall'ipotesi di uniformità della tecnica utilizzata: attributo questo difficilmente sostenibile in una situazione fondata su prodotti differenziati. La conclusione alla quale

approda la teoria della concorrenza imperfetta si specifica in un volume complessivo di offerta e in un prezzo unico e comune, rispettivamente inferiore e superiore a quelli riscontrabili nell'omologa situazione di concorrenza perfetta: il volume di offerta è infatti ancora identificato in corrispondenza del punto in cui si verifica la coincidenza tra costo marginale e incasso marginale e tra costo unitario medio e prezzo. Questa coincidenza si verifica tuttavia lungo il tratto discendente della curva del costo unitario medio, dato che la funzione di domanda presenta un'inclinazione negativa.Il contributo dell'impostazione teorica della concorrenza monopolistica ha avuto il pregio di aprire l'analisi alla realtà della differenziazione e dell'individuazione di aree d'influenza particolare di singole imprese. La preoccupazione di non rompere con l'armonia dell'equilibrio generale ha tuttavia avuto l'effetto di attenuarne fortemente le potenzialità ai fini di un'efficace lettura dell'apparato produttivo. Le forzature del reale sono in effetti tali da suggerire la conclusione secondo la quale la finalità di colmare la separazione tra teoria e pratica ha compiuto pochi passi avanti o forse nessuno (v. Triffin, 1941).Nella consapevolezza di questi limiti si può nondimeno affermare che il tormentato emergere di tale lettura teorica non è stato inutile, in rapporto al maturare dell'economia industriale. Secondo D. A. Hay e D. S. Morris (v., 1979) esso ha aperto la possibilità di meglio conoscere e classificare le strutture di mercato portando l'analisi a un notevole livello di raffinatezza, che più tardi sarà ripreso da E. S. Mason (v., 1939) e da J. S. Bain (v., 1956) per definire ipotesi di collegamento tra assetto del mercato e risultati conseguibili, verificabili empiricamente.Per precisare i contenuti dell'economia industriale si può quindi affermare che la presa di coscienza dell'inadeguatezza del modello di concorrenza perfetta, dei fattori di tipo monopolistico che in qualche grado caratterizzano tutte le imprese, della connessione tra forme di concorrenza e tipologia dei prodotti, da un lato, e il manifestarsi delle curve di domanda e dei costi, dall'altro, ha fatto emergere l'esigenza di analisi fondate sui dati concreti, in grado di rispecchiare efficacemente la struttura dei mercati e dei processi produttivi.

S'è aperta per tale via un'occasione di approfondimento, i cui risultati sono andati oltre le teorizzazioni della concorrenza monopolistica per cogliere il continuum del manifestarsi delle imperfezioni dei mercati, sì da superare le classificazioni esistenti per precisare le interrelazioni tra le singole imprese e i rivali. È maturato su queste basi il filone interpretativo delle moderne teorie dell'oligopolio che hanno tuttora il loro riferimento nello studio di Paolo Sylos Labini (v., 1956) e nelle ricerche che con questo hanno interagito: dagli

scritti di J. S. Bain (v., 1956) a quelli di F. Modigliani (v., 1958). Uno dei pregi di queste letture dell'apparato produttivo sta nella migliore precisazione del concetto di 'struttura' del mercato che va al di là della semplice enumerazione delle presenze operative per considerare simultaneamente le caratteristiche della domanda, in termini di estensione complessiva e di elasticità, e il grado di introduzione e di diffusione del progresso tecnico. In un contesto siffatto, i comportamenti adottati dai singoli operatori possono condurre a stati di quiete interpretabili come equilibri temporanei e limitati, sintesi a un tempo della tecnologia esistente e della distribuzione del potere. Essi non sono più leggibili secondo gli schemi semplificati del tipo stimolo-risposta, ma rispondono a scelte autonome; ogni imprenditore diviene pertanto 'attore' nella costruzione degli stati risolutivi finali, tramite le proprie decisioni e l'interazione con gli operatori rimanenti.

La complessa problematica trova una sintetica espressione nei modelli derivati dalle diverse impostazioni teoriche. Nel segno di una continuità di ricerca esse si identificano inizialmente con quelle proposte da A. Cournot (v., 1838) e da J. Bertrand (v., 1883), nelle quali le variabili oggetto di scelta da parte delle imprese sono prezzi e volumi di produzione (ovvero quelle non soggette a controllo nella concorrenza perfetta), mentre è assunto come dato il comportamento degli operatori concorrenti. Per i maggiori gradi di libertà rispetto all'ipotesi concorrenziale gli schemi alla Cournot-Bertrand offrono una prima base per il successivo sviluppo di impostazioni di tipo strategico; per l'assunzione restrittiva circa l'atteggiamento degli altri imprenditori, il loro contenuto strategico appare limitato. Tale limite si attenua fino a scomparire man mano che ognuno degli operatori assume il ruolo d'attore sullo scenario economico e come tale dà vita a proprie autonome strategie. Come riassume A. Jacquemin (v., 1987), "nei modelli strategici si suppone essenzialmente che gli agenti economici siano in grado di assumere una posizione, finanziaria o psicologica, tale da dissuadere e limitare le azioni e le reazioni delle imprese rivali, sia effettive che potenziali".

Emergono in queste concezioni, nella loro portata risolutiva, le barriere all'entrata, risultato di decisioni adottate in concreto dai diversi operatori, mentre, sul piano metodologico, dai vari modelli descrittivi l'analisi si apre alla teoria dei giochi nelle due fattispecie di quelli cooperativi e di quelli non cooperativi (v. Kreps, 1990). Nell'intento di meglio aderire alla problematica reale, tali schemi interpretativi tengono conto dell'insufficiente o mancata disponibilità di informazioni, abbandonando le analisi sotto certezza e aprendosi a schemi probabilistici o a modelli bayesiani.

Nella fluidità e complessità di questo contesto, partendo dalla differenziazione del prodotto, è venuto precisandosi il concetto di concorrenza spaziale: gruppi di operatori, collegati tra di loro e caratterizzati dalla presenza omogeneizzante in determinate aree, sviluppano rapporti di tipo competitivo. Lo stesso criterio di identificazione delle attività produttive, fondato sul concetto di settore, evolve nell'accettazione di quello di 'filiera', in cui si identifica la rete di rapporti che collegano, in un intreccio di offerta e trasformazione produttivo-distributiva, gli operatori presenti in un contesto determinato. Quando si tenga conto di questi aspetti si può capire come la presa di coscienza dell'imperfezione dei mercati identifichi un'evoluzione degli studi di economia industriale verso una problematica assai ampia, l'analisi della quale ha generato un'integrazione e un affinamento dei risultati utilizzabili.

4. Contributi dell'analisi empirica: nuove linee di studio

Su queste basi s'è sviluppato un approccio di studio strettamente fondato sulle informazioni concretamente acquisibili sulle imprese e sui settori nei quali esse operano. Si è largamente attinto ai dati economico-finanziari, trattandoli mediante metodi econometrici: le sintesi ottenute per questa via hanno consentito la proposta di generalizzazioni, il cui valore sta nella loro forte connessione con la realtà che, in un certo grado, le ha suggerite. In questo senso la linea seguita dall'analisi rivela caratteri induttivi che capovolgono la precedente impostazione, tendente a definire a priori uno schema interpretativo nel quale la realtà osservabile doveva forzatamente rientrare. Il beneficio più rilevante di quest'impostazione consiste nella notevole apertura problematica, alla base della quale sta la rilevanza implicitamente assunta dal fattore discrezionale che muove le scelte delle singole imprese nell'ambito del settore industriale. Si delinea in merito il possibile conflitto sulla rilevanza da attribuire all'andamento economico di un'intera industria o di un mercato, con particolare riferimento alle reazioni ai cambiamenti, colte a livello aggregato o a quello delle specifiche unità produttive. L'interrogativo tende, in ultima analisi, a mettere in luce se sia l'ambiente a determinare le condotte delle imprese o viceversa: domanda non inutile, destinata com'è a sfociare, anche sul piano della politica industriale, nel coordinamento tra la pianificazione individuale e le dinamiche e gli obiettivi di natura generale.
È dato cogliere a questo proposito l'interconnessione tra il precisarsi dell'economia industriale e i contributi più significativi dell'analisi economica dell'impresa. Forte

importanza assume l'esplicitazione della gamma degli obiettivi di breve, medio e lungo termine che animano l'azione imprenditoriale, a sua volta risultante quale momento di sintesi tra le ottiche di quanti forniscono i capitali e quanti operano nell'ambito del mandato e delle competenze manageriali. La concezione dell'impresa neocapitalistica di A. A. Berle e G. C. Means (v., 1932) offre, in questo contesto, la possibilità di analizzare e di approfondire la teoria comportamentistica (v. Cyert e March, 1963) sulla quale si innesta la teoria del capitalismo manageriale, che ha in W. J. Baumol (v., 1959) e O. Williamson (v., 1964) gli antesignani e in R. Marris (v., 1964) il più completo modellizzatore. Merito rilevante di questi studi è stato quello di cogliere nella realtà il frantumarsi della figura monolitica dell'imprenditore in una molteplicità di soggetti, impegnati in funzioni diverse e portatori di obiettivi differenti e, potenzialmente o concretamente, in conflitto tra di loro. In questo contesto l'obiettivo dello sviluppo, misurato da una varietà di parametri individuabili nella concretezza dell'impresa, assume il carattere di una sintesi tra le molteplici finalità manageriali e il raggiungimento di un profitto in grado di soddisfare le attese dei portatori di capitale. I due orientamenti trovano integrazione nella modellistica del capitalismo manageriale tramite la messa a fuoco della strategia della scalata (take-over), resa possibile dall'abbassamento del valore delle azioni dell'impresa, causato dal calo del livello dei profitti (e dei dividendi) provocato dal graduale prevalere dell'orientamento verso la crescita.
Lo sviluppo, inteso anche in senso qualitativo, diviene in questa linea di studio il potenziale elemento di passaggio dall'analisi delle problematiche a livello microeconomico a quella a livello macroeconomico: potrebbe essere l'insieme delle scelte delle imprese, così orientate, a determinare l'andamento delle grandezze rilevabili a livello di ramo produttivo.
L'osservazione non è di breve momento in quanto va a incidere sulla proponibilità dello schema logico-classificatorio proposto da F. M. Scherer (v., 1970) e assunto a paradigma per l'analisi empirica dei settori industriali: l'assunto strutturalista secondo il quale la struttura dei mercati determinerebbe le condotte delle imprese e queste infine condizionerebbero i risultati appare infatti perfettamente reversibile. A determinare tale esito è, in ultima analisi, l'accettazione della discrezionalità che di fatto informa le condotte. In realtà il raccordo tra momento microeconomico e momento macroeconomico rimane in larga parte ancora da definire. Numerosi studi traggono da queste finalità interessanti spunti per proporre nuove linee di approfondimento: possono essere segnalati quelli impegnati sul versante degli effetti di trascinamento che l'evoluzione delle variabili aggregate avrebbe sull'andamento delle

realtà individuali, o quelli centrati sull'effetto propulsivo e operante trasversalmente del progresso tecnologico. Nel fervore delle proposte, della molteplicità degli obiettivi enunciati e delle variabili sotto osservazione è reale il pericolo di smarrirsi sì da non riuscire a ricondurli a unità. Importa in ogni caso sottolineare come gli studi centrati su dati empirici abbiano segnato un avanzamento rispetto all'approfondimento dell'ipotesi della concorrenza monopolistica e abbiano avuto il merito non trascurabile di mettere in rilievo il ruolo dell'impresa nelle dinamiche dei mercati.

5. Le imprese come istituzioni

In realtà lo studio dell'impresa interferisce con quello del sistema al punto che l'economia dell'impresa e l'economia industriale rappresentano due momenti d'uno stesso campo di ricerca. L'evoluzione è tanto più evidente ai tempi attuali, dal momento che l'istituzione impresa tende non tanto a essere un soggetto che concorre a formare il mercato quanto a sovrapporsi a esso e, al limite, a incorporarlo. Con l'affermarsi delle strutture immanenti e capillari dei gruppi, emergono nuclei costituiti da un fitto intreccio di transazioni, sostitutivo del mercato; la stessa interazione tra domanda e offerta (sia dei fattori produttivi che della produzione) tende a essere trasferita all'interno delle istituzioni.Sull'argomento, come bene sintetizza S. Zamagni (v., 1987[2]), si sono espressi diversi autori volti a individuare le motivazioni della tendenza. L'idea originale proviene da R. H. Coase (v., 1937), il quale mette l'accento sui costi d'uso del mercato, ossia sugli oneri che si devono sopportare per operare direttamente su di esso, riconducibili essenzialmente a imperfezioni e ad asimmetrie informative. Tali costi possono rendere preferibile per l'impresa evitare le transazioni dirette sui mercati, tramite l'internalizzazione delle stesse. È quanto avviene in concreto quando ad esempio, mediante joint ventures, investimenti diretti o forme diverse di accordi contrattuali, si formano gruppi, integrati verticalmente, all'interno dei quali fornitori e utilizzatori sono coordinati da un programma centralizzato. K. J. Arrow (v., 1974) sviluppa l'intuizione di Coase, centrata sul ruolo dell'informazione: questa può cioè essere meglio dominata e utilizzata all'interno di un'istituzione, per quanto grande, piuttosto che in un ampio contesto qual è un mercato. L'enorme molteplicità dei rapporti che su quest'ultimo prendono vita, con i relativi costi, può essere di molto ridotta nell'ambito di un'organizzazione al cui interno siano ben precisi le linee di potere e gli schemi di collegamento.

A. Alchian e H. Demsetz (v., 1972), guardando all'impresa-organizzazione secondo uno schema neoclassico, introducono il concetto di sinergia che caratterizza la produzione di squadra, fondata sull'apporto coordinato di una pluralità di fattori: l'istituzione impresa, con i suoi managers e il necessario apparato organizzativo, sarebbe dunque in grado, meglio del mercato, di controllare il contributo dei singoli fattori e di attribuire loro la quota di spettanza in termini di produzione.
In un'ottica più concreta, le linee portanti della tendenza alla sovrapposizione dell'organizzazione imprenditoriale al mercato sono individuabili nei cambiamenti (in senso riduttivo oppure espansivo) dei gradi di integrazione verticale e orizzontale, con la conseguente formazione di unità produttive complesse.
Nella stessa direzione si muovono le proposte del progresso tecnologico: esse hanno offerto negli anni recenti e tuttora offrono l'occasione per introdurre nei processi produttivi gradi più avanzati di meccanizzazione dai quali deriva un irrigidimento delle strutture di costo. Ne possono discendere rinnovate occasioni di economie di scala, acquisibili con il raggiungimento di elevati livelli di fatturato tramite il collocamento di una produzione diversificata. A sua volta, questa è resa possibile dalle innovazioni di prodotto, come risultato combinato dell'attività di ricerca e sviluppo e della flessibilità dell'offerta, consentita da una meccanizzazione sempre più intrecciata con l'elettronica. Nasce l'esigenza di acquisire, in tempi brevi, competenze tecniche, sbocchi di mercato, risorse umane e finanziarie; un risultato possibile tramite processi di tipo aggregativo, dai quali possano derivare sistemi d'impresa integrati, organizzativamente complessi, o forme di collegamento più blande nelle quali le imprese, pur connesse tra di loro, non perdono la loro individualità (joint ventures o interlocking directorates). Sottese a quest'evoluzione sono le potenzialità nascenti dalla mobilità e dalla conseguente possibilità di allocazione ottimale delle risorse finanziarie disponibili, distolte da destinazioni prossime alla maturità o meno promettenti e indirizzate verso quelle suscettibili di maggiore sviluppo. Resta in ogni caso il problema, certo non insignificante, di ricavare i propri spazi vitali rispetto ad altre unità produttive, già presenti sui mercati, e di definire e sostenere i 'valori' dei beni e dei servizi offerti; la risoluzione di questo problema concorre a determinare i già ricordati 'costi di transazione', per contenere i quali le nuove realtà imprenditoriali tendono a inglobare al loro interno le relazioni e i rapporti conflittuali altrimenti presenti sui mercati.Si intuisce da queste considerazioni come la sfida a interpretare correttamente una realtà produttiva complessa, nella cui formazione è

dato cogliere lo sfumare dell'unità di riferimento tecnica (lo stabilimento) nell'unità economica (l'impresa) e, finalmente, nella realtà finanziaria del gruppo, finisca con il dare rilievo a una chiave di lettura di tipo organizzativo-sistematico. Non per nulla nella 'nuova' economia industriale acquista sempre maggiore importanza la 'teoria dell'organizzazione', quasi a riproporre l'originaria espressione industrial organization.
Emerge il ruolo dei comportamenti delle singole imprese nel determinare il proprio cambiamento e, con questo, il mutamento del sistema che le comprende. Anche sotto questo aspetto si assiste a un ridimensionamento dello schema interpretativo legato alla successione logica: strutture, condotte, risultati. Nell'impostazione strutturalista infatti era implicita la conclusione che, essendo le strutture a determinare le condotte e queste i risultati, bastavano le prime a determinare gli ultimi.Ciò equivaleva a minimizzare il peso dei comportamenti nell'evoluzione dell'apparato produttivo. L'approfondimento delle linee lungo le quali il cambiamento avviene sta invece rovesciando tale convinzione e mette decisamente in rilievo l'importanza della connessione tra lo sviluppo delle unità produttive e lo sviluppo del sistema complessivo quale naturale evoluzione del campo di studio dell'economia industriale.

6. Lettura evoluzionista del mutamento economico

La lettura economica dell'impresa con le sue problematiche e dell'industria della quale essa fa parte, tende a dividersi in due momenti di uno stesso corpo di studi. Quanto più l'impresa diviene complessa, tanto più si allontana dalla concezione che ne faceva un soggetto minimale, strutturalmente orientato alla massimizzazione del profitto nel breve termine e funzionale al raggiungimento dell'equilibrio generale. Al suo interno emergono ruoli diversi, potenzialmente decisori, con le loro specifiche aree di responsabilità e relative sfere di autonomia. L'essere e l'evolvere dell'impresa sono la risultante, non senza scontro e tensioni, di decisioni diverse. Il profitto cessa di essere finalità esclusiva per lasciare spazio a una gamma di obiettivi. Parallelamente, scelte e comportamenti non risultano più finalizzati all'ottimizzazione degli obiettivi, bensì alla ricerca di soluzioni ritenute soddisfacenti in rapporto al raggiungimento e mantenimento degli stessi.
Agli equilibri si sostituiscono momenti di quiete alternati a fasi di turbolenza. Le situazioni 'soddisfacenti' non sono peraltro destinate a durare, intaccate come sono dai mutamenti nello scenario, dalle proposte di cambiamento che ne derivano, dall'emergere di nuove finalità. La

ricerca di aderenza tra obiettivi e risultati rappresenta il criterio animatore delle scelte e della gestione dell'impresa: il continuo verificarsi di mutamenti tecnologici e istituzionali è di stimolo al maturare di nuove strutture produttive e organizzative ed è all'origine del succedersi continuo di routines e fasi di ricerca che, mettendo in crisi gli stadi di sviluppo raggiunti, tendono a orientare l'impresa verso nuovi traguardi.
Questa interpretazione del cambiamento economico-strutturale in chiave 'evoluzionistica', nella quale è trasparente il contributo di H. A. Simon (v., 1955) e dei più recenti lavori di R. Nelson e S. Winter (v., 1982), è fondata sull'ipotesi secondo cui lo spostamento dall'una all'altra routine è sollecitato dal conseguimento di risultati insoddisfacenti e, contemporaneamente, dall'aspettativa di poter reagire ad essi tramite l'adozione di nuove tecniche, la promozione di nuovi prodotti, l'uscita da mercati in fase di maturità. Alla base di un simile funzionamento sta l'assoluta flessibilità del modo di concepire l'attività di impresa e delle strutture manageriali e materiali attraverso le quali essa si esprime. La routine innovativa si salda con la programmazione, intesa a fissare gli obiettivi, con il controllo dei risultati conseguiti e con la rilevazione degli scostamenti destinati a provocare nuove occasioni di cambiamento.
L'integrazione tra singole unità produttive e sistema economico si concreta nelle ricadute delle scelte innovative sulla dinamica delle macrograndezze nelle quali le variabili tipiche dell'impresa trovano aggregazione. Teoria dell'organizzazione dell'impresa e sociologia della conoscenza divengono campi conoscitivi rilevanti, nei confronti dei quali la nuova economia industriale si riconosce debitrice, in un approccio in cui è parimenti evidente la matrice schumpeteriana. In questa impostazione emerge un modello d'impresa fondato sul concetto di impresa-organizzazione, rappresentabile appunto attraverso routines e orientata a seguire un comportamento soddisfacente e non massimizzante. Nello stabilire gli obiettivi ai quali tendere e le scelte da adottare ha largo spazio l'esperienza passata, sicché ogni impresa spicca nella propria individualità gestionale e si differenzia dalle rimanenti. Nella concezione di Nelson e Winter il tendere verso nuove routines gioca pertanto un ruolo cruciale: il mancato raggiungimento degli obiettivi stabiliti determina l'inizio del processo di ricerca che terminerà nel momento in cui le imprese avranno adottato nuove linee d'azione.
I cambiamenti che si producono nel sistema produttivo sono dunque la risultante del processo d'integrazione delle eterogenee scelte decisionali compiute dalle imprese. D'altra parte il loro adattamento all'alterarsi delle condizioni esterne si pone come condizione necessaria alla loro

sopravvivenza vitale nell'industria. Ne offrono un esempio gli effetti delle variazioni nei prezzi relativi: la reazione immediata dell'impresa sarà quella di mutare il proprio assetto utilizzando lo stock di capitale e il know-how esistenti nel sistema, non discostandosi in ciò dalla lettura che ne farebbe una modellizzazione di stampo neoclassico. A questo stadio succede tuttavia la ricerca attiva di nuove routines, in specie se non vengono raggiunti gli obiettivi prefissati, nell'intento di superare le limitazioni poste dalle strutture esistenti agli stimoli impressi dalle variazioni dei prezzi. Ne deriva l'avvio di un processo di selezione, interno all'industria, sollecitato dall'interazione tra cambiamento dei prezzi, stock di capitale (e quindi dimensione), risultati economici.

È tuttavia necessario sottolineare come questa concezione venga approfondita e modificata in una lettura moderna della fenomenologia evolutiva. L'attenzione cade in particolar modo sui fattori in grado di influire sull'intensità e sulla velocità dell'innovazione, nonché sulla direzione che essa assume. Ne deriva la messa a fuoco di due diversi tipi di modelli interpretativi, fondati rispettivamente sul ruolo trainante della domanda (demand pull) o sull'autonomia del progresso tecnologico (technology push). Sull'argomento è gradualmente emerso un organico corpo di studi maturati in ambito internazionale e soprattutto presso la SPRU (Science Policy Research Unit dell'Università del Sussex) a opera di C. Freeman (v., 1982[2]) e K. Pavitt (v., 1984), al quale ha significativamente contribuito G. Dosi (v., 1984). A questi si devono la precisazione e la diffusione di nuovi strumenti interpretativi mediati dalla filosofia della scienza, quali il 'paradigma tecnologico' come insieme sistematico delle opportunità offerte all'innovazione, e la 'traiettoria tecnologica' quale linea evolutiva di medio termine lungo la quale esse tendono a collocarsi. In armonia con la letteratura sull'innovazione sono inoltre definiti e utilizzati i concetti di 'cumulabilità del progresso tecnico', come forma di accumulazione immateriale delle conoscenze, di 'opportunità tecnologica' quale offerta concreta dell'occasione per innovare, di 'appropriabilità dell'innovazione', che teorizza ed esprime, per il contesto oligopolistico, la mancata diffusione del cambiamento. Tali strumenti hanno gradualmente costituito un patrimonio conoscitivo per un'analisi che è stata particolarmente attenta e profonda in F. Momigliano (v., 1975; v. Momigliano e Dosi, 1983); in particolare, egli ha posto nella massima evidenza la necessità di individuare le specificità industriali tramite il ricorso a modelli fondati sull'azione differenziata sia delle opportunità tecnologiche, sia del diverso grado di appropriabilità.

Nella complessa intelaiatura di azioni e reazioni tra progresso tecnico e variabili esplicative si enuclea il filone forse più rilevante degli studi di Nelson e Winter, che analizza il collegamento tra progresso tecnico e forme di mercato, e si rifà al pensiero di J. A. Schumpeter (v., 1942) allorché analizza la grande impresa e le economie di scala del processo innovativo.

Secondo questo approccio, sulla struttura di un'industria influiscono numerosi fattori esogeni. In settori nei quali non siano sensibili le prospettive di economie di scala, ma al tempo stesso sia elevato il tasso di progresso tecnologico, il grado di concentrazione è codeterminato dalla politica aggressiva di investimento, dal tasso di crescita della produttività latente, dalla difficoltà del processo imitativo e dalla variabilità dei risultati dell'attività innovativa da parte dell'impresa. In ultima analisi, quindi, in assenza di rilevanti economie di scala, la struttura di un'industria a elevato grado di tecnologia è influenzata da tre fattori esogeni: la curva della domanda, le condizioni di appropriabilità e di convenienza dell'innovazione.

A un livello di maggiore approfondimento e ancora nel solco schumpeteriano, il filone di analisi di Nelson e Winter offre materia di riflessione sul trade-off tra efficienza distributiva e progressività tecnologica. In una concezione statica, una situazione di tipo oligopolistico-monopolistico non presenta caratteri ottimali, dal momento che determina la sottrazione di parte della rendita del consumatore e, di conseguenza, una perdita di efficienza nella distribuzione della ricchezza. Sotto il profilo dinamico, viceversa, un certo grado di potere di mercato appare funzionale a una desiderabile progressività tecnologica, sia per l'entità, in valore assoluto, delle risorse necessarie all'espletazione dell'attività di ricerca e di sviluppo, sia per la possibilità di sfruttare commercialmente un'innovazione. Ne deriva la possibilità di trarre due conclusioni in tema di relazione tra innovazione e forme di mercato. La prima, di netta derivazione schumpeteriana, riguarda la biunivocità della relazione: una data forma di mercato non è solo condizione necessaria per il successo di una politica orientata all'innovazione, ma altresì conseguenza di innovazioni riuscite. La seconda sottolinea la distinzione tra potere di mercato e dimensione: il potere determina una barriera protettrice intorno all'impresa innovativa contro le politiche aggressive delle imprese imitatrici; la dimensione consente viceversa agli imitatori di fruire dei vantaggi delle innovazioni altrui, rese possibili da un rilevante ammontare della produzione.

Si comprende da questo insieme di considerazioni come la nuova economia industriale si fondi su un'ottica di tipo dinamico (evolutivo) e assuma al centro della propria analisi la

singola unità produttiva. Mentre, da un lato, vengono enfatizzati la diversità delle imprese e il ruolo essenziale dell'interazione tra di esse nel determinare l'essenza e la struttura di un'industria, dall'altro l'impresa stessa è posta quale punto di partenza per l'analisi del progresso tecnologico e delle ricadute del medesimo sulla dinamica del settore. Importa infine considerare come nell'ottica di Nelson e Winter possano delinearsi dei 'fallimenti evolutivi' nel garantire un tasso predeterminato di sviluppo del progresso tecnologico in un comparto industriale, sia in caso di concorrenza perfetta sia in caso di monopolio. In una situazione di concorrenza perfetta e in condizioni di bassa appropriabilità, il fallimento deriva dall'impossibilità di impedire il processo di imitazione; in caso di elevata appropriabilità, invece, in cui tutte le unità produttive tendono a cercare le stesse cose, il fallimento trova origine nell'inevitabile duplicazione degli sforzi.In una situazione di monopolio e in condizioni di incertezza il fallimento trova origine nella composizione del portafoglio progetti che, essendo determinato dalle scelte di un singolo operatore, può non essere soddisfacente per l'economia nel suo complesso.

7. Nuove forme di concorrenza nei mercati accessibili

La linea di pensiero secondo la quale sono le scelte a livello individuale, con i comportamenti che ne derivano, a determinare le strutture di mercato, si collega all'approccio teorico dei mercati accessibili, nel quale trova concretezza un altro filone della nuova economia industriale. Esso rifiuta di considerare la 'struttura industriale' come esogenamente data, ma individua in essa la risultante di un processo di determinazione di natura endogena tale da consentire l'ottenimento simultaneo della distribuzione dimensionale delle imprese, il livello dei prezzi e delle quantità. Appartenente all'ambito delle analisi di tipo statico, la teoria dei mercati accessibili mira a stabilire le condizioni in base alle quali si possa pervenire a configurazioni di equilibrio di lungo periodo. Essa si concretizza, nell'ottica dei suoi proponenti, W. J. Baumol, J. C. Panzar e R. D. Willig (v., 1982), in una rivisitazione delle condizioni di forte concorrenzialità, sia pure in termini moderni e quindi lontani dalla concezione cournotiana fondata sul numero delle imprese esistenti sul mercato. Tramite la definizione di opportune condizioni d'entrata e di uscita possono verificarsi esiti di tipo concorrenziale pur in presenza di un numero limitato di operatori. Secondo Baumol (v., 1982) "un mercato è accessibile quando l'entrata è completamente senza costi". Con 'libertà

di entrata' si intende l'assenza di vincoli normativi (monopolio istituzionale) e la possibilità di accesso, per tutti gli entranti potenziali, alla medesima tecnologia della quale dispongono le imprese già presenti. Con 'uscita senza costi' si intende una situazione caratterizzata da una perfetta mobilità del capitale fisso utilizzato nel processo produttivo sì da rendere possibile il disfarsene senza incorrere in oneri di alcun genere. Un esito di questo tipo può verificarsi allorché esista un mercato secondario sul quale il valore dei beni capitali possa essere pienamente realizzato. Se sono valide queste premesse, ovvero, se pur in presenza di costi fissi con relative economie di dimensione, non vi sono costi non recuperabili a fronte dell'eventuale decisione di uscire dal settore ('costi sommersi' - sunk costs), si delinea la possibilità di una concorrenza del tipo 'mordi e fuggi' (hit and run competition). Sulla scena si presentano entranti intesi a realizzare profitti puri temporanei sul mercato (quando il prezzo superi il costo medio) e a uscire dallo stesso prima di subire le reazioni degli operatori già presenti senza sopportare perdite a causa dei costi sommersi collegabili al realizzo dei beni strumentali utilizzati. A complemento si può ancora aggiungere, come ulteriore requisito per definire una perfetta accessibilità, la supposizione che i tempi di reazione delle imprese presenti sui mercati siano maggiori di quelli degli entranti potenziali.

In presenza quindi di perfetta libertà di entrata, di costi fissi ma non di costi sommersi, di tempi di reazione differenziati, nessun produttore operante può praticare un prezzo superiore al costo medio (e quindi conseguire profitti puri). Del pari non può, compatibilmente con la domanda, produrre una quantità alla quale corrisponda un costo medio superiore al valore minimo che la sua funzione può assumere senza provocare l'ingresso di altri operatori (con conseguente hit and run competition). Nel caso in cui il settore industriale risulti al riparo da tale tipo di concorrenza ci si trova in presenza di una 'configurazione industriale sostenibile', per qualche aspetto riecheggiante le situazioni neoclassiche di equilibrio. Essa risulta dal concorso di due condizioni: una produzione complessiva eguale alla domanda globale, senza perdite per alcun operatore (nel qual caso si orienterebbe a uscire); entranti potenziali che non reputano conveniente l'ingresso nel mercato sulla base dei prezzi praticati dalle imprese esistenti (ognuna delle quali produce il volume corrispondente al costo medio minimo e non può praticare un prezzo superiore a quest'ultimo).

La mancanza di una delle due condizioni determina la 'non sostenibilità' di una struttura, ciò che provocherebbe ingressi del tipo 'mordi e fuggi' destinati a mutare l'eventuale configurazione finale del settore.

La lettura offerta dalla teoria dei mercati accessibili è svincolata dal numero più o meno elevato degli operatori presenti nel settore; essa è applicabile anzi all'intera gamma delle strutture industriali, dal monopolio puro alla concorrenza perfetta, incluso l'oligopolio. Ciascuna di queste configurazioni può assumere carattere di accessibilità, tantoché si può affermare che un mercato concorrenziale è sicuramente accessibile mentre non si può sostenere che mercati imperfetti non lo siano affatto.
Nell'iter di formazione delle strutture industriali si possono individuare due fasi: nella prima si determina quali processi produttivi risultano essere i più efficienti nella produzione di un determinato vettore di prodotti nell'ambito di un comparto; nella seconda viene determinata la struttura industriale conseguente e 'sostenibile' nel lungo periodo, tenuto conto delle pressioni esercitate dai meccanismi di mercato. La configurazione finale di uno specifico settore risulta pertanto da un processo di natura endogena fondato: sull'assenza di inefficienze nell'organizzazione delle imprese presenti in un'industria, la produzione delle quali tende a collocarsi in corrispondenza al minimo costo medio lungo l'ipotetica curva a U dei costi medi di lungo periodo; sul gioco combinato degli sforzi a entrare nel settore e a rapidamente uscirne da parte di nuove unità; sul far venire meno le condizioni affinché esso trovi concretezza. La struttura ottima di un'industria dipende dal vettore delle sue produzioni, ma quest'ultimo è a sua volta collegato ai prezzi stabiliti dalle imprese. Questi discendono tuttavia dalla struttura dell'industria, per cui si può concludere affermando che il comportamento in materia di prezzi e la struttura dell'industria in ultima analisi si codeterminano per effetto di scelte di tipo endogeno. Nell'ipotesi di perfetta accessibilità, la struttura più efficiente, definita a livello di tecnologia dei processi produttivi, finisce per coincidere con quella selezionata dai meccanismi di mercato, stante l'impossibilità di allontanarsi dall'eguaglianza tra prezzo e costo unitario medio minimo, determinata dal timore del manifestarsi della concorrenza 'mordi e fuggi'. Al contrario di quanto stabilito nello schema tradizionale di Cournot, la dimensione e il numero delle imprese presenti sul mercato dipendono dalle condizioni di concorrenza esistenti.
Ne discende l'entrata in crisi del collegamento tra tecnologie (economie di scala) ed efficienza degli equilibri settoriali, che per lungo tempo ha ispirato gli interventi di regolamentazione. L'azione pubblica dovrebbe dunque indirizzarsi a rendere possibile l'operare delle condizioni di accessibilità, ossia, in primo luogo, la libertà di entrata e di uscita, potenziandola al massimo. Questa può impedire infatti inefficienze e distorsioni

nell'allocazione, destinate a tradursi in profitti puri positivi. Quanto più si riesce a rendere accessibile un mercato (tramite, ad esempio, la rimozione di barriere normative all'entrata, come nel caso delle telecomunicazioni o del trasporto aereo in taluni paesi), tanto più si può aspirare a creare equilibri potenzialmente concorrenti (con assenza di extraprofitti) seppure in presenza di un esiguo numero di imprese operanti. L'intervento potrebbe essere limitato ai casi di non sostenibilità della configurazione. Nell'intento di evitare l'accessibilità dei mercati le imprese in essi presenti non possono fare altro che comportarsi 'virtuosamente', ossia offrire ai consumatori gli stessi benefici altrimenti ottenibili in una soluzione di perfetta concorrenza.

Malgrado queste conclusioni, il cui contenuto può presentare aspetti operativi, l'approccio dei mercati accessibili mantiene accentuati caratteri di astrazione. Baumol precisa infatti che esso presenta lo stesso grado di astrazione della teoria neoclassica della concorrenza perfetta. In effetti tutto lo schema si regge sull'assenza di costi non recuperabili e sui differenziali nei tempi di reazione dei nuovi entranti rispetto alle imprese già presenti e operanti. L'evidenza empirica, per contro, mostra che in molti settori la produzione richiede un tipo di capitale a bassa o nulla reversibilità (come tale non facilmente smobilizzabile), e che gli stessi consumatori reagiscono a mutamenti di prezzo dopo un certo tempo (per cui nascerebbero problemi nel dar vita a interventi del tipo 'mordi e fuggi'). Possono inoltre esservi costi sommersi non collegati alla tecnologia bensì alle politiche delle imprese; tali ad esempio quelli derivanti da scelte di breve periodo tese a formare situazioni di concorrenza non relativa ai prezzi (non-price competition): costi di distribuzione, promozione, ricerca e sviluppo. Deve infine essere osservato che l'ipotesi di uscite e ingressi rapidi e privi di conseguenze economiche implicherebbe rendimenti di scala costanti e non rendimenti crescenti, come di fatto richiede la teoria dei mercati accessibili, fondata com'è sull'esistenza di livelli minimi nei costi medi. Se ne può inferire che fare riferimento a questa teoria per una descrizione puntuale ed efficace della realtà equivarrebbe a commettere un errore: essa può costituire al più un punto di riferimento per lo sviluppo di schemi interpretativi più flessibili e applicabili di quelli di derivazione neoclassica, per capire e descrivere l'organizzazione della struttura di un settore industriale.

8. Ruolo dell'innovazione nel processo di concorrenza dinamica

A rompere la sostenibilità di posizioni di equilibrio nel quadro dei mercati accessibili è il processo innovativo, là dove questo determina un cambiamento delle curve dei costi unitari medi, e in particolare un loro abbassamento, o la possibilità di vendere i prodotti a prezzi più remunerativi. In realtà la considerazione della sostenibilità nei mercati accessibili non annulla le considerazioni riportate in precedenza sulla teoria evolutiva dell'industria fondata sul manifestarsi delle routines innovative.
La sostenibilità tende a perpetuarsi nelle condizioni di coincidenza tra obiettivi e risultati; allorché tuttavia il cambiamento di scenario induce sia a esprimere nuove finalità, sia a modificare le vecchie, e/o quando diviene evidente che le azioni compiute non consentono di raggiungere i traguardi prefissati, le routines innovative tendono a riportare la mancata convergenza. Il manifestarsi dell'innovazione nella riproposta della schumpeteriana 'distruzione creatrice' diviene pertanto elemento essenziale del gioco concorrenziale e contribuisce anche sotto questo profilo a definire i contenuti della nuova economia industriale. Si definisce in questi termini la concorrenza dinamica, per effetto della quale si possono raggiungere risultati di grande rilievo in termini di abbassamento dei costi e, in ultima analisi, del livello dei prezzi.Legata com'è all'entità assoluta delle spese di ricerca, alla disponibilità di esperienza, alle potenzialità di programmazione e controllo della spesa, all'esistenza dei necessari sbocchi di mercato, la possibilità di innovare (e quindi la concorrenza dinamica) rimane largamente ancorata alla grande dimensione, decisiva per assicurarne i frutti. Essa quindi trova occasioni di fertili applicazioni e terreno ideale per esprimersi nelle strutture di tipo oligopolistico. Ne nasce una profonda revisione dell'atteggiamento verso questa struttura di mercato, dalla quale possono derivare esiti più favorevoli per gli operatori e per i consumatori di quanto possa verificarsi nei regimi di concorrenza perfetta. In questa forma di mercato l'azione di limatura sui livelli di prezzo troverebbe infatti un limite fisiologico nei più elevati costi unitari medi minimi sui quali mancherebbero di esercitarsi i benefici dell'innovazione tecnologica, impedita dai bassi livelli dimensionali.
Nella concezione presentata, impresa e sistema economico si qualificano sempre più come parti interagenti di una fenomenologia in evoluzione, per capire la quale risulta indispensabile il riferimento al momento storico e ai cambiamenti che lo caratterizzano. Si riconferma in questa chiave il ruolo degli studi di 'economia industriale' volti a individuare e descrivere il meccanismo attraverso il quale il sistema produttivo si autodetermina per effetto

delle scelte degli operatori in esso presenti. Da un approccio di tipo statico si passa a una lettura della problematica in termini di sviluppo; l'oggetto di studio si precisa nel divenire dell'apparato produttivo considerato nel suo insieme, delle unità che lo compongono, delle relazioni che reciprocamente le collegano.
La materia che compone l'economia industriale trova pertanto un preciso elemento costitutivo nella misurazione e nella qualificazione del processo di sviluppo delle imprese tramite l'analisi della dinamica di precise variabili elementari, da integrare tuttavia con altri profili temporali afferenti, ad esempio, la produttività del lavoro o il processo di accumulazione in capitale fisso. Fondamentali divengono del pari le analisi delle operazioni di crescita esterna come quelle relative a incorporazioni, scorpori, variazioni nel collocamento entro specifici gruppi di controllo, cambiamenti del grado di integrazione verticale all'interno della 'filiera' di appartenenza. In una progressiva generalizzazione, il processo di studio tende a spostarsi dalla vicenda delle singole imprese all'analisi dei processi di trasformazione in senso lato, quali avvengono nel corso del tempo.In ultima istanza l'obiettivo di un simile indirizzo di analisi del sistema industriale trova la propria consistenza e finalità nella scelta di contribuire a esprimere una teoria dello sviluppo dell'apparato produttivo.

9. Dall'economia alla politica industriale

Un adeguato approfondimento e una solida conoscenza dell'economia industriale dischiudono la possibilità di validi suggerimenti in termini di politica industriale intesa come controllo e indirizzo dell'economia in termini reali. Non si tratta infatti di contrastare o assecondare semplicemente la congiuntura, ma di aprirsi a una logica di interventi modificativi a carattere strutturale.In realtà gli spazi assegnabili alla politica industriale dipendono dal modo con il quale ci si pone di fronte all'operare dei meccanismi di mercato: in altri termini, se si accetta o si rigetta la loro azione. Ove si sia convinti che il loro operare si manifesta in un ambiente economico non distorto, si tende a concepire la politica industriale come un insieme di provvedimenti volti a facilitare o a rendere possibili processi di adattamento, in sé spontanei. La politica industriale ha allora l'obiettivo di assicurare le condizioni affinché il sistema dei prezzi possa emettere segnali corretti, sì da rendere possibili reazioni degli operatori altrettanto corrette. I contenuti riconoscibili per questa branca della politica economica sono individuabili soprattutto in interventi di tipo profondo, orientati a eliminare impedimenti e

ostacoli affinché le forze sane, operanti sui mercati, possano manifestarsi con pienezza. Essi tenderanno, ad esempio, ad assicurare le infrastrutture necessarie qualitativamente adatte, le forze di lavoro dotate di adeguata preparazione, l'eguale possibilità di accesso al credito e la sua valida distribuzione sul piano territoriale, l'operare di meccanismi fiscali coerente con la razionalità economica.

In quest'accezione riduttiva la politica industriale è vista come un insieme di scelte indirizzate a realizzare le condizioni del laissez faire: il loro operare selettivo tramite l'esprimersi delle forze concorrenziali è giudicato sufficiente ad assicurare lo scenario ottimale entro il quale la vita economica può svolgersi. Come conseguenza ne deriva il rifiuto di qualsiasi logica di tipo assistenziale, per lasciare il pieno rinnovamento delle strutture all'innovazione e al superamento del 'vecchio' tramite la sua pura e semplice estinzione; del pari è rifiutato ogni intervento di tipo corporativo.

A conclusioni radicalmente diverse si perviene ove si prenda coscienza delle imperfezioni del mercato che, come già si è visto, stanno ampiamente alla base del sorgere stesso dell'economia industriale. Ci si orienta in tal caso verso linee di intervento che, secondo una schematizzazione di A. Jacquemin (v., 1988), possono essere leggibili su di un piano prevalentemente nazionale e/o internazionale. Con riferimento al primo, assumono importanza le azioni indirizzate a far fronte alle carenze più vistose dei mercati in merito all'utilizzazione di rilevanti esternalità o al mancato riconoscimento di bene pubblico per determinati prodotti o servizi. Sotto altro profilo possono essere disegnati interventi intesi a sopperire alle incompletezze del sistema complessivo della domanda e dell'offerta per particolari presenze in comparti di natura strategica. Ne può conseguire l'individuazione del ruolo dei pubblici poteri nel favorire forme organizzative adatte a trasporre all'interno dei sistemi gli effetti delle grandi scelte tecnologiche e a promuovere la presenza di poli competitivi, o nel fornire sostegno all'attività di ricerca e sviluppo in settori ad alta tecnologia gravati da elevati costi fissi, tramite aiuti finanziari, programmi pubblici specifici, socializzazione dei rischi.

In questo senso la politica industriale tende a occupare gli spazi lasciati aperti dall'insufficienza dell'equilibrio spontaneo e mira a riportare condizioni di sviluppo equilibrato in presenza di mercati imperfetti. A integrazione è da prevedere la necessità di procedere in via disaggregata con l'adeguata considerazione delle specificità dei mercati e della loro interdipendenza.

Nel quadro realistico e rinnovato degli scambi nazionali e internazionali la politica industriale deve generare condizioni tali da favorire il dinamismo industriale. Meritano di essere menzionati in tale direzione gli interventi volti a riequilibrare i rapporti tra le imprese nazionali di uno specifico territorio (regione, paese, ecc.) o tra le imprese nazionali e le grandi multinazionali o, ancora, tra le multinazionali e i poteri pubblici dei paesi con i quali esse entrano in rapporto. Gli esempi al riguardo possono essere numerosi: fornire informazioni sulle partnerships possibili o alternative; utilizzare in via strategica la committenza; negare accordi tra Stati con paesi terzi; finanziare studi esplorativi relativi a processi e a nuove tecnologie; promuovere la diffusione delle innovazioni dalle sedi della ricerca universitaria alla produzione e alla commercializzazione; favorire il trasferimento tecnologico; facilitare la nascita di nuove imprese o la loro riorganizzazione e il loro ampliamento; incoraggiare operazioni congiunte tra industria, università e settore pubblico.
In un quadro interventistico di questo tipo il funzionamento delle economie nazionali e internazionali dipende da relazioni di tipo strategico, protagoniste delle quali sono le aziende e le autorità pubbliche che tendono a crearsi posizioni preferenziali e a condizionare azioni e reazioni di rivali reali e potenziali. Tra imprese e istituzioni si manifesta una serie reciproca di collegamenti che, con riferimento alla realtà italiana, possono identificarsi nelle forme di joint ventures tra gruppi pubblici e privati, nelle direttive emanate dal Ministero delle Partecipazioni Statali alle grandi finanziarie pubbliche o da queste alle imprese da esse coordinate. Rappresentano casi concreti di tali interrelazioni: le operazioni a monte dei processi produttivi pur nel mantenimento del clima concorrenziale a livello di produzione finale; le barriere create contro operatori in grado di penetrare sui mercati; il controllo della gamma dei prodotti e dei servizi succedanei; la modifica del potere di contrattazione a fronte di fornitori e clienti; il cambiamento dell'equilibrio tra le forze in campo con provvedimenti strategici e anticipo dei cambiamenti.
Nei casi in cui le sinergie tra imprese e istituzioni trovano modo di esprimersi con efficacia, chi riesce a goderne e a presentarsi sul mercato con anticipo rispetto a eventuali concorrenti potenziali, difficilmente vedrà intaccata la propria posizione egemonica. Quest'affermazione è tanto più vera quanto più i settori in cui l'interrelazione si verifica sono caratterizzati da una forte differenziazione della produzione, da importanti economie di scala, da processi fondati sull'apprendimento. A questo proposito si può richiamare l'esempio del progetto comune europeo Airbus, grazie al quale l'industria aeronautica europea s'è trovata in posizione di

effettivo e sensibile vantaggio.
Il ruolo della politica industriale non è tuttavia quello di generare sostituti alla libera iniziativa, né quello di limitarsi alla sola regolazione macroeconomica o di istituire una programmazione fortemente strutturata e direttiva, bensì di porre in essere una cornice strategica diversificata, tale da favorire il dinamismo dell'apparato industriale. Questo orientamento ha tanto più valore quanto più lo si riferisce all'evolvere delle economie occidentali nello scenario internazionale. I cambiamenti in atto pongono all'attenzione di chi osservi l'apparato produttivo industriale gli importanti trasferimenti di tecnologie, gli accordi in materia di concessione creditizia, l'importanza o quasi la necessità di un inserimento nel contesto internazionale: si delineano opportunità per l'emergere di rapporti di forza e di alleanza dai quali dipende in ultima analisi l'allargamento o il restringimento degli spazi strategici. Anche in questa prospettiva, il ruolo del potere pubblico può essere individuato nel concorrere a portare equilibrio tra produttori nazionali e internazionali e centri decisionali di altri paesi.
Gli strumenti a disposizione per raggiungere obiettivi di questo tipo possono essere diversi: dall'emanazione di opportune direttive alla destinazione di risorse finanziarie pubbliche, dall'uso appropriato delle commesse pubbliche all'esercizio di una programmazione coerente con il contesto economico e sociale in cui si deve calare. Rimane in ogni caso vero che, in quanto investita di problemi reali, la politica industriale tocca in profondità il funzionamento dei sistemi nelle loro componenti di tipo strutturale. Come tale incontra non di rado opposizioni e ostacoli alla sua attuazione, tanto da vedersi costretta a lasciare spazio a interventi di carattere prettamente monetario. Questo equivale a permettere che scelte di natura monetaria finiscano con l'incidere su problemi di ordine strutturale, il che equivale non tanto a essere assenti nel delineare e nell'attuare una politica industriale, quanto a perseguire, di fatto, quella peggiore.

Economia internazionale
di Giancarlo Gandolfo

sommario: 1. Introduzione. 2. Generalità sulla teoria pura del commercio internazionale. 3.

La teoria dei costi comparati. 4. Il modello neoclassico. 5. La teoria di Heckscher-Ohlin. a) Aspetti teorici. b) Studi empirici. 6. Ulteriori sviluppi. 7. Politiche commerciali. a) Introduzione. b) Dazi e altre misure. c) Il dibattito libero scambio-protezionismo e la teoria del second best. 8. Generalità sull'economia internazionale monetaria. a) Il mercato dei cambi esteri e i regimi di cambi. b) La bilancia dei pagamenti. 9. Gli schemi tradizionali. a) L'approccio delle elasticità. b) Il moltiplicatore in mercato aperto. c) L'approccio integrato. 10. Gli schemi moderni. a) Il modello Mundell-Fleming. b) L'approccio monetario alla bilancia dei pagamenti. c) Sviluppi recenti. 11. La determinazione del tasso di cambio. 12. Il sistema monetario internazionale e il sistema monetario europeo. a) Gli avvenimenti del secondo dopoguerra. b) Il sistema monetario europeo. 13. La liquidità internazionale. a) La domanda di riserve internazionali. b) Le xenovalute. □ Bibliografia.

1. Introduzione

Mentre molte discipline economiche 'specialistiche' sono nate e si sono sviluppate in tempi relativamente recenti dal corpo di base della teoria economica generale, è invece tradizione consolidata nella letteratura economica la presenza di una trattazione specifica della teoria delle transazioni economiche internazionali. Vari motivi possono essere addotti per spiegare l'esigenza di una siffatta trattazione specifica, ma i principali sono due.
Innanzitutto occorre rilevare che i fattori della produzione sono in genere meno mobili fra paesi che all'interno di un paese. Da questa osservazione hanno preso lo spunto gli schemi tradizionali per adottare l'ipotesi estrema di perfetta mobilità interna ed assoluta immobilità internazionale dei fattori, accompagnata dall'ipotesi di perfetta mobilità, sia interna sia internazionale (salvo misure restrittive), dei beni prodotti.Il secondo elemento preso in considerazione è stato il fatto che la semplice presenza di confini politici fra gli Stati dà luogo ad una serie di problematiche economiche che non sorgono nell'economia generale: si pensi ad esempio all'imposizione di dazi all'importazione e di altre misure restrittive, alla presenza di monete nazionali diverse con un valore reciproco variabile nel tempo, ecc.
L'importanza dell'economia internazionale va aumentando data la sempre crescente apertura verso l'estero dei singoli sistemi economici nazionali: basti pensare che oltre il 20% del reddito nazionale (come media mondiale) viene speso nell'acquisto di beni e servizi di provenienza estera. Occorre peraltro osservare che il fenomeno dell'interdipendenza

economica internazionale è un fenomeno relativamente recente dal punto di vista storico. Prima della rivoluzione industriale il commercio internazionale era costituito soprattutto da modeste quantità di beni di lusso (principalmente zucchero e spezie) e da schiavi. La rivoluzione industriale, che dall'Inghilterra si diffuse negli altri paesi europei ed extraeuropei, fu intimamente connessa con la crescita del commercio internazionale, che è stato spesso chiamato il 'motore' dello sviluppo economico del XIX secolo. Verso la metà di tale secolo, la Gran Bretagna esportava oltre il 20% del proprio prodotto nazionale e la sua quota delle esportazioni mondiali era di oltre il 40%. Ma non bisogna pensare che il commercio riguardasse soltanto le merci. Anche il capitale finanziario, sotto forma soprattutto di titoli a reddito fisso a lungo termine destinati a finanziare progetti di sviluppo della allora periferia (Stati Uniti, Russia, ecc.), fluiva fra i vari paesi: all'inizio di questo secolo la Gran Bretagna esportava all'estero circa la metà dei suoi risparmi. Il sistema aureo internazionale (gold standard), che si generalizzò intorno al 1870, facilitava gli scambi internazionali di merci e capitali. Questa specie di 'età dell'oro' ebbe fine con la prima guerra mondiale e i tentativi di ripristinare lo status quo ante vennero definitivamente abbandonati a causa della grande depressione del 1929. Seguì un periodo in cui, come disse Keynes, il commercio internazionale era diventato "un disperato espediente per mantenere l'occupazione interna forzando le vendite sui mercati esteri e restringendo gli acquisti all'estero": guerre commerciali, svalutazioni competitive, ecc. caratterizzano il periodo fra le due guerre mondiali. Dopo la seconda guerra mondiale, passati i primi anni di ricostruzione postbellica, il commercio internazionale riprese con vigore e diventò nuovamente il motore della crescita internazionale: fra il 1955 e il 1973 il volume del commercio mondiale si quadruplicò. A questa crescita contribuirono sia la dichiarazione di convertibilità delle principali valute, avvenuta nel 1958, sia gli altri eventi di cui diremo più diffusamente nel cap. 12. Se volessimo racchiudere in una formula semplificata (e quindi inevitabilmente imprecisa) le caratteristiche degli scambi internazionali attuali rispetto al passato, potremmo dire che gli scambi internazionali di capitali finanziari sono andati crescendo assai più rapidamente degli scambi di merci e servizi, e che tali scambi finanziari, un tempo accessori agli scambi di merci e servizi, hanno sempre più acquisito una valenza autonoma. Questo stato di cose si riflette anche negli schemi teorici, i quali - un tempo concentrati quasi esclusivamente sull'analisi e sulla spiegazione dei flussi commerciali - si sono trovati nella necessità di affrontare anche i problemi degli scambi finanziari internazionali, dando vita a una

specializzazione nella specializzazione: la cosiddetta 'economia internazionale monetaria', di cui diremo fra poco.
Come in ogni disciplina, anche nell'economia internazionale possiamo distinguere una parte teorica e una parte descrittiva. La parte teorica, a sua volta, viene distinta in teoria pura e teoria monetaria. Tutte queste distinzioni hanno soprattutto natura logica e pedagogica, dato che per comprendere le relazioni economiche internazionali nella realtà occorre sia la parte descrittiva sia la parte teorica, pura e monetaria.
La parte descrittiva si occupa della descrizione dei fenomeni relativi agli scambi internazionali quali si verificano nella realtà e nel contesto istituzionale in cui essi avvengono: andamento del commercio internazionale, vari accordi internazionali, organismi internazionali (Fondo Monetario Internazionale, Comunità Economica Europea, ecc.) e così via.
La parte teorica cerca di andare al di là dell'aspetto fenomenico e d'individuare principî generali e schemi logici che servano di guida per la comprensione dei fenomeni concreti (e quindi anche per eventuali interventi di politica economica). A tali fini essa si serve, come ogni teoria economica, di astrazioni e di modelli, spesso esposti in forma matematica.
Nell'ambito della parte teorica si suole distinguere, come dicevamo sopra, la teoria pura e la teoria monetaria, in ciascuna delle quali è possibile individuare sia un aspetto di economia positiva sia un aspetto di economia normativa. I due aspetti sono strettamente connessi nella nostra disciplina (anche se per motivi didattici verranno presentati separatamente), per cui - rifacendoci alle tradizionali distinzioni fra materie presenti da antica data negli ordinamenti universitari - l'economia internazionale comprende sia argomenti di 'economia politica' che argomenti di 'politica economica' e di 'scienza delle finanze'.
Netta è invece la distinzione fra teoria pura e teoria monetaria. La teoria pura (che ha natura essenzialmente microeconomica) si occupa delle cause, della struttura e del volume del commercio internazionale (quali beni cioè, e perché, sono importati e quali esportati da ciascun paese, e in quale ammontare); dei vantaggi del commercio internazionale e della loro ripartizione; della determinazione dei prezzi relativi dei beni in un'economia aperta; della specializzazione internazionale; degli effetti dell'imposizione di dazi e di contingentamenti; degli effetti dello scambio internazionale sulla struttura interna della produzione e del consumo di ciascun paese; degli effetti della crescita economica sul commercio internazionale e viceversa; ecc. Carattere distintivo della teoria pura del commercio

internazionale è l'ipotesi che gli scambi avvengano in forma di baratto o che, se vi è moneta, questa sia soltanto un 'velo', che non ha alcuna influenza sulle sottostanti variabili reali ma serve soltanto come unità di riferimento (numerario). Conseguenza non secondaria di questa ipotesi è che si verifichi una situazione di continuo pareggio nei conti internazionali di ciascun paese nei confronti di tutti gli altri: non esistono, cioè, problemi di bilancia dei pagamenti.La teoria monetaria delle relazioni economiche internazionali (che ha natura essenzialmente macroeconomica e che, come dicevamo sopra, si è venuta sviluppando in tempi relativamente recenti) si occupa dei problemi posti dagli squilibri della bilancia dei pagamenti in un'economia monetaria (e in particolare dei meccanismi automatici e delle politiche di aggiustamento della bilancia dei pagamenti stessa); dei vari regimi di cambi; dei problemi della liquidità internazionale e di altri problemi concernenti il sistema monetario internazionale.In questo articolo ci occuperemo soprattutto della parte teorica (sia pura sia monetaria), in quanto la parte descrittiva è trattata in altri articoli; non mancheranno, tuttavia, nei luoghi opportuni, i riferimenti alla realtà.

2. Generalità sulla teoria pura del commercio internazionale

Sia pure a costo di qualche semplificazione, possiamo distinguere tre schemi teorici principali che mirano a spiegare le determinanti del commercio internazionale e della specializzazione internazionale: 1) la teoria classica (Torrens-Ricardo), che individua tali determinanti nelle differenze tecnologiche; 2) la teoria di Heckscher-Ohlin, che pone l'accento sulle differenze nelle dotazioni dei fattori produttivi; 3) la teoria neoclassica (che ha una gestazione più lunga: se ne trovano spunti in John Stuart Mill, viene ripresa e approfondita da Alfred Marshall e portata ad alto grado di raffinatezza formale da numerosi autori moderni), che tiene conto simultaneamente delle differenze tecnologiche, delle differenze nelle dotazioni dei fattori produttivi e inoltre delle differenze nei gusti. Quest'ultimo elemento è in grado di spiegare l'esistenza del commercio internazionale anche a parità di tecnologia e di dotazioni dei fattori.
Dal punto di vista cronologico la seconda teoria è posteriore alla prima, mentre la terza, come dicevamo, ha avuto una gestazione più lunga e si è quindi venuta evolvendo parallelamente alle altre. A scanso di equivoci è bene precisare che anche la teoria di Heckscher-Ohlin è neoclassica (nel senso in cui la visione neoclassica si contrappone a quella

classica), in quanto accetta tutte le premesse logiche dell'impostazione neoclassica e ne segue la metodologia. In effetti lo schema di Heckscher-Ohlin può essere considerato un caso particolare della teoria neoclassica, in cui si suppone l'identità internazionale della tecnologia e dei gusti. Questo (cioè la minor generalità) è il prezzo da pagare, sostengono alcuni, se si vogliono raggiungere conclusioni definite (ed empiricamente verificabili in modo semplice) sulla struttura del commercio internazionale di un paese. Non intendiamo qui affrontare la questione della superiorità dell'una o dell'altra teoria; quello che ci preme rilevare è che, sebbene la visione neoclassica sia oggi accettata dalla maggior parte degli studiosi che si occupano di teoria pura del commercio internazionale, non mancano critiche e tentativi di ritorno ai classici.

3. La teoria dei costi comparati

Secondo la teoria classica del commercio internazionale (usualmente attribuita a David Ricardo, che la enunciò nel cap. 7 dei Principî, 1817, anche se è possibile trovarne degli elementi in Robert Torrens), la variabile cruciale per spiegare l'esistenza e la struttura del commercio internazionale è la tecnologia. Una differenza nei 'costi comparati' - condizione necessaria perché si verifichi lo scambio internazionale - riflette infatti differenze nelle tecniche produttive. La teoria si propone anche di dimostrare che lo scambio è vantaggioso per tutti i paesi che vi partecipano.
Schematizzando al massimo, supponiamo che vi siano due paesi (Inghilterra e Portogallo nel famoso esempio ricardiano), due merci (panno e vino) e che la produzione di ciascuna merce in entrambi i paesi avvenga a coefficienti fissi, sicché il costo unitario di produzione (espresso in termini di lavoro) di ciascuna merce sia costante.
È chiaro che se un paese è superiore all'altro nella produzione di un bene (ove la 'superiorità' viene misurata da un minor costo unitario) e inferiore nella produzione dell'altro, vi è la base per un fruttuoso scambio internazionale con convenienza per ambedue i paesi, come era già stato chiarito da autori precedenti, ad esempio da Adam Smith. Meno ovvio è invece che possa ugualmente esserci scambio internazionale anche nel caso in cui uno dei due paesi sia superiore all'altro nella produzione di ambedue le merci. Il grande contributo della teoria ricardiana è consistito per l'appunto nel mostrare le condizioni sotto cui anche in tale caso è possibile e conveniente (per entrambi i paesi) lo scambio internazionale. Dunque, la

condizione necessaria per lo scambio internazionale è in ogni caso che vi sia una differenza nei costi comparati. Il costo comparato può essere definito in due modi: come il rapporto fra i costi (unitari) assoluti delle due merci nel medesimo paese, oppure come il rapporto fra i costi (unitari) assoluti della medesima merce nei due paesi. Seguendo la prassi prevalente adotteremo la prima definizione, anche se i risultati non cambierebbero usando la seconda.
La proposizione di base della teoria in esame è che la condizione perché si abbia scambio internazionale è l'esistenza di una differenza fra i costi comparati. Questa è peraltro soltanto condizione necessaria; la condizione sufficiente è che la ragione di scambio internazionale sia compresa fra i costi comparati, senza essere uguale ad alcuno dei due. Verificatesi ambedue le condizioni, ciascun paese avrà convenienza a specializzarsi nella produzione del bene in cui ha il vantaggio relativamente maggiore (o lo svantaggio relativamente minore).
Consideriamo l'esempio mostrato nella tabella.
Come si vede, l'Inghilterra è superiore al Portogallo nella produzione di ambedue i beni. Potrebbe quindi sembrare che non vi sia scopo per lo scambio internazionale, ma non è così.
I costi comparati sono 4/8=0,5 e 6/10=0,6 rispettivamente in Inghilterra e in Portogallo.
L'Inghilterra ha inoltre un vantaggio relativamente maggiore nella produzione di panno: il costo unitario del panno, infatti, è in Inghilterra del 33,3% (2/6) minore rispetto a quello del Portogallo, mentre il costo unitario del vino è solo del 20% (2/10) minore nel primo rispetto al secondo paese. Si può similmente vedere che il Portogallo ha uno svantaggio relativamente minore nella produzione di vino: il costo unitario del vino è infatti nel Portogallo del 25% (2/8) maggiore rispetto a quello dell'Inghilterra, mentre il costo unitario del panno è del 50% (2/4) maggiore.
Pertanto, sempre che la ragione di scambio internazionale risulti maggiore di 0,5 e minore di 0,6, avverrà lo scambio di panno inglese contro vino portoghese con vantaggio di ambedue i paesi. Supponiamo che la ragione di scambio internazionale sia 0,55 (cioè che lo scambio internazionale avvenga in ragione di 0,55 unità di vino per una unità di panno). In Inghilterra, sulla base della tecnologia esistente, una unità di panno si scambia contro 0,5 unità di vino: 0,5 è infatti il costo comparato e l'ipotesi classica è che i prezzi relativi delle merci, cioè i rapporti di scambio fra di esse, siano determinati dai costi. Invece, mediante lo scambio internazionale con una unità di panno si ottengono 0,55 unità di vino, più di quanto è ottenibile all'interno. Similmente, in Portogallo per avere una unità di panno occorrono 0,6 unità di vino (0,6 è il costo comparato portoghese), mentre con lo scambio internazionale

bastano solo 0,55 unità di vino per avere una unità di panno. Lo scambio internazionale è quindi vantaggioso per ambedue i paesi.Che la ragione di scambio internazionale debba essere compresa fra i due costi comparati senza essere uguale ad alcuno dei due è evidente. Se infatti essa fosse uguale a uno dei due, il paese il cui costo comparato risultasse uguale alla ragione di scambio internazionale non avrebbe alcuna convenienza allo scambio, in quanto il rapporto di scambio interno (dato per l'appunto dal costo comparato) sarebbe uguale al rapporto di scambio internazionale. Ciò significa che il paese potrebbe ottenere l'altro bene allo stesso costo interno.

4. Il modello neoclassico

La teoria neoclassica del commercio internazionale estende all'economia aperta il modello neoclassico di equilibrio generale in economia chiusa. Tale modello - date le quantità totali dei vari fattori esistenti nell'economia, la loro distribuzione fra i membri della collettività, lo stato della tecnica (rappresentato dalle funzioni della produzione) e i gusti dei consumatori (rappresentati dalle funzioni di utilità), e supposto un comportamento ottimizzante da parte di tutti gli operatori, nonché una situazione di concorrenza perfetta generalizzata (sui mercati dei beni e dei fattori) - determina i valori di equilibrio (tali cioè che domanda=offerta) sia delle quantità che dei prezzi relativi. Nel caso di economia aperta, in tanto si apre la possibilità di scambio internazionale in quanto il prezzo relativo internazionale o ragione di scambio internazionale sia diversa dalla ragione di scambio di equilibrio interno di economia chiusa. Ciò può essere illustrato nel caso semplice di due beni e due paesi dalla fig. 1.
Nella parte sinistra della figura sono riportate le curve di domanda (D_1AD_1A) e di offerta (S_1AS_1A) di equilibrio generale interne per il bene A del paese 1, mentre nella parte destra sono riportate le analoghe curve del paese 2.
Si può mostrare facilmente che, all'apertura del commercio internazionale, lo scambio si può verificare soltanto se la ragione di scambio internazionale è compresa fra i due prezzi relativi di equilibrio autarchico. Infatti, per ragioni di scambio maggiori di ORE (si noti che con libero commercio, concorrenza perfetta e assenza di costi di trasporto, lo stesso bene deve avere lo stesso prezzo relativo ovunque, sicché i prezzi relativi vigenti nei mercati nazionali e in quello internazionale coincidono), ambedue i paesi offrirebbero il bene A sul mercato internazionale, poiché in entrambi vi sarebbe un eccesso di offerta per tale bene, e dunque

nessun equilibrio internazionale potrebbe verificarsi. Similmente, per ragioni di scambio minori di OPE, ambedue i paesi domanderebbero il bene A sul mercato internazionale, poiché in entrambi vi sarebbe un eccesso di domanda di tale bene.L'equilibrio internazionale si stabilirà in un punto ove l'eccesso di domanda del bene A da parte del paese 1 (domanda di importazioni da parte del paese 1) è in valore assoluto uguale all'eccesso di offerta di questo bene da parte del paese 2 (offerta di esportazioni da parte del paese 2). Questo punto risulta determinato nella fig. 1 in corrispondenza della ragione di scambio OQE, ove $M_1AM_1A=X_2AX_2A$.

Naturalmente, date le condizioni vigenti nell'equilibrio generale e in particolare la legge di Walras, se nel paese 2 v'è un eccesso di offerta del bene A (offerta di esportazioni), nello stesso ci sarà un eccesso di domanda del bene B (domanda d'importazioni) di valore equivalente, e viceversa nel paese 1 (ciascun paese, quindi, pagherà le importazioni con le proprie esportazioni e la bilancia commerciale sarà necessariamente in equilibrio). Un modo alternativo di rappresentare l'equilibrio internazionale, che illustra graficamente come a ogni domanda d'importazioni faccia riscontro un'offerta di esportazioni tramite gli eccessi di domanda e di offerta interni, è quello basato sulle curve marshalliane di domanda-offerta internazionale (chiamate anche curve di domanda reciproca e curve di offerta internazionale). La curva di domanda-offerta internazionale di un paese è per l'appunto definita come il luogo dei punti che rappresentano la quantità del bene importato che il paese è disposto ad accettare in cambio di una data quantità del bene esportato o, il che è lo stesso, la quantità del bene esportato che il paese è disposto a cedere in cambio di una data quantità del bene importato. Equivalentemente essa ci indica le varie ragioni di scambio a cui il paese è disposto a commerciare. La costruzione di tali curve avviene riportando su un grafico gli eccessi di domanda e di offerta interni di ciascun paese. Abbiamo quindi, supponendo che il paese 2 importi il bene B esportando il bene A e che il paese 1 importi il bene A esportando il bene B, il diagramma di cui alla fig. 2, ove OG_1 rappresenta la curva di domanda-offerta internazionale del paese 1 e OG_2 quella del paese 2. Le due curve si intersecano nel punto E, che è il punto di equilibrio internazionale: il paese 1 domanda OEA del bene A, esattamente uguale alla quantità di questo bene che il paese 2 offre, e offre OEB del bene B, esattamente uguale alla quantità di B che il paese 2 domanda. La ragione di scambio internazionale di equilibrio è $tg\alpha$ (pendenza di OE), che è numericamente uguale ad OQE nella fig. 1.

Va osservato, in conclusione, che il commercio internazionale arreca un vantaggio a tutti i

partecipanti, in quanto si può dimostrare che grazie a tale commercio ciascun paese raggiunge un livello di soddisfazione (misurato dalle curve d'indifferenza sociali) maggiore di quello raggiungibile in economia chiusa.Risulta evidente da quanto detto sopra che la teoria neoclassica è del tutto generale, in quanto la domanda d'importazioni (intesa come eccesso della domanda interna sulla produzione interna di un dato bene) e l'offerta di esportazioni (intesa come eccesso della produzione interna sulla domanda interna di un certo bene) sono determinate da tutti gli elementi sottostanti alle curve di domanda e di offerta interne (gusti, dotazioni dei fattori, tecnologia). V'è però un prezzo da pagare per questa generalità, in quanto ci si deve accontentare di sapere che un equilibrio internazionale esiste (e, volendo, analizzarne la stabilità) senza peraltro poter enunciare proposizioni operative che consentano di determinare in modo semplice la struttura e il volume del commercio internazionale. D'altro canto, la teoria neoclassica può essere utilizzata, restringendone la generalità, per fornire spiegazioni semplici della struttura del commercio internazionale. Invero, dal punto di vista puramente analitico, la teoria di Heckscher-Ohlin (con tutti i suoi corollari, come il teorema del pareggiamento dei prezzi dei fattori) può essere considerata come un caso particolare della teoria neoclassica, come vedremo nel prossimo capitolo.

5. La teoria di Heckscher-Ohlin

a) Aspetti teorici

Questa teoria introduce due importanti ipotesi semplificatrici: l'identità internazionale delle funzioni della produzione e dei gusti dei consumatori (i quali, inoltre, sono indipendenti dal reddito, sicché la struttura della domanda dipende soltanto dai prezzi relativi dei beni e non dal livello del reddito); la tecnologia, inoltre, non ammette inversione delle intensità dei fattori, cioè - rimanendo nel caso semplificato di due beni e di due fattori produttivi -, se la produzione di un certo bene utilizza, poniamo, relativamente più capitale che lavoro rispetto all'altro bene a un dato rapporto fra i prezzi dei fattori, anche cambiando tale rapporto (il che provoca una modificazione delle tecniche produttive scelte) il primo bene continuerà a utilizzare relativamente più capitale che lavoro rispetto all'altro.Grazie a queste ipotesi semplificatrici, la teoria di Heckscher-Ohlin è in grado di enunciare una ben precisa proposizione o teorema sulla struttura del commercio internazionale: ciascun paese esporta il bene la cui produzione richiede un impiego relativamente più intenso del fattore di cui il

paese ha una dotazione relativamente più abbondante e importa l'altro bene.
La dimostrazione di questa proposizione è molto complessa ed è possibile soltanto accennarvi. Si comincia col dimostrare che il paese che ha una dotazione relativamente abbondante di un fattore produttivo ha una tendenza a produrre proporzionalmente di più del bene la cui produzione richiede un impiego relativamente più intenso di quel fattore. La struttura della produzione sarà quindi diversa nei due paesi: ad esempio, supponendo che il bene A sia quello a più alta intensità di capitale e che il paese 1 abbia una dotazione relativamente abbondante di capitale, a parità di rapporto fra i prezzi, il paese 1 produce più del bene A rispetto al bene B e il paese 2 più del bene B rispetto al bene A (il rapporto A/B è maggiore nel paese 1 che nel paese 2). Apertosi il commercio internazionale e livellatosi il prezzo relativo dei beni, per l'ipotesi sulla struttura della domanda i due paesi vogliono consumare i due beni nella stessa proporzione: ne segue che il paese 1 esporterà il bene A (importando il bene B) e viceversa farà il paese 2, il che dimostra la proposizione di partenza.
Il modello di Heckscher-Ohlin dà luogo, in aggiunta alla proposizione di base testé esaminata, ad altre tre proposizioni fondamentali: il teorema del livellamento dei prezzi dei fattori, il teorema di Stolper-Samuelson, il teorema di Rybczynski, dei quali ci limitiamo a fornire l'enunciazione. Il teorema del livellamento dei prezzi dei fattori afferma che, nonostante l'ipotizzata immobilità internazionale dei fattori della produzione, il libero scambio internazionale uguaglia i prezzi assoluti dei fattori nei vari paesi. Il teorema di Stolper-Samuelson ci dice che l'aumento del prezzo di un bene (ad esempio per l'imposizione di un dazio) fa aumentare la remunerazione reale del fattore che utilizza quel bene in maniera relativamente intensiva. Infine, il teorema di Rybczynski afferma che, se uno dei fattori della produzione aumenta, fermo restando l'altro, aumenterà la produzione del bene che usa quel fattore in modo relativamente intensivo, mentre diminuirà la produzione dell'altro bene, a parità di prezzi dei beni e dei fattori.Questi tre teoremi, insieme alla proposizione di base del modello Heckscher-Ohlin, vengono abitualmente considerati come il nucleo essenziale della teoria ortodossa del commercio internazionale.

b) Studi empirici
La teoria di Heckscher-Ohlin, proprio perché considerata centrale, è stata sottoposta a numerosissimi studi tendenti ad accertarne la rispondenza alla realtà e quindi l'effettiva valenza esplicativa.Il più noto di questi è indubbiamente quello pionieristico di Leontief

(1953). Questo studioso, servendosi dello schema input-output, da lui stesso elaborato, calcolò - sulla base della matrice intersettoriale dell'economia statunitense riferita al 1947 - il fabbisogno complessivo (diretto e indiretto) di capitale e di lavoro per unità della merce composita "esportazioni statunitensi del 1947" e per unità della merce composita "importazioni sostitutive statunitensi del 1947". I risultati delle complesse elaborazioni di Leontief indicarono che gli Stati Uniti apparentemente esportavano prodotti ad alta intensità di lavoro ed importavano prodotti ad alta intensità di capitale. Dato che gli Stati Uniti erano concordemente ritenuti il paese relativamente più dotato di capitale di tutti gli altri paesi con cui commerciavano, i risultati di Leontief contrastavano palesemente con la teoria di Heckscher-Ohlin (secondo la quale questo paese avrebbe dovuto essere esportatore di beni ad alta intensità di capitale), donde il nome di 'paradosso di Leontief' dato a questi risultati.Il lavoro di Leontief dette luogo a un ampio dibattito, centrato sia sugli aspetti statistici sia su quelli teorici, e a numerosi studi empirici successivi, che sono continuati fino al giorno d'oggi con risultati contrastanti. Non è certo possibile riassumere qui questa sterminata letteratura, per cui ci limiteremo a sottolinearne alcuni aspetti. (Un esauriente studio in materia è quello di Leamer: v., 1984).

Ricordiamo innanzitutto la tesi, avanzata dallo stesso Leontief, secondo cui il lavoro statunitense era all'epoca molto più efficiente del lavoro estero, sicché, in realtà, era proprio il fattore lavoro a essere relativamente più abbondante negli Stati Uniti che negli altri paesi quando il lavoro statunitense fosse stato convertito in unità equivalenti di lavoro estero. Secondo Leontief, il coefficiente di conversione (ipotizzato e non dimostrato) era addirittura 3. Studi successivi non confermarono il coefficiente di conversione. Ad esempio, Kreinin (1965) dimostrò che la superiorità del lavoro statunitense era mediamente del 20 o 25%; il coefficiente di conversione di 1,2 o 1,25 era dunque ben lontano dal coefficiente 3 che, a parere di Leontief, avrebbe consentito di classificare gli Stati Uniti come un paese relativamente abbondante di lavoro.Conviene poi ricordare le osservazioni di vari studiosi secondo cui non ci si può nella realtà limitare a considerare due soli fattori produttivi (il capitale fisico e il lavoro) e si deve prendere in considerazione anche il fattore 'risorse naturali'. Vanek, infatti (1959), in aggiunta ai dati calcolati da Leontief, calcolò il fabbisogno di beni ad alto contenuto di risorse naturali richiesti per la produzione di una unità di esportazioni e di una unità di importazioni sostitutive: tale fabbisogno risultò maggiore per le importazioni. Gli Stati Uniti erano dunque importatori di beni ad alta intensità di risorse

naturali (comunque calcolata, e cioè sia rispetto al capitale sia rispetto al lavoro), fattore relativamente scarso in tale paese, ed esportatori di beni ad alta intensità di capitale e di lavoro rispetto alle risorse naturali (fattori, i primi due, relativamente abbondanti rispetto alle risorse naturali). Il teorema di Heckscher-Ohlin, lungi dall'essere refutato, era confermato in pieno.
Altri autori pongono l'accento sul fattore 'capitale umano', che è quello incorporato nei lavoratori qualificati o specializzati in quanto distinti dai lavoratori comuni o non qualificati. Tralasciando i problemi connessi ai diversi modi di misurare concretamente il capitale umano (capitalizzazione delle differenze salariali, anni di istruzione, qualifiche professionali, ecc.), l'inclusione di questo fattore nelle analisi empiriche consente di suffragare l'ipotesi che le esportazioni statunitensi siano ad alta intensità di capitale umano (fattore relativamente abbondante negli Stati Uniti) rispetto alle importazioni sostitutive, coerentemente con il teorema di Heckscher-Ohlin. Un risultato in questo senso è quello di Stern e Maskus (1981), nel cui testo sono anche citati i risultati analoghi di studi precedenti.Vi sono poi studi i quali attribuiscono la presenza del paradosso al mancato soddisfacimento di alcune delle ipotesi necessarie per la validità del teorema di Heckscher-Ohlin, ad esempio alla non identità internazionale dei gusti dei consumatori o al verificarsi di un'inversione delle intensità fattoriali. Anche questi fenomeni, peraltro, sono tutt'altro che definitivamente accertati.
Abbiamo finora esaminato le spiegazioni del paradosso di Leontief dando per scontato che tale paradosso esista. Ma così non è, poiché studi condotti successivamente, sia per gli Stati Uniti sia per gli altri paesi, non hanno sistematicamente confermato la presenza del paradosso stesso. Per quanto riguarda gli Stati Uniti, lo studio di Stern e Maskus (1981) ha confermato la presenza del paradosso di Leontief utilizzando la matrice intersettoriale del 1958, mentre il paradosso stesso è scomparso utilizzando la matrice intersettoriale del 1972. Si può quindi concludere che il paradosso in questione - e quindi l'eventuale invalidità del teorema di Heckscher-Ohlin - non può essere considerato come una costante, ma va anzi accertato caso per caso con riferimento non soltanto al singolo paese, ma anche al periodo di tempo considerato in uno stesso paese dato.

6. Ulteriori sviluppi

Gli ulteriori sviluppi della teoria pura del commercio internazionale si sono mossi

sostanzialmente lungo due direttrici. La prima consiste nell'introdurre vari perfezionamenti alla teoria stessa mantenendosi però all'interno del suo quadro di riferimento. La seconda consiste invece nell'abbandonare le principali ipotesi semplificatrici (fra cui quella di concorrenza perfetta) dando luogo ad analisi non inquadrabili, o inquadrabili solo parzialmente, nella teoria ortodossa (le cosiddette 'nuove teorie' del commercio internazionale).Tra i principali perfezionamenti introdotti dal primo filone, si può accennare ai costi di trasporto, alla considerazione dei beni intermedi, all'offerta elastica dei fattori produttivi, alla mobilità internazionale dei fattori stessi, alla presenza di fattori produttivi specifici (tipici cioè di un certo settore produttivo e non utilizzabili in altri settori), alla presenza di beni non commerciati accanto a quelli oggetto di commercio internazionale, all'introduzione di condizioni d'incertezza e, infine, all'esame degli effetti economici di transazioni internazionali illegali come il contrabbando. Si tratta di analisi che arricchiscono notevolmente la portata e la rilevanza della teoria ortodossa, ma che non possono essere qui approfondite. Conviene passare invece all'esame del secondo filone, scegliendo per forza di cose soltanto alcuni fra i numerosi schemi proposti.

Un primo approccio è rappresentato dalla teoria della disponibilità (availability), che pone per l'appunto l'accento sulla disponibilità di risorse naturali scarse o di altri elementi che danno al paese un particolare vantaggio (come ad esempio il possesso da parte di un paese di speciali conoscenze tecnologiche in certi rami produttivi). Lo scambio internazionale si spiega col fatto che ciascun paese importa quei beni che non sono disponibili sul proprio mercato: l'indisponibilità può quindi derivare o dal fatto che si tratta di beni la cui produzione richiede risorse naturali inesistenti nel paese (oro, petrolio, ecc.: indisponibilità assoluta) o dal fatto che si tratta di beni che non possono essere prodotti all'interno o potrebbero esserlo solo a costi proibitivi (per ragioni tecnologiche o di altro genere: indisponibilità relativa); per converso il paese esporta quei beni di cui ha la disponibilità nel senso sopra chiarito. Ora, per quanto riguarda il primo aspetto, cioè quello collegato alla disponibilità di risorse naturali, esso potrebbe essere inquadrato nella teoria di Heckscher-Ohlin che, come sappiamo, pone l'accento sulle differenti dotazioni di fattori produttivi, tra cui si possono includere anche le risorse naturali (il fattore 'terra' comprende anche i giacimenti minerari). L'originalità va invece ricercata nel secondo aspetto, cioè nei motivi addotti per spiegare le differenze internazionali di disponibilità del secondo tipo. Questi motivi sono sostanzialmente due: il progresso tecnico e la differenziazione dei prodotti. Per quanto riguarda il primo, esso dà

luogo al cosiddetto ritardo d'imitazione (imitation lag). Come conseguenza della ricerca scientifica (progresso tecnico in senso lato) compaiono nuovi prodotti e il paese innovatore gode di un monopolio fin quando gli altri paesi non hanno imparato a produrli: nel frattempo essi devono importarli. Viene quindi creato commercio internazionale per il tempo necessario all'imitazione. Avvenuta l'imitazione cessa l'importazione; ma, data la continua comparsa di nuovi prodotti, questa componente del commercio internazionale è sempre presente.
Per quanto riguarda la differenziazione dei prodotti, si tratta semplicemente di un'estensione al commercio internazionale dei risultati ben noti della teoria della concorrenza monopolistica. Paesi diversi producono gli stessi beni o, più esattamente, beni non sostanzialmente diversi per lo scopo cui sono destinati (automobili, orologi, liquori, ecc.). Tali beni però, grazie a disegni industriali diversi, pubblicità, caratteristiche secondarie effettive o fittizie, fama acquisita in passato e così via, sono ritenuti diversi dai consumatori. Si vengono quindi a creare, da un lato, un certo potere monopolistico, sia pure ristretto, dei produttori e, dall'altro, una domanda dei consumatori per beni esteri da essi ritenuti diversi dagli analoghi beni nazionali, con il risultato di generare commercio internazionale.
Accenniamo poi alla teoria del ciclo del prodotto di Vernon, secondo cui nell'introduzione e nella diffusione di un nuovo prodotto è possibile distinguere varie 'fasi': la fase della introduzione del nuovo prodotto, quella della sua 'maturazione' e quella della standardizzazione. Quando viene introdotto un nuovo bene, è probabile che la localizzazione della sua produzione sia vicina al mercato su cui esso viene commercializzato. Ciò non tanto per questioni relative ai costi di trasporto e simili, ma soprattutto perché l'incertezza dell'imprenditore innovatore relativa a vari elementi (quali l'accoglienza che i consumatori faranno al nuovo prodotto, le reazioni dei concorrenti, le eventuali modifiche da apportare ai metodi produttivi e al prodotto stesso) richiede una continua, veloce ed efficace comunicazione fra produttore e mercato, che si può ottenere solo localizzando la produzione nello stesso mercato dove viene venduto il prodotto. Nella prima fase, quindi, verrà servito il mercato nazionale del paese in cui si trova l'impresa innovatrice. Quando il prodotto si è affermato sul mercato nazionale e ha quindi raggiunto un certo grado di standardizzazione, il produttore comincerà a inserirsi sui mercati esteri. In questa fase di maturazione i motivi sottostanti alla localizzazione iniziale non sussistono più e può diventare conveniente impiantare unità produttive in paesi esteri sviluppati, non solo allo scopo di realizzare minori costi di produzione e di trasporto, ma anche per cautelarsi contro il pericolo che i governi

esteri, al fine di proteggere le proprie industrie nazionali, adottino misure restrittive nei confronti dell'importazione del nuovo prodotto. Infine, nella terza fase del ciclo, si ha ormai una standardizzazione spinta del prodotto: in questa fase può accadere che convenga impiantare unità produttive in paesi sottosviluppati, soprattutto se si tratta di un bene in cui il costo del lavoro ha un peso rilevante. Nella terza fase, secondo Vernon, la produzione del bene nel paese innovatore comincia a diminuire, mentre la domanda interna continua a crescere, sicché tale paese da esportatore diventa gradualmente importatore (prima da altri paesi sviluppati, poi anche dai paesi sottosviluppati) del bene in questione.

Un ultimo fenomeno a cui non è possibile non accennare è quello del commercio intraindustriale o orizzontale, ovvero la simultanea importazione ed esportazione di beni provenienti dalla stessa industria. Questo tipo di scambio è stato dapprima osservato in sede empirica e poi ne è stata tentata la spiegazione teorica. Si tratta di un fenomeno che a prima vista fa a pugni con la teoria ortodossa, in base alla quale un paese o importa o esporta un determinato bene, ma non può esserne simultaneamente importatore ed esportatore. Si deve però osservare che le classificazioni merceologiche in uso nelle statistiche degli scambi internazionali (basate sulla SITC: Standard International Trade Classification) non possono scendere fino al dettaglio del singolo bene, ma raggruppano i beni in categorie. Non desta quindi sorpresa che beni inclusi nella stessa categoria (ad esempio, automobili a benzina di una certa cilindrata) vengano sia esportati che importati, in quanto, a causa della differenziazione di cui si è già detto in precedenza, si tratta in realtà di beni diversi. Tutte le teorie precedentemente accennate (disponibilità, ciclo del prodotto e altre che non sono state trattate) sono dunque in grado di spiegarci il fenomeno del commercio intraindustriale, che costituisce una grossa fetta del commercio internazionale, in particolare a livello dei paesi industrializzati. Esiste però anche il caso di commercio intraindustriale in beni assolutamente omogenei il quale (anche se quantitativamente meno rilevante rispetto a quello in beni differenziati) non può essere trascurato. È stato dimostrato che questo caso può essere spiegato con opportuni adattamenti della teoria ortodossa (ad esempio come conseguenza di modificazioni periodiche delle curve d'indifferenza o delle curve di trasformazione), per cui in definitiva è soltanto il commercio intraindustriale in beni differenziati a non poter essere inquadrato nella teoria ortodossa.

Questo capitolo può essere concluso con una notazione: lo stato relativamente fluido di questo filone è in realtà dovuto al fatto che, similmente a quanto accade nell'economia chiusa,

non esiste ancora una teoria in grado di analizzare in modo generale il fenomeno delle forme di mercato 'intermedie' (diverse cioè dalla concorrenza perfetta e dal monopolio perfetto).

7. Politiche commerciali

a) Introduzione

La teoria tradizionale dell'intervento dello Stato nel commercio internazionale concentrava l'attenzione sui dazi, partendo da due principî generalmente accettati fino alla prima guerra mondiale: a) che le barriere al commercio internazionale a scopi protettivi dovessero essere confinate ai dazi e b) che non ci dovesse essere discriminazione commerciale nei confronti dei vari paesi, cioè che un eventuale dazio dovesse essere applicato nella stessa misura a tutte le importazioni qualunque ne fosse l'origine. Nonostante che nel periodo fra le due guerre mondiali, e in particolare negli anni successivi alla grande depressione, si fosse assistito a una sistematica violazione di tali principî, essi furono ripresi e posti a base dell'accordo mondiale che avrebbe dovuto regolare il commercio internazionale dopo la seconda guerra mondiale: il GATT (General Agreement on Tariffs and Trade). Numerose riunioni internazionali con lo scopo di negoziare riduzioni dei dazi (i vari rounds) si sono svolte sotto gli auspici del GATT, che, tuttavia, ha dovuto essere tollerante sulle 'violazioni' dei suddetti principî. Gli ultimi decenni hanno quindi visto espandersi sia le barriere al commercio internazionale diverse dai dazi all'importazione sia le discriminazioni commerciali (fra cui la formazione di aree di libero scambio, unioni doganali, ecc.), per cui la teoria tradizionale ha dovuto essere 'ampliata' per poter analizzare rigorosamente anche questi fenomeni.

b) Dazi e altre misure

La teoria tradizionale parte dall'osservazione che un dazio all'importazione, facendo aumentare il prezzo interno del bene importato rispetto alla situazione di libero scambio, causa una riduzione di benessere misurata dalla diminuzione della cosiddetta rendita del consumatore. Accanto a questo effetto negativo vi sono però effetti positivi: l'aumento delle entrate dello Stato e l'aumento dei profitti dei produttori nazionali. Si può tuttavia dimostrare che il saldo rimane negativo, in quanto detti benefici non compensano la riduzione della rendita del consumatore. Analoga conclusione vale per l'altra misura protezionistica classica: il contingentamento. Ma anche le altre misure protezionistiche danno luogo agli stessi

risultati netti negativi: deposito previo sulle importazioni, utilizzo artificioso delle normative a tutela della salute e della sicurezza, cartelli internazionali, sussidi alle esportazioni, discriminazioni nella spesa pubblica sono solo alcune delle pratiche più diffuse. C'è però un caso in cui, date certe condizioni, la politica protezionistica arreca sicuramente un vantaggio a chi la pratica, ed è quello dell'esistenza di disoccupazione (nei ragionamenti precedenti si parte sempre dall'ipotesi di piena occupazione dei fattori produttivi). L'aumento della produzione nazionale consente d'impiegare lavoratori precedentemente disoccupati, con generale beneficio. Ma questo risultato presuppone che il resto del mondo, colpito dal dazio, non effettui rappresaglie: se ciò avvenisse verrebbero danneggiate le industrie d'esportazione del paese in questione, con effetti incerti. Un altro caso - questo valido nel contesto tradizionale di piena occupazione delle risorse - è quello del dazio ottimo, in base al quale, sempre che non vi siano ritorsioni da parte del resto del mondo, è possibile fissare un dazio tale da massimizzare il benessere del paese che lo impone, il quale quindi raggiunge un livello di benessere superiore a quello che aveva in libero scambio. Ma anche qui, se il resto del mondo (che vede diminuire il proprio benessere) effettua rappresaglie, le conclusioni cambiano. Come si vede, entrambi i casi sono basati su un'ipotesi - l'assenza di rappresaglie - tutt'altro che plausibile.C'è allora da chiedersi su che cosa siano fondate le argomentazioni a favore del protezionismo, il che porta a esaminare l'annoso dibattito libero scambio-protezionismo.

c) Il dibattito libero scambio-protezionismo e la teoria del second best

Mentre la teoria tradizionale del commercio internazionale conduce alla proposizione che il libero scambio è meglio del commercio internazionale in presenza di restrizioni e che quest'ultimo è a sua volta meglio della situazione di autarchia, i fautori del protezionismo avanzano una serie di argomentazioni in senso opposto, di cui la principale - a parte quella del dazio ottimo, di cui si è già detto prima - è quella della cosiddetta industria nascente.

Si tratta probabilmente del più antico e noto argomento a favore del protezionismo. Un'industria ai suoi inizi non può ovviamente competere con industrie straniere già da tempo affermate. Essa quindi deve essere protetta, ad esempio mediante l'imposizione di un dazio, per il tempo necessario a crescere e a consolidarsi, in modo da divenire concorrenziale con le industrie straniere; a questo punto la protezione può cessare.È chiaro che per la validità di questa argomentazione occorre che l'industria riesca effettivamente a diventare

concorrenziale con le industrie estere al prezzo internazionale di libero scambio e, inoltre, che i vantaggi derivanti alla comunità dall'attività dell'industria, cessata la protezione, siano tali da compensare abbondantemente le perdite derivanti dalla protezione.
Anche ammesso che tali condizioni siano verificate, è tuttavia possibile vedere che i vantaggi derivanti dal consolidamento dell'industria nascente possono essere ottenuti con costi minori usando altre politiche, diverse dall'imposizione di un dazio: ad esempio mediante la concessione di un sussidio all'industria nascente in modo da non far salire il prezzo interno a livelli superiori al prezzo internazionale.Vi sono poi altre argomentazioni a favore del protezionismo, che ci limitiamo a menzionare, quali la presenza di distorsioni nel mercato dei fattori e nel mercato interno dei prodotti.
Ci si può a questo punto chiedere se non sia possibile raggiungere risultati generali in tema di libero scambio e protezionismo. Molti studiosi affermano che il libero commercio è preferibile al commercio con restrizioni (salvo il caso del dazio ottimo senza ritorsione) e che, se si vogliono aiutare industrie nascenti, correggere distorsioni, ecc., è sempre meglio ricorrere a politiche di sussidi e simili anziché a dazi. Queste affermazioni debbono però essere qualificate, poiché la loro validità è stata dimostrata in una situazione di libera concorrenza perfetta su tutti i mercati (dei fattori e delle merci), sia nazionali che internazionali. Quando tale situazione non si verifica (ed è innegabile che questa è la realtà) le cose cambiano e occorre far ricorso al principio del second best (subottimo). Questo principio - il cui scopo è di rinvenire la situazione migliore quando, a causa di distorsioni o altro, non sia possibile realizzare tutte le condizioni dell'ottimo paretiano - afferma che, una volta che siano violate una o più condizioni, non è necessariamente vero che la situazione migliore (subottimale) sia quella in cui tutte le rimanenti condizioni di ottimo sono verificate. Come corollario di tale principio si ha che non è possibile stabilire a priori se la sostituzione di una violazione delle condizioni paretiane a un'altra migliori o peggiori la situazione; un altro corollario è che l'eliminazione di una di tali violazioni non necessariamente porta a una situazione migliore, così come l'introduzione di una violazione aggiuntiva non necessariamente porta a una situazione peggiore. In altre parole, in un mondo in cui sono presenti situazioni non concorrenziali, distorsioni, restrizioni di vario tipo al libero scambio, l'eliminazione di una o più di tali restrizioni non significa necessariamente il passaggio a una situazione migliore, mentre, paradossalmente, l'introduzione di una ulteriore restrizione potrebbe anche portare a un benessere maggiore, ancorché sempre subottimale.

Applicando la teoria del second best al dibattito libero scambio-protezionismo, segue che non è possibile accertare a priori se una politica protezionistica migliori o peggiori la situazione, né affermare che ogni movimento verso un commercio più libero costituisce senz'altro un miglioramento. Alla stessa stregua, non è nemmeno possibile affermare, come fa la teoria ortodossa, che esistono altre politiche sicuramente migliori della politica protezionistica.
È chiaro che, stando così le cose, diventa impossibile fare affermazioni valide in generale e dedurre prescrizioni di politica economica aprioristiche da pochi principî guida. Nella realtà tutto può succedere e occorre vedere in ogni caso concreto quale sia la politica migliore (liberista o protezionista che sia) senza farsi accecare da preconcetti teorici.
È importante osservare, in conclusione, che la teoria del second best consente anche d'inquadrare il problema delle unioni doganali (e più in generale dell'integrazione economica). Poiché l'unione doganale contempla la rimozione dei dazi fra i paesi membri, essa rappresenta comunque un movimento verso il libero scambio (anche se non perfetto, poiché rimangono i dazi nei confronti dei paesi non appartenenti all'unione) e dunque - così argomentava la teoria tradizionale - essa è senz'altro benefica. Alla luce della teoria del second best sappiamo invece che il risultato è incerto a priori e non è possibile fare affermazioni valide in generale.

8. Generalità sull'economia internazionale monetaria

La teoria monetaria delle relazioni economiche internazionali (che ha natura essenzialmente macroeconomica) si occupa dei problemi posti dagli squilibri della bilancia dei pagamenti in un'economia monetaria (e in particolare dei meccanismi automatici e delle politiche di aggiustamento della bilancia dei pagamenti stessa); dei vari regimi di cambi; dei problemi della liquidità internazionale e di altri problemi concernenti il sistema monetario internazionale.È dunque evidente che nel momento in cui si affrontano i problemi dell'economia monetaria internazionale diventa indispensabile tener conto del fatto che virtualmente ogni paese ha la propria unità monetaria (valuta) e che la maggior parte del commercio internazionale non avviene nella realtà in forma di baratto, ma mediante lo scambio di merci contro l'una o l'altra valuta. Inoltre vi sono transazioni economiche internazionali di carattere puramente finanziario, che coinvolgono quindi necessariamente valute diverse.

Questo ci porta ad affrontare due argomenti: il mercato dei cambi esteri (ed i vari regimi di cambi) e la bilancia dei pagamenti.

a) Il mercato dei cambi esteri e i regimi di cambi

Il mercato dei cambi esteri è il mercato dove le varie valute vengono scambiate (comprate o vendute) fra di loro. Rinviando all'apposita voce per approfondimenti (v. cambio), ci limitiamo a rammentare due concetti: quello di tasso di cambio e quello di regime di cambio. Il tasso di cambio è un prezzo, e precisamente il prezzo di una valuta espresso in termini di un'altra. Dato che entrano in gioco due valute, esistono due modi per quotare il tasso di cambio. Uno, che è quello adottato dall'Italia e dalla maggior parte dei paesi, è detto 'incerto per certo', cioè numero di unità di valuta nazionale per una unità della valuta estera considerata (220 lire per franco francese, ecc.); con questa definizione il tasso di cambio è il prezzo della valuta estera espresso in termini di valuta nazionale. L'altro è detto 'certo per incerto', cioè numero di unità della valuta estera considerata per una unità di valuta nazionale, ed è, ovviamente, il reciproco del precedente; con questa definizione il tasso di cambio è il prezzo della valuta nazionale in termini di valuta estera. Tale modo è adottato dall'Inghilterra e da alcuni paesi già colonie britanniche o legati un tempo alla cosiddetta area della sterlina.Veniamo ora ai vari regimi di cambi, in teoria numerosi, in quanto fra i due estremi di cambi perfettamente rigidi (o fissi) e di cambi perfettamente e liberamente flessibili esiste tutta una gamma di regimi intermedi cosiddetti a flessibilità limitata.Un estremo è dato dai cambi liberamente e perfettamente flessibili. Tale regime è caratterizzato dal fatto che le autorità monetarie del paese che lo adotta non intervengono nella maniera più assoluta sul mercato dei cambi. Il tasso di cambio nei confronti di qualsiasi valuta estera è dunque lasciato a se stesso e può fluttuare liberamente in qualsiasi direzione e in qualsiasi misura in base alle domande e offerte di valuta estera provenienti da tutti gli operatori. L'altro estremo è dato dai cambi rigidamente fissi, come nel gold standard o nel gold exchange standard.Il regime a parità fisse ma aggiustabili posto in essere dopo la seconda guerra mondiale (regime di Bretton Woods), sebbene concettualmente più vicino al gold exchange standard, prevedeva la possibilità, date certe condizioni, di variare la parità e ammetteva quindi una sia pur minima flessibilità. La fluttuazione senza parità dichiarata, ma con interventi delle banche centrali per cercare d'influenzare il cambio in una certa direzione (managed float: fluttuazione manovrata o sporca), è anch'esso un regime 'intermedio', che

sarà più vicino all'uno o all'altro estremo a seconda del grado d'intervento.

b) La bilancia dei pagamenti

Secondo la definizione del Balance of Payments Manual del FMI (Fondo Monetario Internazionale), la bilancia dei pagamenti di un paese è una registrazione sistematica di tutte le transazioni economiche effettuate in un dato periodo di tempo fra i residenti del paese che compie la rilevazione e i residenti degli altri paesi, indicati come stranieri o non residenti.

Le principali voci della bilancia dei pagamenti sono le merci, i servizi (noli, assicurazioni, turismo, ecc.) e i movimenti di capitali (investimenti diretti, investimenti di portafoglio, ecc.).

La 'chiusura' della bilancia dei pagamenti è data dalla variazione delle riserve internazionali del paese, che ha valore assoluto uguale al saldo delle voci precedenti. Considerando vari sottogruppi di voci, è possibile parlare, oltre che della bilancia dei pagamenti globale (che comprende tutte le voci sopra elencate), di una bilancia commerciale, di una bilancia merci e servizi, di una bilancia delle partite correnti (merci, servizi e trasferimenti unilaterali), ecc.

Come abbiamo già detto, uno dei compiti principali dell'economia internazionale monetaria è quello di esaminare i processi di aggiustamento della bilancia dei pagamenti, cioè di quei processi (sia spontanei sia provocati da interventi di politica economica) che conducono al riequilibrio di un deficit o di un surplus nei conti con l'estero. Nei due capitoli che seguono esporremo dapprima i processi tradizionali e poi gli schemi più complessi.

Con processo tradizionale intendiamo quello basato sulle variazioni del tasso di cambio (a parità di reddito nazionale e di tutte le altre condizioni rilevanti) e quello basato sulle variazioni del reddito nazionale (a parità di tasso di cambio e di tutte le altre condizioni rilevanti); anche l'integrazione fra questi due processi può essere considerata come facente parte della teoria tradizionale.

Caratteristica comune dei processi tradizionali è quella di concentrare l'attenzione sulle partite correnti e di considerare gli squilibri della bilancia dei pagamenti come squilibri di flusso, cioè come squilibri che derivano da incoerenza fra le principali grandezze macroeconomiche aventi la natura di flussi (in primo luogo le esportazioni e importazioni di merci e servizi e poi, se del caso, consumi, investimenti, reddito nazionale, ecc.).

Negli schemi più complessi verranno presi in considerazione anche i movimenti di capitale e si esamineranno, oltre agli squilibri di flusso, anche gli squilibri di stock, cioè quegli squilibri che derivano da incoerenza fra le principali grandezze aventi la natura di stock (in primo

luogo lo stock di moneta e poi quelli di altre attività finanziarie), per cui i flussi che si verificano derivano da aggiustamenti degli stocks stessi.

9. Gli schemi tradizionali

Questi schemi, come dicevamo, considerano la bilancia dei pagamenti come un fenomeno di flusso e concentrano l'attenzione sui soli scambi di merci e servizi, anche in conseguenza del fatto che si tratta di schemi elaborati in tempi in cui i movimenti internazionali di capitali finanziari erano del tutto trascurabili rispetto alle transazioni commerciali. In questo capitolo, quindi, con 'bilancia dei pagamenti' s'intenderà fare riferimento alla bilancia merci e servizi.

a) L'approccio delle elasticità

Cominciamo con l'osservare che, siccome i livelli dei prezzi (e tutte le altre variabili rilevanti, tra cui il reddito) sono per ipotesi fissi, l'elemento che influenza le esportazioni e le importazioni di un paese è il tasso di cambio: infatti, a parità di prezzi espressi nelle rispettive valute nazionali, una svalutazione del tasso di cambio rende più a buon mercato i beni nazionali sui mercati esteri e più cari i beni esteri sul mercato nazionale. In altri termini, una svalutazione del tasso di cambio ha lo stesso effetto di una riduzione del prezzo dei beni nazionali esportati e di un aumento del prezzo dei beni esteri importati, beninteso a parità di tasso di cambio. L'opposto vale naturalmente nel caso di una rivalutazione. Possiamo quindi dire che le esportazioni variano nella stessa direzione del tasso di cambio (un aumento del tasso di cambio, cioè una svalutazione, fa aumentare le esportazioni, e una diminuzione del cambio stesso, cioè una rivalutazione, le fa diminuire), mentre le importazioni variano in senso opposto rispetto alla variazione del tasso di cambio.

Naturalmente ciò non è sufficiente per dire che delle opportune variazioni del tasso di cambio (svalutazione nel caso di deficit, rivalutazione nel caso di surplus) riequilibrano la bilancia dei pagamenti. La bilancia dei pagamenti è infatti espressa in valore monetario e non è detto che la variazione delle quantità di esportazioni e importazioni nella direzione giusta assicuri anche una variazione del loro valore nella direzione giusta. Infatti, considerando ad esempio gli effetti di una svalutazione del cambio sulle importazioni, vediamo che la quantità di queste diminuisce, mentre l'esborso per unità d'importazione aumenta (dato che il loro prezzo in termini di valuta nazionale è aumentato in conseguenza della svalutazione del

cambio), sicché il valore complessivo in termini di valuta nazionale può variare in qualsiasi direzione. Più rigorosamente, definendo l'elasticità delle esportazioni, e delle importazioni, rispetto al tasso di cambio come ogni elasticità-prezzo, e cioè come rapporto fra la variazione proporzionale della quantità e la variazione proporzionale del prezzo (qui rappresentato dal tasso di cambio), si può dimostrare matematicamente che la condizione affinché la bilancia dei pagamenti espressa in valuta nazionale migliori a seguito di una svalutazione (e peggiori a seguito di una rivalutazione) è data dalla cosiddetta 'condizione delle elasticità critiche' (di solito chiamata, anche se impropriamente, 'condizione di Marshall-Lerner'), in base alla quale la somma dell'elasticità delle importazioni e dell'elasticità delle esportazioni (entrambe prese in valore assoluto) deve essere maggiore dell'unità. È bene avvertire che siffatta condizione vale nel caso si consideri una situazione iniziale di equilibrio; in tale ipotesi, la medesima condizione assicura che anche la bilancia dei pagamenti espressa in termini di valuta estera migliorerà a seguito della svalutazione. Nel caso invece in cui la situazione di partenza non sia di equilibrio, la condizione si fa più complessa e differisce a seconda che si consideri la bilancia dei pagamenti espressa in valuta nazionale oppure in valuta estera. In tale caso sembra preferibile riferirsi alla bilancia dei pagamenti espressa in valuta estera, dato che quest'ultima è la risorsa scarsa nel paese in difficoltà di bilancia dei pagamenti. Occorre notare, in conclusione, che secondo questo schema, verificate le opportune condizioni, lo squilibrio della bilancia dei pagamenti potrà venire completamente eliminato grazie ad adeguate variazioni del tasso di cambio.

b) Il moltiplicatore in mercato aperto

Il moltiplicatore keynesiano, originariamente concepito in relazione a un'economia chiusa (anche se nella Teoria generale di Keynes non mancano cenni ai problemi dell'economia aperta), venne successivamente esteso all'economia aperta e applicato all'analisi dell'aggiustamento della bilancia dei pagamenti. Il problema che ci interessa è quello di vedere se e in quale misura eventuali squilibri della bilancia dei pagamenti possano essere corretti da variazioni di reddito. Se, partendo da una situazione di equilibrio, si verifica, poniamo, un aumento delle esportazioni, la bilancia dei pagamenti passa in surplus. Quali sono, nell'ambito della presente analisi - che assume prezzi e tassi di cambio rigidi -, le forze correttive automatiche che tendono al riequilibrio della bilancia dei pagamenti? La risposta è semplice: l'aumento delle esportazioni genera, tramite il moltiplicatore, un aumento di

reddito, il quale a sua volta genera, tramite la propensione marginale all'importazione, un aumento delle importazioni. Quest'ultimo aumento tende ovviamente a controbilanciare l'iniziale aumento delle esportazioni: si tratta di vedere se lo controbilancia esattamente (nel qual caso il riequilibrio è completo) oppure no (nel qual caso la bilancia dei pagamenti rimane in surplus, sia pur minore di quello iniziale); vedremo inoltre come non sia possibile escludere a priori il caso in cui l'aumento indotto delle importazioni è tale da controbilanciare ampiamente l'iniziale aumento esogeno delle esportazioni, per cui la bilancia dei pagamenti nella situazione finale passa in deficit. Il caso di un deficit iniziale della bilancia dei pagamenti dovuto a una riduzione dell'export è del tutto simmetrico.
La situazione di deficit dovuta a un aumento esogeno delle importazioni è più complessa da esaminare. Per analizzare rigorosamente questo e altri simili problemi occorre basarsi sulle formule dei moltiplicatori in mercato aperto, di cui ci limitiamo a riportare le conclusioni principali. Secondo l'opinione tradizionale, il meccanismo moltiplicativo non riuscirà mai a riequilibrare completamente la bilancia dei pagamenti, ma essa è basata su un'ipotesi che non ha validità generale. Tale ipotesi è che la propensione marginale alla spesa globale da parte dei residenti sia minore dell'unità. Per evitare confusioni occorre sottolineare che tale propensione è data dalla somma della propensione marginale al consumo e della propensione marginale all'investimento; entrambe sono sicuramente minori dell'unità, ma la loro somma potrebbe essere maggiore dell'unità (purché sempre minore di un certo valore critico che non può essere oltrepassato pena l'instabilità del sistema). Soltanto nel caso di investimenti interamente esogeni (per cui la propensione marginale all'investimento risulterebbe nulla) l'ipotesi tradizionale è sicuramente valida. In caso contrario, l'aggiustamento della bilancia dei pagamenti può essere completo o addirittura sovrabbondante a seconda che, rispettivamente, la propensione marginale alla spesa globale da parte dei residenti sia uguale o maggiore dell'unità.Nella trattazione precedente, il complesso dei paesi costituenti il resto del mondo - da cui proviene la domanda di esportazioni e verso cui si dirige la domanda d'importazioni del paese in questione (d'ora in poi indicato come paese 1) - è stato considerato esogeno. Ma, così facendo, si trascurano alcuni fenomeni: le importazioni del paese 1, infatti, sono le esportazioni di uno o più paesi del resto del mondo e quindi entrano nel reddito nazionale di questi, da cui dipendono le loro importazioni e quindi anche le esportazioni del paese 1. Pertanto, ad esempio, un aumento del reddito nel paese 1 provoca, tramite l'aumento delle importazioni del paese 1 stesso, cioè delle esportazioni di uno o più

paesi del resto del mondo, un aumento del reddito di questi, donde un aumento delle loro importazioni e quindi anche un aumento delle esportazioni del paese 1. Dal che segue un effetto espansivo sul reddito del paese 1 e via seguitando con una catena di ripercussioni la quale, una volta esauritasi (si suppone che il processo sia convergente), darà luogo a un ben definito effetto finale, certamente diverso da quello che si sarebbe verificato in assenza di tali ripercussioni.
Un'analisi compiuta di tali fenomeni - che vanno sotto il nome di 'ripercussioni internazionali' - richiede un modello a n paesi, matematicamente assai complesso. È però possibile affermare che le principali conclusioni del modello semplificato rimangono confermate da questo modello più generale.

c) L'approccio integrato

Abbiamo visto nei paragrafi precedenti il ruolo del tasso di cambio e del reddito nel processo di aggiustamento della bilancia dei pagamenti. I tentativi di fusione tra i due approcci, in un contesto più ampio in cui l'aggiustamento può avvenire simultaneamente - sia dal lato del tasso di cambio che dal lato del reddito -, rappresentano il successivo passo avanti della teoria tradizionale. Riteniamo che in questo contesto possano essere inquadrati sia il cosiddetto 'approccio-assorbimento' (absorption approach) sia il modello di Laursen e Metzler, nei quali entrano in gioco in modo essenziale sia le elasticità che i moltiplicatori. La conclusione principale di questa analisi è che la condizione delle elasticità critiche non è più sufficiente ad assicurare il riequilibrio della bilancia dei pagamenti, ma occorrono condizioni più complicate. Concludiamo questo capitolo sui processi tradizionali osservando che gli schemi di interazione fra tasso di cambio e reddito nei processi di aggiustamento della bilancia dei pagamenti possono anche essere applicati all'analisi di un fenomeno assai complesso, quello della cosiddetta 'curva a J'. Supponiamo che tutte le opportune condizioni perché una svalutazione migliori la bilancia dei pagamenti (tenuto conto degli effetti di assorbimento) siano verificate, e che invece nella realtà si osservi che una svalutazione provoca inizialmente un peggioramento della bilancia dei pagamenti, la quale soltanto dopo un intervallo di tempo più o meno lungo comincia a migliorare (immaginando di riportare l'andamento nel tempo della bilancia dei pagamenti in un sistema di riferimento cartesiano, con il tempo in ascisse e la bilancia dei pagamenti in ordinate si ha per l'appunto un andamento dapprima decrescente e poi crescente, che somiglia alla lettera J). L'apparente

contraddizione può essere sanata ricorrendo a un'analisi dinamica, in cui si tiene conto delle velocità d'aggiustamento delle diverse variabili, che reagiscono non già istantaneamente, ma con ritardi di diversa entità al mutare delle condizioni e in particolare alle variazioni del tasso di cambio, come dimostrano le vicende della bilancia dei pagamenti statunitense a seguito della massiccia rivalutazione del dollaro nella prima metà degli anni ottanta e dell'altrettanto massiccia svalutazione nella seconda metà.

10. Gli schemi moderni

Come abbiamo accennato (v. § 8b), esamineremo qui i processi di aggiustamento della bilancia dei pagamenti in cui la moneta entra in gioco in maniera essenziale e in cui vengono presi in considerazione anche gli squilibri degli stocks.
L'introduzione della moneta e del saggio dell'interesse nello schema keynesiano tradizionale - che avviene estendendo a un'economia aperta il modello macroeconomico IS-LM di economia chiusa (v. § 10a) - non altera invece la visione degli squilibri della bilancia dei pagamenti come squilibri di flusso e dei suoi aggiustamenti come aggiustamenti dei flussi. Il ruolo del mercato monetario è infatti quello di determinare il saggio dell'interesse da cui dipendono, tra l'altro, i movimenti di capitale, intesi come flussi puri e non come flussi derivanti da aggiustamenti di stock.Il superamento delle varie visioni avviene mediante l'analisi delle relazioni fra equilibrio di portafoglio ed equilibrio macroeconomico in economia aperta.Nel corso dell'esposizione affronteremo, oltre ai problemi dei meccanismi di aggiustamento 'spontanei', anche quelli connessi agli interventi di politica economica per conseguire l'equilibrio interno (piena occupazione) ed esterno (equilibrio della bilancia dei pagamenti).

a) Il modello Mundell-Fleming

Questo schema, così chiamato perché venne elaborato in modo indipendente da Robert Mundell e Marcus Fleming all'inizio degli anni sessanta (esso ebbe grande diffusione soprattutto a opera del primo), estende al mercato aperto lo schema IS-LM elaborato da John Hicks per presentare il modello macroeconomico keynesiano completo di economia chiusa. Com'è noto, la scheda IS indica tutte le combinazioni di tasso dell'interesse e di reddito nazionale che assicurano l'equilibrio sul mercato reale, equilibrio determinato per l'appunto

dall'uguaglianza fra investimento e risparmio; in economia aperta tale scheda viene arricchita con l'introduzione dell'import e dell'export di merci. La scheda LM denota tutte le combinazioni di tasso dell'interesse e di reddito nazionale che assicurano l'equilibrio sul mercato monetario, determinato per l'appunto dall'uguaglianza fra domanda di liquidità monetaria e stock di moneta esistente. A queste relazioni Mundell aggiunge la relazione che rappresenta l'equilibrio della bilancia dei pagamenti, intesa ora come bilancia dei pagamenti globale, comprensiva cioè anche dei movimenti di capitale privati. Questi ultimi sono posti in funzione del differenziale d'interesse: un tasso dell'interesse nazionale maggiore di quello estero provocherà, ceteris paribus, un afflusso di capitali dall'estero, mentre vi sarà un deflusso di capitali se il tasso dell'interesse nazionale è inferiore a quello estero. Abbiamo detto ceteris paribus, perché le aspettative di variazione del cambio possono modificare la direzione dei flussi di capitali rispetto a quella dianzi indicata, ma il problema delle aspettative non viene considerato nello schema in esame.Il problema dell'aggiustamento della bilancia dei pagamenti e più in generale quello dell'equilibrio macroeconomico interno ed esterno si pone in maniera diversa a seconda che i cambi siano fissi o flessibili. Cominciamo dal primo caso e vediamo come funziona il meccanismo di aggiustamento in assenza d'interventi di politica economica, considerando ad esempio un deficit della bilancia dei pagamenti. Tale deficit provoca, data l'ipotizzata assenza d'interventi di sterilizzazione da parte delle autorità monetarie, una riduzione della quantità di moneta e quindi, data la relazione che esprime l'equilibrio monetario, un aumento del saggio dell'interesse. Questo avrà un duplice effetto favorevole sulla bilancia dei pagamenti: da un lato il maggior tasso dell'interesse nazionale rispetto a quello estero (che è per ipotesi rimasto fermo) attirerà capitali, dall'altro questo aumento del tasso dell'interesse deprimerà gli investimenti e quindi il reddito, donde minori importazioni di beni. Come si vede da questo semplice caso, il riequilibrio della bilancia dei pagamenti viene ottenuto a spese del reddito e quindi dell'occupazione (conclusione questa a cui ci portava anche la teoria tradizionale del moltiplicatore in mercato aperto). Quindi il cosiddetto dilemma fra piena occupazione ed equilibrio della bilancia dei pagamenti sembra rimanere privo di soluzione. Senonché Mundell mostrò che, mediante una combinazione appropriata di politica fiscale e di politica monetaria, è in linea di principio possibile perseguire entrambi gli obiettivi, purché si adotti il criterio generale dell'assegnazione di ciascuno strumento di politica economica al perseguimento dell'obiettivo per il quale lo strumento stesso abbia l'efficacia relativa

maggiore. Poiché il tasso dell'interesse, e dunque la politica monetaria, ha un'efficacia sulla bilancia dei pagamenti comparativamente maggiore di quella della politica fiscale (grazie all'effetto del saggio dell'interesse stesso sui movimenti di capitale), la 'ricetta' mundelliana per risolvere l'anzidetto dilemma consiste nel combinare una politica fiscale espansiva (che stimola la produzione a spese della bilancia dei pagamenti) con una politica monetaria restrittiva (che attira capitali consentendo di riequilibrare la bilancia dei pagamenti: pur avendo gli alti tassi dell'interesse effetti negativi sulla produzione, questi non saranno mai tali da compensare gli effetti positivi della politica fiscale espansiva).
Veniamo ora al caso dei cambi flessibili. In questa situazione la flessibilità del cambio s'incarica di mantenere in equilibrio la bilancia dei pagamenti (posto naturalmente che le opportune condizioni riguardanti le elasticità siano verificate), per cui il problema del perseguimento dell'obiettivo occupazionale con politiche espansive è di più facile soluzione. Tuttavia, nel caso di perfetta mobilità dei capitali, solo uno dei due strumenti di politica economica risulta efficace, e precisamente la politica monetaria. Infatti, un'espansione monetaria tende a far diminuire il tasso dell'interesse, il che provocherà un deflusso di capitali potenzialmente illimitato e quindi un deficit della bilancia dei pagamenti; a questo punto il cambio si svaluta automaticamente, il che da un lato stimola la produzione (grazie all'effetto espansivo sulle esportazioni) e dall'altro riequilibra la bilancia dei pagamenti. Nel caso di un'espansione fiscale, invece, l'aumento del reddito, a parità di quantità di moneta, tende a far aumentare il saggio dell'interesse e dunque a far affluire un ammontare potenzialmente illimitato di capitali con conseguente rivalutazione del cambio, il che annulla (tramite l'effetto depressivo sulle esportazioni) l'iniziale effetto espansivo della politica fiscale. Si può anche dimostrare, con ragionamenti simili, che nel caso di cambi fissi e perfetta mobilità dei capitali (il caso di cambi fissi precedentemente esaminato presupponeva una mobilità dei capitali non perfetta) vale il risultato opposto (piena efficacia della politica fiscale, assoluta inefficacia della politica monetaria per quanto riguarda l'espansione del prodotto nazionale).Il modello Mundell-Fleming, pur con tutte le sue ipotesi semplificatrici e nonostante le critiche cui può essere sottoposto, rimane un utile punto di partenza per l'esame degli schemi più raffinati di cui si dirà avanti (v. § 10c). Prima di passare a questi, tuttavia, occorre accennare al cosiddetto approccio monetario alla bilancia dei pagamenti.

b) L'approccio monetario alla bilancia dei pagamenti

La prima formulazione compiuta della teoria classica del meccanismo di aggiustamento della bilancia dei pagamenti basato sui flussi di moneta (oro) è comunemente associata al nome di Hume (1752), sebbene i più importanti elementi costitutivi della teoria classica fossero già stati enunciati da autori precedenti. In sintesi, tale teoria può essere così descritta: un saldo positivo della bilancia dei pagamenti provoca un ingresso di oro nel paese, e quindi - vigente un legame rigido tra riserve auree e quantità di moneta - un aumento dei prezzi (si suppone valida la teoria quantitativa della moneta). Tale aumento da un lato tende a ridurre le esportazioni, dato che i beni del paese in questione - essendo fisso il cambio - divengono relativamente più cari sul mercato internazionale, e dall'altro tende a stimolare le importazioni, dato che i beni esteri divengono relativamente più a buon mercato. Si verifica, perciò, una graduale riduzione del saldo positivo della bilancia dei pagamenti. Un ragionamento analogo spiega l'aggiustamento nel caso di un disavanzo: si ha un'uscita di oro che causa una diminuzione della quantità di moneta e una riduzione dei prezzi interni, con conseguente stimolo sulle esportazioni e riduzione delle importazioni, e porta quindi a una graduale eliminazione del disavanzo stesso.
Con un salto di oltre duecento anni, i fautori dell'approccio monetario alla bilancia dei pagamenti (che ha avuto origine negli anni sessanta e grande diffusione negli anni settanta) dichiarano di ricollegarsi a Hume. Tale affermazione è in parte esatta e in parte sbagliata. È esatta nella misura in cui tale approccio individua la causa ultima degli squilibri della bilancia dei pagamenti in uno squilibrio fra stocks monetari e cioè in una divergenza fra la quantità di moneta esistente e una quantità 'ottima' o desiderata (da definire opportunamente). È sbagliata, invece, in quanto secondo tale approccio il meccanismo di aggiustamento è basato su di un effetto diretto degli squilibri monetari sulle funzioni di spesa (domanda di beni e servizi) senza alcun intervento dei prezzi relativi (che anzi ritiene fissati a un livello determinato dalla parità dei poteri di acquisto): abbiamo infatti visto che per Hume il meccanismo di aggiustamento agisce proprio tramite variazioni dei prezzi relativi.
L'idea di base dell'approccio monetario alla bilancia dei pagamenti è che eventuali squilibri monetari si ripercuotono sulla spesa aggregata per beni e servizi (assorbimento), nel senso che un eccesso di offerta di moneta provoca - ceteris paribus - un aumento dell'assorbimento e, viceversa, un eccesso di domanda di moneta provoca una riduzione dell'assorbimento stesso. Il collegamento della domanda di moneta con le altre variabili macroeconomiche è a sua volta immediato, data l'ipotesi che essa sia una funzione stabile di poche di tali variabili;

risulta quindi semplice determinare, noto lo stock di moneta esistente, l'eccesso di domanda e di offerta di moneta.
D'altra parte, la divergenza fra reddito e assorbimento che si viene così a creare si traduce necessariamente in acquisizione o cessione di attività finanziarie da parte del pubblico: infatti, tale divergenza equivale a una divergenza fra risparmio e investimento e dunque, dato il vincolo di bilancio del settore privato, a una variazione degli stocks di attività finanziarie detenute da questo settore. Se introduciamo l'ipotesi semplificatrice che l'unica attività finanziaria sia la moneta, si viene a verificare una variazione dello stock di moneta che a sua volta coincide con il saldo globale della bilancia dei pagamenti.
Ciò che è in definitiva accaduto, attraverso questa concatenazione di effetti, è che un eccesso di offerta o di domanda di moneta si è 'scaricato' sulla bilancia dei pagamenti causando un eccesso o difetto di assorbimento rispetto al reddito (prodotto) nazionale: un eccesso di assorbimento significa un deficit di bilancia dei pagamenti (l'unico modo di assorbire più di quanto si produce è infatti ricevere dall'estero più di quanto si fornisce all'estero) e un difetto di assorbimento significa un surplus di bilancia dei pagamenti. In altri termini, se la collettività si trova con più moneta rispetto allo stock desiderato, se ne sbarazza cedendo tale eccesso all'estero in cambio di beni e servizi (deficit di bilancia dei pagamenti); se invece la collettività desidera più moneta rispetto allo stock esistente, se la procura riducendo l'assorbimento e in definitiva cedendo all'estero beni e servizi in cambio di moneta (surplus di bilancia dei pagamenti).

c) Sviluppi recenti

Gli studiosi più avveduti sono sempre stati consapevoli del fatto che la contrapposizione tra squilibri di stock e squilibri di flusso e quella fra squilibri monetari e squilibri reali non servono a interpretare la realtà, nella quale interagiscono fenomeni reali e monetari, di flusso e di stock. Ciò comincia a essere generalmente riconosciuto negli sviluppi più recenti, i quali si basano sull'integrazione fra equilibrio macroeconomico ed equilibrio di portafoglio nell'ambito di uno schema più generale di quelli precedentemente considerati. Inoltre, nel caso in cui i cambi siano flessibili, non si possono non considerare le aspettative sull'andamento futuro del cambio, che entrano in maniera essenziale nella determinazione dei movimenti di capitali. Infatti, la decisione di convenienza sull'investimento finanziario nell'uno o nell'altro paese dipende non soltanto dai differenziali d'interesse, ma anche - a

parità di rischio - dalle previsioni sulla svalutazione o rivalutazione futura delle diverse valute. Si tratta, in conclusione, di modelli molto complessi, per la cui trattazione rinviamo alla bibliografia (v. Gandolfo, 1986, vol. II, capp. 15 e 16).

11. La determinazione del tasso di cambio

La più antica teoria del tasso di cambio è forse la teoria della 'parità del potere d'acquisto' (PPA; viene anche usata la sigla PPP dalle iniziali della dizione inglese - Purchasing Power Parity - attribuita a Karl Gustav Cassel, 1918). Si distinguono comunemente due versioni della PPA: quella assoluta, secondo cui il tasso di cambio fra le valute di due paesi è uguale al rapporto fra i livelli dei prezzi dei due paesi stessi, e quella relativa, secondo cui le variazioni percentuali del tasso di cambio in questione sono uguali alle variazioni percentuali degli indici generali dei prezzi dei due paesi considerati.In entrambe le versioni, la PPA si propone come una teoria di lungo periodo del tasso di cambio di equilibrio, nel senso che nel breve periodo vi possono essere deviazioni anche consistenti da essa, che però mettono in moto forze tali da riportare, nel lungo periodo, il tasso di cambio verso il suo valore di equilibrio. Tali deviazioni, che rendono inutile la PPA per spiegare l'andamento del tasso di cambio nel breve periodo (studi recenti hanno peraltro constatato deviazioni anche nel periodo meno breve), indussero gli studiosi ad abbandonarla a favore dell'approccio del mercato dei cambi; occorre però osservare che essa è stata recentemente ripresa dall'approccio monetario, di cui diremo oltre.
L'approccio del mercato dei cambi, detto anche 'della bilancia dei pagamenti', parte dall'osservazione che il tasso di cambio viene a determinarsi sul mercato dei cambi esteri in base alle domande e alle offerte delle varie valute e si muove per equilibrare tali domande e offerte e quindi, supponendo assenza d'interventi, la bilancia dei pagamenti. Secondo la definizione di Ragnar Nurkse (1945), infatti, il tasso di cambio di equilibrio è quello che "mantiene la bilancia dei pagamenti in equilibrio in un certo arco di tempo".Che il tasso di cambio si determini sul mercato dei cambi in base alle domande e alle offerte delle valute estere è un dato di fatto irrefutabile, ma è proprio nel determinare tali domande e offerte che sorgono i problemi.L'approccio tradizionale o di flusso vede tali domande e offerte come puri flussi, derivanti, nella visione più antica, dalle esportazioni e importazioni di beni, a loro volta dipendenti dal tasso di cambio e, dopo la modellistica keynesiana, anche dal reddito.

Tale approccio è già stato descritto in precedenza, in particolare nel cap. 9, al quale rinviamo. L'introduzione dei movimenti di capitale come ulteriore componente delle domande e offerte di valute sul mercato dei cambi non altera questa visione ove tali movimenti siano anch'essi visti come puri flussi: tale è il caso del modello descritto nel § 10a.
Questo approccio è criticabile per una serie di motivi, in particolare perché esso trascura gli aggiustamenti di stock. Desideriamo però sottolineare un punto importante, e cioè che queste critiche possono indurre a ritenere inadeguato l'approccio tradizionale nella sua individuazione delle determinanti delle domande e offerte di valuta estera, ma non toccano il fatto che sono queste domande e offerte a determinare il tasso di cambio. Comunque, le critiche suddette hanno portato alla formulazione del cosiddetto approccio 'moderno' degli stocks di attività finanziarie.
In tale approccio (detto anche asset market approach, cioè approccio basato sui mercati delle attività finanziarie) il tasso di cambio viene visto come il prezzo relativo di due stocks di moneta (approccio monetario) o come il prezzo relativo di due stocks di titoli (approccio di portafoglio). I due schemi si differenziano in base alle ipotesi fatte sulla sostituibilità dei titoli, data però la comune ipotesi della perfetta mobilità dei capitali. Nell'approccio monetario si assume perfetta sostituibilità fra titoli nazionali e titoli esteri (per cui diventa irrilevante per gli investitori detenere gli uni o gli altri e altrettanto irrilevante diventa l'offerta di tali titoli), mentre nell'approccio di portafoglio si assume imperfetta sostituibilità fra detti titoli, la cui quantità emessa diventa quindi rilevante.
L'approccio moderno conduce a conclusioni spesso diametralmente opposte a quelle dell'approccio tradizionale. Considerando ad esempio l'approccio monetario, è possibile dimostrare che un aumento dello stock di moneta nazionale conduce a una svalutazione del cambio, mentre un aumento del reddito nazionale ne provoca una rivalutazione, e ancora che un aumento del tasso dell'interesse nazionale provoca una svalutazione. Si tratta di conclusioni (specie le ultime due) in netto contrasto con l'approccio tradizionale, ove un aumento del reddito, facendo aumentare le importazioni, tende a far svalutare il cambio, mentre un aumento del tasso dell'interesse, attirando capitali, ha effetti rivalutativi. Tali diverse conclusioni sono peraltro del tutto ovvie secondo la visione dell'approccio monetario alla bilancia dei pagamenti già chiarita nel § 10b. Un aumento del reddito, ad esempio, fa aumentare la domanda di moneta; dati lo stock di moneta e il livello dei prezzi, la collettività cercherà di procurarsi la desiderata liquidità aggiuntiva riducendo l'assorbimento, il che tende

a far andare in surplus la bilancia dei pagamenti, donde la rivalutazione del cambio. Tale rivalutazione, riducendo simultaneamente il livello dei prezzi data la PPA, fa aumentare il valore reale del dato stock di moneta e quindi ripristina l'equilibrio monetario. Un ragionamento simile spiega la svalutazione nel caso di un aumento del saggio dell'interesse.Il cosiddetto approccio 'moderno' al tasso di cambio è, a detta dei sostenitori, sicuramente superiore all'approccio tradizionale, sia sotto il profilo teorico che sotto il profilo empirico. Senonché, se è vero per definizione che il tasso di cambio è il prezzo relativo di due monete, è vacuo dire che esso è determinato dalle quantità relative offerte e domandate delle due monete. Per fare un paragone, è anche vero che il prezzo dell'acciaio in dollari è il prezzo relativo tra acciaio e dollari, ma nessuno penserebbe di analizzare la determinazione del prezzo dell'acciaio in termini di domanda e di offerta di moneta. Ciò che si dovrebbe fare è cercare di capire le determinanti della domanda e dell'offerta di acciaio, nonché il meccanismo tramite cui esse vengono equilibrate nel mercato dell'acciaio, determinandone il prezzo. Esattamente le stesse considerazioni valgono per il tasso di cambio, il quale viene in concreto determinato nel mercato dei cambi esteri tramite le domande e offerte di valuta estera. Questa mancata o scarsa considerazione del mercato dei cambi esteri costituisce il principale difetto delle teorie 'moderne', poiché nessuna teoria del tasso di cambio può essere ritenuta soddisfacente se non spiega come le variabili che essa ritiene cruciali (siano queste le quantità di moneta o le aspettative o le attività finanziarie) si traducano in domande e offerte di valuta estera sul mercato dei cambi, in concomitanza con le domande e offerte provenienti da altre fonti.

Ci sembra che questa posizione sia da condividere in pieno. In effetti, a nostro avviso, né la teoria tradizionale né la teoria moderna possono essere ritenute del tutto soddisfacenti, in quanto - come abbiamo già avuto modo di osservare - le determinanti che cerchiamo sono sia reali che finanziarie, derivano sia da puri flussi che da aggiustamenti di stock, in un contesto di interrelazioni reciproche in situazione di squilibrio. Ne segue che soltanto un approccio eclettico può essere ritenuto soddisfacente. In effetti, le verifiche empiriche hanno dimostrato la scarsa o nulla valenza esplicativa dei modelli che presentano una visione unilaterale della determinazione del cambio, mentre l'uso di un modello eclettico in cui il tasso di cambio viene determinato nel contesto di un modello dinamico dell'intero sistema economico ha dato risultati sicuramente migliori (v. Gandolfo, Padoan e Paladino, 1990).

12. Il sistema monetario internazionale e il sistema monetario europeo

a) Gli avvenimenti del secondo dopoguerra

Il sistema monetario internazionale del secondo dopoguerra fu ricostruito sulla base degli accordi di Bretton Woods (1944), che dettero vita tra l'altro al FMI (Fondo Monetario Internazionale) e gettarono le basi per la cooperazione monetaria internazionale. Senza entrare nei dettagli del funzionamento del Fondo, basterà ricordare che i paesi aderenti si accordarono per vincolare le loro valute all'oro o al dollaro (ciò era equivalente, essendo il dollaro convertibile in oro al prezzo ufficiale di 35 dollari per oncia di fino) dichiarando una parità monetaria e impegnandosi a mantenere il corso del cambio entro ristretti margini (%1%) della stessa. Queste caratteristiche qualificavano il sistema come un regime di cambi fissi del tipo gold exchange standard (ancorché temperato dalla possibilità di variare la parità, di cui diremo appresso). Per vari motivi, tuttavia, in parte anche politici, i paesi occidentali utilizzarono in misura molto limitata la facoltà di convertire in oro i dollari detenuti nelle proprie riserve internazionali, sicché il sistema divenne di fatto un dollar standard. La inconvertibilità di fatto del dollaro divenne inconvertibilità de jure nel 1971.

I paesi aderenti al FMI si impegnarono altresì a variare la parità soltanto in caso di 'squilibrio fondamentale' e anche in tal caso seguendo determinate regole: per variazioni fino al 10% vi era piena discrezionalità, mentre per variazioni maggiori occorreva informare preventivamente il Fondo e ottenerne il consenso. Il Fondo, dal canto suo, con la dotazione a sua disposizione, assumeva anche funzioni di sostegno delle valute in difficoltà. Un altro elemento degli accordi fu l'impegno a rendere convertibili le valute eliminando i controlli sui cambi dopo un periodo di transizione (la piena convertibilità fu ripristinata verso la fine degli anni cinquanta). L'impegno a mantenere la parità dichiarata (tipico dei cambi fissi) salvo variarla con le modalità descritte fece denominare il sistema come un regime di adjustable pegs (parità fisse ma aggiustabili). Si trattava di un compromesso fra i cambi rigidamente e permanentemente fissi e i cambi flessibili, ed è chiaro che la maggiore o minore vicinanza all'uno o all'altro regime dipendeva in modo essenziale dall'interpretazione data alle regole per variare la parità. L'interpretazione che venne data nella prassi fu restrittiva, nel senso che la parità andava difesa a oltranza e variata soltanto quando ogni difesa risultava impossibile. Ciò ebbe come conseguenza non secondaria l'applicazione continuata di misure restrittive interne in situazioni di difficoltà della bilancia dei pagamenti e il sorgere di movimenti di

fondi speculativi destabilizzanti. Occorre inoltre segnalare che, mentre negli intendimenti degli accordi di Bretton Woods l'onere dell'aggiustamento degli squilibri della bilancia dei pagamenti doveva essere sostenuto sia dai paesi in deficit sia dai paesi in surplus, nella prassi tale onere è stato di solito sostenuto in larga misura dai paesi deficitari.
Tra gli avvenimenti di maggiore rilievo verificatisi in campo monetario nel dopoguerra sono a nostro avviso da segnalare (in ordine cronologico): a) il ripristino della convertibilità delle valute dei maggiori paesi industrializzati (1958); b) la formazione del mercato dell'eurodollaro (anni cinquanta-sessanta); c) la creazione dei 'diritti speciali di prelievo' in seno al FMI (1968); d) la inconvertibilità de jure del dollaro statunitense (15 agosto 1971) e il successivo abbandono della parità per passare a una fluttuazione da parte di vari paesi (fra cui l'Italia): avvenimenti a proposito dei quali si è parlato di crollo del sistema di Bretton Woods; e) l'acquisizione di enormi surplus (denominati in dollari) da parte dei paesi produttori di petrolio (anni settanta); f) la 'demonetizzazione' dell'oro e la 'legalizzazione' della fluttuazione dei cambi (1976); g) la costituzione del sistema monetario europeo (13 marzo 1979).
Cade qui opportuna un'osservazione: il sistema di Bretton Woods non è stato sostituito da alcun altro sistema, se con tale termine s'intende un insieme preciso di regole, diritti e doveri universalmente accettato. Al momento attuale, infatti, ciascun paese è libero di scegliere il regime di cambi che preferisce, nel rispetto di alcuni obblighi molto vaghi. Abbiamo così paesi o gruppi di paesi che agganciano la loro moneta a una qualche valuta chiave, altri che lasciano fluttuare più o meno liberamente la loro valuta, altri ancora che aderiscono ad accordi di cambio quale il Sistema Monetario Europeo (SME). È su quest'ultimo che concentreremo la nostra attenzione, rinviando alla bibliografia per la descrizione degli altri avvenimenti elencati sopra.

b) Il sistema monetario europeo

Nel 1979 (13 marzo) i paesi della Comunità Economica Europea (con l'esclusione iniziale dell'Inghilterra, poi entrata nell'ottobre 1990) dettero vita, in applicazione degli accordi di Brema dell'anno precedente (7 luglio 1978), ad un'area valutaria, denominata Sistema Monetario Europeo (SME) e basata su un numerario costituito da una moneta paniere chiamata ECU, termine che è simultaneamente un acronimo inglese (dalle iniziali di European Currency Unit) e una parola che indica un'antica moneta francese (lo scudo).Tale

sistema appartiene alla categoria dei regimi a flessibilità limitata con parità dichiarate, fisse ma aggiustabili in modo discreto, e con ampliamento dei margini di oscillazione (che sono del ± 2,25% salvo eccezioni). L'elemento caratterizzante del sistema è proprio l'ECU e il suo ruolo nello SME. Rammentiamo infatti che è soltanto quando si vuole tendere a una vera e propria unione monetaria che sorge l'esigenza di definire un'unità monetaria comune. L'introduzione dell'ECU fu dunque, se si vuole usare un'espressione oramai abusata, una manifestazione di 'volontà politica'. In effetti nelle trattative che condussero allo SME le maggiori controversie si ebbero a proposito delle regole operative che avrebbero dovuto guidare gli interventi delle banche centrali per mantenere le rispettive monete alle parità stabilite. Il primo meccanismo considerato era semplicemente la fissazione di una griglia di parità bilaterali con l'obbligo da parte di ciascun paese d'intervenire quando la propria valuta fosse giunta al limite di oscillazione consentito nei confronti di una qualsiasi altra valuta, in pratica una riedizione del vecchio 'serpente'. Il secondo meccanismo considerato consisteva invece nella costituzione di una moneta 'paniere' (cioè di una moneta composita, costituita da quantità prefissate delle varie monete nazionali dei paesi membri), l'ECU appunto, rispetto alla quale fissare i tassi centrali delle singole valute e i rispettivi margini di fluttuazione consentiti, e quindi anche gli obblighi d'intervento in caso di deviazione. Il compromesso che venne raggiunto consisteva nel mantenere il 'meccanismo griglia' come base per gli obblighi d'intervento, utilizzando tuttavia l'ECU (a) come numerario per definire i tassi centrali della griglia e (b) come base per la costruzione di un 'indicatore di divergenza' e cioè per una specie di campanello d'allarme che avrebbe dovuto segnalare in tempo il fatto che la valuta di un paese stesse cominciando a deviare troppo, creando quindi per quel paese una 'presunzione' (non un obbligo) a intervenire (sul mercato dei cambi o con politiche interne) per frenare la deviazione.Oltre agli accordi di cambio finora accennati, lo SME prevede anche misure di cooperazione monetaria e di sostegno monetario a favore delle valute in difficoltà. Nel momento in cui scatta la presunzione d'intervento o si raggiunge un punto d'intervento obbligatorio, le misure poste in atto sono quelle usuali di un regime di cambi fissi: intervento sui mercati dei cambi da parte delle banche centrali coinvolte, che cederanno o acquisteranno valute estere, politiche monetarie e fiscali interne. È proprio nell'ottica degli interventi sui mercati dei cambi che negli accordi costitutivi dello SME sono previsti anche accordi di cooperazione monetaria (e in particolare di sostegno monetario ai paesi membri che si trovino in difficoltà: a brevissimo, a breve e a medio termine). In tali accordi, inoltre,

era prevista - tra l'altro - la formazione, entro due anni, di un Fondo Monetario Europeo (FME), nel quale avrebbe dovuto confluire l'allora (e tuttora) vigente Fondo Europeo di Cooperazione Monetaria (FECOM). Al momento, tuttavia, il FME non è stato costituito dati i disaccordi tra gli stati membri concernenti la natura del FME stesso. Secondo una proposta 'massimale' esso dovrebbe essere una vera e propria banca centrale sovranazionale in grado di emettere una propria moneta (l'ECU), banca nella quale confluirebbero tutte le riserve internazionali dei paesi membri che verrebbero da essa gestite in modo autonomo nei confronti dei paesi terzi. Secondo la proposta 'minimale' esso dovrebbe essere praticamente simile all'attuale FECOM, che ha un ruolo poco più che contabile, di gestione di una parte delle riserve a esso cedute dalle banche centrali dei paesi membri. Fra i due estremi esiste tutta una gamma di proposte intermedie, ma è chiaro che soltanto muovendosi verso la proposta 'massimale' si potrà realizzare la completa unione monetaria europea.

13. La liquidità internazionale

Con il termine 'liquidità internazionale' si indicano di solito due fenomeni, uno più antico e l'altro più recente. Il primo è costituito dai mezzi di pagamento internazionali a disposizione delle autorità monetarie dei singoli paesi per far fronte a squilibri della bilancia dei pagamenti e cioè alle riserve internazionali (valute convertibili, posizione di riserva sul FMI, diritti speciali di prelievo, ECU, ecc.). Il secondo è costituito dalle cosiddette 'xenovalute'.

a) La domanda di riserve internazionali

Il problema della 'adeguatezza' della liquidità internazionale e quello connesso della domanda di riserve internazionali hanno fatto versare fiumi d'inchiostro e sono ancora ben lungi dall'essere risolti. In questa sede ci limitiamo ad osservare che, in prima approssimazione, la liquidità internazionale si può definire 'adeguata' quando consente il finanziamento dei deficit di bilancia dei pagamenti da parte dei paesi che ne siano colpiti senza che questi siano costretti a misure di aggiustamento 'indesiderabili' per lo sviluppo della loro economia e del commercio internazionale. Questa definizione è stata giustamente criticata, perché condurrebbe inevitabilmente a dire che la liquidità internazionale è sempre inadeguata, poiché nessun ragionevole ammontare di essa sarà sufficiente a consentire il finanziamento senza aggiustamento.

Poiché chi detiene ed utilizza le riserve internazionali sono le banche centrali, il problema diventa allora quello di individuare una funzione di domanda di riserve internazionali: la liquidità internazionale sarà da ritenersi adeguata quando il suo ammontare sarà uguale all'ammontare 'desiderato' in base a tale funzione di domanda. Ciò, come osservò Machlup, è cosa diversa dall'individuazione del fabbisogno di riserve in base a parametri oggettivi (quali il rapporto con le importazioni, ecc.), come si riteneva un tempo, in quanto la domanda, derivando da un desiderio, può non essere giustificabile sulla base di elementi 'oggettivi'. Divertente è a questo proposito la 'regola del guardaroba della signora Machlup', enunciata dallo stesso Machlup, il quale dapprima paragona il desiderio di possedere riserve da parte dei banchieri centrali al desiderio di possedere abiti della propria consorte, emendando poi questa regola nel senso di applicarla alle variazioni: "Ella [la signora Machlup] non dà poi molta importanza al fatto di possedere 25 o 52 abiti, purché ottenga alcuni nuovi abiti ogni anno. Questa ambizione è perfettamente analoga all'ambizione del banchiere centrale. Egli non si preoccupa poi molto del fatto che le sue riserve siano il 47 o il 74% (delle sue passività o dell'offerta di moneta complessiva), purché esse aumentino, sia pur modestamente, e non diminuiscano". La tesi di Machlup è che, per quanto grande sia lo stock di riserve possedute da una banca centrale, essa interverrà con misure di aggiustamento se vede che questo stock sta diminuendo. Se, dunque, si vogliono evitare tali misure, l'unico modo è far aumentare lo stock di riserve.
La tesi di Machlup non è stata però universalmente accettata e gli studiosi si sono mossi nella direzione d'individuare le determinanti della domanda di riserve internazionali da parte delle banche centrali. In tale contesto è possibile individuare due modi principali di affrontare il problema: l'approccio interpretativo o descrittivo e l'approccio ottimizzante. Il primo si propone d'individuare, anche mediante analisi empiriche, le determinanti che spiegano la domanda di riserve effettivamente promanante dalle autorità monetarie. Il secondo si propone invece d'individuare, in base a un qualche criterio di ottimalità, il livello 'ottimo' delle riserve internazionali, a prescindere dal fatto se poi in concreto tale livello 'ottimo' sia quello domandato. Volendo ricorrere a una distinzione ben nota in teoria economica generale, potremmo dire che il primo approccio è sostanzialmente di economia positiva, il secondo sostanzialmente di economia normativa.

b) Le xenovalute

Punto di partenza è il mercato dell'eurodollaro, che ebbe origine negli anni cinquanta a seguito di depositi in dollari raccolti da banche europee e da esse utilizzati per concedere crediti anch'essi denominati in dollari. Rammentiamo che, in generale, banche di un dato paese europeo possono raccogliere depositi e concedere prestiti anche in valute diverse dal dollaro (e, ovviamente, diverse dalla valuta del paese in cui la banca è residente); si parla in tal caso di eurovalute. Ancora più in generale, dato che operazioni consimili (raccolta di depositi e concessione di crediti in valute diverse da quella del paese in cui la banca è residente) possono essere fatte anche da banche non europee, è stata proposta da Machlup la denominazione di xenovalute.

La spiegazione analitica del grande sviluppo dei depositi di eurodollari e più in generale degli xenomercati, le cui dimensioni quantitative sono mostruose (secondo le stime della Morgan Guaranty Trust Co. a fine 1988 si aggiravano intorno ai 3.000 miliardi di dollari), ha dato luogo a un acceso dibattito teorico. Le teorie dell'espansione degli xenodepositi si possono classificare, pur con tutte le peculiarità di ciascuna, in due categorie. La prima è la cosiddetta scuola tradizionale, che spiega gli xenodepositi mediante un meccanismo moltiplicativo analogo a quello che viene utilizzato nella teoria monetaria tradizionale per spiegare la 'creazione' di depositi bancari nazionali in un'economia chiusa. La seconda rifiuta l'approccio dei moltiplicatori fissi e si rifà invece all'approccio di portafoglio, similmente a quanto viene fatto dalla teoria 'moderna' dell'offerta di moneta nazionale in un'economia chiusa.

La diversità di visione teorica ha profonde implicazioni sulla valutazione dei costi e benefici degli xenomercati. Gli aderenti alla scuola tradizionale, infatti, ritengono che la crescita di tali mercati e quindi della liquidità internazionale sia stata responsabile, secondo una visione monetarista, dell'accelerazione dell'inflazione internazionale avvenuta a partire dai primi anni settanta e durata fino ai primi anni ottanta. Inoltre, essi continuano, tali mercati sono basati su una 'piramide di carta' che potrebbe crollare da un momento all'altro con conseguenze catastrofiche, anche perché, a differenza dei sistemi bancari nazionali, non esiste un 'prestatore di ultima istanza'. Se invece si accetta la visione 'moderna', si deve ammettere che la capacità dell'euromercato di 'creare' liquidità è invero limitata, per cui le cause dell'aumento della liquidità internazionale vanno cercate altrove. È comunque generale la preoccupazione che gli xenomercati possano alimentare flussi di 'moneta calda' e cioè movimenti internazionali di capitale aventi natura destabilizzante: date le dimensioni di

questi mercati nessuna banca centrale (o anche pool di banche centrali) sarebbe in grado di farvi fronte. Accanto agli aspetti negativi vengono naturalmente individuati anche degli aspetti positivi, quali l'aumento dell'efficienza dei mercati internazionali dei capitali, la funzione di intermediazione del mercato, che va a favore non solo degli operatori privati, ma anche delle banche centrali, ecc.In definitiva, come in ogni caso di fenomeni che presentano costi e benefici, è difficile, se non impossibile, determinare se il saldo sia positivo o negativo.

Economia-mondo
di Michele Rostan

sommario: 1. Il termine e la sua origine. 2. Il concetto di economia-mondo nell'opera di Braudel. 3. Il concetto di economia-mondo nell'opera di Wallerstein. 4. Conclusioni. □ Bibliografia.

1. Il termine e la sua origine

Il termine 'economia-mondo' è la traduzione del termine inglese world-economy e di quello francese économie-monde. Quest'ultimo è tratto, a sua volta, dall'espressione tedesca Weltwirtschaft. Fernand Braudel fu il primo a utilizzare il termine économie-monde (v. Braudel, 1949), prendendolo a prestito da un libro dello storico ed economista Fritz Rörig (v., 1933). Adottato "in mancanza di meglio" per tradurre un uso particolare dell'espressione tedesca, il termine ha accompagnato lo storico francese nella sua pluridecennale attività. Negli anni settanta esso è stato ripreso dal sociologo americano Immanuel M. Wallerstein che lo ha utilizzato nella sua opera The modern world-system. Da questo momento il termine world-economy è al centro della ricerca e dell'elaborazione di Wallerstein e del Fernand Braudel Center for the Study of Economies, Historical Systems and Civilizations da lui fondato nel 1976.

2. Il concetto di economia-mondo nell'opera di Braudel

Braudel (v., 1949) ha dedicato vent'anni della sua vita, grosso modo tra il 1930 e il 1950, a una ricerca sul mondo mediterraneo, e i successivi trent'anni a una seconda impresa, di grande portata, che ha dato come frutto un'opera in tre volumi (v. Braudel, 1979). Prima della pubblicazione definitiva di quest'opera, egli ne presentò e ne riassunse i contenuti in tre conferenze, tenute alla Johns Hopkins University di Baltimora, i cui testi sono raccolti in un volumetto più agile (v. Braudel, 1977). Analizzando queste tre opere è possibile capire che cosa Braudel intenda per economia-mondo e a quali fini egli abbia utilizzato tale concetto.
Nella prima edizione del libro sul Mediterraneo, Braudel espone il suo programma scientifico. Egli si batte in favore della convergenza di due scienze sociali, la geografia e la storia, che hanno uno scopo comune, studiare l'uomo, la società, gli Stati, utilizzando due mezzi, lo spazio e il tempo. Tale programma è realizzabile obbligando i geografi a prestare maggiore attenzione al divenire della storia, e gli storici a tener adeguato conto della formidabile permanenza dello spazio.
Per scrivere la storia del Mediterraneo nella seconda metà del Cinquecento, Braudel si allontana dagli schemi della storia diplomatica cercando di costruire una storia diversa, in cui si possano distinguere un tempo geografico, un tempo sociale e un tempo individuale. Il tempo della storia viene, così, scomposto in tre piani sovrapposti. A livello più profondo, c'è una storia quasi immobile, della quale lo spazio, inteso come ambiente geografico, è l'attore principale. Al di sopra di questa storia, in cui le trasformazioni avvengono assai lentamente, si svolge una storia sociale o strutturale. Infine, c'è il livello più superficiale, quello della storia tradizionale, la storia degli avvenimenti.
Ai nostri fini, è necessario rivolgere l'attenzione al piano intermedio. Qui, lo storico affronta la storia dei gruppi, delle strutture e dei destini collettivi, cioè i movimenti d'insieme che coinvolgono tutte le forme della vita. Il suo difficile compito consiste nel dividerle e nel capire in quale modo esse agiscano e reagiscano le une sulle altre.Le forme di vita prese in considerazione nello studio del mondo mediterraneo sono, nell'ordine, le economie, gli Stati, le società, le civiltà e, infine, le forme della guerra. Tale suddivisione è uno strumento di lavoro utilizzato nella consapevolezza che nessuna forma determina pienamente le altre. La vita rimane una, ma per illuminarla è necessario procedere all'operazione, "forzatamente convenzionale e che non appaga mai perfettamente", di "ritagliarla a nostro talento" (v. Braudel, 1949; tr. it., p. 412).
Il punto di partenza della storia sociale del mondo mediterraneo è costituito dall'analisi di

come lo spazio influenzi la vita degli uomini del XVI secolo. A costoro il Mediterraneo si presenta come uno spazio immenso e smisurato in cui il problema principale è rappresentato dalle distanze, dal tempo e dai mezzi necessari a coprirle, sia che a muoversi siano viaggiatori e informazioni, oppure merci e denaro. Tale problema riguarda sia l'attività politica, il governo dell'Impero, sia l'attività economica. "Il problema delle distanze - scrive Braudel - non si pone soltanto agli Stati, ma anche alle economie. L'economia moderna tende sempre più a divenire mondiale, via via che le distanze diminuiscono, con una ripartizione più o meno perfetta dello spazio in zone economiche complementari, legate tra loro, gerarchizzate in un ordine geograficamente disegnato [...]. I Tedeschi chiamano Weltwirtschaft lo spazio economico organizzato alla maniera del mondo odierno, per concludere che il Mediterraneo del secolo XVI ne costituiva da solo uno. Al che si potrebbe rispondere: sì e no, [...] sì, perché, per l'essenziale, viveva anche su se stesso; perché aveva le sue zone particolari adattate alla vita generale che circolava a fianco, al di sopra, attraverso quei piccoli universi economici, mai completamente chiusi." (ibid., pp. 434-435).
Le caratteristiche di un'economia-mondo sono, dunque, rintracciabili già nella prima edizione de La Méditerranée. Innanzitutto, l'economia-mondo è un 'ritaglio' nello spazio, un modo per 'render sensibile' allo storico lo spazio smisurato del Mediterraneo. L'economia-mondo è un universo in sé, con propri confini, distinguibile da altri universi. Tali confini sono identificabili in base al fatto che un'economia-mondo vive su se stessa e intrattiene scambi commerciali con altre economie-mondo solo per il superfluo. In secondo luogo, l'economia-mondo si presenta come uno spazio economico organizzato in zone complementari e dotato di un centro. Per Braudel, il cuore industriale ed economico dell'economia-mondo mediterranea è costituito da quattro città, Genova, Milano, Venezia, Firenze, e da una serie di loro satelliti urbani. "Il quadrilatero dominava la vita del mare ancora nel secolo XVI. La grande economia era organizzata, nella misura in cui poteva esserlo, proprio in funzione di quella esigua zona." (ibid., p. 454).
L'idea che l'economia-mondo sia organizzabile in zone concentriche complementari è tratta da un'opera dell'economista tedesco Johann H. von Thünen, alla quale Braudel fa riferimento indiretto in una nota del libro sul Mediterraneo per poi riprenderla in modo più esteso nel terzo volume della Civilisation.
In conclusione, ne La Méditerranée il concetto di economia-mondo ha sia una funzione analitica sia una funzione descrittiva. Ha una funzione descrittiva perché l'economia-mondo

mediterranea è un oggetto concreto di cui è possibile scrivere la storia, quella delle strutture di più lunga durata come quella degli avvenimenti politici dell'età di Filippo II. Esso, però, ha anche una funzione analitica perché permette allo storico di costruire l'oggetto della sua ricerca, di effettuare comparazioni e di tracciare grandi quadri d'insieme di durata secolare.
La seconda opera dello storico francese (v. Braudel, 1979) è stata realizzata con lo scopo di scrivere una storia economica universale del periodo che va dal Quattrocento all'inizio dell'Ottocento. Per scrivere questa storia, Braudel ricorre a uno schema tripartito secondo il quale la realtà economica è costituita dalla sovrapposizione di tre livelli distinti: la vita materiale, l'economia di mercato e il capitalismo. I tre livelli corrispondono, rispettivamente, alla sfera dell'autoconsumo familiare o di villaggio, e a due distinte sfere di circolazione. Nella prima prevalgono i traffici locali o a breve distanza, i mercati sono spazi aperti tanto ai piccoli quanto ai grandi mercanti, gli scambi sono trasparenti e i profitti bassi. Nella seconda, invece, prevale il commercio a lunga distanza; la concorrenza, regola essenziale dell'economia di mercato, ha poco spazio; tra produttori e consumatori si frappongono lunghe catene mercantili grazie alle quali mercanti/intermediari impongono l'effettuazione di scambi ineguali realizzando alti profitti.Il concetto di economia-mondo viene ripreso e precisato sia nel terzo volume della Civilisation sia negli Afterthoughts. L'economia-mondo - intesa come "l'economia di una parte del nostro pianeta, a condizione che essa formi una totalità, un insieme" (v. Braudel, 1977; tr. it., p. 90) - viene distinta dall'economia mondiale in quanto "mercato dell'universo". Secondo Braudel, un'economia-mondo occupa un dato spazio geografico e i limiti che la individuano variano lentamente. Essa presuppone sempre un polo, un centro, rappresentato da una città dominante. Intorno a questo polo si collocano delle aree intermedie e poi delle zone periferiche che, nella mappa della divisione del lavoro che caratterizza l'economia-mondo, si trovano in una posizione subordinata e dipendente, proprio a causa della loro posizione geografica.
Queste sono, in sintesi, le tre regole tendenziali che definiscono i rapporti delle economie-mondo con lo spazio. Esse sono discusse ampiamente nel terzo volume della Civilisation, dove vengono fornite alcune precisazioni. In primo luogo, una città dominante non occupa la posizione centrale illimitatamente, ma viene rimpiazzata in questo ruolo da altre città. In secondo luogo, le zone in cui è suddivisa un'economia-mondo sono gerarchizzate in modo tale che la disuguaglianza tra le diverse zone assicuri il funzionamento dell'economia-mondo nel suo insieme. Tali zone formano, sulla base di legami commerciali, un insieme coerente: i

mercati locali e regionali dell'economia-mondo sono periodicamente integrati e riorganizzati razionalmente a profitto di una zona o di una città dominanti. In terzo luogo, lo schema spaziale dell'economia-mondo definisce i caratteri peculiari di ciascuna zona: le caratteristiche della società, dell'economia, della tecnica, della cultura e dell'ordine politico mutano passando da una zona all'altra.

Alle tre regole formulate in base alla dimensione spaziale, è necessario affiancare un'unità temporale di riferimento. A partire dal XIII secolo, l'economia-mondo europea ha cambiato più volte forma, ricollocando il suo centro e riorganizzando le sue periferie. Occorre, quindi, trovare la misura che permetta di conservarne l'identità. Poiché il tempo delle economie-mondo è un tempo che scorre lentamente, un tempo storico di lunga durata, il 'ritaglio' temporale che più si addice al loro studio è quello dei trend o dei cicli secolari, che sono caratterizzati da tre fasi principali: l'ascesa, la crisi e la decadenza.

Nella Civilisation e negli Afterthoughts, il concetto di economia-mondo è posto in relazione, da un lato, con l'origine e lo sviluppo del capitalismo, e, dall'altro, con la possibilità di scrivere una storia universale.

Braudel ritiene che la crescita e il successo del capitalismo richiedano una serie di condizioni, tra le quali c'è anche l'esistenza di una gerarchia. Il capitalismo nascente non inventa una nuova gerarchia, ma si serve di quella secondo la quale è ordinato il mondo. Questo mondo è un'unità che si delinea tra il XV e il XVIII secolo e che si afferma sotto il segno della disuguaglianza. L'attuale divisione tra paesi ricchi e paesi sottosviluppati viene ritenuta una realtà già operante tra il XV e il XVIII secolo. "Il capitalismo - scrive Braudel - vive, in effetti, di questa regolare suddivisione in piani verticali: le zone periferiche nutrono quelle intermedie e, soprattutto, le aree intorno al centro. Ma cos'è il centro se non la punta estrema della piramide, la superstruttura capitalistica dell'intera costruzione? E siccome esiste una reciprocità di prospettive, se il centro dipende dai rifornimenti della periferia, quest'ultima, a sua volta, dipende dai bisogni del centro che le impone la sua legge." (ibid., pp. 101-102).

Per stabilire il nesso tra sviluppo del capitalismo ed economia-mondo, Braudel si rifà al pensiero del secondo autore che esaminiamo in queste pagine. La tesi che il capitalismo sia una creazione dell'ineguaglianza del mondo è tratta, infatti, da Wallerstein. Secondo le parole dello storico francese, affinché il capitalismo "possa svilupparsi gli è necessaria la connivenza dell'economia internazionale. Il capitalismo è figlio dell'organizzazione di uno

spazio sicuramente smisurato. Non sarebbe divenuto così forte in uno spazio limitato, forse non si sarebbe sviluppato affatto, senza la possibilità di utilizzare il lavoro ancillare di altri." (ibid., p. 102).
L'obiettivo di Braudel, come è detto nell'ultima conferenza di Baltimora, intitolata Capitalismo e divisione del mondo, è quello di "verificare in che modo le successive economie-mondo a base europea, cioè costruite attraverso gli itinerari della espansione dell'Europa, spiegano, o meno, i giochi del capitalismo e la sua diffusione. Non esiterei a sostenere in anticipo che queste tipiche economie-mondo sono state le matrici del capitalismo europeo ed in seguito di quello mondiale." (ibid., p. 94).
Il concetto di economia-mondo, quindi, non solo è utile per tracciare la dinamica del capitalismo, ma risulta essere la chiave per poter scrivere una storia universale. La storia del mondo, infatti, si presenta come una successione di economie-mondo generata dal meccanismo di ricollocazione del loro centro. "Ogni volta che si ha un décentrage si opera una polarizzazione attorno ad un nuovo centro, come se ogni economia-mondo non potesse vivere senza un centro di gravità, senza un polo. Questi processi di décentrage e récentrage sono comunque rari e, per questo, tanto più importanti." (ibid., p. 95).La storia europea ha visto il centro dell'economia-mondo trasferirsi da Venezia ad Anversa, poi a Genova, ad Amsterdam e, infine, a Londra. L'ascesa di Londra chiude una fase della storia europea e mondiale, perché, per la prima volta, l'economia-mondo europea aspira al controllo dell'economia mondiale. Dopo il 1914 avverrà un'ulteriore ricollocazione e il centro dell'economia-mondo capitalistica si situerà a New York.

3. Il concetto di economia-mondo nell'opera di Wallerstein

L'opera principale di Wallerstein è The modern world-system, di cui sono apparsi i primi due volumi, rispettivamente nel 1974 e nel 1980. Altri saggi sull'economia-mondo sono raccolti in due libri (v. Wallerstein, 1979 e 1984). Per chiarire il concetto di economia-mondo nell'opera del sociologo americano è necessario ripercorrere il suo itinerario intellettuale. Alla fine del periodo coloniale Wallerstein vive in Africa e assiste ad avvenimenti che lo costringono a misurarsi con il problema del mutamento sociale e, in particolare, della modernizzazione. Alcuni problemi irrisolti legati allo studio della modernizzazione e l'insoddisfazione per un apparato concettuale ereditato dalla sociologia americana provocano

una 'conversione intellettuale' che porta Wallerstein a promuovere un nuovo filone di studi. A questi studi e alla visione teorica che li guida è stato dato il nome di world-system perspective o di world-system analysis.Wallerstein ha proposto la sua visione del mondo moderno in opposizione alle teorie dello sviluppo economico prevalenti nei primi due decenni del secondo dopoguerra. La world-system perspective è particolarmente critica sia verso le teorie unilineari dello sviluppo 'per stadi' di ispirazione liberale, sia verso quelle del marxismo 'evoluzionistico'. Questi due filoni di pensiero condividono alcuni presupposti comuni. L'unità d'analisi è sempre la singola società, in genere identificata con uno Stato nazionale. Il mondo è visto come una molteplicità di società collegate tra loro, ma sostanzialmente autonome. Ciascun paese può seguire un analogo percorso di sviluppo. Esiste un paese che, grazie a sue caratteristiche particolari, ha percorso il cammino dello sviluppo per primo e che può fare da modello agli altri. Nella versione liberale un paese passa dall'arretratezza alla modernità attraverso un cammino formato da stadi di sviluppo economico. Il modello da seguire è rappresentato, prima, dalla Gran Bretagna e, poi, dagli Stati Uniti. Nella versione marxista la storia delle società umane è intesa come una sequenza di modi di produzione: la schiavitù, il feudalesimo, il capitalismo e il socialismo. Questi sono gli stadi che segnano il cammino dello sviluppo e il paese preso a modello è l'Unione Sovietica.
L'opposizione a queste teorie guida Wallerstein nella ricerca di termini di riferimento teorici alternativi sui quali fondare la sua analisi. Il dibattito sulla transizione dal feudalesimo al capitalismo, che oppose Paul Sweezy a Maurice Dobb sulle pagine di "Science and Society", viene utilizzato come riferimento nella discussione sulle origini del sistema-mondo moderno.Dal dibattito sul sottosviluppo, cominciato negli anni cinquanta, sono tratti numerosi concetti che, rielaborati, troveranno posto nell'analisi del sistema-mondo moderno. Wallerstein concorda con la tesi degli studiosi, soprattutto marxisti, che collegano il problema delle cause del sottosviluppo a quello delle origini del capitalismo. Da un filone del dibattito è tratta la coppia di termini 'centro-periferia', utilizzata per la prima volta all'inizio degli anni cinquanta da Raul Prebisch e da altri economisti della Commissione economica sull'America Latina dell'ONU e successivamente impiegata anche dai teorici della 'dipendenza'. L'uso del concetto di surplus, riferito all'economia mondiale, viene da Paul Baran, mentre da studiosi quali Gunnar Myrdal, André Gunder Frank e Arghiri Emmanuel sono tratti il concetto di 'dipendenza', che indica il legame di funzionalità tra lo sviluppo del capitalismo in alcune aree del mondo e il sottosviluppo di altre, e i concetti di 'causazione circolare e cumulativa

del sottosviluppo' e di 'scambio ineguale', utili per descrivere i meccanismi che provocano l'aumento cumulativo del gap tra queste zone del mondo.Un altro riferimento teorico fondamentale è costituito dall'opera di Fernand Braudel e dalla scuola delle "Annales". Il dialogo tra storia e scienze sociali, la pluralità del tempo storico, l'importanza della dimensione spaziale nello studio della realtà sociale e delle sue grandi strutture, le spiegazioni globali contro quelle parziali, la longue durée contro l'évenementiel, e l'elaborazione di una storiografia quantitativa in grado di individuare cicli economici e tendenze di lungo periodo, sono gli argomenti che interessano Wallerstein.
Il punto di partenza della nuova prospettiva teorica è la risposta alla domanda seguente: "qual è l'unità d'analisi corretta nello studio del mutamento sociale?". Wallerstein ritiene che entità politico-culturali come lo Stato, la nazione, il popolo, o - nel caso africano - la tribù, la colonia, non possano essere considerate unità d'analisi adeguate, perché l'azione sociale che si svolge al loro interno dipende sempre dal funzionamento di entità più vaste. L'unica unità d'analisi utilizzabile nello studio del mutamento sociale è il sistema sociale, un'entità caratterizzata dal "fatto che la sua vita interna è in gran parte autonoma, e che la dinamica del suo sviluppo è in gran parte interna." (v. Wallerstein, The modern world-system I..., 1974; tr. it., p. 474). I sistemi sociali sono entità economico-materiali la cui autonomia è fondata sull'esistenza al loro interno di un'unica struttura della divisione del lavoro che non va intesa soltanto in termini socio-professionali, ma anche spaziali. I sistemi sociali sono entità più o meno estese nello spazio, sono cioè 'mondi' dei quali è possibile tracciare i confini. Questi confini, che mutano nel tempo e nello spazio, "corrispondono ai limiti (rilevabili empiricamente) di un reticolato di processi produttivi più o meno strettamente interdipendenti." (v. Wallerstein, 1981, p. 305).
Sulla base di questi criteri i sistemi sociali sono individuati in entità molto piccole o, al contrario, molto vaste. Gli unici sistemi sociali reali sono le 'economie di sussistenza' e i 'sistemi-mondo'. I sistemi-mondo, come le economie di sussistenza, si fondano su un'unica struttura portante della divisione del lavoro, ma comprendono una molteplicità di culture, e possono essere caratterizzati dall'esistenza oppure dall'assenza di un unico sistema politico. Nel primo caso i sistemi-mondo sono chiamati 'imperi-mondo', nel secondo 'economie-mondo'. I destini di questi sistemi sociali sono collegati alle due grandi rivoluzioni che hanno segnato la storia dell'umanità. Le piccole economie di sussistenza tendono a sparire in seguito alla rivoluzione neolitica. Si creano invece sia imperi-mondo sia economie-mondo. I

primi sono più stabili e, per un lungo periodo, risultano la forma prevalente di sistema sociale. Soltanto nel XVI secolo si forma un nuovo tipo di economia-mondo che, a differenza delle precedenti, non si disgrega né si trasforma in un impero-mondo. Questo sistema sociale, che è riuscito a sopravvivere per cinquecento anni, è chiamato sistema-mondo moderno; esso è un'economia-mondo capitalistica ed è l'unico sistema sociale reale esistente nel secolo XX. Secondo Wallerstein, l'unica possibiltà di 'maneggiare' un'unità di analisi temporalmente e spazialmente tanto estesa quanto il sistema-mondo moderno, è quella di scriverne, o meglio, di riscriverne la storia, descrivendo il sistema-mondo a un certo livello di astrazione, quello dell'evoluzione delle strutture dell'intero sistema.Tra la fine del XV e l'inizio del XVI secolo, comincia a formarsi in Europa un sistema sociale di tipo nuovo, che viene definito 'economia-mondo europea'. All'epoca, l'Europa non è l'unica economia-mondo esistente, è però l'unica a imboccare la strada dello sviluppo capitalistico. L'economia-mondo europea diventa, quindi, un'economia-mondo capitalistica.
La nascente economia-mondo capitalistica si afferma grazie a tre fattori: l'espansione geografica; lo sviluppo di svariati metodi di controllo della forza lavoro per prodotti differenti e differenti zone geografiche; la creazione di organizzazioni statali relativamente forti in quello che sarebbe diventato il suo centro. L'espansione geografica dell'economia-mondo europea porta alla creazione di una struttura sociale a livello mondiale stratificata in tre zone - centro, periferia e semiperiferia - ciascuna caratterizzata da specifici tipi di produzione, da specifici modi di controllo del lavoro e da specifiche forme di proprietà fondiaria. La nuova divisione 'europea' del lavoro, oltre ad assegnare a certe aree determinati compiti economici e a collocare in esse determinate attività produttive, instaura, grazie agli scambi commerciali, una forte interdipendenza, di ampiezza mai riscontrata prima, tra le diverse zone.Nel XVII secolo, caratterizzato da una fase di stagnazione economica, la struttura dell'economia-mondo capitalistica si consolida definitivamente, gettando le basi del decollo industriale successivo al 1750. Durante questo periodo si assiste a un processo di mobilità interna sia ascendente sia discendente: alcune aree geografiche passano da una zona all'altra del sistema-mondo.
Nell'opera di Wallerstein il concetto di economia-mondo è opposto a quello di impero-mondo ed è collegato a quello di capitalismo.
Un primo tratto distintivo dell'economia-mondo è illustrato discutendo le origini del capitalismo. Per spiegare perché esso si affermi nell'economia-mondo europea e non altrove,

Wallerstein propone di paragonare Europa e Cina. Richiamandosi a Weber, egli sostiene che il feudalesimo portò in Europa allo smantellamento della struttura imperiale, permettendo la creazione di una molteplicità di unità politiche, mentre in Cina il sistema delle prebende favorì la sua sopravvivenza. A lungo andare la struttura del sistema sociale europeo si dimostrò più favorevole allo sviluppo capitalistico che non quella imperiale della Cina.
La tesi di Wallerstein è che il capitalismo sia realizzabile soltanto all'interno della struttura di un'economia-mondo, composta, come quella europea, di piccoli imperi, di Stati nazionali e di città-Stato, e non in quella di un impero-mondo (v. Wallerstein, The modern world system I..., 1974, tr. it., p. 52). In un impero, come quello cinese, le decisioni di natura economica sono vincolate da necessità di natura politica, quali l'amministrazione e la difesa di un territorio immenso abitato da una popolazione numerosa. In un'economia-mondo, invece, le decisioni di natura economica sono rivolte all'intera economia-mondo, mentre quelle di natura politica si indirizzano prevalentemente a strutture più piccole che detengono il controllo legale, gli Stati che si trovano al suo interno (ibid., p. 67). Il fatto che i fattori economici operino in un'area più vasta di quella controllabile da un'entità politica fornisce ai capitalisti la possibilità di manovrare liberamente a favore dell'accumulazione di capitale.
Il concetto di economia-mondo è collegato alle origini del capitalismo anche in un altro modo. Per Wallerstein l'aspetto più importante del rapporto fra le tre zone che formano l'economia-mondo europea è la destinazione alla zona centrale del surplus estratto nelle periferie.
Riprendendo quanto scritto da Marx nel XXIV capitolo del Capitale a proposito dei "momenti fondamentali dell'accumulazione originaria", della formazione di un mercato mondiale sin dall'inizio dell'era capitalistica e del ruolo che in essi giocarono le aree periferiche, egli afferma che "L'economia-mondo si basava proprio sull'assunto che esistevano di fatto queste tre zone, e che di fatto avevano modi diversi di controllo del lavoro. Non fosse stato così, non sarebbe stato possibile garantire quel flusso di surplus che permise al capitalismo di nascere." (ibid., p. 119).
Durante il "lungo XVI secolo" - secondo un'espressione che Wallerstein mutua da Braudel - si afferma una tendenza al dominio delle aree periferiche da parte di quelle centrali; questa forma di dominio, da cui deriva lo sviluppo diseguale delle diverse zone dell'economia-mondo, è intrinseca al capitalismo. Esso, infatti, si basa su un doppio meccanismo di sfruttamento: l'appropriazione del plusvalore del lavoratore da parte del proprietario e quella

del surplus dell'intera economia-mondo da parte delle aree centrali (v. Wallerstein, The rise and..., 1974, p. 400).In conclusione, nell'opera di Wallerstein, il concetto di economia-mondo è fondamentale perché su di esso si basa la possibilità di condurre lo studio del mutamento sociale. Esso è stato applicato allo studio del mondo moderno attraverso l'analisi delle strutture e dei processi che, nel tempo e nello spazio, ne hanno caratterizzato l'esistenza e l'evoluzione. Il sistema-mondo moderno è considerato un'economia-mondo capitalistica la cui principale caratteristica è la divisione in tre zone interdipendenti e stratificate. Tale divisione non è un tratto temporaneo del sistema-mondo, ma ne costituisce una struttura. Le aree in cui si collocano i processi produttivi centrali o periferici cambiano, ma la divisione rimane. L'economia-mondo è intesa come una divisione 'spaziale' del lavoro. I termini 'centro', 'periferia' e 'semiperiferia' sono un modo di descrivere un ordine dello spazio economico che ha reso possibile la genesi, la sopravvivenza e lo sviluppo del sistema-mondo moderno. Il concetto di economia-mondo capitalistica indica che nel sistema-mondo moderno la complementarità delle parti si accompagna alla loro disuguaglianza. La prima si rafforza con l'approfondirsi della seconda. L'opera di Wallerstein è collocabile nell'ambito della cosiddetta sociologia storica (v. Skocpol, 1984). Come altri studiosi, Wallerstein si muove in un campo di ricerca che tenta di mettere in relazione storia e teoria sociale. Egli vuole, inoltre, combinare l'applicazione di un modello teorico generale alla storia con un'interpretazione storica politicamente significativa. Il suo obiettivo, infatti, è quello di fornire un'interpretazione del mondo moderno e della sua storia che colleghi avvenimenti significativi del passato a problemi politici contemporanei, basandosi sull'assunto che disuguaglianza e sfruttamento ne abbiano caratterizzato l'esistenza sin dall'inizio.Nell'ambito delle scienze sociali, al concetto di economia-mondo elaborato da Wallerstein sono state rivolte molte critiche (v. Lane, 1976; v. Brenner, 1977; v. Skocpol, 1977; v. Hunt, 1978; v. Gourevitch, 1978; v. Aronowitz, 1981; v. Kimmel, 1982). Tali critiche riguardano: la possibilità di considerare i rapporti di scambio come il fattore che determina l'organizzazione sociale e come il motore dello sviluppo capitalistico; la validità del concetto di economia-mondo, in quanto sistema, come strumento di analisi storica o di studio del mutamento sociale; la tendenza a utilizzare spiegazioni funzionali o teleologiche in cui le unità che si analizzano si comportano in un certo modo a causa delle conseguenze di questo comportamento per il sistema nel suo insieme.

4. Conclusioni

Tra il concetto formulato da Braudel e quello elaborato da Wallerstein esistono differenze esplicitamente riconosciute (v. Braudel, 1977; tr. it., p. 92). Il concetto di Wallerstein è più specifico di quello di Braudel. Per il sociologo americano l'unica economia-mondo che è possibile prendere in considerazione è quella capitalistica formatasi nel XVI secolo, perché quelle precedenti sono considerate sistemi sociali instabili, incapaci di sopravvivere. Per lo storico francese, al contrario, è possibile suddividere il mondo in parecchie economie-mondo coesistenti a partire dal medioevo se non dall'antichità (ibid., pp. 92-93).

In secondo luogo, per Wallerstein, gli imperi-mondo e le economie-mondo sono sistemi sociali differenti e mutuamente esclusivi. Il capitalismo può sorgere e affermarsi solo in un'economia-mondo e non in un impero-mondo. Per Braudel, invece, un'economia-mondo può vivere e organizzarsi anche sotto i vincoli imposti da un impero (v. Braudel, 1979, vol. III, p. 42).Queste differenze non hanno pregiudicato la collaborazione tra i due studiosi e tra i due istituti a cui essi hanno dedicato gran parte del loro impegno.

Il concetto di economia-mondo è stato usato soprattutto nelle ricerche e negli studi promossi dalla Maison des Sciences de l'Homme di Parigi e dal Fernand Braudel Center for the Study of Economies, Historical Systems and Civilization di Binghamton nello Stato di New York. Al primo istituto è collegata la rivista "Annales. Economies, Sociétés, Civilisations", mentre il Fernand Braudel Center pubblica, dal 1977, la rivista "Review".Sebbene il concetto di economia-mondo sia strettamente collegato al programma scientifico dei due autori che abbiamo considerato, esso può interessare anche chi non condivide gli scopi ultimi di coloro che lo hanno proposto.Esso, soprattutto nella versione di Wallerstein, può essere utile a coloro che occupandosi di mutamento sociale: 1) scelgono come unità d'analisi il mondo intero, intendendolo come un insieme di parti interconnesse e interdipendenti; 2) vogliono studiare le strutture e i processi di grandi dimensioni che hanno trasformato e stanno trasformando il mondo; 3) prediligono nella loro attività scientifica l'approccio storico-comparativistico (v. Tilly, 1984).

L'analisi di strutture e di processi di grandi dimensioni può essere condotta a diversi livelli. A un livello molto generale, si specificano le caratteristiche peculiari di un'era della storia umana, considerando quali strutture i sistemi-mondo e quali processi le loro reciproche relazioni, le loro successioni o le loro trasformazioni. A questo livello si comparano sistemi-

mondo: economie-mondo e imperi-mondo.
A un livello immediatamente inferiore, i sistemi-mondo rimangono unità d'analisi significative, ma lo diventano anche le grandi strutture che li compongono. Si studiano, quindi, le relazioni tra Stati, aree geografiche, modi di produzione regionali, esaminando processi di subordinazione, di produzione, di distribuzione su larga scala. Gli oggetti che si comparano sono strutture, tra loro interdipendenti, interne a un sistema-mondo.Il concetto di economia-mondo capitalistica si applica a questi livelli di analisi. Per Wallerstein le strutture che compongono un sistema-mondo non possono essere spiegate senza spiegare il sistema nel suo insieme. Lo sviluppo delle tre zone costituenti l'economia-mondo capitalistica, ma anche altri oggetti, quali la storia degli Stati, l'azione delle classi sociali o di altri gruppi, non possono essere compresi e spiegati senza riferirli allo sviluppo del sistema-mondo nel suo complesso. Nell'opera di Wallerstein la comparazione avviene tra parti di un'unica economia-mondo capitalistica, le cui caratteristiche distintive dipendono dalla diversa collocazione di una parte rispetto a un'altra all'interno di una rete di interdipendenze.È stato sostenuto che a tali livelli di generalità le indagini storiche e sociologiche sono assai controverse e rischiose. Tuttavia, in attesa di nuove idee, l'elaborazione del concetto di economia-mondo rimane un tentativo di risposta alla necessità di formulare delle ipotesi sul funzionamento del sistema-mondo moderno, senza le quali lo studio di strutture e di processi a livelli più bassi di generalità ha poco senso (v. Tilly, 1984, p. 63).

Economia e Società
di Richard Swedberg e George Dalton

Sociologia economica
di Richard Swedberg

sommario: 1. Introduzione. 2. Le prime concezioni sui rapporti tra economia e società (l'economia politica inglese). 3. La polarizzazione tra economia e società (1880-1920). 4. La

sociologia economica: nascita e primi sviluppi. 5. Il revival degli anni cinquanta (Parsons, Polanyi). 6. Le tendenze attuali. a) L'imperialismo economico e la nuova economia istituzionale. b) La nuova sociologia economica. □ Bibliografia.

1. Introduzione

Intorno al 1800 il concetto di 'economia' divenne estremamente importante nell'ambito della riflessione teorica in Occidente e acquisì anche un nuovo significato. Dopo aver a lungo denotato più che altro la gestione degli affari domestici (οἰκονομία), il termine 'economia' cominciò a essere usato anche per indicare una parte della società. Questo mutamento concettuale dipese da una pluralità di ragioni. Proprio in quel periodo, per esempio, grazie all'apporto di Adam Smith e di altri studiosi, nacque la scienza economica moderna. Contemporaneamente prese l'avvio anche il processo di industrializzazione, prima in Inghilterra e poi nel resto dell'Europa e negli Stati Uniti. Nella società sempre più secolarizzata che cominciava a emergere si scatenò ben presto un vivace dibattito su quale fosse il ruolo che questo nuovo settore - l'economia, appunto - era chiamato a svolgere nella società. Secondo alcuni il settore economico doveva essere lasciato completamente libero e la società non doveva mai limitarne l'azione. Altri, invece, erano dell'avviso che l'economia dovesse essere subordinata alla società e/o allo Stato. In questo articolo tratteremo solo occasionalmente degli aspetti direttamente politici di questo dibattito, anche se bisogna comunque tener presente che il dibattito sui rapporti tra economia e società - che vi partecipino uomini politici o sociologi - comporta spesso una componente ideologica. Il problema che affronteremo, però, è un altro: in che modo va inteso il rapporto tra economia e società dal punto di vista delle scienze sociali? La risposta a questa domanda verrà data contestualmente alla presentazione delle diverse tesi che economisti e sociologi hanno avanzato in proposito dalla fine del secolo XVIII ai giorni nostri.
Nelle scienze sociali il dibattito su economia e società è ruotato essenzialmente attorno a tre posizioni principali. Secondo la prima l'economia costituisce un universo a sé nella società e può essere analizzata soltanto in termini puramente economici o non sociali. La società è considerata estranea all'ambito economico e, al massimo, riveste per l'economista un interesse puramente marginale. Poiché esclude completamente l'analisi di tipo sociale, questa posizione non verrà analizzata dettagliatamente nel presente articolo (per una sua esposizione

esauriente cfr., per esempio, Mark Blaug, Economic theory in retrospect, 1978). Per la seconda posizione l'economia è una parte integrante della società e funziona come il resto della società. Pochi studiosi hanno sostenuto questo punto di vista con convinzione, perché esso implica, in definitiva, un totale rifiuto dell'analisi economica standard. Alcuni studiosi, tuttavia, condividono questo punto di vista proprio perché esso si contrappone in modo radicale alla teoria economica dominante. Il terzo punto di vista, infine, vanta un maggior numero di aderenti, in quanto è meno estremistico; secondo questo punto di vista, le parti più importanti dell'economia dovrebbero essere studiate servendosi dei normali strumenti (non sociali) del pensiero economico, ma, dato che all'interno del settore economico vi sono anche istituzioni sociali, queste andrebbero analizzate in termini più sociologici. Molti economisti e altri esponenti delle scienze sociali hanno oscillato tra queste diverse posizioni, anche all'interno di un medesimo lavoro. Tuttavia, per ragioni di chiarezza, tali posizioni vanno tenute separate e così il lettore potrà tenerle presenti nel corso di questo articolo.

2. Le prime concezioni sui rapporti tra economia e società (l'economia politica inglese)

Si è spesso osservato che i primi economisti - gli esponenti dell''economia politica' - erano molto interessati alle istituzioni sociali e al comportamento sociale. È un'osservazione sostanzialmente giusta, ma bisognerebbe aggiungere immediatamente che la tendenza a considerare l'economia e la società come due sfere distinte era già presente nelle prime teorie economiche. Ciò è vero non soltanto per Ricardo o per John Stuart Mill, ma anche per Adam Smith. L'unico pensatore che abbia davvero cercato di fondere l'analisi dell'economia con quella della società è stato Karl Marx, che, però, intrecciò i due ambiti in un modo troppo stretto, dando origine a una serie di nuovi problemi (v. sotto). Quando, durante la prima metà del nostro secolo, la separazione tra teoria economica e teoria della società raggiunse il suo culmine, l'economia era considerata un settore autonomo e distinto della società, soggetto a leggi proprie; inoltre, dato che le categorie della teoria economica erano considerate atemporali e universalmente valide, non si vedeva la necessità di introdurre dati storici nell'analisi economica. L'immagine dell'economia presentata nella Ricchezza delle nazioni (1776) è invece completamente diversa: innanzitutto, in quest'opera, l'economia è considerata parte integrante della società. Probabilmente il motivo principale di questa concezione sta nel fatto che, per Smith, l'economia politica era una disciplina 'pratica', cioè destinata, in primo

luogo, a essere concretamente utilizzata dagli uomini di Stato e dai legislatori. Secondo Smith, quando si parla di economia, si intende sempre l'economia di una nazione in particolare. Diversamente da molti altri economisti successivi, inoltre, Smith aveva un vivo senso della storia e spesso nelle sue analisi osserva come le stesse cose vengano fatte in un certo modo in Inghilterra e in modo diverso in altri paesi. Smith sottolinea continuamente il ruolo delle 'istituzioni' nell'economia e parla spesso dei 'costumi' e dei 'caratteri' delle persone, facendo così ricorso a categorie eminentemente sociali. È chiaro, comunque, che anche per Smith l'economia costituisce, fino a un certo punto, una realtà a sé stante, extrasociale. Questo punto di vista emerge con particolare chiarezza quando egli parla della propensione naturale degli uomini al "traffico, al baratto e allo scambio" (v. Smith, 1776). Invece di concepire le attività di scambio e affini come diverse forme di comportamento sociale (come avrebbero fatto più tardi sociologi e antropologi), Smith le considera caratteristiche biologiche. La tendenza allo scambio, al baratto, ecc. è - sostiene Smith - ciò che distingue l'uomo dagli animali. Tale tendenza diventa addirittura il motore della società, il fattore che permette alla società di passare da uno stadio al successivo della sua evoluzione. Anche le considerazioni svolte nella Ricchezza delle nazioni circa il fatto che le "istituzioni" spesso sopprimono le "inclinazioni naturali" delle persone, traggono origine da questa concezione 'biologica' dell'azione economica.

La tendenza a separare il comportamento economico dalle altre forme di comportamento sociale è ancor più accentuata nella successiva opera fondamentale dell'economia politica, i Principles of political economy and taxation (1817) di David Ricardo. Il tono dell'opera di Ricardo è molto diverso da quello della Ricchezza delle nazioni: i riferimenti alla storia e ad altri paesi sono estremamente rari; l'intera analisi segue piuttosto un'impostazione astorica e atemporale, ed è chiaro che l'ambizione di Ricardo era proprio quella di formulare un insieme di teoremi economici universalmente validi, indipendentemente dalle variabili di tempo e di spazio.Il modo di procedere di Ricardo, altamente astratto e del tutto indipendente da considerazioni di ordine sociale, è ben illustrato dalla sua celebre analisi del commercio estero. Ricardo presenta il suo tipo di analisi nel modo seguente: "Per semplificare il problema ho fatto l'ipotesi che il commercio tra due paesi sia limitato a due sole merci" (v. Ricardo, 1817). Nell'esempio di Ricardo i due paesi erano l'Inghilterra e il Portogallo e le due merci panno e vino, ma il discorso sarebbe stato egualmente valido per qualsiasi altra coppia di paesi e di merci. La società non viene presa in considerazione nell'ambito dell'analisi

principale: essa compare soltanto come fattore esogeno, alla stregua della guerra, per esempio. L'idea di fondo è, quindi, quella di trattare l'economia come un'area autonoma e autosufficiente, circondata da varie istituzioni sociali cui non viene dedicata molta attenzione.

Anche a proposito degli elementi sociali presenti nell'economia Ricardo si discosta da Smith. Dopo aver sottolineato che per il Portogallo è vantaggioso, in certe circostanze, importare panno dall'Inghilterra (anche se produrre abiti direttamente in Portogallo costa di meno), Ricardo osserva, incidentalmente, che, se tutto funzionasse secondo le regole dell'economia politica, dovrebbe verificarsi un trasferimento di capitali dall'Inghilterra al Portogallo. Ciò, viceversa, non accade: gli imprenditori inglesi preferiscono un tasso di profitto più basso in Inghilterra a quello più alto che potrebbero realizzare in Portogallo. Per Ricardo le ragioni di questo particolare comportamento sono tutte di natura sociale: i capitalisti inglesi non vogliono abbandonare il proprio paese; essi hanno in Inghilterra tutte le proprie 'conoscenze' e le loro abitudini sono già fissate (v. Ricardo, 1817). Ciò che questo esempio sui capitalisti inglesi dimostra, comunque, è che Ricardo tende a considerare il comportamento sociale come un fattore che impedisce all'economia di funzionare efficientemente come dovrebbe. Questo ci riconduce a una delle posizioni standard nel dibattito su economia e società, e cioè alla tesi secondo cui per comprendere nel modo migliore il comportamento economico è necessario eliminarne la dimensione sociale.

Anche nella successiva grande opera dell'economia politica inglese - i Principles of political economy (1848) di John Stuart Mill - si riscontra una separazione fra economia e società. A prima vista il lettore può avere l'impressione di essere tornato indietro ai tempi di Adam Smith, poiché Mill inizia la sua esposizione in modo simile a Smith con una storia del genere umano, e fa diversi riferimenti alle usanze economiche di altri paesi. Ma si capisce presto che questa somiglianza è più apparente che reale. L'economia politica, secondo Mill, deve occuparsi della produzione della ricchezza, lungo un processo che si articola in due parti: la produzione vera e propria e la distribuzione. La prima è governata da leggi che hanno "il carattere di verità fisiche" (v. Mill, 1848). Gli individui sono costretti a obbedire a tali leggi, "che lo vogliano o meno". Le leggi della distribuzione, invece, sono diverse, poiché "dipendono dalle leggi del costume e della società". Si tratta dunque di leggi di ordine sociale e in linea di principio possono essere modificate, anche se non in modo arbitrario.La sezione sulla proprietà, nei Principles, contiene la più ampia analisi di un'istituzione economica mai

condotta da Mill. La proprietà, secondo Mill, rientra nell'ambito della distribuzione della ricchezza e ha pertanto natura sociale; ma anziché cercare di studiare in che modo 'le leggi e i costumi' di un dato paese abbiano influenzato e modellato l'istituzione della proprietà, Mill si perde in una lunga disamina dei diversi progetti politici per modificare le vigenti leggi sulla proprietà. Mill, in altre parole, sostituisce a un'analisi sociale dell'economia una rassegna di diversi progetti di riforma sociale. Secondo Schumpeter questo modo di considerare le istituzioni sociali, come oggetto di riforme più che di analisi sociologica, ha esercitato una notevole influenza su altri economisti (v. Schumpeter, 1954).

Nei Principles c'è solo un punto in cui Mill supera l'artificiale separazione tra teoria economica e teoria della società, ed è là dove - nel capitolo intitolato Of competition and custom - egli osserva che l'economista può definire le leggi dei salari, della rendita, del profitto, ecc. solo in condizioni di concorrenza perfetta. "L'economia politica può assurgere al rango di scienza solo grazie al principio della concorrenza" (v. Mill, 1848). Il problema, però, come dice Mill, è che in realtà i prezzi vengono spesso decisi più dal costume che dalla concorrenza. Non è raro, per esempio, che una determinata merce abbia due prezzi nello stesso mercato. Tuttavia, dato che l'economia politica deve essere "una scienza astratta e ipotetica", è lecito trascurare tutti i fatti di questo genere e continuare a basarsi sull'assunto fittizio che esistano condizioni di concorrenza perfetta.

L'opera di Karl Marx rappresenta una critica feroce dell'economia politica inglese. Anche il punto di vista di Marx sui rapporti tra economia e società è radicalmente diverso da quelli di Smith, Ricardo e Mill. Secondo Marx la società è un'unità organica, e l'economia ne costituisce una componente. Inoltre Marx introduce per la prima volta il concetto di classe nell'analisi economica. Vero è che anche gli economisti inglesi parlano spesso di classi sociali, ma per loro il termine 'classe' indica solo una categoria economica astratta, mentre per Marx la classe è una distinta realtà sociale. In verità è impossibile in linea di principio distinguere nelle categorie di Marx gli elementi sociali da quelli economici. Al microlivello del suo sistema si trova l'individuo, che appartiene a una specifica classe sociale e produce plusvalore con il suo lavoro. Di questo plusvalore si appropria quindi la classe dirigente, che in diversi modi domina le altre classi. Al macrolivello si trova un sistema di classi, che fa parte di una specifica formazione sociale. Ogni formazione sociale si costituisce storicamente e fa ricorso a più di un modo di produzione. La società progredisce costantemente attraverso un processo dinamico che ha origine all'interno della sfera economica e si esprime a livello

politico e sociale tramite la lotta di classe.
È chiaro che la sua posizione critica nei confronti dell'economia politica inglese condusse Marx a trattare in modo completamente nuovo i rapporti tra economia e società. Alla separazione piuttosto artificiale sostenuta dal pensiero borghese Marx contrappose un punto di vista decisamente più organico, secondo cui la società consiste in una 'struttura economica' e una 'sovrastruttura'. La prima include i 'rapporti di produzione' determinati dalle 'forze produttive' e costituisce - dice Marx - "la base reale su cui sorge la sovrastruttura giuridica e politica e a cui corrispondono determinate forme di coscienza sociale" (v. Marx, 1867-1894; v. anche Marx ed Engels, 1845-1846). Tra le forme di coscienza sociale Marx annovera essenzialmente la religione, la filosofia e il pensiero economico borghese (v. Marx, 1867-1894). L'eventuale impressione che questo schema basato sulla distinzione tra struttura e sovrastruttura sia semplicistico sfuma rapidamente all'atto di condurre uno studio concreto basato sui principî di Marx. Per far ciò si deve analizzare una 'formazione socioeconomica', che consiste in una mescolanza di modi di produzione. Un classico lavoro di questo genere è il saggio di Lenin Lo sviluppo del capitalismo in Russia, dove Lenin osserva che la Russia ha "due sistemi economici" contrariamente a quanto sostenuto dai populisti, per i quali la struttura socioeconomica della campagna russa era già essenzialmente borghese (v. Lenin, 1899). La presenza di questa duplice struttura economica, come Lenin mette in evidenza, genera un'infinità di complicate contraddizioni (v. anche Godelier, 1971).
A Marx va riconosciuto il merito di aver tentato di colmare il divario fra teoria economica e teoria della società, ma il tentativo fu prematuro, come ha chiarito, in particolare, Schumpeter in Capitalism, socialism and democracy (1942). Secondo Schumpeter il nesso istituito da Marx fra categorie economiche e categorie sociologiche è troppo stretto. Ciò che per necessità presenta un carattere analitico e astratto, sottolinea Schumpeter, non è facilmente traducibile in qualcosa di empirico e concreto: "Un valido teorema economico, tradotto in termini sociologici, anziché acquisire un più profondo significato, può rivelarsi erroneo, e viceversa" (p. 46). La sola precisazione che va aggiunta alla critica di Schumpeter è che l'intenzione di Marx era quella di stabilire delle leggi economiche analoghe a quelle fisiche, proprio come Ricardo e Mill. Ciò è particolarmente evidente nel suo capolavoro, Il capitale, dove Marx proclama esplicitamente che la sua ambizione è scoprire "la legge del moto economico della società moderna" (v. Marx, 1867-1894). In Marx, in altre parole, le categorie economiche tendono in ultima analisi a prendere il sopravvento sulle categorie

sociali e a sostituirsi ad esse.

3. La polarizzazione tra economia e società (1880-1920)

Il periodo intorno alla fine del secolo fu decisivo per il dibattito sui rapporti tra economia e società. Le due sfere vennero separate radicalmente. Gli economisti eliminarono completamente la componente sociale dalle loro analisi, e qualsiasi elemento fosse considerato appartenente alla 'società' fu dichiarato di competenza di altre discipline. È ancora presto per dire perché tutto ciò accadde proprio in quel particolare momento storico, anche se due cose sono abbastanza chiare. In primo luogo la teoria economica dominante continuò a occuparsi di problemi - come quelli affrontati dalla teoria dei prezzi - a proposito dei quali era molto facile escludere influenze di ordine sociale. In secondo luogo, in questo periodo, tra teorici puri ed economisti di orientamento sociologico scoppiò una disputa di vaste proporzioni: il cosiddetto Methodenstreit ('disputa sul metodo'). Questo conflitto metodologico si concluse con un vero e proprio trionfo dei teorici puri sui loro avversari. Può anche darsi che il dibattito su economia e società abbia risentito dell'influenza del processo che proprio in quel periodo si stava svolgendo nel mondo accademico: l'economia, la storia economica e la sociologia cercavano di guadagnarsi l'accesso all'università come discipline ufficiali e ciascuna di esse avvertiva l'esigenza di accentuare la propria specificità in rapporto alle altre scienze sociali; nessuna di esse voleva sovrapporsi alle altre, dato che un'eventuale sovrapposizione avrebbe pregiudicato l'ammissione fra le materie universitarie.
Prima di passare a illustrare brevemente il Methodenstreit e la scuola storica in economia è necessario ricordare che la scienza economica inglese aveva cominciato a perdere interesse per le istituzioni sociali anche prima del disastroso conflitto sul metodo. I riferimenti che si continuavano talvolta a fare alle classi sociali e a realtà consimili erano più che altro formalità di rito. A questo riguardo, presentando per grandi linee la situazione della scienza economica dominante nel periodo 1870-1914, Schumpeter nella sua Storia dell'analisi economica scrive: "[Secondo l'economia politica inglese] le nazioni continuavano ad essere amorfi agglomerati di individui. Le classi sociali non erano realtà vive e in conflitto tra loro, ma semplici etichette che venivano affibbiate a determinate funzioni economiche (o categorie funzionali). Neanche gli individui erano esseri vivi e in lotta fra loro: continuavano ad essere semplici 'corde da bucato' su cui stendere gli enunciati della logica economica. Conseguito

un maggior rigore formale, questa situazione si fece ancor più evidente di quanto non fosse nelle opere del periodo precedente. I critici sogghignavano: vedevano bene che si trattava di una sociologia da quattro soldi e di una psicologia ancor più povera" (v. Schumpeter, 1954).
Fin qui si è presentato soltanto il contributo degli inglesi alla scienza economica, in quanto furono proprio gli economisti inglesi che promossero lo sviluppo della teoria economica nei primi anni del XIX secolo. Un altro motivo per rivolgere l'attenzione a questo tipo di teoria economica è che durante il XX secolo la teoria economica dominante ha continuato a percorrere la strada segnata da Adam Smith, da Ricardo e da Mill. Tuttavia nel XIX secolo il panorama delle teorie economiche era piuttosto vario: fiorivano concezioni della vita economica alternative e le critiche al modello inglese erano abbastanza diffuse, soprattutto nei paesi europei di lingua tedesca, dove, nell'ambito della scuola storica, si sviluppò un tipo diverso di teoria economica. Questa corrente di pensiero è oggi in larga misura dimenticata, ma avrebbe meritato una sorte migliore, dato che costituì un importante tentativo di introdurre una prospettiva autenticamente sociale nell'economia.
La nascita della scuola storica è strettamente legata alla situazione economica e sociale della Germania dei primi anni del XIX secolo. Il nazionalismo, il romanticismo e lo storicismo nascenti si contrapponevano frontalmente all'universalismo che caratterizzava il pensiero economico inglese. Le giovani generazioni tedesche sapevano inoltre che il processo di industrializzazione non avrebbe seguito sul continente il medesimo corso che aveva seguito in Inghilterra. In poche parole esisteva una netta discrepanza tra la concreta realtà europea e il pensiero 'disincarnato' e universalistico che permeava le opere degli economisti britannici. Per colmare questo divario sarebbe stato necessario o creare un nuovo tipo di teoria economica oppure attenersi in modo più rigoroso alla realtà storica. Entrambe queste strade furono battute.
Le idee di Friedrich List anticipano le posizioni della scuola storica. Nella sua opera List rimprovera ad Adam Smith e agli altri economisti inglesi di non aver capito che ciascun paese si sviluppa in modo diverso e che la teoria economica deve innanzitutto occuparsi delle singole nazioni. Molte delle idee di List furono presto superate dalla scuola storica, che cominciò ad affermarsi negli anni quaranta dell'Ottocento. Si suole distinguere una 'vecchia' scuola storica, che fa capo a Wilhelm Roscher, Bruno Hildebrand e Karl Knies, e una 'giovane' scuola storica, che fa capo a Gustav von Schmoller e ai suoi discepoli. Si è osservato talvolta che questi economisti costituiscono un gruppo alquanto eterogeneo e non

una vera e propria scuola. Ciò è vero; tuttavia le loro idee presentano una certa unità, soprattutto nei confronti di quelle degli economisti inglesi. Tutti gli economisti tedeschi criticavano Adam Smith e gli altri, in quanto questi sostenevano il laissez faire e avevano elaborato una teoria economica universalistica non aderente alla realtà. In contrasto con gli economisti inglesi, gli esponenti della scuola storica miravano a individuare le leggi specifiche dello sviluppo economico di diversi paesi. Per quanto estremamente attenti ai dettagli storici, comunque, questi economisti non rifiutavano la dimensione teorica. La principale ambizione della scuola storica era quella di "analizzare modelli o processi particolari [della vita economica] nei loro dettagli concreti" (v. Schumpeter, 1954). Considerato l'oblio in cui è caduta ai giorni nostri la scuola storica (soprattutto in confronto all'attenzione riservata dagli studiosi alla teoria economica inglese) è opportuno spendere qualche parola su Roscher, Knies e Hildebrand. A Roscher va riconosciuto il merito di aver intuito per primo la possibilità di creare un nuovo tipo di economia introducendo la storia nel quadro della teoria economica. "Come si vedrà - dice Roscher - questo metodo si prefigge di raggiungere nel campo dell'economia politica i medesimi risultati ottenuti da Savigny e da Eichorn nell'ambito della giurisprudenza" (v. Roscher, 1895). Nella famosa prefazione al Grundriss zu Vorlesungen über die Staatswirtschaft nach geschichtlicher Methode (1843), Roscher elenca le ragioni principali dell'introduzione del metodo storico in economia: 1) l'economia deve studiare la nazione in quanto entità vivente; 2) bisogna sempre tenere conto del passato; 3) a meno che le nazioni non siano messe a confronto le une con le altre non è possibile stabilire se qualcosa è importante o meno; 4) anziché elogiare o criticare le istituzioni economiche bisognerebbe studiarle attentamente. Hildebrand, come Roscher, definisce l'economia come "la dottrina che studia lo sviluppo economico delle diverse popolazioni" (v. Hildebrand, 1848). Nella sua opera principale, Nationalökonomie der Gegenwart und Zukunft, egli spiega che cosa intenda per 'leggi economiche'. Poiché l'economia è una Kulturwissenschaft e non una Naturwissenschaft è evidentemente impossibile considerare le leggi economiche alla stregua di leggi naturali. Secondo Hildebrand l'economia è un prodotto dello spirito e della mente dell'uomo; di conseguenza egli mette in risalto il ruolo della moralità e della libertà nella vita economica.Secondo i membri della scuola storica il miglior testo di metodologia storicistica è Die politische Ökonomie vom Standpunkt der geschichtlichen Methode (1853) di Karl Knies. In quest'opera Knies sviluppa un interessante ragionamento sulle leggi economiche stabilendo che nel

campo dell'economia politica non è possibile individuare leggi causali ma solo 'leggi (rapporti) di analogia'. Per giungere a scoprire tali leggi sono necessari accurati studi comparativi: Knies fu anche il primo ad applicare il metodo storico al pensiero economico stesso. Nel suo trattato metodologico egli scrive: "In contrasto con chi concepisce la teoria in termini assoluti, la concezione storica dell'economia politica si basa sul principio che la stessa teoria economica [...], proprio come l'economia, è un risultato dello sviluppo storico" (v. Knies, 1853).

Si è osservato spesso che gli economisti della vecchia scuola storica non seguirono i criteri metodologici che essi stessi avevano stabilito. Nelle sue opere più tarde, per esempio, Roscher continuò a percorrere il cammino tracciato da Adam Smith e da Ricardo, e Knies deve la sua fama a un libro sulla moneta e sul credito che non si ispira affatto ai criteri dello storicismo. Max Weber (v., 1903-1906), inoltre, rimprovera alla vecchia scuola storica di essersi precipitata a definire le leggi economiche senza rendersi conto del ruolo che il significato (Verstehen) deve svolgere in un qualsiasi modello storico di teoria economica. Ma per quanto queste critiche siano fondate, agli economisti della vecchia scuola storica va riconosciuto il merito di aver concepito un nuovo tipo di teoria economica che prevedeva la coesistenza armoniosa di economia e società.

A partire dagli anni settanta dell'Ottocento divenne evidente che la vecchia scuola storica stava per essere rimpiazzata da una nuova generazione di economisti storici. L'esponente di maggior spicco della cosiddetta 'giovane scuola storica' fu, come si è detto, Gustav von Schmoller, che divenne ben presto il più importante economista tedesco. Schmoller si prodigò per dare un orientamento riformista all'economia storica e nel 1872 fu tra i fondatori del "Verein für Sozialpolitik". Egli mise anche a punto un particolare programma di ricerca nel campo dell'economia storica. Secondo Schmoller si deve sempre partire da un'indagine storiografica ed empirica su un argomento ristretto; si può studiare, per esempio, l'attività dei fabbricanti di stoffe e dei tessitori di certe città (Schmoller in effetti dedicò un libro a questo argomento). Quante più monografie di questo genere si fossero realizzate tanto più facile sarebbe stato, in un secondo tempo, procedere a una sintesi generale. Ma questo voleva dire che prima di poter cominciare a generalizzare sarebbe stata necessaria un'enorme mole di lavoro. Verso la fine della sua vita Schmoller pose mano a tale sintesi generale e scrisse un'imponente opera in due volumi, il Grundriss der allgemeinen Volkswirtschaftslehre (1900-1904). Secondo i critici, l'opera di Schmoller, ancorché di portata veramente

enciclopedica, non è però molto ricca di idee.Quando il Grundriss di Schmoller vide la luce la scuola storica era al culmine della sua influenza. In pratica tutte le facoltà universitarie di economia e tutte le principali riviste di economia tedesche erano sotto il suo controllo. Nell'arco di appena vent'anni la situazione cambiò completamente: dei contributi della scuola storica scomparve ogni traccia ed essa perse tutto il suo prestigio. Ma per quale motivo? Cosa era accaduto nel frattempo di così importante da mutare in modo tanto radicale le sorti della scuola storica? La risposta è: il Methodenstreit. Scoppiata negli anni ottanta, questa famosa guerra tra i metodi avrebbe avuto le conseguenze più disastrose per l'impostazione storicistica della scienza economica. La disputa esplose nel 1883, quando Carl Menger, un brillante economista austriaco - uno degli ideatori dell'analisi dell'utilità marginale - pubblicò un saggio dal titolo pomposo Untersuchungen über die Methode der Sozialwissenschaften und der politischen Ökonomie insbesondere. Secondo Menger gli economisti della scuola storica avevano completamente frainteso la natura della scienza economica. La scienza economica, spiega Menger, si suddivide in tre parti - una parte storica, una teorica e una pratica - e ciascuna di esse ha un suo specifico ambito di competenza. L'errore degli economisti storici tedeschi - un errore imperdonabile per Menger - è stato l'aver confuso il proprio campo di studi con quello dell'economia teorica. Secondo Menger l'economia storica può occuparsi soltanto di eventi economici particolari, collegati tra loro in modo particolare. Rientrano perciò nell'ambito di competenza dell'economia storica argomenti quali le istituzioni economiche, lo sviluppo economico e il ruolo dello Stato nell'economia. L'economia teorica, invece, si occupa dei fenomeni economici generali, collegati tra loro in modo generale. Tra gli argomenti propri dell'economia teorica Menger cita i prezzi, la domanda e l'offerta e lo scambio. L'economia teorica può essere o empirica ('realistico-empirica') o esclusivamente teorica ('esatta'). Tramite la versione empirica si possono individuare regolarità e leggi empiriche, ma non è possibile determinare leggi analoghe a quelle delle scienze naturali: per far ciò è indispensabile l'economia teorica 'esatta'.
Chiaramente questo modo di vedere l'economia non escludeva il punto di vista storico e sociale, anche se per Menger la versione 'esatta' dell'economia teorica rappresentava la forma più alta di scienza economica. Ma per quale motivo, allora, il Methodenstreit finì con l'eliminare, in pratica, l'economia di tipo storico anziché limitarsi a relegarla in una posizione di secondo piano? La risposta a questa domanda è che l'animosità personale tra le parti in causa trasformò il dibattito sull'importanza relativa dell'economia teorica rispetto a quella

storica in uno scontro per decidere se dovesse sopravvivere l'economia teorica o quella storica. Così, per esempio, quando recensì il libro di Menger, Schmoller osservò con asprezza che non vi era più alcun bisogno dell'economia teorica. Questo tipo di economia, a suo avviso, non avrebbe fatto altro che "continuare a distillare teoremi astratti intrisi di vecchio dogmatismo già ricavati centinaia di volte" (v. Schmoller, 1883). Per Schmoller il tentativo di Menger di elaborare un'economia teorica si riduceva a "esercitazioni intellettuali scolastiche false quanto le loro premesse".
La recensione di Schmoller mandò su tutte le furie Menger, che replicò immediatamente con un libello intitolato Die Irrthümer des Historismus in der deutschen Nationalökonomie (1884), dove Schmoller veniva accusato di perdersi in "ridicoli dettagli" (Kleinmalerei). Di lì a poco Menger e Schmoller smisero di rivolgersi la parola e lo stesso fecero molti dei loro discepoli. Di conseguenza il mondo accademico di lingua tedesca si spaccò in due fazioni rivali che si detestavano profondamente. Gli storicisti venivano descritti dai loro avversari come perdigiorno interessati esclusivamente a ridicoli dettagli storici, mentre gli economisti teorici erano a loro volta accusati di dedicarsi soltanto a teorie astruse. Nessuno di questi due ritratti era vero, ovviamente, ma entrambi vennero accettati con passione.Il tipo di scienza economica che emerse vittoriosamente dal Methodenstreit era molto diverso da quello per cui si era battuto Menger col suo libro del 1883. Secondo Menger la storia e la società dovevano rientrare nell'ambito dell'analisi economica, anche se in posizione subordinata. Nella scienza economica degli anni venti del Novecento, invece, non c'era più posto né per la storia né per la società. La storia e la società furono dichiarate di competenza di altre discipline (la storia economica e la sociologia), mentre la scienza economica divenne economia 'pura'. Lo stesso tipo di situazione - va detto - si verificò, oltre che in Austria e in Germania, anche in altri paesi, come l'Inghilterra e gli Stati Uniti. Anche questi due paesi ebbero il loro Methodenstreit, e con gli stessi disastrosi risultati.

4. La sociologia economica: nascita e primi sviluppi

Proprio mentre il Methodenstreit raggiungeva il suo culmine, nel campo delle scienze sociali, quasi inosservato, si verificò un evento destinato a segnare profondamente il dibattito sui rapporti tra economia e società: la nascita della 'sociologia economica' (economic sociology; sociologie économique; Wirtschaftssoziologie). Come si è visto, economisti inglesi e

tedeschi avevano già cercato di studiare alcuni fenomeni economici avvalendosi di analisi di tipo sociale. Per autori come Adam Smith o John Stuart Mill, comunque, il fulcro dell'analisi era sempre altrove, ed essi stessi si consideravano studiosi dell'economia molto più che della società. Economisti come Roscher e Schmoller erano ovviamente molto più favorevoli dei loro colleghi britannici a un'impostazione di tipo sociale. D'altra parte essi fondamentalmente intendevano introdurre metodi storici nell'economia, ma non erano particolarmente interessati alla teoria sociale in sé. Di conseguenza solo alla fine del XIX secolo poté vedere la luce una sociologia economica ben elaborata e consapevole.La sociologia economica nacque nell'ultimo decennio del secolo scorso essenzialmente ad opera di due autori: Max Weber ed Émile Durkheim. Dato che tra i due non vi fu mai alcun rapporto diretto, si è trattato di un caso di scoperta multipla. Tanto Weber che Durkheim hanno scritto sia opere di taglio programmatico e metodologico sia lavori di sociologia economica applicata a casi concreti. La loro influenza ha anche indotto alcuni loro allievi e colleghi - tra cui studiosi illustri quali Marcel Mauss, François Simiand, Georg Simmel, Werner Sombart e Joseph Schumpeter - a dedicarsi alla nuova disciplina. Ciò che bisogna sottolineare è che la sociologia economica creata da questi studiosi rappresenta il tentativo più elaborato che sia mai stato fatto per mettere l'analisi sociale in relazione con quella economica.
La sociologia economica francese affonda le sue radici nella critica di Auguste Comte all'economia politica classica. Nel suo Cours de philosophie positive (18301842), Comte aveva rinfacciato agli economisti di non aver capito che la scienza economica deve rientrare nell'ambito della scienza sociale, e Durkheim più tardi riprese le tesi di Comte a questo proposito. Meno facile da stabilire, invece, è il momento esatto in cui Durkheim concepì la sociologia economica: probabilmente verso la metà degli anni ottanta, anche se il termine 'sociologie économique' compare per la prima volta nei lavori di Durkheim alla fine degli anni novanta. La sociologia economica, secondo Durkheim, è quella branca della sociologia che si occupa delle istituzioni economiche. Il suo compito principale coincide dunque con quello della sociologia in generale, cioè consiste nell'individuare le origini delle istituzioni, i rispettivi sviluppi e le funzioni che svolgono nella società. Sulla base di questa definizione si potrebbe pensare che Durkheim fosse un istituzionalista, come Veblen o Mitchell. Ciò non è vero: anche nell'analisi dei fenomeni economici, infatti, Durkheim ricorre al proprio modello sociologico originale e ai propri concetti più tipici: 'rappresentazione collettiva', 'anomia', ecc.Il primo contributo di Durkheim alla sociologia economica è rappresentato dalla sua tesi

di laurea, De la division du travail social (1893). Durkheim riconosce ad Adam Smith il merito di aver introdotto il concetto di divisione del lavoro nelle scienze sociali. Ma la concezione degli economisti classici è troppo ristretta per Durkheim, secondo il quale è decisamente sbagliato concepire la divisione del lavoro in termini esclusivamente economici e come un fattore il cui ruolo si esaurisce nel far aumentare la ricchezza della società. Il vero ruolo della divisione del lavoro è, per Durkheim, molto diverso: consiste nel creare solidarietà e coesione sociali. Nella società primitiva - sostiene Durkheim - la convivenza era garantita dal fatto che gli uomini erano simili e tutti facevano e pensavano le stesse cose ('solidarietà meccanica'). Nella società industriale, invece, i vari individui si occupano di cose diverse e hanno opinioni diverse a causa della divisione del lavoro molto spinta. Ma ciò pone anche le basi per un nuovo tipo di coesione sociale: ciascuno ora ha bisogno degli altri, esattamente nello stesso modo in cui i diversi organi del corpo necessitano l'uno dell'altro ('solidarietà organica'). L'utilizzazione da parte di Durkheim di concetti quali quelli di funzione, organo e corpo sociale mostra anche come egli abbia ampliato l'idea di divisione del lavoro sino a includervi anche quella forma di divisione del lavoro (dei ruoli) che sussiste, nella società in generale, fra differenti 'settori'.
Nell'opera De la division du travail social, Durkheim critica aspramente l'idea degli economisti secondo cui il fatto che ognuno persegua i propri interessi si risolve in sostanza nel bene della società. Secondo Durkheim gli interessi non sono in grado di unire le persone per la semplice ragione che non possono produrre nient'altro che contatti brevi e superficiali: "Se anche gli interessi comuni avvicinano gli uomini fra loro, ciò avviene solo per brevissimi periodi di tempo. Gli interessi possono dar luogo solo a legami artificiali" (v. Durkheim, 1893). Durkheim era scettico anche riguardo all'idea che le azioni basate sull'interesse conducessero comunque al bene della società o, per dirla con Mandeville, che quelli che sul piano individuale erano "vizi privati" si trasformassero in "pubbliche virtù" sul piano collettivo. Per Durkheim un grave problema della società moderna è che le attività economiche non producono automaticamente coesione sociale. Proprio per questa sua incapacità di dar luogo a un sistema coesivo di regole l'economia moderna è "costantemente in una condizione di instabilità" (ibid.). Il termine usato da Durkheim per designare questa mancanza di regole nella società è 'anomia'. Successivamente, nel suo lavoro dedicato al suicidio, Durkheim cercò di mostrare in che modo l''anomia economica' possa addirittura, in determinate circostanze, spingere le persone a togliersi la vita.

Dall'analisi condotta nel libro sulla divisione del lavoro Durkheim trasse la conclusione pratica che fosse necessario fare qualcosa per regolare l'economia e per porre fine a questa condizione di anomia. Per rendere stabile la società bisognava introdurre a tutti i costi delle norme nella vita economica. Come soluzione di questi problemi Durkheim suggerì di introdurre nel sistema economico una versione aggiornata del modello corporativo. All'interno di queste corporazioni o organizzazioni professionali gli individui avrebbero potuto consolidare i rapporti reciproci in un modo altrimenti impossibile nella società moderna.La soluzione corporativistica venne presentata da Durkheim nel 1902, nella celebre prefazione alla seconda edizione dell'opera De la division du travail social, ma le idee al riguardo erano state sviluppate molto prima e avevano costituito il fulcro di un corso tenuto da Durkheim nel 1894. Le dispense del corso furono successivamente pubblicate con il titolo Leçons de sociologie: physique des moeurs et du droit (1950). Questo testo contiene molti interessanti contributi alla sociologia economica: oltre alla disamina dei gruppi professionali, esso presenta anche un tentativo di analizzare la proprietà privata e il rispetto che questo tipo di proprietà ispira nella società. Benché troppo complesso per essere illustrato in dettaglio, il ragionamento di Durkheim merita comunque di venir sinteticamente riassunto, in quanto dà un'idea del modo in cui, a suo avviso, bisognerebbe analizzare le istituzioni economiche. Secondo Durkheim l'aura di rispetto che circonda la proprietà privata ricorda i tabù relativi agli oggetti sacri, per cui, probabilmente, il rispetto per la proprietà ha origini religiose. In modo estremamente ingegnoso egli descrive quindi la genesi della proprietà privata e come in un primo tempo essa avesse un significato religioso, che è poi venuto meno, lasciando però dietro di sé una sorta di fascia protettiva intorno alla proprietà privata, il 'rispetto' appunto. Anche la terza importante opera di sociologia economica di Durkheim - Le socialisme (1928) - trae origine da un corso: quello tenuto nell'anno accademico 1895-1896. Il libro, che contiene un'analisi molto originale del socialismo e un importante esame dell'opera dell'economista svizzero Sismondi, raggiunge comunque il suo punto più alto là dove Durkheim interpreta e critica brillantemente le tesi di Saint-Simon. Questa parte del lavoro di Durkheim è così originale e stimolante che si sarebbe tentati di sostenere che rappresenta l'analogo francese del testo di Max Weber Die protestantische Ethik und der Geist des Kapitalismus. In breve Saint-Simon aveva sostenuto che, in Europa, alla società feudale e bellicosa si stava progressivamente sostituendo una società industriale e pacifica. In questa nuova società, secondo Saint-Simon, l'industria sarebbe diventata la principale attività e lo

scopo primario delle persone. L'industrializzazione avrebbe anche permeato il sistema di valori della società. Secondo Saint-Simon, conclude in sintesi Durkheim, "la società deve diventare un'enorme impresa produttiva" (v. Durkheim, 1928). Nel nuovo stato industriale, inoltre, la politica sarebbe stata dedicata esclusivamente a questioni di ordine economico: "La politica è la scienza della produzione" (Saint-Simon, cit. in Durkheim, 1928).
Personalmente Durkheim trovava ripugnante la società industriale preconizzata da Saint-Simon e sosteneva che proprio come gli economisti anche Saint-Simon aveva erroneamente concepito l'economia come la parte più importante della società. Che Saint-Simon sostenesse questa posizione non sorprende, dato che - secondo Durkheim - il socialismo e l'economia politica condividevano gli stessi valori: "In entrambi i sistemi le persone non hanno altro scopo al di fuori del proprio benessere materiale e la società è fine a se stessa" (v. Durkheim, 1928). Per Durkheim la società di tipo industriale o economico sarebbe stata, comunque, una calamità per i cittadini, perché individui interessati esclusivamente al proprio benessere materiale avrebbero cercato di conseguire obiettivi sempre nuovi e non sarebbero mai stati felici; avrebbero voluto comprare sempre di più nella vana speranza di arrivare ad avere abbastanza. L'uomo moderno, quindi, deve, per Durkheim, superare la prospettiva di una 'società industriale' - comune all'economia politica e al socialismo - e sostituirla con quella di una società la cui economia si basi sulla moralità e sulla solidarietà. Senza moralità, conclude Durkheim, non è possibile alcuna società; una 'società industriale' è una contraddizione in termini.
Sembra che verso la fine della sua attività Durkheim abbia perso interesse per le questioni economiche e si sia dedicato sempre di più a problemi morali e religiosi. Tuttavia che continuasse a considerare la sociologia economica molto importante risulta chiaro dal fatto che dedicò sempre a questa disciplina una sezione della rivista da lui diretta, "L'année sociologique" (1896-1912). Durkheim inoltre continuò a incoraggiare i suoi collaboratori a lavorare su temi economici, e ciò condusse alla realizzazione, in Francia, di alcune importanti ricerche. Tra queste le più interessanti sono dovute a Marcel Mauss e a François Simiand. Mauss si occupava prevalentemente di antropologia mentre Simiand era uno specialista di temi di attualità e storici. La più importante opera di argomento economico di Mauss è una breve monografia intitolata Essai sur le don: forme et raison de l'échange dans le sociétés archaïques (1923-1924). Come risulta chiaro dal titolo, per Mauss lo scambio di doni non è una questione di generosità ma una forma di scambio. Il dono implica tutta una serie di

obblighi: l'obbligo di dare, l'obbligo di ricevere e l'obbligo di ricambiare. Lo studio di Mauss è pieno di brillanti digressioni sui più svariati argomenti economici. Molte di queste digressioni servono a dimostrare la tesi di Mauss secondo cui per comprendere le istituzioni economiche contemporanee è necessario conoscere le forme precedenti. La forma arcaica del dono, per esempio, non implica soltanto le nozioni di vendita e di acquisto, ma anche quelle di credito e di elemosina.

L'opera di Simiand è profondamente diversa da quella di Mauss. Agli inizi della sua attività Simiand pubblicò un lavoro metodologico che ebbe una certa risonanza: La méthode positive en science économique (1912). Simiand vi criticava radicalmente la teoria economica ortodossa ("uno scandalo metodologico") e avanzava quindi una propria alternativa di stampo sociologico. Ma l'opera più importante di Simiand è una monumentale ricerca in tre volumi sui salari in Francia tra la fine del XVIII e gli inizi del XX secolo: Le salaire, l'évolution sociale et la monnaie (1932). Questo lavoro rappresenta una novità: Simiand fu il primo sociologo a studiare un fenomeno economico nei suoi dettagli empirici durante un periodo di tempo tanto lungo. Per spiegare l'aumento e la diminuzione dei salari nei diversi periodi Simiand prese in considerazione varie possibili teorie. Tra queste una in particolare sottolineava il ruolo svolto dalla moneta, un argomento cui Simiand dedicò anche un importante articolo, La monnaie, réalité sociale. Come Durkheim, Simiand insiste sul fatto che la moneta è un fenomeno eminentemente sociale. La moneta - scrive Simiand - è una realtà oggettiva perché esprime "una credenza sociale e una fede sociale, ed è perciò una realtà sociale" (v. Simiand, 1934).

Tra la scuola di sociologia economica tedesca e quella francese sussistono alcuni evidenti parallelismi. Entrambe sorsero negli anni novanta dell'Ottocento e si spensero nei primi anni trenta; entrambe furono dominate da una figura di spicco (Durkheim in Francia e Weber in Germania); entrambe erano basate sul presupposto che la sociologia economica dovesse essere una disciplina storica e comparativa. Vi sono tuttavia anche alcune interessanti differenze fra le due scuole. Mentre la sociologia economica francese era decisamente positivistica, per gli studiosi tedeschi dell'epoca era scontato che l'economia rientrasse nell'ambito delle scienze della cultura (Kulturwissenschaften) e non di quelle della natura (Naturwissenschaften). Intorno al 1900, per esempio, Friedrich Gottl diede avvio al suo tentativo sistematico di introdurre il Verstehen in economia, e d'altra parte la sociologia economica di Weber poggia saldamente sui concetti di significato e di comprensione.

L'atteggiamento della sociologia economica tedesca nei confronti della teoria economica fu inoltre condizionato dal Methodenstreit. Alla luce della radicale polarizzazione in atto tra teoria astratta e conoscenza storica, fu relativamente facile per i sociologi tedeschi rendersi conto che entrambi gli schieramenti in lizza nella guerra dei metodi avevano torto. In effetti la sociologia economica tedesca nacque in larga misura come un tentativo di superare il Methodenstreit. Ciò è particolarmente vero per quel che riguarda proprio Weber, che nel suo lavoro cercò sempre, con estrema consapevolezza, di giungere a un equilibrio tra materiale storico e teoria economica. Il celebre concetto di 'tipo ideale', per esempio, è il risultato di una siffatta sintesi. In generale i sociologi dell'economia tedeschi ritenevano di poter trarre utili insegnamenti dalla teoria economica, benché si sentissero anche scettici nei suoi confronti. In definitiva, però, non le erano ostili, contrariamente a Durkheim e ai suoi seguaci.

La sociologia economica attrasse molti brillanti studiosi in Germania, tra cui Weber, Simmel, Schumpeter, Bücher e Sombart. I due autori che si dedicarono più sistematicamente alla nuova disciplina furono Weber e Schumpeter. Simmel scrisse un'opera interdisciplinare, intitolata Philosophie des Geldes (1900), che contiene anche alcune brillanti osservazioni sociologiche sulla moneta. Simmel, per esempio, fu il primo a mettere in evidenza il ruolo della fiducia nella vita economica. Bücher e Sombart non si consideravano sociologi, ma facevano piuttosto parte di una generazione di economisti storici molto influenzati dalla sociologia. Bücher è noto soprattutto per la sua stimolante concezione istituzionalista esposta in Die Entstehung der Volkswirtschaft (1893), un testo che divenne molto popolare e non soltanto in Germania. Sombart fu un pensatore audace e imprevedibile, che cambiò spesso posizione su questioni teoriche. Ai fini del presente articolo basta ricordare il suo interessante tentativo di sviluppare una "verstehende Nationalökonomie" in uno dei suoi ultimi lavori (v. Sombart, 1930). Precedentemente Sombart aveva dedicato diversi anni alla stesura di Der moderne Kapitalismus (1902), un grandioso tentativo alla Marx di tracciare tutto lo sviluppo del capitalismo.

La sociologia economica di Weber è stata relativamente trascurata. Se si volesse studiare a fondo questo aspetto della sua opera, comunque, bisognerebbe prendere le mosse dalla sua dissertazione di laurea sulle compagnie commerciali nel Medioevo (1889) e dall'Habilitationsschrift sulla storia agraria romana del 1891. Prima dell'esaurimento nervoso di cui soffrì nel 1897 Weber portò anche a termine due studi importanti per la sociologia

economica: una ricerca sulla condizione dei contadini a est dell'Elba, pubblicata in un unico, grosso volume nel 1892, e una ricerca sulla borsa valori, pubblicata in diversi scritti minori (1894-1897). In entrambi i casi Weber tratta argomenti economici da un punto di vista decisamente sociale. Nel lavoro sui contadini, per esempio, Weber rileva come i contadini tedeschi e i contadini polacchi avessero opinioni assai diverse su ciò che costituisce un accettabile standard di vita, ragione per cui i lavoratori polacchi stavano progressivamente costringendo i tedeschi ad abbandonare la zona a est dell'Elba. Egli sottolinea inoltre che i lavoratori tedeschi della zona preferivano essere liberi in altre parti della Germania che restare nel loro luogo di origine dove godevano di una certa sicurezza economica, ma avevano poche possibilità di migliorare la propria situazione. Il punto essenziale sostenuto da Weber nel suo studio sulla borsa valori è che - contrariamente all'opinione prevalente in Germania all'epoca - le operazioni finanziarie che si concludono in ogni borsa valori moderna non sono di per sé fraudolente, ma costituiscono una forma di speculazione razionale necessaria in un regime di capitalismo avanzato. L'impostazione sociologica di Weber trapela anche dal suo esame dei diversi modi in cui si può evitare che il codice deontologico relativo alle attività di borsa venga violato.

L'intenzione di Weber di continuare questi studi sull'economia è testimoniata dal fatto che appena si fu ripreso dalla malattia egli cominciò immediatamente a lavorare a un saggio critico sulla 'vecchia scuola storica'. Due anni più tardi vide la luce la sua celebre opera Die protestantische Ethik und der Geist des Kapitalismus (1904-1905). Criticando Roscher e Knies Weber mette in evidenza il fatto che per condurre uno studio storico dell'economia è indispensabile abbandonare il concetto di legge economica: l'approccio storico e il concetto di legge economica sono semplicemente incompatibili. Il testo sull'etica protestante è il primo di una serie di straordinari lavori dedicati alle diverse religioni del mondo, in cui Weber analizza i rapporti tra religione ed 'etica economica' (Wirtschaftsethik). Con quest'ultimo concetto Weber intende designare non tanto il punto di vista della religione su questioni economiche quanto "gli impulsi pratici all'azione che si basano sulla psicologia e sui contesti pragmatici della religione". Va detto che il concetto di etica economica di Weber ha dato luogo a una sociologia tuttora insuperata per ricchezza e sottigliezza.

Weber cominciò a definirsi sociologo (anziché economista) intorno al 1910; pochi anni più tardi apparve per la prima volta nei suoi scritti il termine Wirtschaftssoziologie. La ricerca sui lavoratori dell'industria (1908-1909) e quella sulla condizione agraria nel mondo antico

(1909) rientrano nell'ambito della sua sociologia economica. Ma i due più importanti contributi di Weber a questa disciplina risalgono a qualche anno più tardi; si tratta di Economia e società (pubblicato postumo) e di una serie di lezioni tenute nell'anno accademico 1919-1920, dal titolo Lineamenti di storia economica e sociale universale (anche i Lineamenti, basati sugli appunti degli studenti, furono pubblicati postumi). Economia e società presenta tutta una serie di osservazioni e di analisi attinenti alla sociologia economica. La parte più importante del testo è, comunque, il secondo capitolo, dove Weber cerca di delineare i fondamenti teorici della sociologia economica. Il capitolo in questione è intitolato Categorie sociologiche dell'azione economica e con le sue 150 pagine ha le dimensioni di un piccolo libro autonomo. Esso contiene una rassegna scarna e molto sintetica delle categorie fondamentali della sociologia economica. In effetti quando Weber tenne le lezioni relative a questo capitolo gli studenti lo pregarono di tenere un corso di storia economica per agevolare la comprensione di tutta la materia: fu questa l'origine del corso sulla storia economica e sociale universale. Il secondo capitolo di Economia e società e il corso del 19191920 sono stati concepiti come complementari e pertanto andrebbero letti insieme.Nel secondo capitolo di Economia e società Weber chiarisce che il suo intento non è quello di proporre una teoria economica, ma di "analizzare alcune delle più elementari relazioni sociologiche presenti nella sfera economica" (v. Weber, Wirtschaft..., 1922). A ciò Weber aggiunge che anche se la sociologia economica era destinata a poggiare sui fondamenti della teoria economica, tuttavia avrebbe dovuto elaborare "i propri costrutti teorici" (ibid.). Per sapere a quali costrutti alluda Weber basta leggere attentamente il secondo capitolo, che abbonda di definizioni, tra cui quelle di 'azione economica', 'organizzazione economica', 'moneta', 'commercio', 'impresa', ecc. A proposito di tali definizioni vi sono due osservazioni da fare, particolarmente degne di nota. La prima è che nel fornire tutte queste definizioni Weber rivendica la natura interpretativa della sociologia economica: per il sociologo, egli scrive, i fenomeni economici sono "caratterizzati [...] completamente dal significato che hanno" (ibid.). La seconda è che, nel piano delineato da Weber, la sociologia economica, contrariamente alla teoria economica, deve sempre tener conto del potere economico. Il termine esatto usato da Weber per indicare il potere economico è 'Verfügungsgewalt', che letteralmente significa 'poteri di controllo e gestione sanciti legalmente'. Per Weber categorie quali quelle di 'prezzo', 'scambio', 'profitto', 'impresa' implicano sempre conflitti e compromessi tra interessi diversi ovvero ciò che egli chiama "la

lotta dell'uomo contro l'uomo".
Le lezioni weberiane sulla storia economica e sociale universale occupano all'incirca 250 pagine e coprono il periodo che va dalla prima forma di agricoltura in Germania fino agli inizi dell'Ottocento. L'esposizione è estremamente sintetica e il libro è, a modo suo, altrettanto arduo da leggere quanto il secondo capitolo di Economia e società. Weber passa in rassegna diverse forme di agricoltura, studia le prime forme di attività mineraria e industriale, analizza il commercio e il ruolo della moneta nel mondo antico e nel Medioevo e infine riassume la propria teoria sulla nascita del capitalismo. Se il secondo capitolo di Economia e società getta le basi concettuali della sociologia economica, la serie di lezioni sulla storia economica e sociale universale svolge il medesimo ruolo per quanto riguarda la teoria delle istituzioni economiche. General economic history (è questo il titolo con cui la serie di lezioni è stata pubblicata in inglese, tradotta da Frank Knight) è a tutt'oggi il più importante lavoro che sia mai stato scritto sulle istituzioni economiche.
Prima di abbandonare Weber è necessario aggiungere qualche considerazione sulle sue tesi a proposito dei rapporti tra economia e società nel contesto del processo di modernizzazione. È chiaro che l'idea di razionalità costituisce il grande tema unificante di tutta l'opera weberiana e che Weber era particolarmente affascinato da ciò che egli definiva "lo specifico e peculiare razionalismo della cultura occidentale" (v. Weber, 1904-1905). Secondo Weber si possono riscontrare elementi di razionalità in tutte le culture e in molte diverse aree della società, quali il diritto, la scienza, la burocrazia, l'arte e la musica; ma soltanto nella civiltà occidentale il processo di razionalizzazione in generale e nella sfera economica in particolare si è spinto tanto avanti. Per Weber il mercato rappresenta "l'archetipo di ogni azione sociale razionale" e i mercati sono stati ovunque presenti nel mondo occidentale (v. Weber, Wirtschaft..., 1922). Ma la causa prima e più importante dell'esplosivo sviluppo dell'economia in Europa e negli Stati Uniti è un'altra, e cioè la calcolabilità. In Occidente il processo produttivo è completamente calcolabile: si può prevedere facilmente quanto rapidamente e con quanta precisione un operaio o una macchina possono produrre un determinato oggetto. Ed ugualmente calcolabile è l'ambiente in cui il sistema economico è inserito: si può infatti prevedere molto facilmente anche che cosa faranno i giudici e i politici in diverse circostanze. Analizzando la razionalità nella sfera economica Weber sottolinea la differenza tra "razionalità formale" e "razionalità sostanziale" (ibid.). Può darsi il caso che una situazione sia perfettamente razionale da un punto di vista formale (per esempio la

separazione degli operai dai mezzi di produzione), ma irrazionale dal punto di vista sostanziale (per esempio da quello del socialismo o del comunismo). Dal canto suo Weber era restio a inserire i propri ideali nei propri scritti sociologici. Sembra comunque chiaro che egli pensasse che il capitalismo fosse, in ultima analisi, privo di qualsiasi valore umano, fosse, secondo la celebre metafora contenuta nell'Etica protestante, una "gabbia di ferro" (v. Weber, 1904-1905). Schumpeter viene generalmente presentato come un economista neoclassico dotato di un'ampia prospettiva storico-sociale; ma questa definizione non coglie appieno l'essenza della sua opera di economista: la parte principale della sua attività si inquadra infatti nell'ambito di un paradigma economico di marca decisamente weberiana (v. Swedberg, 1991). Come Weber, Schumpeter considera l'economia una disciplina vasta e comprensiva, articolata in parecchi campi distinti. Il termine tedesco adottato da Schumpeter (sulla scia di Weber) per designare questo tipo di economia è Sozialökonomik. Questo termine dimostra chiaramente che Schumpeter considera l'economia una scienza sociale. I differenti campi che costituiscono la Sozialökonomik sono, secondo Schumpeter, la teoria economica, la storia economica, la sociologia economica e la statistica economica (v. Schumpeter, 1954). Ciascuno di questi quattro campi ha la propria specifica area di ricerca, anche se i confini tra loro non sono rigidi. Mentre la teoria economica, per esempio, si occupa dei meccanismi economici, ovvero del perché in economia le cose funzionano come funzionano, la sociologia economica si occupa delle istituzioni economiche, ovvero del perché le cose sono diventate ciò che sono nel corso della storia.

Il modo esemplare in cui Schumpeter ha espresso e ampliato la concezione weberiana della Sozialökonomik costituisce uno dei suoi contributi fondamentali al dibattito sui rapporti tra economia e società. Un altro contributo altrettanto importante è costituito dai tre saggi sociologici sull'imperialismo, sulle classi sociali e sullo Stato fiscale. Tra questi quello forse più suggestivo è The crisis of the tax State (1918), dove Schumpeter cerca di sviluppare una sociologia della finanza e analizza in modo assai originale il ruolo dello Stato nell'economia. Il terzo contributo fondamentale di Schumpeter alla sociologia economica è rappresentato dai suoi lavori di teoria economica, a partire almeno dalla Theorie der wirtschaftlichen Entwicklung (1911). Per lo stesso Schumpeter la sua teoria sull'imprenditore rientra nell'ambito della teoria economica pura. È chiaro, comunque, che determinati aspetti della sua teoria sono sociologici: per esempio, l'analisi della resistenza che ogni imprenditore generalmente incontra nell'attuare una nuova concentrazione. Anche Capitalism, socialism

and democracy (1942) contiene molte acute riflessioni sociologiche su argomenti quali il ruolo della concorrenza nel capitalismo avanzato, i vantaggi di cui godono le grandi imprese nel processo di innovazione, ecc.

5. Il revival degli anni cinquanta (Parsons, Polanyi)

Nel ventennio 1930-1950 il dibattito sui rapporti tra economia e società praticamente cessò. Gli economisti volevano studiare soltanto l'economia (esclusa la sua dimensione sociale), mentre i sociologi si dedicavano unicamente all'analisi della società (economia esclusa). Nei primi anni cinquanta Schumpeter sintetizzò malinconicamente lo stato dei rapporti tra economisti e sociologi nel modo seguente: "il fatto è che fin dal XVIII secolo i due gruppi hanno proceduto sempre separati finché ora il tipico economista e il tipico sociologo sanno poco e si interessano ancor meno l'uno del lavoro dell'altro, e ciascuno dei due preferisce usare, rispettivamente, una propria sociologia primitiva e una propria economia primitiva anziché accettare i risultati professionali dell'altro; questo stato di cose non viene certo attenuato dalla persistente tendenza al reciproco scambio di insulti" (v. Schumpeter, 1954).
I motivi per cui in tale periodo il dibattito sui rapporti tra economia e società praticamente cessò sono molteplici. Per quel che riguarda la sociologia, sembra che gli studi avviati da Durkheim e da Weber si fossero esauriti verso il 1930. I durkheimiani non erano più in grado di rinnovarsi, mentre la comunità dei sociologi tedeschi era stata decimata dalla prima guerra mondiale (e lo sarà, poi, dal nazismo). Negli Stati Uniti, dove la posizione della sociologia era molto più solida, i sociologi non erano interessati alle questioni economiche, e anche gli economisti, dal canto loro, avevano, per diverse ragioni, perso interesse per la questione dei rapporti fra economia e società. Nel frattempo l'economia istituzionalista era caduta in discredito e se si voleva essere considerati studiosi seri si doveva tenersene alla larga. L'economia neoclassica andava intanto assumendo un ruolo sempre più dominante ed era estremamente difficile introdurre variabili sociali in tale tipo di analisi. Ciò è vero in particolar modo per quel che riguarda la versione matematica dell'economia neoclassica, che si stava affermando negli anni trenta e quaranta. Vero è che Keynes cercò di introdurre un certo realismo nell'economia accademica, ma questo tentativo fu rapidamente soffocato quando le sue idee furono assorbite nell'ambito della nuova ortodossia vigente negli anni quaranta e cinquanta.

Sarebbe decisamente sbagliato dare l'impressione che il dibattito sul tema 'economia e società' si sia riacceso con veemenza negli anni cinquanta. In realtà, infatti, la netta separazione tra economia e società non sarebbe stata posta seriamente in discussione se non negli anni ottanta. Negli anni cinquanta, comunque, furono fatti due interessanti tentativi per avviare tale dibattito. Il primo fu promosso da un gruppo di sociologi, fra cui, in particolare, Talcott Parsons, Neil Smelser e Wilbert E. Moore; il secondo, di ispirazione antropologica, da Karl Polanyi. Entrambi questi tentativi si concretizzarono in alcuni libri interessanti e suscitarono vivaci polemiche.Tra i sociologi il giovane Neil Smelser fu quello che più si dedicò alla sociologia economica, intesa come specifico sottosettore della sociologia. Dopo essersene già occupato nella sua tesi di laurea, Smelser scrisse il primo manuale di sociologia economica e il primo testo introduttivo all'argomento. Inoltre Smelser fu coautore, con Talcott Parsons, del famoso Economy and society (1956). Insieme a Trade and market in the early empires (pubblicato a cura di Karl Polanyi e altri nel 1957), Economy and society è la più importante opera di sociologia economica scritta dagli anni cinquanta in poi. Inoltre ancor oggi questo libro si rivela importante per le molte idee sui rapporti tra teoria economica e teoria sociologica in esso contenute.L'obiettivo principale di Economy and society è favorire l'integrazione fra teoria economica e teoria sociologica. Secondo Smelser e Parsons questi due tipi di teorie sociali erano stati tenuti separati troppo a lungo per una serie di motivi legati alla tradizione e a fattori di ordine istituzionale. Le cose tuttavia non erano sempre andate così. All'inizio del secolo sia Weber che Marshall avevano ravvicinato fra loro la teoria economica e la teoria sociologica (Marshall soprattutto con la sua tesi secondo cui l'organizzazione è un fattore di produzione indipendente). Dopo questo primo passo, tuttavia, il processo di integrazione si era bloccato per svariati motivi, tra cui il fatto che gli economisti avevano incominciato a occuparsi di questioni tecniche e di applicazioni politiche. I sociologi, dal canto loro, si erano tenuti alla larga dall'economia sia perché risentivano del predominio degli economisti sia perché avvertivano la necessità di elaborare il proprio tipo di analisi. In ogni caso, concludono Parsons e Smelser, il risultato fu che, a metà degli anni cinquanta, quasi tutti gli economisti ignoravano il lavoro dei sociologi e viceversa.
Questa separazione tra teoria economica e teoria sociologica è, per gli autori di Economy and society, qualcosa di artificioso, perché tra le due discipline vi è un"ntrinseca intimità'.
Secondo Parsons e Smelser un modo per uscire dalla situazione di stallo è il seguente. Da un punto di vista teorico la teoria economica può essere concepita come un caso particolare

della teoria generale dei sistemi sociali. In effetti, se si considera la teoria economica da questa prospettiva - sostengono i due autori - diventa chiaro che esistono sorprendenti parallelismi logici, per esempio, tra i concetti di domanda e offerta e quelli di prestazione e sanzione. L'unica differenza tra la teoria economica e altri tipi di teoria sociale è che la teoria economica opera con un proprio specifico insieme di parametri. La teoria economica, in altre parole, ha esattamente la stessa struttura di base della scienza politica o della sociologia; solo i suoi argomenti sono diversi.Il fatto che la stessa economia possa essere vantaggiosamente trattata facendo ricorso alla teoria generale dei sistemi sociali viene spiegato nel modo seguente. Per Parsons e Smelser qualsiasi forma di interazione costituisce un sistema sociale. Per poter raggiungere un equilibrio e per continuare a esistere, però, un sistema di questo genere deve risolvere quattro problemi: 1) deve adattarsi all'ambiente ('adattamento'); 2) deve conseguire un obiettivo specifico ('raggiungimento degli scopi'); 3) deve integrare l'una con l'altra le sue diverse componenti ('integrazione'); 4) deve infine difendere e stabilizzare i propri valori attraverso la conservazione della struttura ('latenza'). Secondo Parsons e Smelser l'economia è quella componente del sistema sociale che garantisce l'adattamento all'ambiente; inoltre costituisce un distinto 'sottosistema' del sistema sociale. Parsons e Smelser definiscono così l'economia: "l'economia è quel sottosistema della società che si contraddistingue fondamentalmente in quanto preposto alla funzione adattiva della società nel suo complesso" (v. Parsons e Smelser, 1956).
Secondo Parsons e Smelser, si può considerare anche l'economia, e non solo la società, come un sistema sociale con determinati problemi da risolvere. Anche il sottosistema economico, pertanto, ha una propria funzione adattiva, una propria funzione integrativa, ecc. La teoria dei sistemi sociali permette a Parsons e a Smelser, cosa ancor più importante, di avanzare l'idea dell'interazione tra i diversi sottosistemi. Ciò che contraddistingue la teoria economica, come si è detto, è che essa dispone di un proprio insieme di parametri. Muovendo dall'idea di fondo che i vari sottosistemi interagiscono sempre tra loro attraverso inputs e outputs, Parsons e Smelser analizzano tali parametri in termini di "scambi di confine" (boundary exchanges). Il sottosistema economico, per esempio, interagisce con il sottosistema di conservazione della struttura (soprattutto con la famiglia) scambiando lavoro con beni di consumo e servizi. Procedendo in tal modo, attraverso la descrizione sistematica di un gran numero di questi "scambi di confine", Parsons e Smelser ritengono di aver sviluppato la teoria economica in senso sociologico, e quindi di aver contribuito a integrarla con la teoria

sociologica.
L'importanza di Economy and society sta nell'aver riproposto in termini nuovi e creativi la questione del rapporto tra teoria economica e teoria sociologica. Quanto all'integrazione tra le due discipline, tuttavia, l'opera di Parsons e Smelser presenta una certa ambiguità. Da un lato l'economia è chiaramente considerata parte integrante della società: un sottosistema di un più ampio sistema sociale, che è la società; dall'altro la teoria economica detiene il monopolio dell'analisi di tutti i fenomeni che rientrano in questo sottosistema. Alla sociologia viene lasciato un ruolo da svolgere soltanto quando si tratta di discutere i parametri della teoria economica o i confini tra l'economia e gli altri sottosistemi. Tutto ciò che gli autori chiamavano 'istituzioni economiche' deve, per esempio, essere studiato dal punto di vista di questi parametri e confini. Questo approccio apre la strada ad alcune idee interessanti: diversi tipi di mercato, per esempio, possono essere distinti l'uno dall'altro a seconda dei diversi sottosistemi cui sono collegati. Il mercato del lavoro, per esempio, è collegato al sistema di conservazione della struttura, il mercato dei capitali al sistema politico, ecc. Resta tuttavia aperto il problema se Parsons e Smelser non abbiano, da un lato, indebitamente ristretto l'ambito della sociologia economica e, dall'altro, concesso troppo alla teoria economica.L'altro importante tentativo di riesumare il dibattito sui rapporti fra economia e società negli anni cinquanta è quello di Karl Polanyi. Studioso interdisciplinare, nato nell'Impero austroungarico, Polanyi lavorò in vari paesi. La sua prima opera di rilievo, The great transformation (1944), fu scritta in parte in Inghilterra e in parte negli Stati Uniti. Nonostante la sua notevole originalità, The great transformation non ebbe lo stesso successo della seconda grande opera di Polanyi, un lavoro collettivo intitolato Trade and market in the early empires (1957). La tesi principale di Polanyi è che una società governata da liberi mercati ('mercati autoregolati') non è in grado di sopravvivere, perché la società umana, per esistere, ha bisogno di unità e di coesione, due caratteristiche che i mercati non regolati non sono in grado di garantire. Polanyi illustra così lo scenario ipotetico di una società governata esclusivamente dai mercati: "Privi della protezione di istituzioni culturali, gli esseri umani perirebbero per gli effetti dell'esposizione alle insidie della società; vittime di un acuto sconvolgimento sociale, morirebbero decimati dal vizio, dalla perversione, dal crimine e dalla fame. La natura verrebbe sconvolta, i centri urbani e rurali contaminati, i fiumi inquinati, la sicurezza militare messa a repentaglio, la capacità di produrre alimenti e materie prime distrutta" (v. Polanyi, 1944).

Per illustrare la sua tesi circa il potenziale distruttivo del mercato autoregolato, Polanyi fa l'esempio dell'Inghilterra del XIX secolo. Nei primi anni dell'Ottocento l'introduzione di un libero mercato del lavoro ebbe conseguenze assolutamente catastrofiche per i lavoratori. Secondo Polanyi l'idea che ogni cosa si possa trasformare in merce costituisce un'utopia estremamente pericolosa; inoltre fa sì che la società, per proteggersi, adotti contromisure preventive che turbano il funzionamento del mercato. Quando tutto questo processo interagisce, a sua volta, con la lotta di classe, accade che si scatenino delle forze tremende, come il fascismo.

Anche nella seconda grande opera di Polanyi, Trade and market, l'accento è posto sul mercato; ma in questo lavoro Polanyi si prefigge un obiettivo diverso da quello di The great transformation: cioè elaborare un nuovo tipo di analisi economica, non centrata sul mercato. La scienza economica, leggiamo in Trade and market, esemplifica 'la fallacia economistica', cioè il fatto che l'economia riguardi soltanto i mercati. Secondo Polanyi il fatto di concentrarsi esclusivamente sui mercati è accettabile nel caso della teoria economica moderna, dato che questa disciplina si occupa fondamentalmente della società contemporanea in cui il mercato svolge un ruolo centrale. Ma identificare economia e mercato è del tutto sbagliato quando si tratta di storia economica, sociologia economica e antropologia economica. Queste discipline, invece, dovrebbero muovere dall'assunto che l'azione economica è "incorporata" nella vita sociale. Secondo un passaggio chiave del saggio programmatico The economy as instituted process di Polanyi, "l'economia umana [...] è incorporata e inglobata nelle istituzioni, sia economiche che non economiche" (v. Polanyi e altri, 1957).

Trade and market rappresenta innanzitutto un tentativo di sviluppare un'impostazione completamente nuova dell'economia. La nuova scienza economica dovrebbe basarsi, per Polanyi, su un concetto 'sostanziale' (contrapposto a 'formale') di economia. Secondo l'approccio 'sostanzialistico', la questione cruciale per l'economia è il concreto rapporto dell'uomo con la natura. L'economia 'formale', invece, considera l'economia solo in termini di scelta e di relazioni mezzi-fini. Secondo Polanyi "le due concezioni dell'economia - quella sostanziale e quella formale - non hanno nessun punto in comune. La seconda si basa sulla logica, la prima sui fatti" (ibid.).In contrasto con i sostenitori dell'economia 'formale', che interpretano tutti i fenomeni economici in termini di mercato, Polanyi sostiene che nel corso della storia si sono avute diverse forme di interazione in campo economico, che Polanyi

chiama "reciprocità", "redistribuzione" e "scambio". La reciprocità presuppone una struttura sociale simmetrica e si riscontra, per esempio, nelle società tribali. La redistribuzione presuppone un potere centrale, come in Unione Sovietica, per esempio. Lo scambio, infine, è la principale forma di interazione nelle moderne società di mercato.Secondo Polanyi un'altra manifestazione della 'fallacia economistica' è la tendenza a far rientrare automaticamente nella categoria del mercato molti fenomeni economici indipendenti. Secondo il pensiero economico moderno, per esempio, la moneta e il commercio appaiono sempre in concomitanza con il mercato; ma ciò non è vero. Inoltre si ignora il fatto che il 'mercato autoregolato' è un'invenzione molto recente e che nel corso della storia sono esistite altre forme di mercato. Per trattare questa problematica Polanyi introduce alcuni concetti nuovi, come quelli di "gruppo di offerta" (supply crowd) e "gruppo di domanda" (demand crowd), e sostiene che questi due tipi di gruppo, che coesistono sempre nel mercato moderno, raramente sono coesistiti in passato. Questa terminologia, però, non ha attecchito.
Quella che ha attecchito è stata invece la distinzione tra economia 'formale' e 'sostanziale'. Nel campo dell'antropologia economica la pubblicazione di Trade and market ha dato luogo a un acceso dibattito tra 'formalisti' e 'sostanzialisti'. Il dibattito si è trascinato per parecchi anni senza che nessuna delle due parti sia riuscita a prevalere nettamente sull'altra. Riassumendo si può dire che Polanyi ha fatto un deciso tentativo per rianimare la discussione sui rapporti tra economia e società, ottenendo un certo successo, almeno nell'ambito dell'antropologia. I sociologi e gli economisti, d'altra parte, hanno ignorato la sua opera e hanno continuato a pensare che economia e società abbiano, fondamentalmente, poche cose in comune.

6. Le tendenze attuali

Oggi - per la prima volta da oltre mezzo secolo a questa parte - tra economisti e sociologi si è aperta una discussione nuova e interessante. Pur non coinvolgendo, finora, tutti i principali esponenti di entrambe le discipline, questa discussione si sviluppa costantemente e potrebbe forse portare a una sorta di rinascita del dibattito sui rapporti tra economia e società. Solo il futuro, naturalmente, potrà dire se assisteremo realmente a questa rinascita. Nel frattempo, però, osserviamo che tale discussione, cominciata tra la fine degli anni settanta e i primi anni ottanta, annovera tra i suoi principali partecipanti economisti e sociologi interessati al ruolo delle strutture e delle istituzioni sociali nell'economia. Il tipo di analisi messo a punto da

questi economisti è noto come 'nuova economia istituzionale', mentre i sociologi hanno creato la cosiddetta 'nuova sociologia economica'. Questo nuovo dibattito su economia e società coinvolge anche altri partecipanti, soprattutto economisti secondo i quali l'approccio economico è applicabile non solo alle questioni economiche ma a tutti i problemi oltre all'economia in senso stretto. Questa impostazione, nota generalmente come 'imperialismo economico', si sovrappone in parte alla 'nuova economia istituzionale'. Va menzionato inoltre il tentativo di alcuni sociologi di lanciare un nuovo tipo di sociologia, detto 'sociologia delle scelte razionali'.

Non è facile determinare per quale motivo il dibattito su economia e società sia tornato alla ribalta dopo tanti anni di silenzio. Le ragioni probabilmente sono varie: alcune relative alle vicende della società nel suo complesso, altre inerenti all'economia e alla sociologia. L'embargo sui prodotti petroliferi del 1973, per esempio, scosse l'economia mondiale nonché la fiducia di molti economisti, e lasciò la sensazione che la scienza economica dominante non fosse in grado di spiegare tutto ciò che accadeva nel mondo economico. Gli anni settanta, inoltre, videro la fine del pensiero keynesiano come ideologia economica guida e l'ascesa di ideologie alternative, come il monetarismo. Con le elezioni di Reagan e della Thatcher, nel biennio 1979-1980, l'ideologia del mercato divenne estremamente popolare: ciò che restava dell'ideologia economica riformista fu sostituito dall'idea che il mercato potesse risolvere tutti i problemi. Il diffondersi dell'ideologia del mercato influenzò indubbiamente molti studiosi e li spinse a interessarsi di questioni economiche, ma è chiaro che a suscitare il nuovo dibattito su economia e società hanno contribuito anche iniziative prese nell'ambito delle scienze sociali, come si vedrà nel prossimo paragrafo, dedicato ad alcuni recenti sviluppi della teoria economica.

a) L'imperialismo economico e la nuova economia istituzionale

L'approccio noto come 'imperialismo economico' (o 'approccio economico') ha svolto un ruolo molto importante nel dibattito in corso, in quanto è stato il primo a mettere decisamente in discussione e a infrangere il vecchio dogma della divisione del lavoro tra l'economia e le altre scienze sociali, secondo cui l'economia dovrebbe occuparsi soltanto di argomenti economici (senza considerarne la dimensione sociale), mentre le altre scienze sociali dovrebbero occuparsi del resto della società (senza considerarne la dimensione economica). Va anche detto, comunque, che, dal punto di vista del dibattito su economia e società,

l'imperialismo economico presenta una certa ambiguità. Ciò dipende dal fatto che questo tipo di analisi non è interessato al dialogo con le altre scienze sociali, ma si prefigge di sostituirle con l'analisi microeconomica. Tuttavia, anche se il risultato finale dell'imperialismo economico - da un punto di vista teorico - dovrebbe essere l'annullamento di tutte le altre scienze sociali, è anche vero che buona parte delle ricerche ispirate a questa concezione si sono rivelate preziose per il dibattito su economia e società. È pertanto opportuno analizzare più in dettaglio l'imperialismo economico.Il termine 'imperialismo economico' è stato usato per la prima volta all'inizio degli anni trenta da un giovane economista statunitense, Ralph William Souter, nel libro Prolegomena to relativity economics. Secondo Souter "la salvezza della scienza economica nel XX secolo risiede in un 'imperialismo economico' illuminato e democratico, che invada i territori dei suoi vicini" (v. Souter, 1933, pp. 94-95). La tesi di Souter sollevò subito qualche polemica. Talcott Parsons, per esempio, avvertì che l'imperialismo economico avrebbe costretto le altre scienze sociali in "una camicia di forza di 'categorie economiche"' (v. Parsons, 1934). Nel complesso, comunque, le idee di Souter non ebbero un gran seguito tra gli economisti e l'intera questione fu accantonata. Se ne tornò a parlare dopo più di venti anni, quando alcuni giovani economisti incominciarono ad avanzare tesi simili a quelle di Souter. Negli anni cinquanta furono pubblicati tre importanti lavori che introdussero l"approccio economico' nella sociologia, nella scienza politica e nella storia economica: The economics of discrimination (1957) di Gary Becker; An economic theory of democracy (1957) di Anthony Downs; e The economics of slavery in antebellum South (1958) di A. H. Conrad e J. R. Myers. Negli anni sessanta apparve una gran quantità di opere nuove e importanti su altri argomenti di sociologia, scienza politica e storia economica. L'approccio economico venne esteso anche ad altri settori, quali la demografia, il diritto (criminologia inclusa) e la teoria dell'organizzazione.
Negli anni settanta questa tendenza cominciò ad acquisire un'identità più definita e l'espressione 'imperialismo economico' divenne popolare. Oltre al primo manuale sull'argomento venne pubblicato anche un articolo che può essere considerato il manifesto intellettuale del nuovo approccio. Scritto da Gary Becker, questo articolo era intitolato The economic approach to human behavior (1976). Becker così descriveva il modo di condurre un'analisi: "il nocciolo dell'approccio economico è costituito da tre ipotesi combinate - quella del comportamento massimizzante, quella dell'equilibrio di mercato e quella delle preferenze stabili - usate sistematicamente e senza eccezioni". In sé e per sé questa definizione non è

particolarmente degna di nota, ma le parole "usate sistematicamente e senza eccezioni" danno un'idea abbastanza chiara della nuova ortodossia. Secondo Becker, l'approccio economico è applicabile anche a "qualsiasi comportamento umano". Durante gli anni ottanta apparvero parecchi altri lavori ispirati a questa impostazione e l'imperialismo economico venne accettato nell'ambito della teoria economica dominante, come dimostrano l'assegnazione nel 1984 del premio Nobel per l'economia a James Buchanan e l'elezione, tre anni più tardi, di Gary Becker alla presidenza dell'Associazione americana degli economisti.Molti degli economisti partecipanti al dibattito in corso su economia e società appartengono alla cosiddetta 'nuova economia istituzionale'. Questa corrente è piuttosto diffusa sia negli Stati Uniti che in Europa, e particolarmente in Germania, dove viene pubblicata la rivista "Journal of institutional and theoretical economics". La nuova economia istituzionale in parte si sovrappone all'imperialismo economico, ma tra le due correnti vi sono anche alcune differenze, e bisogna riconoscere che la nuova economia istituzionale ha una propria specifica identità. Il neoistituzionalismo, per esempio, è molto più eclettico dell'imperialismo economico. Mentre quest'ultimo si basa esclusivamente sulla microeconomia - usata "sistematicamente e senza eccezioni" (Becker) - il neoistituzionalismo ricorre anche alla teoria dell'organizzazione, all'economia schumpeteriana e persino al vecchio istituzionalismo di Commons e Schmoller (ma non a quello di Veblen, che continua ad essere escluso dal pensiero economico dominante). Questa capacità di attingere a tradizioni diverse conferisce alla nuova economia istituzionale una flessibilità che all'imperialismo economico manca completamente. Inoltre il neoistituzionalismo si pone un obiettivo diverso da quello dell'imperialismo economico. Mentre quest'ultimo si prefigge soprattutto di estendere l'approccio microeconomico a tutti gli altri tipi di analisi sociale e anche di sostituirsi alle altre scienze sociali, alla nuova economia istituzionale interessa soltanto analizzare determinati fenomeni che, pur facendo parte della vita economica, non sono mai stati studiati dagli economisti: l'organizzazione interna delle imprese, l'interazione tra imprese e diritto, ecc.

Per quanto concerne invece le teorie usate dagli esponenti della nuova economia istituzionale, è evidente che esse talvolta coincidono con quelle dell'imperialismo economico. Questo è il caso, per esempio, della teoria della rappresentanza (agency) e di quella dei diritti di proprietà. Il concetto di rappresentanza proviene dal pensiero giuridico. La teoria della rappresentanza si occupa essenzialmente di rapporti che coinvolgono un rappresentato (un

datore di lavoro, per esempio) e un rappresentante (un lavoratore). Il problema consiste nell'analizzare per quale motivo il rappresentante fa o non fa ciò che il rappresentato desidera e quali possono essere le conseguenze economiche di ciò. La teoria dei diritti di proprietà studia l'interazione tra economia e struttura giuridica, collegandole in modo più stretto di quanto facciano di solito le analisi economiche o giuridiche convenzionali. Una terza teoria adottata dalla nuova economia istituzionale (ma definita ufficialmente come una forma di neoistituzionalismo) è quella del costo di transazione. La teoria del costo di transazione è stata sviluppata soprattutto da Oliver Williamson e parte dal presupposto che lo studio dei diversi costi delle transazioni economiche può rivelarsi molto utile per risolvere una serie di difficili problemi della teoria economica. In Market and hierarchies (1975), per esempio, Williamson sostiene che la decisione di far ricorso a un'impresa o a un mercato dipende dai costi di transazione che si devono affrontare. Altre opere che, secondo Williamson, rientrano, benché marginalmente, nell'ambito della nuova economia istituzionale sono Internal labor markets and manpower analysis (1971) di Peter Doeringer e Michael Piore e An evolutionary theory of economic change (1982) di Richard Nelson e Sidney Winter (v. Williamson, 1975). Il primo di questi due lavori è di natura prevalentemente sociologica, mentre il secondo è un misto di idee schumpeteriane e di modelli matematici.

Una critica che è stata rivolta a più riprese alla nuova economia istituzionale riguarda il suo modo di usare il concetto di 'efficienza'. Secondo la nuova economia istituzionale una istituzione economica si afferma e sopravvive semplicemente perché rappresenta una soluzione efficiente di qualche problema che il mercato non è in grado di risolvere. Per i critici questa tesi non è affatto convincente, perché conduce nella stessa trappola in cui cade la sociologia funzionalista: se tutto ciò che esiste è 'efficiente', allora il concetto di efficienza dev'essere esteso a tal punto che l'intera analisi diventa tautologica. Alcuni esponenti della nuova economia istituzionale hanno reagito positivamente a queste critiche, e non mancano buoni motivi per pensare che essi continueranno a sviluppare le proprie concezioni in direzioni nuove e interessanti, pur senza troncare del tutto i legami con la scienza economica dominante. Inoltre alcuni autori - tra cui lo stesso Williamson - hanno dato prova di una certa disponibilità al dialogo con la sociologia. In definitiva la nuova economia istituzionale è caratterizzata da una certa flessibilità, che invece manca all'imperialismo economico e che lascia ben sperare per quanto riguarda i futuri sviluppi di questo nuovo tipo di istituzionalismo.

b) La nuova sociologia economica

Il rinnovato interesse che oggi si avverte nei confronti del dibattito sui rapporti tra economia e società dipende in larga misura dai più recenti sviluppi della sociologia economica. La sociologia economica comprende un ampio spettro di concezioni, che vanno dalla 'sociologia delle scelte razionali' alla 'socioeconomia'. La tesi di fondo della sociologia delle scelte razionali, come dice, per esempio, James Coleman nel suo Foundations of social theory (1990), è che la sociologia dovrebbe abbandonare buona parte dei suoi interessi tradizionali e fare invece assegnamento sull'individualismo metodologico e su altri concetti della microeconomia. L'idea chiave della socioeconomia, come spiega il suo fondatore Amitai Etzioni in The moral dimension: towards a new economics (1988), è che l'economia neoclassica è troppo angusta e andrebbe perciò sostituita con una nuova scienza economica interdisciplinare molto più ampia. La nuova sociologia economica si colloca tra queste due posizioni e sostanzialmente si prefigge lo scopo di analizzare l'economia da un punto di vista strettamente sociologico. In altri termini la nuova sociologia economica rappresenta un tentativo di studiare i fenomeni economici tradizionali attraverso i concetti tipici della sociologia, quali quelli di 'struttura sociale', 'ruolo', 'reticolo' (network), ecc.

La nuova sociologia economica è sorta negli Stati Uniti, dove essa ha anche le sue basi più solide; comunque un marcato interesse nei suoi confronti è riscontrabile anche in vari paesi europei, come l'Italia e la Germania. Inoltre i sociologi dei paesi dell'Est e della ex Unione Sovietica sembrano particolarmente ansiosi di assimilare la nuova sociologia economica, spinti sia dal desiderio di trovare un sostituto dell'ideologia marxista sia da pura curiosità intellettuale. In breve in diversi paesi c'è un grande interesse per la nuova sociologia economica, interesse che sembra tutt'ora in crescita. La nuova sociologia economica è nata nei primi anni ottanta, anche se le sue radici risalgono al periodo a cavallo tra gli anni sessanta e settanta. La contestazione studentesca, per esempio, contribuì alla diffusione nei circoli accademici del marxismo e dell'idea che economia e società siano strettamente connesse. Nel corso degli anni settanta tornarono in auge la sociologia weberiana e la teoria sociale femminista, che lasciarono la loro impronta sulla nuova sociologia economica. Nello stesso periodo si riaccese anche l'interesse per la teoria dei reticoli, che avrebbe avuto conseguenze importanti per la nuova sociologia economica. Parecchi importanti esponenti della nuova sociologia economica americana, per esempio, studiarono la teoria dei reticoli

che poi applicarono ad argomenti economici.
L'ispiratore della nuova sociologia economica è stato Harrison C. White, un fisico dedicatosi alla sociologia, che lavorava alla Harvard University. Negli anni settanta, White cominciò a tenere lezioni di teoria dei reticoli; qualche anno dopo prese a studiare i mercati e cercò di sviluppare una teoria sociologica dei mercati. Secondo White l'economia neoclassica non aveva una vera e propria teoria dei mercati, ma soltanto una teoria dello scambio puro e semplice. Nei primi anni ottanta, in un articolo pubblicato sull'"American journal of sociology" e intitolato Where do markets come from? (1981), White propose una teoria sociologica dei mercati. L'idea di fondo era che il mercato nasce da una forma di controllo reciproco tra compratori e venditori, particolarmente per quanto riguarda il volume, il prezzo e la qualità delle merci. White presentò anche una versione matematica della sua teoria, precisando di riferirsi soltanto a mercati con pochi venditori e con pochi compratori. Anche se White deve essere considerato in un certo senso il padre della nuova sociologia economica, il merito di aver divulgato la tesi che i sociologi possono e debbono occuparsi dei problemi economici spetta a Mark Granovetter. Tale tesi costituisce l'argomento di un articolo, molto citato, intitolato Economic action and social structure: the problem of embeddedness, comparso nel 1985 su "The American journal of sociology". Come chiarisce il titolo, Granovetter si ispira all'analisi di Polanyi, e in particolare alla tesi che l'azione economica è 'incorporata' nei rapporti sociali. Il punto di vista di Granovetter, comunque, differisce da quello di Polanyi per due aspetti. In primo luogo, mentre Polanyi pensava che la teoria economica dominante potesse essere usata benissimo per analizzare la società contemporanea (ma non la società preindustriale), per Granovetter anche l'economia moderna è 'incorporata' nei rapporti sociali. Nella stessa ottica, Granovetter critica anche Oliver Williamson e altri esponenti della nuova economia istituzionale rimproverandoli di far troppo affidamento sulla teoria economica dominante. La critica avanzata nei confronti di Williamson e di altri autori perché utilizzano il concetto di efficienza come chiave interpretativa principale in ogni tipo di analisi (v. sopra) proviene proprio da questo articolo di Granovetter. Il secondo aspetto che differenzia la concezione di Granovetter da quella di Polanyi concerne il ruolo dei reticoli in economia. Mentre Polanyi era stato piuttosto vago a proposito del modo in cui l'economia è effettivamente incorporata nei rapporti sociali, Granovetter sostiene invece che tale modalità può essere mostrata esattamente mediante l'analisi dei reticoli.

A partire dai primi anni ottanta i sociologi americani hanno fatto ricorso alla teoria dei reticoli per analizzare tutta una serie di fenomeni economici, tra cui i mercati, i fondi di investimento e l'interazione tra le grandi imprese. Anche la sociologia che si ispira al marxismo può del resto avvalersi di questo tipo di approccio, come dimostra, per esempio, il saggio di Beth Mintz e di Michael Schwartz The power structure of American business (1985). L'idea di fondo di questo studio è che per localizzare i centri del potere economico basta esaminare la composizione dei consigli di amministrazione delle grandi imprese. Se un membro di una di queste imprese è rappresentato nel consiglio di amministrazione di un'altra, allora tra le due società esiste una connessione che può essere considerata alla stregua di un collegamento in un reticolo. I limiti di questo tipo di analisi sono stati messi in evidenza, per esempio, da Donald Palmer nel suo studio sui 'legami spezzati', dove egli analizza cosa accade quando un membro di un consiglio di amministrazione di una grande impresa si dimette.

Oltre che dello studio dei reticoli, la nuova sociologia economica si è occupata anche di varie forme di organizzazione economica e del ruolo della cultura nella vita economica. L'analisi proposta da Williamson in Market and hierarchies, per esempio, è stata prontamente messa in dubbio sulla base del fatto che i mercati e le strutture gerarchiche sono spesso simili. Analogamente sono state messe in discussione anche le tesi sulla nascita delle imprese 'multidivisionali' esposte in The visible hand (1977) da Alfred Chandler. Si è osservato, per esempio, che nello spiegare l'adozione di tale tipo di imprese Chandler sottolinea troppo il ruolo della tecnologia e degli incentivi economici, mentre ignora, per esempio, il ruolo dello Stato nel modellare l'organizzazione delle imprese. Si è fatto inoltre notare che Chandler ritiene che una certa struttura organizzativa sia adottata semplicemente a causa della sua 'efficienza'. Secondo alcune recenti ricerche sociologiche, tuttavia, le grandi imprese si imitano fra loro sulla base della convinzione che sia necessario avere un certo tipo di organizzazione.

Sul ruolo della cultura nella vita economica i pareri degli esponenti della nuova sociologia economica sono discordi. Da un lato c'è chi sostiene che se si fa ricorso a un concetto come quello di 'cultura' per spiegare, poniamo, il successo dell'economia giapponese, in realtà non si spiega assolutamente niente. Dall'altro vi sono quelli secondo cui la sociologia economica odierna tende a essere troppo limitata e a trascurare molti importanti e sottili influssi che la cultura esercita sull'economia. Una sostenitrice di questa impostazione è Viviana Zelizer, che

ha analizzato svariati fenomeni quali l'assicurazione sulla vita, il valore economico dei bambini e il ruolo del denaro nella società moderna.
In conclusione la sociologia economica sta attraversando una fase di grande vitalità. Ciò non significa, ovviamente, che essa non abbia le sue difficoltà. Tra queste va senz'altro annoverato il fatto che essa è scarsamente istituzionalizzata: nelle università non vi sono molte cattedre di sociologia economica; in pratica non esistono riviste di settore; non è facile ottenere finanziamenti per la ricerca in questa disciplina. Queste carenze dipendono in fondo dal persistente scetticismo di buona parte degli esponenti delle scienze sociali nei confronti della sociologia economica. In un certo senso si continua a pensare che soltanto gli economisti debbano occuparsi di economia. In ultima analisi, naturalmente, è la scienza economica che dovrebbe essere al centro del dibattito sui rapporti tra economia e società. Nel corso di questo secolo, però, la teoria economica dominante si è isolata sempre più dalle altre scienze sociali, e ciò ha impoverito e per un lungo periodo di tempo ha quasi completamente interrotto ogni discussione su economia e società. C'è da augurarsi che questa situazione si modifichi e che possa nascere una nuova scienza economica più aperta e più duttile. Non possiamo ancora sapere che aspetto avrà questa economia del futuro; già oggi, tuttavia, disponiamo di un modello estremamente valido cui la nuova economia potrebbe ispirarsi, vale a dire la Sozialökonomik di Weber e Schumpeter. Secondo Schumpeter, come abbiamo già detto, la scienza economica dovrebbe comprendere diversi campi disciplinari, tra cui la teoria economica, la sociologia economica e la storia economica. È abbastanza evidente che una tale Sozialökonomik sarebbe molto più confacente al dibattito su economia e società di quanto non lo sia la teoria economica oggi dominante. (V. anche Economia; Sociologia).

bibliografia

Becker, G., The economic approach to human behavior, Chicago-London 1976.
Bücher, C., Die Entstehung der Volkswirtschaft, Tübingen 1893.
Durkheim, É., De la division du travail social: étude sur l'organisation des sociétés superieures, Paris 1893 (tr. it.: La divisione del lavoro sociale, Milano 1962).
Durkheim, É., Le socialisme, Paris 1928 (tr. it.: Il socialismo, Milano 1973).
Durkheim, É., Leçons de sociologie: physique des moeurs et du droit, Paris 1950 (tr. it.: Lezioni di sociologia: fisica dei costumi e del diritto, Milano 1973).

Godelier, M., Qu'est-ce que définir une 'formation économique et sociale'? L'exemple des Incas, in "Pensée", 1971, n. 159, pp. 99-115.
Granovetter, M., Economic action and social structure: the problem of embeddedness, in "American journal of sociology", 1985, XCI, pp. 481-510.
Hildebrand, B., Die Nationalökonomie der Gegenwart und Zukunft, Frankfurt a. M. 1848.
Knies, K., Die politische Ökonomie vom Standpunkt der geschichtlichen Methode, Braunschweig 1853.
Lenin, N., Razvitie kapitalizma v Rossii, Sankt-Peterburg 1899 (tr. it.: Lo sviluppo del capitalismo in Russia, in Opere complete, vol. III, Roma 1956).
Martinelli, A., Economia e società: Marx, Weber, Schumpeter, Polanyi, Parsons e Smelser, Milano 1986.
Marx, K., Zur Kritik der politischen Ökonomie, Berlin 1859 (tr. it.: Per la critica dell'economia politica, Roma 1957).
Marx, K., Das Kapital: Kritik der politischen Ökonomie, 3 voll., Hamburg 1867-1894 (tr. it.: Il capitale: critica dell'economia politica, 3 voll., Roma 1970).
Marx, K., Engels, F., Die deutsche Ideologie (1845-1846), Berlin 1932 (tr. it.: L'ideologia tedesca, in Opere complete, vol. V, Roma 1972).
Mauss, M., Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques (1923-1924), in Sociologie et anthropologie, Paris 1950, pp. 145-279 (tr. it.: Saggio sul dono: forma e motivo dello scambio nelle società arcaiche, in Teoria generale della magia e altri saggi, Torino 1965).
Menger, C., Untersuchungen über die Methode der Sozialwissenschaften und der politischen Ökonomie insbesondere, Leipzig 1883.
Mill, J. S., Principles of political economy, 2 voll., London 1848 (tr. it.: Principî di economia politica, Torino 1953).
Nelson, R. R., Winter, S. G., An evolutionary theory of economic change, Cambridge, Mass.-London 1982.
Palmer, D., Broken ties, in "Administrative science quarterly", 1983, XXVIII, pp. 40-55.
Parsons, T., Some reflections on the nature and significance of economics, in "Quarterly journal of economics", 1934, XLVIII, pp. 511-545.
Parsons, T., Smelser, N., Economy and society: a study in the integration of economic and social theory, London 1956 (tr. it.: Economia e società: uno studio sull'integrazione della

teoria economica e sociale, Milano 1970).
Polanyi, K., The great transformation: the political and economic origins of our time, New York 1944 (tr. it.: La grande trasformazione: le origini economiche e politiche della nostra epoca, Torino 1974).
Polanyi, K., Arensberg, C. M., Pearson, H. W. (a cura di), Trade and market in the early empires: economies in history and theory, Glencoe, Ill., 1957 (tr. it.: Traffici e mercati negli antichi imperi: le economie nella storia e nella teoria, Torino 1978).
Ricardo, D., On the principles of political economy and taxation (1817), in The works and correspondence of David Ricardo (ed. critica a cura di P. Sraffa e M. Dobb), vol. I, Cambridge 1951 (tr. it.: Sui principî dell'economia politica e della tassazione, Milano 1976).
Roscher, W., Roscher's programme of 1843, in "Quarterly journal of economics", 1895, IX, pp. 99-105.
Schmoller, G. von, Zur Methodologie der Staats- und Sozialwissenschaften, in "Schmollers Jahrbuch", 1883, VII, pp. 975-984.
Schmoller, G. von, Grundriss der allgemeinen Volkswirtschaftslehre, 2 voll., Leipzig 1900-1904.
Schumpeter, J. A., Capitalism, socialism and democracy, New York 1942, 1950³ (tr. it.: Capitalismo, socialismo e democrazia, Milano 1955).
Schumpeter, J. A., History of economic analysis, New York 1954 (tr. it.: Storia dell'analisi economica, 3 voll., Milano 1959-1960).
Schumpeter, J. A., The economics and sociology of capitalism (a cura di R. Swedberg), Princeton 1990.
Simiand, F., La monnaie, réalité sociale, in "Annales sociologiques", série D, 1934, pp. 1-58.
Simmel, G., Philosophie des Geldes, Leipzig 1900 (tr. it.: Filosofia del denaro, Torino 1984).
Smelser, N., The sociology of economic life, Englewood Cliffs, N. J., 1963 (tr. it.: Sociologia della vita economica, Bologna 1967).
Smith, A., An inquiry into the nature and causes of the wealth of nations (1776), in The works and correspondence of Adam Smith (ed. critica a cura di R.H. Campbell, A.S. Skinner e W.B. Todd), vol. II, Oxford 1976 (tr. it.: Indagine sulla natura e le cause della ricchezza delle nazioni, Milano 1973).
Sombart, W., Der moderne Kapitalismus, 2 voll., Leipzig 1902; 3 voll., München-Leipzig 1916-1927² (tr. it. parziale: Il capitalismo moderno, Torino 1967).

Sombart, W., Die drei Nationalökonomien, München-Leipzig 1930.
Souter, R. W., Prolegomena to relativity economics, New York 1933.
Swedberg, R., Economics and sociology: on reinterpreting their boundaries, Princeton 1990.
Swedberg, R., Schumpeter: a biography, Princeton 1991.
Weber, M., Roscher und Knies und die logischen Probleme der historischen Nationalökonomie (1903-1906), in Gesammelte Aufsätze zur Wissenschaftslehre, Tübingen 1922, pp. 1-145.
Weber, M., Die protestantische Ethik und der Geist des Kapitalismus (1904-1905), in Gesammelte Aufsätze zur Religionssoziologie, vol. I, Tübingen 1922 (tr. it.: L'Etica protestante e lo spirito del capitalismo, in Sociologia della religione, vol. I, Milano 1982, pp. 17-194).
Weber, M., Wirtschaft und Gesellschaft, 2 voll., Tübingen 1922, 1925² (tr. it.: Economia e società, 2 voll., Milano 1961).
White, H. C., Where do markets come from?, in "American journal of sociology", 1981, LXXXVII, pp. 514-547.
Williamson, O., Markets and hierarchies: analysis and antitrust implications, New York 1975.
Zelizer, V., Morals and markets: the development of life insurance in the United States, New York 1979.

Antropologia economica

di George Dalton

sommario: 1. Introduzione. 2. L'oggetto dell'antropologia economica. a) L'antropologia economica come parte dell'antropologia sociale. b) La differenza tra bande, tribù, società contadine e regni. c) Differenze teoriche. 3. Tipologie dei sistemi economico-sociali precoloniali, coloniali e postcoloniali. a) Economie aborigene nelle società tribali senza Stato: tipo I. b) Economie aborigene nei regni tribali: tipo II. c) Il mutamento determinato dal colonialismo nelle società tribali: 1500-1960. d) Società contadine di antico insediamento all'interno di Stati-regno: tipo III. e) Società contadine create dal colonialismo spagnolo in America Latina: tipo IV. □ Bibliografia.

1. Introduzione

Prima di indicare i fattori che caratterizzano, differenziandolo, il vasto e variegato insieme di realtà economiche studiate dagli antropologi, è necessario dar conto di alcune peculiarità dell'antropologia economica, la disciplina che si occupa delle forme che l'economia assume nel contesto di società organizzate in bande, tribù e comunità contadine. Bisognerà poi spiegare quali siano i tratti comuni a tutte le economie, passate e presenti, e quali, invece, le differenze, per giungere poi a delineare le caratteristiche specifiche del capitalismo industriale. Tutte queste operazioni preliminari sono indispensabili perché è innanzitutto necessario individuare due serie di quelli che Ludwig Wittgenstein avrebbe chiamato "ostacoli posti sul cammino della comprensione". Una prima serie di ostacoli scaturisce dalla complessità e dalla diversità delle società concrete studiate dagli antropologi e dal mutamento nel tempo di tali società. La seconda dipende invece dalla struttura estremamente diversificata delle teorie adottate dagli studiosi di antropologia economica - il formalismo, il marxismo e il sostanzialismo di Polanyi - ognuna delle quali conduce ad analisi (e conclusioni) profondamente diverse delle caratteristiche specifiche dell'economia e della società strutturate in bande, tribù e comunità contadine.

2. L'oggetto dell'antropologia economica

a) L'antropologia economica come parte dell'antropologia sociale

L'antropologia economica condivide molte caratteristiche della disciplina da cui proviene, vale a dire l'antropologia sociale. Questa disciplina, che è usualmente classificata nel novero delle scienze sociali insieme all'economia, alla sociologia e alla scienza politica, se ne differenzia significativamente per molti aspetti: gli studiosi impegnati in essa non sono molti, mentre la gamma delle problematiche, dei tipi di società e dei periodi storici che essi studiano è molto ampia e le loro analisi riguardando società diverse sparse in tutto il mondo, al di fuori dell'area delle nazioni industriali avanzate dell'Europa occidentale e del Nordamerica. A differenza dell'economia, l'antropologia sociale non si è mai occupata, fino a tempi recentissimi, di formulare specifiche proposte politiche finalizzate al progresso, allo sviluppo e alla modernizzazione dei diversi sistemi economici analizzati. Malinowski non studiava la coltivazione dell'igname nelle isole Trobriand (1935) o gli scambi cerimoniali kula (1922) per trovare il modo di incrementare la produttività agricola e il commercio estero delle

Trobriand. Agli antropologi interessava semplicemente comprendere la realtà, non migliorarla. Da questo punto di vista l'atteggiamento degli antropologi sociali è molto più affine a quello degli archeologi o degli storici del Medioevo che non a quello degli economisti, il cui interesse per la conoscenza dei fenomeni è sempre finalizzato alla possibilità di individuare concrete strategie politiche atte a favorire un miglior rendimento economico. Gli antropologi, per altro, non si interessano esclusivamente a quelle che definirei 'conclusioni analitiche', ma anche di trasmettere ai lettori le concezioni dei popoli indigeni circa l'economia, la cultura e la società in cui vivono: per esempio i miti sull'origine del loro clan, la spiegazione che essi danno della crescita delle piante, del destino che li attende dopo la morte, o della siccità dell'anno passato. Ma ciò che soprattutto differenzia l'antropologia sociale dall'economia è il metodo di cui si servono gli antropologi per raccogliere informazioni e dati: la ricerca sul campo attuata vivendo per un certo periodo di tempo in piccole comunità di villaggio. In realtà gli antropologi non studiano le economie o le società nazionali dell'India, del Perù o della Nuova Guinea; piuttosto essi si occupano di un campione ristretto, costituito da piccoli villaggi rurali in India, in Perù o in Nuova Guinea. Così la principale fonte di informazioni concrete è costituita per l'antropologo sociale dalla ricerca sul campo, dal lavoro solitario che, per un anno o due, egli svolge vivendo a contatto diretto con le popolazioni che vuole studiare, inserito nelle comunità di villaggio, costituite da poche decine o centinaia di persone. Queste ricerche sono condotte di solito nelle zone rurali delle nazioni in via di sviluppo del Terzo Mondo, nelle piccole comunità residue di Indiani nordamericani o tra gli attuali discendenti degli Aborigeni australiani. Abbiamo tuttavia anche libri di antropologi che utilizzano la ricerca storica: lavori sull'organizzazione della società inca in Perù prima dell'arrivo degli spagnoli nel 1532 (v. Murra, 1980), sulla reazione degli Inca nei primi cinquanta anni di dominio coloniale dopo la conquista spagnola (v. Wachtel, 1971), sui cambiamenti provocati da 300 anni di incursioni effettuate dagli Spagnoli prima e dagli Americani poi tra le tribù indiane degli Stati Uniti sudoccidentali (v. Spicer, 1962). Quando studiano le economie e le società del periodo precoloniale, o i cambiamenti provocati dal dominio coloniale europeo, americano o giapponese, gli antropologi, da ricercatori sul campo, si trasformano così in veri e propri storici.

b) La differenza tra bande, tribù, società contadine e regni

La caratteristica più rilevante dei milioni di comunità di villaggio sparse in tutto il mondo è

costituita dalla loro straordinaria varietà e dal loro costante modificarsi nel tempo. Dagli inizi del Novecento (con l'opera di Boas, Thurnwald, Malinowski) ai giorni nostri, queste comunità sono state ripetutamente oggetto di ricerche sul campo. Fino alla seconda guerra mondiale, le ricerche sono state prevalentemente condotte da antropologi provenienti dalla Francia, dalla Gran Bretagna (cioè dai paesi che avevano il maggior numero di possedimenti coloniali) e dagli Stati Uniti, sul cui territorio si trovavano gli Indiani indigeni e dove, a partire dagli anni trenta, furono fatte oggetto di studio le comunità agricole caraibiche o dell'America Latina. Dopo il secondo conflitto mondiale, in concomitanza con il rapido sviluppo di tutte le scienze sociali e con il sorgere dell'interesse nei confronti dei nuovi paesi indipendenti del Terzo Mondo in Africa o in Asia, antropologi di diverse nazioni, del vecchio continente o extraeuropee, pubblicarono svariati lavori di ricerca sul campo. Una parte significativa di questi studi era dedicata alla situazione delle comunità contadine di villaggio in nazioni o regioni in via di sviluppo e industrializzazione, come il Perù, la Turchia, la Grecia, Haiti, la Sicilia, la Spagna e la Iugoslavia.

La storia e le condizioni geografiche hanno di per sé grande importanza nel determinare le enormi differenze riscontrabili tra le diverse comunità di villaggio analizzate dall'antropologia economica. Questa disciplina studia comunità sparse su tutti i continenti e su tutte le isole abitate del mondo, la loro struttura sociale originaria e le modifiche introdotte da quando gli Europei, cinquecento anni fa, scoprirono e colonizzarono il Nuovo Mondo prima, e poi l'Australia e la Nuova Zelanda, e inviarono spedizioni commerciali in Asia, in Africa e altrove. La maggior parte delle comunità studiate dagli antropologi (ma non tutte) si trova nelle zone del mondo colonizzate dagli imperi europei, nordamericani e giapponesi tra il 1500 e il 1900. Ciò consente di distinguere tra periodo precoloniale, coloniale e postcoloniale. Il periodo precoloniale è quello delle economie e società cosiddette 'aborigene', 'indigene' o 'tradizionali'; il periodo del dominio coloniale europeo, americano e giapponese è dal canto suo estremamente diversificato sia per quanto riguarda la durata, sia per la profondità dell'impatto che ha determinato (il Messico, l'America centrale e meridionale e i Caraibi sono stati sotto il dominio spagnolo per più di tre secoli, la Corea invece è stata una colonia giapponese solo per 35 anni); il periodo postcoloniale, infine, può essere fatto iniziare, in genere, con il conseguimento dell'indipendenza da parte dell'India nel 1947. Anche in questo caso ci sono differenze significative: basti pensare che le colonie spagnole del Nuovo Mondo erano per la maggior parte Stati indipendenti già nel 1825.

Anche a limitarsi esclusivamente alle strutture precoloniali, il lavoro degli antropologi prende in esame economie e società che variano grandemente quanto a popolazione, territorio, tecnologie, fonti alimentari e tipi di organizzazione politica. Si va dalle bande di cacciatori e raccoglitori del deserto del Kalahary - piccole unità autosufficienti composte da una ventina di persone che non conoscono né Stato, né mercato, né città, né scrittura - ai raggruppamenti tribali presenti all'interno dei regni africani, che comprendono centinaia di migliaia di persone, ai villaggi di contadini dell'India, dell'Indonesia e di altre zone asiatiche, i cui abitanti condividono lingua e religione con milioni di altri abitanti dello stesso Stato e hanno rapporti stabili con i mercati e con le città.Queste notevoli differenze ci costringono a inventare tutta una varietà di raggruppamenti, classificazioni, tipologie, distinzioni e concetti tecnici, al fine di poter formulare giudizi analitici sull'economia e sulla struttura sociale che risultino attendibili e pertinenti riguardo a particolari periodi storici, a specifici tipi di economia e a specifici ambiti e argomenti all'interno dell'universo antropologico nel suo complesso (v. tab. I).
Le classificazioni e i concetti tecnici che possono essere adottati da uno studioso di antropologia economica variano in modo estremamente significativo a seconda del sistema teorico di riferimento, o 'paradigma', in cui egli si colloca, vale a dire a seconda che sia marxista (v. Godelier, 1980), formalista (v. Schneider, 1974) o, come chi scrive, sostanzialista (v. Polanyi, 1957; v. Valensi e altri, 1981; v. Dalton, 1981). Questo fatto è di particolare rilevanza, dato che i diversi sistemi teorici ai quali possono essere ricondotti i concetti tecnici e le classificazioni portano a conclusioni diverse riguardo a quelle che sono le caratteristiche più autentiche, più importanti e più peculiari della società o dell'economia che si prende in esame.

c) Differenze teoriche

I formalisti si servono dei concetti base della teoria dei prezzi elaborati dai neoclassici fondatori della microeconomia tra il 1870 e il 1890, W. Jevons, C. Menger, J.B. Clark e A. Marshall. A questi studiosi in realtà interessava esclusivamente analizzare il funzionamento dei mercati degli inputs e degli outputs nel nascente capitalismo industriale dell'Inghilterra vittoriana, dell'Europa occidentale e degli Stati Uniti. Questi padri fondatori non pensavano affatto alle economie degli Inca, dei Nuer o delle isole Trobriand. Gli antropologi formalisti si servono invece della teoria del mercato, ideata per spiegare il caso dell'Inghilterra

vittoriana, per analizzare proprio le economie Inca, Nuer o delle isole Trobriand, perché pensano che tutte le economie, in ogni epoca e in ogni luogo, presentino una affinità fondamentale con il capitalismo industriale: la condizione di 'scarsità', cioè una richiesta di beni e servizi sempre eccedente rispetto a ciò che si riesce a produrre. Tale scarsità dà luogo necessariamente a dinamiche di 'massimizzazione' ed 'economizzazione', che sono proprio i processi analizzati dalla microeconomia. Per i formalisti le economie Inca, Nuer e delle Trobriand erano solo delle varianti preindustriali di quel capitalismo di mercato dell'Inghilterra vittoriana per analizzare il quale era stata concepita la teoria della domanda e dell'offerta. Essi ritengono, pertanto, che le analogie tra il capitalismo industriale e le economie contadine o tribali studiate dagli antropologi siano più importanti delle differenze (v. Hopkins, 1973; v. Dalton, 1976).

Anche la teoria economica marxista, che risale alla seconda metà del XIX secolo, venne messa a punto per analizzare il capitalismo industriale europeo e americano. Un secolo dopo, negli anni sessanta del Novecento, gli antropologi marxisti giunsero alla conclusione che il sistema teorico di Marx forniva concetti e sistemi di classificazione atti a definire anche le caratteristiche peculiari delle economie e delle società precapitalistiche, preidustriali e precoloniali (quali quelle Inca, Nuer e delle Trobriand) come pure le trasformazioni delle società contadine e lo sviluppo postcoloniale. E come i marxisti interpretano la struttura e il funzionamento del capitalismo industriale in Italia o negli Stati Uniti in modo molto diverso da quello degli economisti non marxisti, lo stesso vale anche per l'interpretazione marxista delle economie dell'Africa precoloniale o degli Inca. Questi studiosi riconducono tutti i dati fattuali delle economie primitive nell'ambito della teoria di Marx - concetti, classificazioni, principi - nella quale sono già individuate, a loro avviso, le strutture, le istituzioni e le relazioni che devono considerarsi fondamentali nello studio di qualsiasi economia: i modi e i rapporti di produzione, le forze produttive, le classi, il surplus e lo sfruttamento.

Il terzo quadro teorico è stato elaborato da Karl Polanyi (1886-1964) in tre libri pubblicati tra il 1944 e il 1966 e in una dozzina di articoli (v. Polanyi, 1968). Molti dei suoi concetti tecnici, dei suoi principi fondamentali e delle sue classificazioni provengono dal lavoro di autori precedenti, quali Bücher, Maine, Tönnies, Weber, Mauss, Thurnwald e Malinowski.Scrivendo un centinaio di anni dopo Marx e circa sessanta anni dopo Alfred Marshall, Polanyi aveva sugli economisti del secolo precedente il considerevole vantaggio di poter tener conto anche degli avvenimenti significativi del XX secolo (la Rivoluzione

bolscevica, il sistema economico stalinista, i due conflitti mondiali, la grande depressione degli anni trenta, la differenza tra le risposte politiche di Hitler e di Roosevelt al tracollo del capitalismo negli anni trenta). A questi eventi è dedicato La grande trasformazione (1944). Polanyi comprese anche l'importanza che il lavoro e la ricerca sul campo effettuati dagli antropologi potevano avere ai fini della comprensione delle economie preindustriali, e di ciò parla nel capitolo quarto - Società e sistemi economici - e nella relativa appendice di La grande trasformazione. Il punto di vista di Polanyi sull'antropologia economica, le sue analisi della storia economica preindustriale europea ed extraeuropea, le sue tesi sulle origini, sulla crescita e sullo sviluppo del capitalismo industriale contribuirono a dar vita al modello teorico relativo alle economie precapitalistiche esposto in Traffici e mercati negli antichi imperi (1957).

Il completo disaccordo di Polanyi dalle posizioni dei formalisti (v. Polanyi, 1957, cap. 13, e 1971) può essere spiegato brevemente. Polanyi concorda con Marx su due punti soltanto: 1) che la disciplina che noi oggi definiamo come antropologia economica dovrebbe, insieme alla storia economica primitiva, costituire un unico campo di ricerche sulla struttura e sul funzionamento delle economie e delle società precapitalistiche, preindustriali e precoloniali; 2) che lo studio di qualsiasi economia non può essere separato dall'analisi dell'organizzazione politica e sociale entro la quale tale economia si colloca.I tratti principali del pensiero di Polanyi possono venir così riassunti: l'organizzazione nazionale e internazionale del mercato, che è giunta a dominare e a integrare il capitalismo industriale del XIX secolo in Europa e in America, è qualcosa di unico a livello storico e antropologico; essa ha reso le società nazionali basate sul mercato molto diverse - sia dal punto di vista dell'organizzazione e del funzionamento dell'economia, sia per le conseguenze a livello economico, sociale e politico - da qualsiasi altra società precedente.

La microeconomia - la tradizionale analisi della domanda e dell'offerta ideata per analizzare le scelte di mercato delle famiglie e delle imprese nel capitalismo industriale - è fuorviante, inappropriata e distorce la realtà delle cose ove i suoi concetti, le sue premesse e le sue conclusioni analitiche siano applicati a sistemi economici che ignorano i mercati o le macchine: per esempio quando è applicata a economie come quelle dei Nuer, delle Trobriand o degli Inca, le cui caratteristiche salienti e peculiari non coincidono con le caratteristiche principali delle economie e delle società del capitalismo industriale come quelle degli Stati Uniti o dell'Italia.Per comprendere il funzionamento delle economie non di mercato sono

dunque necessari uno specifico vocabolario concettuale e un insieme di principi -un paradigma specifico insomma - molto diversi da quelli della microeconomia o del marxismo. Le caratteristiche strutturali di queste società sono infatti sistematicamente diverse dalle caratteristiche del capitalismo industriale, e cioè di quel sistema economico per analizzare il quale Marx e Marshall misero a punto le loro teorie. La differenza strutturale di fondo delle economie precoloniali - Nuer, delle Trobriand, Inca - non è data semplicemente dall'assenza di tecnologia industriale, ma anche dall'assenza di quella capillare rete di mercati interdipendenti necessaria per effettuare scambi e transazioni su base monetaria, tipici dell'economia capitalistica. Nelle società precapitalistiche gli individui non dipendono dal mercato per vendere il proprio lavoro e per guadagnare il reddito monetario con cui procacciarsi i beni di consumo: i contadini e i pastori non comprano lavoro, terra o strumenti per produrre, né vendono i loro raccolti o il loro bestiame. Le occasionali transazioni di mercato indigene, come gli scambi gimwali studiati da Malinowski nelle Trobriand, sono fenomeni limitati e sporadici: il mercato costituisce un piccolo settore marginale all'interno di un sistema economico più ampio governato da quei principi di transazione che oggi usiamo definire di 'reciprocità'. (Analogamente, anche nel sistema non di mercato staliniano esisteva un limitato settore di mercato per alcuni generi alimentari prodotti su appezzamenti privati nell'ambito di aziende agricole collettive).

L'assenza di un'economia di mercato globale, che risalta con tanta evidenza nei sistemi precoloniali, caratterizza anche i sistemi economici, greci e romani, dell'età classica. Scrive per esempio Moses I. Finley: "La società antica non possedeva un sistema economico costituito da un enorme agglomerato di mercati interdipendenti [...] i 'principî' [della microeconomia], siano quelli di Alfred Marshall o di Paul Samuelson, e i modelli [di mercato] che adottiamo [per analizzare il capitalismo industriale] tendono a darci un'idea sbagliata delle cose" (v. Finley, 1973, pp. 22-23).

Un giudizio analogo è espresso anche da W. Arthur Lewis, un economista che si occupa dei paesi in via di sviluppo. Nella sua introduzione a un ottimo lavoro di ricerca antropologica sul campo, dedicato allo sviluppo rurale postcoloniale in due villaggi dell'India meridionale, Lewis scrive: "Gli economisti si sono occupati prevalentemente dell'economia di mercato e hanno lasciato lo studio dell'economia non di mercato agli antropologi [...] l'economista che studia l'economia non di mercato dovrà perciò mettere da parte la maggior parte delle sue conoscenze e adottare le tecniche degli antropologi" (v. Epstein, 1962; v. anche Epstein,

1973).
In assenza di un'organizzazione economica basata largamente su meccanismi di mercato, che regoli le transazioni di terra, lavoro e strumenti, l'economia è 'incorporata' nella società. Sia la sussistenza ordinaria (il procacciamento e la produzione di risorse alimentari), sia l'economia politica e sociale (il 'prezzo-della-sposa', il risarcimento per uccisioni o ferite, gli scambi rituali come kula, potlatch e moka nelle società senza Stato, i beni del re o i monopoli reali, i traffici con l'estero amministrati politicamente e i rapporti di alleanza sanciti da tributi nei regni primitivi) sono socialmente e politicamente controllate (v. Dalton, 1981). Per quanto riguarda le società precapitalistiche, Polanyi pertanto prende le distanze dal determinismo economico dell'organizzazione politica e sociale che caratterizza l'analisi marxiana.In società quali quelle Nuer, delle Trobriand o Inca non è possibile contraddistinguere l''economia' come un sottosistema separato e nettamente distinto dal lignaggio, dal matrimonio, dalle alleanze o dalle attività economiche del re. Gli organismi e le istituzioni che provvedono a garantire con sistematicità i rifornimenti materiali sono strutture tanto politiche e sociali quanto economiche: il lignaggio, il matrimonio, l'alleanza militare, la figura e la funzione del re. Economia, società, sistema politico e cultura sono inestricabilmente connessi in tutte quelle società, del passato o del presente, nelle quali la rete capitalistica di mercato non costituisce la matrice fondamentale dell'organizzazione.
"L'antropologia sociale dà ormai per scontato che le istituzioni economiche, soprattutto nelle società più semplici e ristrette, siano strettamente connesse ad altre istituzioni e ad altri valori generalmente considerati non di tipo economico, e che anzi siano addirittura prive di significato al di fuori di questi" (v. Beattie, 1961).
Ciò tuttavia non vale solo per le società 'più semplici'. Ad esempio è ovvio che nell'economia dirigista staliniana il sistema economico delle industrie nazionalizzate, della pianificazione centrale e dell'agricoltura collettivizzata era 'incorporato' nel sistema politico di Stalin, che controllava direttamente il Partito e lo Stato comunista. E proprio come gli economisti hanno dovuto inventare una speciale terminologia tecnica e concettuale per analizzare il sistema economico staliniano ('economia dirigista', 'pianificazione', 'indici del successo di un'impresa'), del momento che il linguaggio concettuale dell'economia di mercato non era utilizzabile, anche Polanyi e i suoi seguaci hanno creato termini come 'reciprocità', 'redistribuzione', 'porti di scambio', 'scambi amministrati politicamente', 'moneta adatta a scopi speciali'.

Una delle principali conclusioni analitiche di Polanyi è che le differenze tra le economie Nuer, delle Trobriand e Inca da una parte e il capitalismo industriale dall'altra sono più importanti delle affinità. Proprio tali differenze rendono necessario mettere a punto uno specifico insieme di concetti analitici e di classificazioni, diversi da quelli adottati per studiare l'economia di mercato. Tuttavia tra le varie economie, passate e presenti, esistono anche analogie di cui bisogna render conto (v. Polanyi, 1971). Quali sono i tratti comuni a tutte le economie? Perché tutte le società hanno una economia o un sistema economico di qualche tipo? Le bande, le tribù, i regni, i moderni Stati capitalisti o comunisti sono tutti costituiti da esseri umani - uomini, donne, bambini - riuniti in un qualche tipo di raggruppamento sociale, di comunità o di società. Sia che studiamo una banda di cacciatori dell'età della pietra, o l'Impero romano o l'Unione Sovietica, siamo comunque sempre in presenza di gruppi di esseri umani, che, in quanto entità biologiche, hanno bisogno di approvvigionamenti costanti per sopravvivere fisicamente (cibo), e di società (bande, clan, nazioni), che necessitano di un analogo approvvigionamento per continuare a sussistere come entità autonome e integrate. Tali gruppi devono, cioè, essere approvvigionati di tutti i beni e i servizi indispensabili a soddisfare i bisogni e le attività che la cultura e il tipo di governo della loro società richiedono, e delle tecnologie e delle risorse che la stessa società consente (armamenti, finanziamenti per usi di pace o di guerra, prezzo-della-sposa per potersi sposare, mezzi per il sostentamento degli sciamani, dei sacerdoti oppure dei membri del Politburo). Ovunque insomma vi sia una società di uomini capace di perpetuarsi, possiamo aspettarci di trovare dei principi fondamentali che regolano la sua organizzazione 'economica': norme per l'allocazione e lo scambio di terra, lavoro, strumenti tecnici e prassi istituzionalizzate (come, per esempio, il commercio estero) che rendono possibile l'approvvigionamento continuativo delle persone e dei gruppi sociali. Questo non è mai lasciato al caso, ma è sempre strutturato, perché le carenze materiali conducono alla morte sia degli individui, sia delle società in quanto entità autonome e integrate. Per 'sistema economico' intendiamo dunque l'organizzazione, le regole e le istituzioni attraverso le quali viene strutturato questo costante processo di approvvigionamento materiale. Ma come studiando le economie contemporanee dell'Italia, del Giappone, della Svezia o dell'Unione Sovietica constatiamo che le rispettive forme organizzative, regole e istituzioni possono essere molto diverse tra loro, allo stesso modo possiamo riscontrare notevoli differenze tra le economie precoloniali Inca o Nuer, o anche tra l'economia della Cina del secolo scorso e quella di oggi.

Per approvvigionarsi, tutte le società utilizzano risorse naturali, lavoro umano, tecnologia (utensili, macchine e conoscenze relative ai processi di acquisizione e di produzione), una serie di procedimenti e di meccanismi istituzionalizzati e un qualche tipo di commercio estero, di moneta e di mercati. Dal punto di vista dell'antropologia economica e della storia economica antica questi fattori sono molto importanti, perché la conoscenza del modo in cui il commercio estero, il denaro e i mercati funzionano concretamente nelle singole economie (degli Stati Uniti, dell'Unione Sovietica, o delle isole Trobriand) ci fornisce importanti indizi strutturali sulla realtà di quelle società. Una delle intuizioni più importanti di Polanyi è che l'uso degli oggetti-moneta e l'organizzazione degli scambi con l'estero in una determinata economia sono indici estremamente efficaci e rivelatori dei principi fondamentali che regolano l'approvvigionamento materiale in quel tipo di economia. Se i porti adibiti al commercio estero nell'Inghilterra anglosassone (v. Hodges, 1982) o nella Cina dei Qing (v. Mancall, 1968) funzionavano diversamente dal modo in cui è organizzato il commercio estero britannico o cinese ai giorni nostri, questo dipende dal fatto più generale che la struttura politica ed economica interna dell'Inghilterra anglosassone o della Cina dei Qing è diversa da quella della Gran Bretagna o della Cina contemporanee. Lo stesso vale anche per quanto riguarda la moneta (v. Grierson, 1978).

3. Tipologie dei sistemi economico-sociali precoloniali, coloniali e postcoloniali

Per sistematizzare i dati che abbiamo a disposizione e per giungere a conclusioni analitiche sulle economie e sulle società studiate dagli antropologi, si adottano oggi diversi tipi di classificazione e di concetti tecnici. La tab. II propone una classificazione duplice, secondo le tipologie economico-sociali e secondo i periodi storici. Per motivi che saranno spiegati più avanti, le società contadine richiedono uno specifico tipo di analisi e debbono essere tenute distinte dalle bande e dalle tribù.

a) Economie aborigene nelle società tribali senza Stato: tipo I

Questo primo gruppo di economie e società è di particolare interesse per varie ragioni. Ad esso si rivolse l'attenzione dei primi antropologi - americani, inglesi e tedeschi - che si dedicarono alla ricerca sul campo e introdussero l'antropologia nel novero delle discipline accademiche. L'opera e le pubblicazioni di questi pionieri (Boas, Malinowski, Evans-

Pritchard) definirono l'ambito e l'oggetto dell'antropologia sociale. In alcune di queste società senza Stato era in uso una pratica di grande rilevanza, quella dello scambio cerimoniale (il potlatch tra i Kwakiutl, il kula delle isole Trobriand, o il moka delle Highlands in Nuova Guinea) che alcuni studi recenti hanno ulteriormente chiarito e illustrato (v. Strathern, 1971; v. Rosman e Rubel, 1971; v. Dalton, 1978). Proprio queste società (e altre) vennero analizzate nel 1924 da Marcel Mauss nel suo studio comparativo sullo scambio di doni (v. Mauss, 1925).Prima di cadere sotto il dominio coloniale queste economie, questi sistemi politici, queste società e culture erano quanto di più lontano si possa immaginare dall'Europa o dall'America contemporanee (si tratta in effetti delle società più 'primitive' tra tutte quelle studiate dagli antropologi). Esse non presentavano nessuna delle caratteristiche chiave del moderno capitalismo industriale: non avevano tecnologia delle macchine, né scienza applicata; non esisteva lo Stato, e quindi non vi erano né re, né aristocrazia, né moneta coniata; non c'erano città; la scrittura era sconosciuta; non esisteva alcuna religione universale (come il cattolicesimo o l'Islam); non c'erano mercati del lavoro o della terra, e la sussistenza non dipendeva da essi. Per di più, queste società erano costituite da popolazioni poco numerose e da piccole comunità politiche. Erano pressoché cronicamente in guerra e quasi tutte avevano sperimentato periodi di carestia o di fame.Proprio come l'impiego su vasta scala della tecnologia delle macchine costituisce lo spartiacque che separa l'economia di paesi come l'Italia o il Giappone da una situazione preindustriale, analogamente nelle economie precoloniali studiate dagli antropologi la presenza o l'assenza di un governo centrale (un regno o uno Stato) rappresenta un cruciale fattore di demarcazione. Tutte le società senza Stato adottano principî molto peculiari di organizzazione e di transazione a livello sia economico che sociale: cioè, sia per quanto riguarda la sussistenza (approvvigionamento giornaliero di cibo tramite caccia, raccolta e pastorizia, agricoltura basata sulla tecnica del 'tagliae-brucia'), sia nel settore dell'economia politica e sociale (prezzo-della-sposa, risarcimento per uccisioni o ferite, scambi cerimoniali: kula, potlatch, moka). In entrambi i settori, le istituzioni socioeconomiche di queste popolazioni sono diretta espressione delle loro particolari condizioni demografiche, ecologiche, tecnologiche, politico-sociali e religiose: una rudimentale tecnologia che non riesce a risolvere problemi come la siccità o le malattie delle piante; l'importanza centrale del lignaggio, del clan o della comunità locale in tutte le attività, incluse quelle economiche; le speciali forme di scambio di doni, al fine di stabilire quattro tipi di alleanze con segmenti di clan esterni (alleanze

matrimoniali, alleanze belliche, alleanze basate sullo scambio cerimoniale o sul commercio); i sistemi di credenze religiose in cui gli antenati assurgono al ruolo di spiriti protettori e vegliano sui successi e sugli insuccessi delle comunità; il ruolo e le iniziative volte a promuovere l'ordine interno e i rapporti con gli altri, la guerra e la pace, dei big-men e dei capi del lignaggio; le minacce croniche di carestia e di conflitto. Tutte queste cose concorrono a stabilire una situazione di dipendenza ecologica dall'ambiente fisico (dalla terra, dal clima, dai corsi d'acqua e dalle foreste); di dipendenza soprannaturale dagli spiriti degli antenati; infine di dipendenza dal lignaggio, dal clan e dagli alleati esterni, per quanto riguarda la terra, il lavoro, il sostentamento materiale dei malati e degli anziani, i matrimoni, la soluzione delle dispute, la protezione politica contro i nemici esterni, e la disponibilità di rifugi e di scorte alimentari di emergenza.

A parte l'espressione, assai vaga, 'economie tribali', non esiste un modo convenzionale per designare con precisione questo tipo di 'sistema economico', anche perché le istituzioni della comunità (Gemeinschaft) che provvedono alla sussistenza ordinaria e orientano l'economia politica e sociale - il lignaggio, il clan, la guida politica in guerra e in pace e la formazione di alleanze - sono tutte istituzioni 'non economiche'. In queste società senza Stato il principio della 'reciprocità' svolge un ruolo egemone e generale sia per l'ambito della sussistenza ordinaria che per quello dell'economia politica e sociale. Per reciprocità si intendono le allocazioni e le transazioni, interne ed esterne, di terra, lavoro, prodotti alimentari, spose e oggetti di valore, nonché determinate forme coercitive di scambio di doni finalizzate alla creazione e al mantenimento di stabili relazioni sociali e politiche tra individui e tra gruppi (v. Sahlins, 1972). Presso queste società senza Stato lo scambio di 'oggetti di valore primitivi' e lo scambio cerimoniale rappresentano istituzioni socioeconomiche peculiari e di grande importanza (v. tabb. III e IV).

Il dominio coloniale europeo modificò drasticamente queste caratteristiche, le relazioni di dipendenza, i rapporti di alleanza e di ostilità: venne posto termine alla guerra tra i vari gruppi indigeni e gli Europei introdussero i loro beni e i loro strumenti, il cristianesimo, il governo centralizzato, e quelle potenti istituzioni di mercato che sono la valuta, la produzione agricola per il mercato e il lavoro salariato (v. Dalton, 1978 e 1984).

b) Economie aborigene nei regni tribali: tipo II

Solo nell'Africa subsahariana continuarono ad esistere, fin oltre l'inizio della dominazione

coloniale europea alla fine del XIX secolo, più di una cinquantina di regni tribali. Di questa categoria (tipo II) fanno parte anche l'impero Inca (v. Murra, 1980) e l'Inghilterra anglosassone del periodo compreso tra l'arrivo dei Germani, intorno al 400 d.C., e la conversione al cristianesimo intorno al 600. Per quanto riguarda i regni africani, disponiamo - oltre che di due dozzine e più di ricerche antropologiche sul campo - di una serie ancor più ampia di analisi a opera di storici; tra queste va menzionato lo studio di Polanyi sul Dahomey e il traffico degli schiavi (Dahomey and the slave trade, 1966), che spiega e illustra in modo dettagliato ciò che egli intende con "redistribuzione" e "scambio amministrato" e descrive la "sagra annua dei sacrifizi" tenuta dal re.

Non è possibile far rientrare in un unico modello unitario ed esauriente le diverse realtà dei regni africani precoloniali. La maggior parte di questi ignorava la scrittura e pertanto gli antropologi non poterono far ricorso a testimonianze scritte quando giunsero in quei paesi, almeno una generazione dopo l'inizio del dominio coloniale (francese, inglese, tedesco o belga) che ne aveva già modificato la realtà in modo significativo. Ci manca perciò una gran quantità di informazioni. "Quando uno Stato soccombe alla conquista di una potenza coloniale, il sistema di governo subisce immediatamente tutta una serie di cambiamenti bruschi e di vasta portata [...]. Pertanto il lavoro di ricerca sul campo effettuato in uno Stato in tali condizioni [...] ci darà un'immagine di quel regno molto diversa dalla situazione precedente" (v. Goody, 1967, p. 179). Tra l'arrivo in Africa dei primi mercanti europei (portoghesi, inglesi, francesi, ecc.) nel 1500 e l'insediamento ufficiale del dominio coloniale verso il 1890 passano più di trecento anni. In questo periodo il traffico degli schiavi oltre Atlantico e il commercio dell'oro e dell'avorio tra Europei e Africani trasformarono alcuni regni, come l'Ashante e il Dahomey, in imperi di mercanti di schiavi. Un'altra profonda trasformazione, precedente il dominio coloniale, fu causata dalla penetrazione dell'Islam nell'Africa Nera e dalla conquista del potere da parte dei musulmani in parecchi regni africani. In nessun momento della loro esistenza precoloniale i regni africani hanno presentato strutture organizzative simili, e in molti di essi si sono verificati profondi cambiamenti nei secoli che hanno preceduto il dominio coloniale europeo. Schematicamente, possiamo classificare i regni africani in quattro gruppi distinti, evidenziandone così le differenze strutturali.

Il primo gruppo è quello che chiamerò dei regni 'feudali', ad esempio il regno Nkole o il Rwanda: qui un'élite dominante di guerrieri e allevatori di bestiame si differenzia tanto

nettamente dai suoi clientiservi, fisicamente diversi, più numerosi e dediti all'agricoltura, che per descrivere questi sistemi gli antropologi e gli storici parlano di caste, di feudalesimo, di servitù e di Stati di conquista (v. Oberg, 1940; v. Maquet, 1961). Ci sono poi i regni 'tribali' che non commerciano con gli Europei, come il Lozi o il Baganda: vengono definiti tribali perché le loro unità costitutive sono i clan e perché la loro tecnologia non è più progredita - né a livello militare né per quanto riguarda la produzione del cibo, ecc. - di quella dei clan e delle altre tribù africane senza Stato (v. Gluckman, 1983).
Il terzo gruppo è quello dei regni tribali attivamente impegnati nel commercio con l'Europa, come il regno degli Ashanti o il Dahomey. Il commercio permise ai re o all'élite dominante di avere accesso alle armi e ai beni di lusso europei: le armi servivano per favorire un'espansione imperialista, finalizzata ad ottenere tributi e schiavi; i beni di lusso per pagare i servizi di un'élite burocratica in via di espansione (v. Wilks, 1975; v. Polanyi, 1966). Infine ci sono gli Stati-regno musulmani, molto più simili alle civiltà contadine dell'Europa, del Giappone e della Cina del periodo preindustriale che non agli altri regni africani che abbiamo appena descritto: le analogie si riscontrano riguardo alla religione, alla più sviluppata tecnologia, all'organizzazione militare, al livello di urbanizzazione, al commercio e all'alfabetizzazione. L'Islam implicava una religione universale, la conoscenza della scrittura, la presenza di mercanti di professione, contatti continui - religiosi, culturali, commerciali - verso nord, attraverso il Sahara settentrionale, con le coste del Mediterraneo, e verso nord-est con l'Egitto. I regni musulmani avevano una milizia a cavallo e artigiani riuniti in corporazioni (v. Nadel, 1942; v. Smith, 1967). La tab. V colloca i regni tribali precoloniali tra le società precoloniali senza Stato e le economie preindustriali di Stati-regno che includono comunità contadine. La caratteristica più rilevante dei regni tribali africani è che essi condividevano significative caratteristiche strutturali con entrambi gli altri modelli: per quanto riguarda la vita di villaggio, la tecnologia della produzione alimentare, il possesso della terra, l'assenza di alfabetizzazione, l'importanza delle affiliazioni di lignaggio, ecc., questi regni erano molto simili alle società senza Stato, come quella dei Tiv o dei Tallensi. Ma la presenza di un re e di una élite di governo e le loro importanti funzioni - economiche, politiche, militari e religiose - mostrano che essi avevano anche parecchie affinità con le società contadine degli Stati-regno dell'Europa, del Giappone e della Cina preindustriali: re sacri che svolgevano diverse funzioni; un settore redistributivo di pagamenti obbligatori a favore del governo centrale e di esborsi da parte di quest'ultimo; una stratificazione sociale,

leggi suntuarie, monopoli reali, un settore del commercio estero amministrato politicamente e un sistema di alleanze tributarie (v. Dalton, 1981).

c) Il mutamento determinato dal colonialismo nelle società tribali: 1500-1960

Un sottogruppo di società tribali - non tutte, solo alcune - subì un'esperienza brutale e traumatica che ne causò la decimazione; per indicare il profondo sconvolgimento, lo shock, la trasformazione forzata e la catastrofe demografica causati dal colonialismo si devono usare termini come 'distruttivo' e 'degenerativo'. I casi meglio documentati sono quelli degli Aborigeni australiani (v. Rowley, 1970) e degli Indiani d'America e delle isole caraibiche (v. Cook, 1976; v. Wolf, 1959, capp. 8-11; v. Wachtel, 1971). La conquista violenta, il drastico declino demografico causato dalle malattie, dalla guerra e dalla fame, l'assoggettamento, il trasferimento e la trasformazione forzati ridussero sistematicamente gli Indiani del Nuovo Mondo e gli Aborigeni australiani al livello di inferiori rispetto ai loro colonizzatori europei e americani, che alla fine giunsero a sopraffarli anche numericamente.

Un secondo gruppo di società tribali sfuggì del tutto all'effetto distruttivo del mutamento imposto dal colonialismo. I casi meglio documentati sono quelli delle colonie britanniche in Africa, grazie all'opera di molti antropologi che lavorarono sul campo nel contesto africano nel periodo tra le due guerre; disponiamo inoltre di alcune penetranti indagini analitiche sulle conseguenze del dominio coloniale britannico in Africa (v. Perham, 1963; v. Mair, 1963).

Le intenzioni degli inglesi nei confronti delle loro colonie africane sono abbastanza chiare: stabilire l'ordine e la legge, ossia porre fine alle guerre tribali e alla schiavitù; imporre le leggi scritte e le procedure giudiziarie europee; introdurre tecniche agricole e adottare misure sanitarie per migliorare le condizioni di vita dei nativi, combattendo la carestia cronica e le epidemie; dare avvio, infine, a un processo di 'civilizzazione' a lungo termine, con l'introduzione del cristianesimo, dell'alfabetizzazione e dell'istruzione elementare, e promuovendo la commercializzazione (introduzione del lavoro salariato e produzione agricola per il mercato interno ed internazionale).

Ma ciò che gli Inglesi non furono in grado di creare in Africa durante i quasi settanta anni del loro dominio coloniale (1890-1960) sono quelli che Lewis (v., 1970, cap. 1) considera i fondamenti dello sviluppo: una trasformazione del settore agricolo capace di garantire un incremento annuo di produttività; la creazione di infrastrutture (trasporti, elettricità, scuole, ecc.); la formazione di una forza lavoro professionalmente qualificata, sia sul piano tecnico

che imprenditoriale; tasse e risparmi in misura sufficiente per finanziare i servizi pubblici e la formazione del capitale. Il dominio coloniale britannico ridusse così l'incidenza della fame e delle malattie, ma non fu in grado di avviare lo sviluppo economico.Infine, una conseguenza non intenzionale dei cambiamenti causati dal colonialismo inglese fu l'avvio di un processo di trasformazione dei membri delle tribù africane in 'contadini'. Il dominio coloniale inglese introdusse lo Stato e il governo centrale, la pacifica risoluzione delle dispute (proibendo gli scontri e i conflitti cruenti all'interno del territorio dello Stato), una religione universale (il cristianesimo), l'alfabetizzazione e soprattutto il mercato del lavoro e un grande aumento della produzione agricola per il mercato (v. United Nations, 1979). L'indipendenza e la fine del colonialismo dovevano accelerare questo processo di trasformazione degli indigeni africani in contadini.

d) Società contadine di antico insediamento all'interno di Stati-regno: tipo III

Le economie e le società contadine sono particolarmente importanti almeno per due motivi: in qualsiasi momento, negli ultimi 2000 anni, la popolazione mondiale è stata composta da un maggior numero di contadini che non da appartenenti a tribù senza Stato o a regni tribali; disponiamo inoltre di un'approfondita letteratura storica relativa alle società contadine dell'Europa e dell'Asia. Possiamo così studiare la struttura delle società contadine in diversi momenti storici e il loro progressivo cambiamento nel corso di diversi secoli, mentre ciò non è invece possibile per quanto riguarda le società tribali.

Ma le società contadine sono assai numerose e diffuse, sono presenti in epoche storiche molto diverse e alcune di esse sono dotate di istituzioni specifiche (il sistema delle caste in India, il feudalesimo in Europa); inoltre non tutte hanno vissuto le stesse esperienze storiche, determinanti per la loro trasformazione (pensiamo alle numerose conquiste della Sicilia ad opera di conquistatori così diversi, alla conquista della Spagna da parte dei Mori, all'espansione dei Turchi nei Balcani e in Nordafrica, alla Rivoluzione francese, alla carestia che colpì l'Irlanda negli anni quaranta del secolo scorso); queste società sono state analizzate in modo molto particolare dagli studiosi marxisti (v. Shanin, 1971). Ad ogni modo, la complessità delle società contadine e le differenze che le caratterizzano sono enormi, sicché anche la controversia teorica intorno a ciò che le distingue dagli altri tipi di economia e di società è molto serrata (v. Dalton, 1974).

Attualmente vi è almeno una settantina di nazioni in cui la maggior parte della popolazione

rurale è composta o da contadini (l'India, il Messico) oppure da agricoltori che discendono da contadini, ma che si trovano ormai in una situazione 'post-contadina' (Francia, Giappone, Germania). Per motivi che saranno spiegati in seguito, è particolarmente rivelatore lo studio di due gruppi di società contadine. L'analisi di questi modelli ci metterà in grado di apprendere alcune cose essenziali sulla struttura e sulle trasformazioni di tutte le economie e società contadine. Il primo gruppo è composto dalle società contadine presenti un tempo in Francia, Germania e Giappone, che fanno parte delle società del tipo III; il secondo, di cui fanno parte le società contadine dell'America Latina e dei Caraibi, sarà classificato separatamente, come società del tipo IV.

Le economie contadine della Francia, della Germania e del Giappone ci offrono un modello di trasformazione molto istruttivo, in quanto possiamo studiarle lungo un periodo di circa un millennio, dal feudalesimo medievale fino a oggi. La lezione che se ne trae è che il mondo contadino tramonta in generale in due maniere diverse: con l'abbandono del lavoro agricolo per la città e il lavoro industriale; oppure, pur rimanendo nel settore agricolo, con il passaggio a una condizione 'post-contadina': questi agricoltori-cittadini perdono la loro vecchia identità contadina perché il loro reddito, il loro livello di partecipazione al mercato e di dipendenza da esso, nonché il loro grado di istruzione, di partecipazione politica e di espressione culturale non sono più inferiori o sostanzialmente diversi da quelli dei lavoratori dell'industria urbanizzati (v. Dalton, 1986). I contadini in effetti restano contadini in senso stretto solo finché restano, rispetto agli altri membri della società nazionale cui appartengono, in una condizione di sensibile inferiorità e di diversità culturale che si rispecchia in tutte le dimensioni più importanti dell'esistenza - economica e tecnologica, politica e giuridica, sociale e culturale. Le caratteristiche che contraddistinguono i contadini sono la loro arretratezza per ciò che riguarda i metodi della produzione agricola, la loro inferiorità politica, economica e sociale rispetto agli altri membri della società nazionale cui appartengono e le loro usanze e tradizioni culturali peculiari (dialetto, abitudini alimentari, modo di vestire, usanze matrimoniali, ecc.; v. Weber, 1976, capp. 1-11).

Il passaggio dei contadini dell'Europa occidentale (e del Giappone) dallo stato di servitù feudale (in cui vivevano circa mille anni fa nel Medioevo) alla condizione post-contadina dei giorni nostri ha implicato mutamenti generali dell'economia e cambiamenti tecnologici, politici, giuridici, sociali e culturali, i quali hanno contribuito a ridurre o a eliminare l'inferiorità e la diversità dei contadini rispetto agli altri membri della società, e ad abbassare

a circa il 10% o anche meno la quota nazionale di forza lavoro impiegata nel settore agricolo. Le società contadine dei paesi come la Francia e la Germania sono sopravvissute più di un migliaio di anni perché, in tali paesi, la trasformazione delle tribù senza Stato e dei regni tribali primitivi in società contadine all'interno di Stati-regno avvenne molto presto; e inoltre perché i fattori che contribuirono a trasformare i contadini in non contadini o in post-contadini (sviluppo dei mercati nazionali e internazionali, sviluppo industriale sostenuto, progresso economico, moderno Welfare State, democrazia politica) impiegarono circa 150 anni (1820-1970) ad affermarsi e a ridurre il settore agricolo dal 50 al 10% della forza lavoro, garantendo al tempo stesso a chi continuava a dedicarsi all'agricoltura redditi, istruzione, risorse tecnologiche, diritti politici e pratiche culturali di tipo urbano sostanzialmente equivalenti a quelli degli impiegati nell'industria (v. Priebe, 1976; v. Franklin, 1969). In Giappone le medesime trasformazioni si verificarono in circa 100 anni (1870-1970), mentre a Taiwan e nella Corea del Sud il processo sarà ancora più rapido. Comunque tutte le economie contadine di lungo insediamento hanno conosciuto il medesimo processo di trasformazione.

e) Società contadine create dal colonialismo spagnolo in America Latina: tipo IV
Prima dell'arrivo degli spagnoli nel XVI secolo, in Messico, nell'America centrale e meridionale e nei Caraibi non esistevano società contadine, ma solo tribù senza Stato (ad esempio i Caribi o gli indiani Yanomamo) e regni tribali (Inca). Le comunità contadine dell'America Latina furono create durante i trecento anni di dominio coloniale spagnolo con la trasformazione e con il trasferimento di popolazioni appartenenti a ceppi diversi: gli indigeni indiani furono trasformati in contadini 'ibridi' (conservarono la lingua e le abitudini alimentari indiane, ma si convertirono al cattolicesimo e furono ridotti allo stato di servi degli Spagnoli); gli schiavi importati dall'Africa e messi al lavoro nelle piantagioni di canna da zucchero divennero anch'essi contadini 'ibridi', conservando alcuni elementi della cultura africana, come dimostra il caso di Haiti; ci fu poi l'immigrazione dei contadini asiatici ed europei in America Latina e nei Caraibi (v. Mintz, 1968).
I trecento anni di colonialismo spagnolo modellarono sia queste comunità contadine, sia il più ampio insieme delle economie e delle società latinoamericane mediante l'introduzione di tre tipi di istituzioni tipicamente europee: il cattolicesimo, alcune varianti della servitù feudale cui vennero assoggettati gli Indiani (e l'assegnazione di grandi latifondi ai signori spagnoli) e il mercantilismo, vale a dire un tipo preindustriale di esportazione di materie

prime per la vendita sul mercato (minerali e prodotti agricoli).
L'indipendenza politica dalla Spagna, ottenuta relativamente presto, non pose tuttavia le basi dello sviluppo economico, non portò con sé la democrazia politica, né spinse i governi a impegnarsi a favore di una certa equità nella distribuzione del reddito o a migliorare le condizioni sanitarie e dell'istruzione.Le realtà contadine dell'America Latina e dei Caraibi vanno perciò distinte da quelle di lungo insediamento del tipo III (la Francia e il Giappone) per due ordini di motivi: in primo luogo perché esse vennero create dopo il 1500 dalle istituzioni imposte dal regime coloniale spagnolo; in secondo luogo perché, pur a un secolo e mezzo dall'indipendenza dagli spagnoli, l'America Latina non ha ancora conosciuto un grande sviluppo economico e per questo motivo i suoi molti contadini, sempre più poveri, non hanno ancora raggiunto la condizione non contadina o post-contadina raggiunta invece in Francia e in Giappone. Tutti gli indicatori che permettono di misurare il basso livello di sviluppo dell'America Latina contribuiscono del resto a spiegare le ragioni del persistere di questa condizione di povertà contadina: essenzialmente esse vanno individuate nella distribuzione estremamente diseguale del reddito e della proprietà terriera; nell'ampiezza delle zone di povertà assoluta; negli alti livelli di mortalità infantile e di crescita demografica; in un'instabilità politica cronica; in problemi cronici di indebitamento con l'estero; nell'inflazione, nella crisi della bilancia dei pagamenti; nella disoccupazione urbana e nella presenza di baraccopoli alle periferie delle principali città; in un'agricoltura stagnante (v. Griffin, 1974; v. Barraclough, 1976).

Equilibrio economico

di Bruna Ingrao, Giorgio Israel

sommario: 1. Il concetto di equilibrio economico. 2. La teoria dell'equilibrio economico generale nel suo profilo storico. 3. Il modello di Arrow e Debreu: l'equilibrio intertemporale con informazione perfetta. 4. Una valutazione critica del modello di Arrow e Debreu: struttura, risultati analitici, interpretazioni, sviluppi recenti. 5. Conclusioni. □ Bibliografia.

1. Il concetto di equilibrio economico

Fin dalla seconda metà del Settecento gli studiosi che si sono occupati di economia hanno fatto uso del concetto di equilibrio economico per analizzare i fenomeni della produzione, del consumo e dello scambio. Nel corso dell'Ottocento questo concetto ha man mano acquisito un ruolo centrale nei tentativi di trasformare la giovane scienza economica in una scienza quantitativa avente le stesse caratteristiche di rigore e oggettività e metodi analoghi a quelli della fisica-matematica. La teoria dell'equilibrio economico generale, la cui prima formulazione compiuta è dovuta all'economista francese Léon Walras (1834-1910), è considerata da molti interpreti come il fondamento della teoria economica contemporanea in cui il concetto di equilibrio riveste un ruolo fondamentale.

L'indirizzo di pensiero ispirato dalla teoria dell'equilibrio economico generale si differenzia da altre correnti teoriche contemporanee (come la teoria che riprende la tradizione classica inglese, la teoria evolutiva di Alfred Marshall, la teoria della disoccupazione involontaria di John Maynard Keynes e della scuola keynesiana, la teoria di Schumpeter), ma non perché sia l'unico a utilizzare il concetto di equilibrio. Questo concetto, infatti, è stato usato ed è tuttora correntemente usato anche in accezioni diverse e secondo definizioni diverse da quella adottata nella teoria walrasiana. Nei capitoli che seguono accenneremo ai diversi concetti di equilibrio ma ci soffermeremo sulle idee e sui problemi della teoria dell'equilibrio economico generale.

Quello di equilibrio economico è un concetto che descrive uno stato possibile del sistema economico, definito idealmente da particolari proprietà. Questo stato ideale del sistema economico è l'oggetto di una riflessione teorica tesa a definire modelli capaci di rappresentarlo con chiarezza e precisione e a ottenere risultati interessanti circa le sue proprietà.

I problemi che nascono in questa riflessione possono così riassumersi: come è descritto il sistema economico di cui si vuol studiare questo stato ideale? come sono definite le proprietà che caratterizzano tale stato ideale? su quale arco di tempo si osserva il sistema economico quando se ne considera uno stato di equilibrio? perché - infine - viene attribuita tanta importanza allo stato di equilibrio?

Iniziamo col descrivere brevemente il contesto teorico entro cui vengono esaminati questi problemi e precisamente la nozione di sistema economico. Il sistema economico è caratterizzato, in questa teoria, da tre elementi: le merci, i prezzi e l'attività degli agenti

economici. Una merce è da intendere come un bene (presente in natura o risultato dell'attività produttiva) o una prestazione di attività lavorativa o di servizi ed è caratterizzata da tre elementi: la sua quantità fisica, la sua collocazione nello spazio, la sua collocazione nel tempo. Il prezzo è un numero che definisce ciò che deve essere pagato ora, in termini della quantità di una data merce, per ottenerne un'altra: esso definisce quindi i rapporti di scambio ed è anche un indicatore della 'scarsità' delle merci. Gli agenti economici sono i produttori e i consumatori. I primi sono mossi nella loro attività (la produzione di merci) dall'intento di rendere massimo il proprio profitto. I secondi si scambiano le merci nell'intento di 'massimizzarne l'utilità', ovvero di soddisfare, al livello massimo consentito, un sistema di 'preferenze' che guida le loro scelte. Gli uni e gli altri sono mossi dalla spinta a ottenere il massimo vantaggio possibile (insaziabilità). Si suppone che essi agiscano sulla base di un piano perfettamente determinato (il che presuppone una capacità di previsione perfetta): questa ipotesi ha come conseguenza il carattere strettamente deterministico del sistema economico.

Tre proprietà sono state usate, in numerose varianti, per definire lo stato di equilibrio del mercato. In base alla sua prima e più intuitiva caratterizzazione lo stato di equilibrio è uno stato dell'economia che, una volta raggiunto, permane, o, più precisamente, è uno stato che il sistema economico tende spontaneamente a raggiungere e a conservare invariato. Questa nozione intuitiva, secondo cui l'equilibrio economico è uno stato di invarianza del mercato e allo stesso tempo è un punto di arrivo cui tende l'evoluzione del mercato, è stata oggetto di sofisticate elaborazioni formali che ne hanno di volta in volta reso preciso il significato mediante diversi modelli matematici.

In secondo luogo, l'equilibrio economico è concepito come uno stato in cui ciascun agente economico si trova in una posizione ottimale, tenuto conto delle limitazioni ('vincoli') che gli sono imposte dalle scelte degli altri agenti e dalle risorse di cui egli stesso dispone o che sono globalmente disponibili nell'economia. Nello stato di equilibrio ogni scambista sceglie i comportamenti e i piani che più gli convengono secondo criteri di ottimalità.

Infine lo stato di equilibrio è inteso come uno stato di coerenza dei piani e di compatibilità delle scelte dei singoli operatori, scelte che possono pertanto essere tradotte in un sistema di scambi coerente e realizzabile, nel senso che nessuno si troverà nell'impossibilità di scambiare ciò che ha programmato in quanto non compatibile con le decisioni degli altri agenti. Questa proprietà dello stato di equilibrio è variamente rappresentata da condizioni di

uguaglianza tra domanda e offerta sui mercati.
È spontaneo chiedersi perché il concetto di equilibrio abbia avuto un ruolo tanto centrale nella teoria economica. La risposta è che dietro la scelta di questo concetto vi è una 'visione', per usare un termine caro a Schumpeter, una concezione del mondo dei fenomeni economici che ha indirizzato la scelta dei modelli e dei termini utilizzati. La concezione è quella secondo cui il comportamento degli agenti economici porta in modo naturale l'economia verso una condizione di compatibilità dei differenti fini. Data questa concezione, che costituisce una premessa 'metafisica' e indimostrabile della teoria, obiettivi dello studio diventano la caratterizzazione dell'evoluzione spontanea postulata e la scoperta delle condizioni sotto le quali si manifesta lo stato di compatibilità.
In questa 'visione' la compatibilità dei fini nello stato di equilibrio del mercato è caratterizzata dalla proprietà dell'ottimalità: gli operatori si trovano nello stato di equilibrio 'meglio' che in altri stati, in quanto in esso realizzano le loro scelte preferite. Intervengono così nozioni come quella di maggior benessere e di migliori condizioni del sistema, la cui introduzione ha dato luogo a lunghi dibattiti per pervenire alla definizione rigorosa di cosa debba intendersi per ottimalità a livello delle scelte di una collettività e non di un singolo. I problemi che si ponevano possono essere riassunti nelle seguenti domande: è possibile misurare il benessere di una collettività? si può in alternativa adottare un criterio di confronto tra stati del sistema economico per decidere quale sia ottimale ed entro quali limiti?
Una definizione esatta della nozione di ottimalità fu data da Vilfredo Pareto (1848-1923). Secondo tale definizione, uno stato ottimale (o 'efficiente nel senso di Pareto') è uno stato in cui non è possibile migliorare la soddisfazione di un agente senza diminuire quella di almeno un altro agente. Questa nozione precisa ha inaugurato un filone di ricerca tendente a mostrare l'esistenza di ottimi nel senso di Pareto e a studiare la relazione di questa nozione con quella di equilibrio. Si è così mostrato non soltanto che un ottimo di Pareto esiste, sotto condizioni piuttosto generali, ma anche che un equilibrio economico relativo a un dato sistema di prezzi è un ottimo di Pareto e viceversa (v. Debreu, 1959; v. Arrow e Hahn, 1971).
Diamo ora una sommaria idea delle diverse nozioni di equilibrio che sono state introdotte nella teoria economica.
Un ruolo importante (e anticipatore) ha avuto l'analisi dell'equilibrio parziale: l'equilibrio tra la domanda e l'offerta nel mercato di un solo bene considerato isolatamente. I primi modelli matematici di equilibrio parziale furono costruiti da Antoine-Augustin Cournot (1801-1877),

matematico e filosofo francese, nel testo Recherches sur les principes mathématiques de la théorie des richesses pubblicato nel 1838. In questo testo Cournot distingueva le situazioni di mercato in base al numero delle imprese presenti dal lato dell'offerta e individuava tre principali forme di mercato: il monopolio (un unico produttore), il duopolio (due produttori), la concorrenza indefinita (molti produttori ciascuno di dimensioni modeste rispetto al mercato, così da non essere in grado di influire individualmente con la propria produzione sulle condizioni generali dell'offerta). L'analisi parziale di Cournot permetteva di studiare le proprietà dell'equilibrio nelle diverse forme di mercato.

Nell'analisi di equilibrio parziale si considerano la domanda dei consumatori e l'offerta dei produttori come funzioni del prezzo della merce scambiata nel mercato sul quale si vuole concentrare l'attenzione. L'equilibrio è lo stato del mercato caratterizzato da un prezzo della merce (il prezzo di equilibrio), al quale la domanda espressa dai consumatori e l'offerta espressa dai produttori sono eguali. Tutti i modelli di equilibrio parziale si fondano, quindi, sull'ipotesi che sia possibile, per semplicità, isolare ciò che accade su un mercato da ciò che accade sugli altri mercati del sistema economico, senza che l'indagine perda rilevanza teorica. I modelli di equilibrio economico generale vogliono invece sottolineare le relazioni tra gli scambi che avvengono sui diversi mercati di un'economia. L'equilibrio generale è lo stato dei mercati (anch'esso definito da un sistema di prezzi di equilibrio) in cui, su tutti i mercati presenti nel sistema economico, la domanda globale del mercato e l'offerta globale del mercato - ovvero le somme rispettive di tutte le domande e di tutte le offerte individuali su ciascun mercato - sono eguali. I modelli di equilibrio generale, come vedremo, vanno incontro a grandi difficoltà formali, per la complessità delle interrelazioni tra i mercati che intendono porre in evidenza. I modelli di equilibrio parziale, benché più elementari, sono stati preferiti da molti economisti, proprio perché ritenuti più maneggevoli e tuttavia efficaci per descrivere situazioni di mercato circoscritte, per analizzare i problemi particolari di un settore industriale, per definire le forme di mercato (monopolio, oligopolio, concorrenza perfetta), per studiare gli scambi di determinati tipi di beni.

Il metodo dell'equilibrio parziale fu adottato in particolare dall'economista inglese Alfred Marshall (1842-1924) nel suo trattato Principles of economics, pubblicato nel 1890. Marshall studiava l'equilibrio di ciascun mercato isolatamente come equilibrio dell'industria, ponendo l'accento sull'evoluzione e l'adattamento della struttura produttiva (imprese, impianti, manodopera, reti di vendita) alle condizioni della domanda finale espressa dai consumatori.

A seconda dell'arco di tempo entro il quale questo adattamento poteva aver luogo, Marshall distingueva tre diversi equilibri del mercato: l'equilibrio del giorno di mercato, l'equilibrio di breve periodo, l'equilibrio di luogo periodo. L'equilibrio del giorno di mercato è l'equilibrio tra domanda e offerta, che si realizza spontaneamente in un mercato in condizioni di concorrenza, quando gli scambi si svolgano in un arco di tempo molto breve: la struttura produttiva dell'industria è data, le imprese hanno già deciso i livelli di produzione e portato al mercato la propria offerta, il prezzo di mercato si fissa al livello in corrispondenza del quale l'intera offerta sul mercato viene assorbita dalla domanda dei consumatori. Nell'equilibrio di breve periodo, che si determina in un arco di tempo più lungo, i livelli di produzione possono essere adattati dalle imprese per sfruttare le opportunità di profitto offerte dalle condizioni della domanda. In un arco di tempo ancora più lungo anche la struttura produttiva dell'intera industria si potrà modificare fino a raggiungere le condizioni di efficienza e minimizzazione dei costi (equilibrio di lungo periodo) che, secondo la teoria di Marshall, caratterizzano le imprese in condizioni di concorrenza.
L'idea che si possa isolare idealmente uno stato del sistema economico in cui la struttura produttiva sia data e le imprese adattino la propria offerta alle condizioni della domanda sulla base delle aspettative di profitto fu ripresa da J.M. Keynes nella General theory of employment, interest and money, pubblicata nel 1936, una riflessione teorica sulle gravi difficoltà delle economie di mercato nella crisi degli anni trenta. Keynes introdusse la nozione di equilibrio di sottoccupazione: l'equilibrio che si determina in un'economia nel breve periodo quando l'attività produttiva delle imprese ristagna a livelli di sottoutilizzo degli impianti e una parte delle forze di lavoro non trova occupazione, pur essendo disponibile a lavorare accettando le retribuzioni correnti di mercato. Questa nozione estendeva all'intero sistema economico l'idea marshalliana dell'equilibrio di breve periodo, con la particolarità di considerare come equilibrio, ovvero come stato nel quale il sistema economico tende a persistere, una situazione dei mercati caratterizzata dalla disoccupazione delle risorse e cioè, in primo luogo, dallo squilibrio tra la domanda e l'offerta sul mercato del lavoro.
Per questa singolare combinazione di analisi di equilibrio ed enfasi sull'eccesso di offerta nel mercato del lavoro, la definizione di Keynes è stata molto criticata nel dibattito successivo alla pubblicazione della General theory of employment, interest and money e ancora nel dibattito recente. Ciò nonostante, la nozione di equilibrio macroeconomico di breve periodo con disoccupazione delle risorse (in particolare della risorsa lavoro) è entrata a far parte della

teoria economica contemporanea ed è correntemente usata nella letteratura.Molte controversie ha suscitato anche l'idea di equilibrio di lungo periodo, per la difficoltà di combinare in modo coerente lo studio del cambiamento dell'apparato produttivo nel tempo e l'ipotesi dell'invarianza dei dati e del contesto, che caratterizza i modelli di equilibrio parziale.

Il fatto di considerare il tempo esplicitamente (e quindi non soltanto come una variabile formale) complica enormemente il concetto di equilibrio anche nei modelli di equilibrio generale, perché introduce elementi di variabilità nel comportamento di mercato degli operatori e implica la considerazione di un contesto concettuale non deterministico. Ciò comporta sottili problemi logici nella definizione dell'ottimalità e della compatibilità delle scelte. La coerenza dei piani individuali e la compatibilità delle scelte degli operatori possono riguardare soltanto le scelte presenti oppure le scelte che interessano un arco di tempo più lungo o persino le scelte che interessano tutto il futuro, nei modelli più astratti. Le scelte economiche che si proiettano nel futuro comportano delle aspettative, sia perché impegnano oggi risorse e redditi che saranno disponibili nel futuro, sia perché promettono consegne di beni non ancora effettuate e quindi sono espressione di piani di produzione e scambio non ristretti al presente: sono scelte dettate dall'aspettativa che nel futuro prevarranno certe condizioni attese, che un evento o un insieme di eventi avranno o non avranno luogo. Se esplicitiamo il dato che gli agenti economici fanno scelte e svolgono la loro attività nel corso di successivi periodi di tempo, allora dobbiamo anche tener conto esplicitamente del fatto che anche i piani di scambio formulati nel periodo presente per essere immediatamente realizzati sono parte di scelte e progetti che si fondano su aspettative circa eventi futuri, scelte e progetti che gli agenti potranno rivedere e correggere alla luce delle conferme che riceveranno dai fatti o degli errori di previsione. La scelta di piani ottimali è una scelta intertemporale, che investe una successione di periodi e comporta un complicato processo di formazione e verifica delle aspettative.

Queste considerazioni fanno intendere che sono possibili diverse nozioni di equilibrio, per descrivere l'intricata rete di relazioni che si stabiliscono sui mercati quando si vogliano rappresentare l'ottimalità e la compatibilità di piani intertemporali, che hanno attuazione nel corso del tempo. Le difficoltà logiche e formali che emergono nella trattazione del tempo, nell'ambito della teoria dell'equilibrio economico generale, sono state oggetto di dibattito e riflessione soprattutto durante gli anni venti e trenta del XX secolo. In quel periodo tre

studiosi, l'austriaco Friedrich A. von Hayek, lo svedese Erik Lindahl, l'inglese John Hicks, hanno contribuito a definire il concetto di equilibrio temporaneo, rappresentando gli scambi sui mercati come una successione di stati di equilibrio condizionati dalle aspettative circa il futuro degli agenti economici e dalle risorse e dall'apparato produttivo che l'economia ha ereditato dai periodi precedenti. Nei modelli di equilibrio temporaneo gli scambi sono o contrattati e attuati in ciascun periodo (scambi a pronti, cioè a consegna immediata) o contrattati per essere attuati più tardi (scambi a termine, cioè a consegna differita). La compatibilità delle scelte, ovvero lo stato di equilibrio, che si realizza in ognuno di questi periodi non è però garanzia di compatibilità e ottimalità dei piani per tutta la successione dei periodi, vale a dire anche relativamente ai periodi futuri, quando gli agenti potranno rivedere le proprie scelte, correggere le proprie aspettative, contrattare e attuare nuovi scambi. Una successione di stati dell'economia, nella quale gli agenti economici non debbano mai rivedere le proprie aspettative sugli eventi futuri, perché l'andamento del mercato conferma le loro previsioni, prende il nome di equilibrio con aspettative confermate.Per aggirare le difficoltà poste dal fatto di considerare il tempo, si è preferito adottare un modello artificioso, che rappresenta il caso particolare di un sistema economico nel quale tutti gli scambi sono contrattati in un unico momento iniziale, per essere poi attuati, secondo i contratti stabiliti, nel corso dei successivi periodi di tempo. Lo stato di equilibrio che si determina in questa particolarissima economia di mercati a termine completi, nella quale sono progettati e contrattati nel periodo iniziale tutti gli atti di scambio e di produzione che saranno realizzati in futuro, prende il nome di equilibrio intertemporale, perché rappresenta appunto uno stato di coerenza e compatibilità delle scelte ottimali degli agenti economici che si estende a tutto il futuro. A questo modello dedicheremo il seguito della nostra trattazione.

2. La teoria dell'equilibrio economico generale nel suo profilo storico

L'ambizione della teoria dell'equilibrio economico generale, dalle origini fino ai suoi sviluppi recentissimi, e quindi l'aspirazione degli autori che l'hanno sviluppata, è stata quella di produrre una teoria scientifica rigorosa, formulata mediante una struttura matematica internamente coerente, che permettesse di ottenere delle dimostrazioni dei risultati previsti. Benché la teoria sia stata formulata agli inizi in un linguaggio matematico molto elementare, se non primitivo, i problemi che affrontava erano i medesimi che costituiscono ancora

l'oggetto delle ricerche enormemente più sofisticate di oggi.
L'idea fondamentale della teoria è quella dell'interdipendenza dei mercati: i prezzi relativi sono determinati dall'interazione tra domanda e offerta su tutti i mercati. Poiché uno stato di equilibrio è definito da un sistema di prezzi di equilibrio, l'oggetto principale della teoria è sempre stato la determinazione dei prezzi in un mercato di concorrenza perfetta. L'assunto metodologico fondamentale è che ciò che avviene sul mercato nell'aggregato è determinato dalla somma delle scelte dei singoli produttori e consumatori e che queste scelte possono essere rappresentate in modo chiaro e completo a partire dalle preferenze degli individui e dalle tecnologie produttive delle imprese. I prezzi di equilibrio sono determinati dall'interazione tra le scelte razionali dei singoli e i vincoli posti dalla tecnologia e dalla disponibilità delle risorse. L'ipotesi della concorrenza perfetta esprime l'idea che i singoli agenti, consumatori o imprese, non siano in grado di influire con interventi consapevoli sul risultato di questa interazione e nelle loro decisioni debbano assumere come un dato i prezzi relativi espressi dal mercato. Questa ipotesi si enuncia dicendo che gli agenti sono price-takers.
Proprio in quanto così determinati, i prezzi sono gli unici segnali che permettono di condurre il sistema dei mercati interdipendenti (di concorrenza perfetta) all'equilibrio. Un dato sistema di prezzi relativi determina l'eccesso di domanda dei vari mercati; se tale eccesso non è nullo, il sistema dei prezzi si modificherà in modo che la domanda e l'offerta si avvicinino a uno stato di eguaglianza. La modificazione seguirà la legge della domanda e dell'offerta: il prezzo in un mercato cresce in corrispondenza di un eccesso positivo della domanda e decresce in corrispondenza di un eccesso di domanda negativo. Tale processo di aggiustamento dei prezzi (chiamato da Walras tâtonnement) non deve essere concepito come un processo consapevolmente guidato (come accadrà nella reinterpretazione moderna del concetto di equilibrio nell'ambito delle teorie della pianificazione), bensì come un meccanismo di autoregolazione dei mercati, che, guidati da una sorta di 'mano invisibile' (secondo l'espressione di Adam Smith), vengono spontaneamente condotti all'equilibrio. Un mercato di concorrenza perfetta tende all'equilibrio tra domanda e offerta e lo raggiunge attraverso aggiustamenti dei prezzi relativi, che cessano solo quando si raggiunge appunto un sistema di prezzi di equilibrio.
Nonostante le formulazioni precedenti si ispirino alle versioni più moderne della teoria dell'equilibrio economico, non c'è dubbio che i concetti sottostanti operano anche nelle prime

versioni di tale teoria. I problemi storiografici che si pongono sono allora quelli di analizzare come si sia formato il concetto di equilibrio economico, come si sia perfezionato e affinato e, allo stesso tempo, come gli economisti che hanno sviluppato questo filone di pensiero abbiano elaborato i criteri di scientificità che intervengono nella costruzione della teoria. Questi problemi storiografici si intrecciano strettamente ai problemi di carattere analitico: la dimostrazione della coerenza interna della teoria, la verifica empirica dei suoi risultati, l'applicabilità dell'apparato teorico ai fenomeni reali.

Nell'evoluzione della teoria dell'equilibrio economico generale possiamo distinguere quattro grandi tappe: 1) una prima fase di gestazione del concetto di equilibrio economico, che prende forma nella cultura francese del tardo illuminismo e si ispira direttamente al concetto di equilibrio della meccanica; 2) una seconda fase di formazione e consolidamento della teoria, nell'opera di Walras e Pareto, in cui si definiscono il nucleo paradigmatico della teoria e i suoi concetti fondamentali sotto l'evidente influsso dei concetti e dei metodi della fisica matematica classica; 3) una terza fase in cui la teoria, dopo la crisi della versione precedente, rinasce negli anni trenta del nostro secolo nell'ambiente scientifico di lingua tedesca (a Vienna, in particolare) e vengono poste le basi di una linea di sviluppo assiomatica; 4) la fioritura negli Stati Uniti, a partire dagli anni quaranta, dei programmi di assiomatizzazione della teoria e quindi degli sviluppi che la teoria ha conosciuto fino a oggi, attraverso varie peripezie metodologiche e concettuali.

La prima formulazione del concetto di equilibrio economico è un prodotto del pensiero filosofico e scientifico del XVIII secolo nell'ambito della nascente scienza della società. Questo concetto e le sue prime definizioni scientifiche sono ispirati da un tema centrale della cultura illuminista (e soprattutto tardoilluminista): la centralità e il valore universale della scienza fisico-matematica newtoniana, le enormi speranze suscitate dalle sue conquiste in ogni campo del sapere, l'impulso che tale scienza trasmette a costruire nuove regulae philosophandi nella storia, nel diritto, nella teoria politica. Uno dei risultati più evidenti di tale clima culturale è l'emergere nelle scienze sociali della nozione di 'legge', intesa come 'legge naturale'. Il modello da imitare, per determinare le leggi scientifiche della società, è la meccanica newtoniana, di cui uno dei concetti centrali è quello di equilibrio meccanico (statico o dinamico che sia). È possibile seguire con precisione (v. Ingrao e Israel, 1987) l'emergere progressivo di questo concetto nell'opera dei fisiocratici e in particolare nel pensiero di Anne-Robert-Jacques Turgot (1727-1781) e di Pierre-Samuel Dupont de

Nemours (1739-1817).
Non va trascurato, inoltre, il clima generale di sviluppo delle applicazioni della matematica a tematiche di carattere sociale e biologico, che è bene espresso dalle ricerche nel campo della arithmétique politique e che spingerà il marchese di Condorcet (1743-1794) alla formulazione di un ambizioso programma di mathématique sociale, di cui l'economia politica costituisce soltanto un aspetto. Tali ricerche saranno proseguite nel contesto delle attività della Classe de Sciences Morales et Politiques dell'Institut de France (v. Israel, 1991), che è il centro di attività del gruppo degli idéologues. Sarà proprio la Classe a premiare, nel 1801, quello che può essere considerato il primo testo contenente una formulazione esplicita del concetto di equilibrio economico e un primo tentativo coerente di matematizzazione della teoria dello scambio e della determinazione dei prezzi. Si tratta di un libro del matematico francese Nicolas-François Canard (1750-1833). In questo testo (v. Canard, 1801) l'equilibrio parziale è paragonato all'equilibrio statico di una leva, mentre l'equilibrio generale del mercato è rappresentato attraverso una metafora tratta dalla meccanica dei fluidi: il sistema economico è assimilato a un sistema di canali intercomunicanti in cui le acque raggiungono uno stato di equilibrio. Si noti che l'analogia idrodinamica è ispirata a sua volta dal parallelismo fra il processo di circolazione del sangue e l'equilibrio economico (inteso in forma dinamica), che è un tema classico del pensiero fisiocratico. Da questa fonte Canard lo ha ripreso. Il tentativo di Canard è per molti versi assai ingenuo, ma contiene in nuce parecchi concetti importanti che saranno ripresi nell'opera di Cournot e Walras. Fra di essi troviamo, in forma embrionale, un'idea del processo di aggiustamento dei prezzi (tâtonnement). Il lavoro di Canard e la parallela fioritura di lavori di matematica sociale, che spaziano dalla demografia alla teoria giuridica e politica, ai problemi di debito pubblico, alla statistica sanitaria, cadono infine nel discredito per un processo complesso di mutamento che si verifica nella cultura francese e che ha anche risvolti direttamente politici: il dispotismo napoleonico, l'esaurirsi del razionalismo utopistico riformatore dell'epoca rivoluzionaria, la separazione disciplinare che tende ad allontanare le scienze esatte dalle altre scienze. Anche nel gruppo degli idéologues prevale la tendenza a perseguire un progetto di costruzione della scienza sociale che rifiuta radicalmente la matematizzazione. L'economista Jean-Baptiste Say (1767-1832) è una figura di spicco di questa tendenza.
Si determina così una svolta che renderà l'ambiente della scienza sociale francese ostile alla matematizzazione, un'ostilità che sarà poi amaramente sperimentata da Cournot e da Walras.

Questa ostilità si sviluppa anche negli ambienti delle scienze fisico-matematiche, che pure erano stati così sensibili alle tematiche dell'arithmétique politique e della mathématique sociale e avevano visto manifestarsi l'interesse diretto per queste discipline da parte di scienziati di primo piano come Bernoulli, Condorcet, Lagrange, Vandermonde e Laplace. In estrema sintesi ciò è conseguenza del prevalere di una visione romantica, che nega la possibilità di esprimere 'in numeri', e cioè di studiare quantitativamente, questioni che toccano la libertà del soggetto sociale, la quale appare (in pieno conflitto con la visione prevalente nel materialismo settecentesco) come irriducibile a uno studio scientifico esatto. Da un lato ciò si esprime, nel mondo delle scienze fisico-matematiche, nel prevalere di una 'ripugnanza' per l'applicazione del calcolo alle cose dell'ordine morale, per l''aberrazione' consistente nell'"assimilare l'uomo a un dado" (per usare le parole del fisico-matematico Louis Poinsot, allievo di Laplace). E dall'altro lato, nel mondo delle scienze sociali, nel prevalere del metodo storico. Questa sostanziale convergenza stronca lo sviluppo scientifico di quelle discipline che trovano uno spazio di sviluppo in ambienti marginali rispetto alla scienza 'ufficiale', come quelli degli ingegneri o dei cultori della statistica. (Di particolare interesse, in questo contesto, è il contributo di Jules Dupuit, 1804-1866). Poiché la polemica si appunta soprattutto sull'uso del calcolo delle probabilità, cui viene negato uno status epistemologico di pari dignità con il punto di vista del determinismo e del calcolo infinitesimale, un margine sottilissimo viene lasciato all'uso di un punto di vista direttamente plasmato su quello della meccanica.
È proprio in questa stretta fessura che si fa strada l'opera fondamentale di Cournot. Cournot è ormai del tutto lontano dai grandi progetti riformatori, finalizzati alla felicità del genere umano, tipici del pensiero illuministico ed è anche molto più prudente di quanto sarà poi Walras nel valutare la portata del progetto di matematizzazione dell'economia: egli si limita alla formulazione di teorie parziali, con molti distinguo sulla loro applicabilità. In coerenza con questo atteggiamento egli prende energicamente le distanze dal tentativo pioneristico di Canard, pur ispirandosi con ogni evidenza a diverse idee dell'ingenuo professore francese di matematica e mostrando a più riprese di trarre spunto dall'analogia meccanica. L'originalità del progetto di Cournot (v., 1838) consiste nell'aver indicato la via dell'uso della teoria delle funzioni nello studio dei problemi economici, senza la necessità di specificare la forma delle funzioni che descrivono le leggi studiate, se non in termini di proprietà analitiche molto generali. Non c'è dubbio che questo indirizzo astratto e assolutamente alieno da ogni

applicazione numerica diretta ha plasmato gli sviluppi di tutta la teoria dell'equilibrio economico generale fino ai nostri giorni. Oltre a questo contributo metodologico fondamentale, i contributi specifici più importanti di Cournot sono la chiara enunciazione della legge che regola in un singolo mercato la relazione fra la quantità domandata del bene scambiato e il prezzo in moneta del bene stesso (loi du débit) e, come abbiamo già ricordato, la definizione delle 'forme' fondamentali del mercato, che nella teoria del duopolio anticipa e imposta l'analisi contemporanea dei mercati di oligopolio.

Tuttavia, per le ragioni sopra illustrate, il mondo scientifico e il mondo degli economisti erano ormai poco disposti a interessarsi a un siffatto approccio e l'opera di Cournot cadde in un silenzio quasi totale, al punto che nella sua opera successiva (v. Cournot, 1863), che riproponeva i contenuti della precedente, Cournot espunse tutte le formule matematiche. Con questo silenzio si chiuse la prima fase di gestazione del concetto di equilibrio, ma ormai erano stati gettati i semi di un progetto scientifico che continuò a sopravvivere in forma sotterranea.

Tale progetto si basava su un'idea che riemerse in Walras come in Pareto (e in molti altri autori, tra i quali ricordiamo William Stanley Jevons, Irving Fisher e Paul Samuelson): l'idea che i concetti della meccanica possano, una volta spogliati del riferimento specifico ai fenomeni fisici, costituire un repertorio di concetti astratti utilizzabili per una rappresentazione efficace dei fenomeni di mercato, e che a questo repertorio si debba attingere a piene mani, perché esso offre una sorta di materia prima concettuale già elaborata e sperimentata in una scienza solida come la fisica. Quest'idea è la premessa per il pieno sviluppo del concetto di equilibrio nell'opera di Walras, più precisamente nel testo che raccoglie le sue lezioni universitarie, pubblicato col titolo Éléments d'économie politique pure ou théorie de la richesse sociale (v. Walras, 1874-1877).

Nell'opera di Walras il progetto di creare un'economia politica pura in forma matematica, secondo il modello scientifico della meccanica classica e dell'astronomia newtoniana, diviene un progetto articolato e compiuto. Walras mutuò lo scheletro delle equazioni dell'equilibrio dal trattato di statica di Poinsot e attinse dalla meccanica classica non solo, e non tanto, gli strumenti matematici, ma anche, e soprattutto, un sistema di concetti per descrivere e interpretare i fenomeni di mercato. Ciò emerge con chiarezza nella discussione metodologica che Walras premise alla sua costruzione teorica: la distinzione tra tipi reali e tipi ideali di mercato e la descrizione di come si passi dalla definizione dei primi all'analisi formale per

teoremi e dimostrazioni. Il mercato di concorrenza perfetta è concepito da Walras come la rappresentazione ideale, senza 'attriti', del tipo reale del 'mercato d'asta', che è scelto a sua volta come la forma tipica del mercato tout court. Il tâtonnement è la rappresentazione astratta e depurata di aspetti accessori ('attriti') del meccanismo di funzionamento di un mercato d'asta, costruita per mostrare, attraverso i concetti puri della teoria, come i mercati di concorrenza convergano verso lo stato di equilibrio. Per Walras il tâtonnement è una sorta di simulazione ideale di ciò che accade effettivamente nei mercati, una rappresentazione ideale di come si svolgerebbero i fenomeni dello scambio in mercati che si avvicinassero alle condizioni ideali della concorrenza perfetta.

L'ultimo scritto di Walras (v., 1909) è un breve saggio tutto dedicato a convincere i "contraddittori matematici" che vi è perfetta analogia tra la teoria dell'equilibrio economico walrasiana e due temi centrali della meccanica: l'equilibrio della bilancia e le equazioni della gravitazione universale. Walras arriva a concludere che l'economia è una scienza 'psichico-matematica' del tutto analoga alle scienze fisico-matematiche, salvo il fatto che si occupa appunto di grandezze psichiche anziché fisiche. Nella corrispondenza di Walras (v. Jaffé, 1965) un intero carteggio è dedicato alla discussione puntuale e precisa di questa analogia, che egli considera fondamentale al fine di convincere i fisici e i matematici della solidità scientifica della nuova disciplina. Con insistenza accanita Walras si rivolse, infatti, ai fisici e ai matematici dell'epoca per ricevere approvazione e incoraggiamento al suo programma scientifico.

Nel programma di Walras era presente un problema aperto e irrisolto: la relazione tra l'anima normativa e l'anima descrittiva della teoria dell'equilibrio economico. Fin dal XVIII secolo il progetto di trasferire il concetto di legge dalle scienze fisiche alle scienze sociali aveva sollevato il problema della compresenza, nelle scienze sociali, di due facce del concetto di legge: da un lato la regolarità dei fenomeni e la loro causalità indipendente dal volere dell'uomo, dall'altro la norma della ragione e quindi il precetto dell'azione sociale. Questo conflitto - o duplicità di significato - insito nel concetto di 'legge naturale' esteso al mondo dei fenomeni umani è presente in Montesquieu, in Condorcet, nella fisiocrazia. Lo ritroviamo nell'opera di Walras, che discende, per aspetti importanti, dalla teoria del diritto naturale. Per i teorici del diritto naturale la legge naturale ha il carattere di una verità certa ed evidente di tipo deduttivo, come le verità della geometria, e non è indotta a partire da regolarità osservate nel mondo umano, né smentita dalle assurdità della storia. Per Walras l'equilibrio di

concorrenza perfetta, prima ancora di essere il modello astratto di un'organizzazione sociale esistente, è una norma della ragione, un ideale di organizzazione sociale cui tendere. L'equilibrio economico generale è un ideale normativo più che una fotografia della realtà dei mercati. Come si concili questa concezione normativa dell'equilibrio con la concezione meccanicistica mutuata dalla fisica classica è il problema interpretativo più delicato che si presenti allo studioso dell'opera di Walras.

Nell'opera di Vilfredo Pareto (1848-1923) la difficile convivenza divenne un'aperta contraddizione, che portò in definitiva Pareto a spostarsi sul terreno della sociologia. Pareto, un ingegnere italiano, era anche più radicato di Walras nella cultura scientifica del suo tempo. Contrariamente a Walras, aveva avuto una formazione scientifica regolare presso il Politecnico di Torino, che era allora un centro molto vivo di dibattito scientifico e un'ottima scuola di meccanica. Il suo programma scientifico fu subito informato, a differenza di quello di Walras, a una esigenza preminente di fondazione empirica della teoria, perché Pareto considerava il metodo sperimentale cardine e fondamento di ogni discorso scientifico. Nella costruzione della teoria dell'equilibrio Pareto fu guidato ancor più di Walras dall'idea dell'analogia meccanica. Nel suo Cours (v. Pareto, 1896-1897) egli inserì una tabella a doppia entrata delle corrispondenze tra fisica e teoria economica, ribadendo sia in quest'opera che nel Manuale (v. Pareto, 1906) che l'equilibrio di un sistema di punti materiali della meccanica offre alla teoria economica un sistema di concetti già ben costruito per ragionare sull'equilibrio economico.

Nella costruzione della teoria dell'equilibrio Pareto si scontrò con quella che ritenne una mancata corrispondenza tra teoria e fatti, e in particolare con quello che egli chiamò il "fenomeno soggettivo", cioè l'impatto sui comportamenti e quindi sui fenomeni sociali delle convinzioni soggettive, per irrazionali o false che esse possano essere. In definitiva, pur essendo molto più immerso di Walras nella problematica della matematizzazione dei processi reali, Pareto considerava tuttavia l'uso della matematica come un aspetto di un processo conoscitivo che deve essere saldamente radicato nel metodo sperimentale. Non essendo riuscito a dare un fondamento, un sostegno empirico alla teoria, Pareto entrò in quella crisi che lo spinse alla fine ad abbandonare il terreno dell'economia pura.

Della sua svolta vanno ricordati due aspetti: il distacco dalla razionalità come movente prevalente delle azioni umane e l'esigenza di ritrovare nel campo allargato della sociologia la verifica sperimentale che gli risultava impossibile trovare nell'ambito ristretto della teoria

economica. Per il primo aspetto Pareto, che aveva introdotto il concetto di homo oeconomicus come astrazione analoga al 'punto materiale' in meccanica, giunse a collocare la razionalità in una posizione di secondo piano nella spiegazione dei moventi delle azioni umane, quei moventi che forniscono la chiave per interpretare i grandi fatti sociali. Costruì invece una teoria delle azioni 'non logiche', attribuendo a tali azioni un'importanza sempre maggiore per comprendere le regolarità del mondo sociale. Per il secondo aspetto, pur senza rinnegare la teoria dell'equilibrio economico generale, Pareto giunse a sottolinearne la limitatissima applicabilità a fini sia normativi sia descrittivi.
La povertà dei contenuti empirici della teoria era stata lamentata anche dai fisici e matematici in risposta ai tentativi di diffondere la teoria stessa fatti soprattutto da Walras (v. Ingrao e Israel, 1987, cap. 6). È questo il problema centrale che emerge dal carteggio tra Walras e Poincaré (sollecitato da Walras a dare il suo appoggio alla teoria): il problema che Poincaré solleva è appunto quello del procedimento di astrazione adottato nella costruzione della teoria e dei "giusti limiti" di questa procedura. Secondo Poincaré, "all'inizio di ogni speculazione matematica vi sono delle ipotesi e [...] affinché questa speculazione sia fruttuosa occorre (come nelle applicazioni alla fisica, del resto) che ci si renda conto di queste ipotesi. Se si dimenticasse questa condizione si supererebbero i giusti limiti". Anche il fisico matematico Paul Painlevé muove, nella sostanza, la stessa critica, nell'introduzione all'edizione francese del 1909 della Theory of political economy di Stanley William Jevons (1835-1882). Tale critica circola in molti ambienti di economisti, manifestandosi, in particolare, nel rifiuto da parte dell'ambiente di Cambridge di accettare la teoria in tutte le sue ambizioni di generalità e astrattezza.
Alle soglie degli anni trenta, dopo circa cinquant'anni di vita, la teoria walrasiana era ancora quasi inesplorata sotto il profilo matematico: il suo nucleo concettuale era già chiaramente espresso negli Eléments, ma esso non aveva ancora trovato una chiara definizione formale, in particolare attraverso una precisa e soddisfacente formulazione matematica dei problemi dell'esistenza, dell'unicità e della stabilità dinamica dell'equilibrio, né tanto meno erano stati conseguiti risultati concernenti tali problemi.
Ha inizio a questo punto quella che abbiamo indicato come la terza fase nello sviluppo storico della teoria e che è legata alla diffusione della modellizzazione matematica e quindi di un rapporto più elastico fra matematica e realtà, che non subisce in modo stretto i vincoli del riduzionismo meccanicista. Non a caso questa fase ebbe luogo in un diverso contesto

geografico, trovando fertile terreno nei nuovi sviluppi assiomatici e astratti della scienza in Germania e in Austria. Questa nuova fase, legata soprattutto ai nomi di Abraham Wald (1902-1950) e John von Neumann (1903-1957), è caratterizzata da un cambiamento profondo dell'approccio matematico, che non è più quello dell'analisi infinitesimale, secondo le vedute della scuola di Losanna, bensì quello dell'analisi convessa. In verità i primi teoremi di esistenza dell'equilibrio, dimostrati da Wald nel quadro delle attività del Seminario Menger a Vienna, utilizzano ancora strumenti classici (ma con risultati limitati, proprio per le tecniche usate), mentre von Neumann apre la via, con l'uso delle tecniche di analisi convessa e dei teoremi di punto fisso, agli sviluppi che condurranno ai risultati degli anni cinquanta.
Questa fase (che è trattata in dettaglio in Ingrao e Israel, 1987, e in Weintraub, 1986) è tanto breve quanto decisiva per lo sviluppo della teoria e ha il suo culmine nel trattato scritto da von Neumann in collaborazione con Morgenstern (v. von Neumann e Morgenstern, 1944), in cui l'originale visione walrasiana viene rimpiazzata da una descrizione dei processi economici nel quadro della teoria dei giochi strategici. In tale impostazione l'adesione a un punto di vista modellistico astratto è la chiave che permetterà finalmente di tradurre il nucleo paradigmatico della teoria (le questioni dell'esistenza, dell'unicità e della stabilità dinamica dell'equilibrio) in una fruttuosa struttura matematica, ma implica anche un indebolimento dell'adesione al riduzionismo fisico-matematico classico. Ma soltanto in parte, perché più tardi la necessità di riferirsi - in assenza di altri validi riferimenti concettuali - al modello della fisica si presenterà ancora nel trattato di von Neumann e Morgenstern. L'asse appare tuttavia spostato dal riferimento alla fisica-matematica classica a quello allora dominante della fisica teorica, in pieno e impetuoso sviluppo.
La quarta fase può essere caratterizzata come quella in cui ha luogo la matematizzazione completa della teoria dell'equilibrio nella sua veste classica, entro un punto di vista assiomatico integrale di orientamento 'bourbakista': la sua più chiara manifestazione è la Theory of value di Debreu (v., 1959). Il primo periodo di questa fase si caratterizza per una prosecuzione del programma di uso sistematico della teoria dei giochi e dell'analisi convessa iniziato da von Neumann, che conduce, nel 1954, alla dimostrazione di un teorema molto generale di esistenza dell'equilibrio da parte di Kenneth Arrow e Debreu (entrambi nati nel 1921). Tuttavia, in questo lavoro di ricerca, il modello matematico viene man mano 'depurato' dei riferimenti alla teoria dei giochi attraverso un recupero dei concetti e del linguaggio della teoria walrasiana classica, sia pure entro il nuovo quadro formale e astratto. Tale

impostazione ha consentito non soltanto di rispondere in modo positivo al problema dell'esistenza dell'equilibrio, ma anche di mostrare le insormontabili difficoltà che esistono sulla via di una dimostrazione della sua unicità.
In un secondo periodo, inaugurato dal contributo di Stephen Smale (nato nel 1930), si è avuto un ritorno ai metodi del calcolo differenziale classico, sia pure nell'ottica della moderna topologia differenziale: questo approccio ha consentito di ridimostrare tutti i risultati ottenuti in precedenza. In questa cornice differenziale è stato inoltre affrontato il problema della stabilità dinamica (e in particolare quello del tâtonnement) mostrando che, anche in questo ambito, è impossibile conseguire sul piano analitico i risultati sperati: i processi di aggiustamento dei prezzi non sono quasi mai convergenti, se non sotto ipotesi molto restrittive e di significato economico molto particolare.Questi risultati conducono allo stato presente della teoria e ai difficili problemi che essa è costretta ad affrontare per individuare possibili vie di sviluppo.

3. Il modello di Arrow e Debreu: l'equilibrio intertemporale con informazione perfetta

Ci proponiamo ora di fornire una sintesi molto semplificata della struttura del modello matematico dell'equilibrio economico generale, seguendo l'impostazione di Arrow e Debreu. A nostro avviso è del tutto legittimo scegliere il modello di Arrow e Debreu come riferimento fondamentale per dare un'immagine della struttura formale della teoria dell'equilibrio e per valutare i suoi risultati. Se è vero che il modello di Arrow e Debreu è diverso dalla costruzione di Walras, non è meno vero che la logica interna dello sviluppo della teoria porta in modo coerente dall'impostazione di Walras a quella dell'equilibrio intertemporale di Arrow e Debreu. Il modello di Arrow e Debreu può essere considerato un punto di arrivo dello sviluppo storico della teoria, che ne consente la 'radiografia' e una piena valutazione della coerenza fra intenti e risultati effettivamente conseguiti (per maggiori dettagli, v. Debreu, 1959).
Nel modello dell'equilibrio generale di Arrow e Debreu hanno un ruolo centrale i concetti di equilibrio, merci e prezzi. Del primo si è già detto. Le merci sono l'oggetto dell'attività economica, che consiste nella produzione e nello scambio di beni. La funzione dei prezzi è di operare come indicatori che informano gli agenti della scarsità dei beni.
Una merce è un bene o un servizio completamente specificato dai punti di vista fisico,

spaziale e temporale, la cui quantità viene misurata mediante un numero reale e quindi con un punto x sulla retta dei numeri reali, IR. Se le merci presenti sul mercato sono in numero di n, un punto, o vettore, dello spazio cartesiano IRn delle n-ple di numeri reali indicherà la quantità in cui ciascuna delle merci è presente sul mercato. Tale vettore (x_1, ..., xn) sarà detto paniere di merci e può essere anche inteso come il piano d'azione di un agente economico, formulato ora per tutto l'orizzonte futuro (da cui il carattere deterministico del modello). IRn, ovvero lo spazio di tutte le possibili allocazioni (o panieri), sarà detto spazio delle merci.
A ogni merce è associato un numero reale, detto prezzo, che misura il suo 'valore' ovvero ciò che deve essere dato per ottenerla. Data la merce h, il numero ph indica il prezzo di un'unità di h, per cui data la quantità xh, il prezzo di xh è phxh. Dato un sistema di prezzi (p_1, ..., pn), il valore di un paniere di merci x=(x_1, ..., xn) è p_1x_1+...+pnxn=p·x.
Gli agenti economici si dividono in due categorie: i produttori e i consumatori. Il ruolo dei primi è di formulare un piano di produzione, mentre i secondi formulano un piano di consumo. Entrambi i piani sono espressi in termini di un paniere di merci; inoltre hanno validità illimitata, per cui la dinamica temporale è interamente conseguenza della scelta effettuata nel presente. Quest'ipotesi deterministica è giustificata dall'assunto che gli agenti abbiano una capacità di 'preveggenza illimitata'. L'altro postulato fondamentale è che gli agenti siano mossi dall'intento di realizzare il massimo vantaggio possibile, ovvero da un 'infinito egoismo'. L'attività di un consumatore è definita da due aspetti: le limitazioni fisiche e di mercato e i criteri di scelta del piano di consumo. Non tutti i piani sono, com'è ovvio, fisicamente possibili: quelli fisicamente possibili appartengono a un ben definito sottoinsieme di IRn, detto insieme di consumo. Non ci soffermeremo sulle ipotesi matematiche che permettono di caratterizzare gli insiemi di consumo.
La descrizione matematica dei criteri di scelta si consegue ammettendo che il consumatore i, di fronte a due panieri x e y, sia in grado di decidere quale preferisce. Si scriverà x≺y per indicare che i preferisce y a x, xĩy per indicare che x e y sono equivalenti per i e x≾iy per indicare che y è preferibile o equivalente a x per ĩ. È facile constatare che la relazione i è riflessiva e transitiva ma non antisimmetrica (come lo sono le relazioni d'ordine), mentre ĩ è una relazione d'equivalenza. Le classi di equivalenza rispetto alla relazione d'indifferenza ĩ si dicono classi d'indifferenza e curve d'indifferenza se si riducono a curve. Si possono fare numerose ipotesi circa le relazioni di preferenza. La più famosa è l'ipotesi di convessità forte, che si esprime dicendo che, se xĩy, allora tx+(1-t)y>ix(0<t<1), ovvero che, se x è indifferente

a y per i, allora questi preferisce a entrambi una media ponderata di essi. Non è difficile constatare che tale ipotesi implica il fatto che le curve d'indifferenza siano convesse, il che semplifica molte dimostrazioni.
L'ipotesi di 'infinito egoismo' si traduce asserendo che non sono possibili 'consumi di sazietà' per il consumatore, ovvero che non esistono panieri tali che non ne esistano altri ad essi preferibili.Passiamo ora all'introduzione dei vincoli di mercato. Si assume che all'inizio dell'attività economica il consumatore i possieda una certa 'ricchezza', espressa in termini di un paniere di merci ('risorse iniziali'): $\omega i=(\omega i_1, ..., \omega in)$. Il valore di ωi (o 'ricchezza' di i) è determinato dal vettore di prezzi p vigente sul mercato ed è quindi

formula.

Dato un vettore di prezzi p, l'agente i selezionerà un piano di consumo 'ammissibile', rappresentato da un vettore xi, il cui valore non potrà superare la sua ricchezza, cioè tale che $p \cdot xi \leq p \cdot \omega i = wi$.
L'insieme di queste scelte ammissibili $\gamma i(p, wi)=\{xi | p \cdot xi \leq wi\}$ è detto insieme di bilancio. L'iperpiano $p \cdot xi = wi$ è detto vincolo di bilancio di i. A causa dell'ipotesi di 'insaziabilità', i sceglierà un paniere appartenente a tale vincolo e all'insieme di preferenza 'migliore' per lui. In termini matematici questa scelta xi dovrà essere massimale rispetto alla relazione $\precsim i$. Resta così definita una corrispondenza ξi che associa al prezzo di mercato p e alla ricchezza wi di i un insieme $\xi i(p, wi)$ di piani di consumo, detto insieme di domanda. Tale corrispondenza non è in generale una funzione univoca: lo è se la relazione $\precsim i$ è fortemente convessa; ove si faccia a meno di questa ipotesi restrittiva è quindi necessario far ricorso alla complessa teoria delle funzioni multivoche.
Passiamo ora al settore della produzione. Come si è visto, il compito del produttore j sarà quello di formulare un piano di produzione espresso mediante un paniere di merci yj (vettore di IRn). La scelta dei piani di produzione è limitata da vincoli fisici e tecnologici. Dato un vettore di prezzi p e un piano di produzione $yj=(yj_1,...,yjn)$, la quantità

formula

è detta profitto di j e la denoteremo con $\pi j(p)$. Il produttore cercherà di ottenere il massimo

profitto possibile al dato prezzo p; resterà così definita una funzione (in generale multivoca) ηj che associa a ogni vettore dei prezzi p l'insieme $\eta j(p)$ delle produzioni che massimizzano il profitto di j e che sarà detta corrispondenza di produzione.
Possiamo ora dare la definizione formale di un'economia. Essa è definita dagli elementi seguenti.
1. Sono dati r consumatori e s produttori, cioè r+s agenti che operano sul mercato.
2. Per ogni consumatore i (i=1,...,r) sono definiti i vincoli fisici delle sue attività e la relazione di preferenza $\precsim i$.
3. Per ogni produttore j (j=1,...,s) sono definiti i vincoli tecnologici delle sue attività.
4. Per ogni consumatore i è dato l'insieme delle risorse iniziali $\omega i=(\omega i_1,...,\omega in)$.
Si noti ora che le risorse iniziali dell'intera economia in termini della merce t (t=1,...,n) sono date dalla somma

formula,

per cui il vettore $\omega=(\omega_1,...,\omega n)$ { IRn sarà il vettore delle risorse iniziali dell'economia. Uno stato dell'economia è definito da un piano di consumo completo per ogni consumatore e un piano di produzione completo per ogni produttore, cioè da una (r+s)-pla di vettori di IRn: $(x^1 ,..., xr, y^1 ,..., ys)=(x, y)$. Dato un siffatto stato dell'economia, il vettore

formula

è la domanda totale netta. Sottraendo le risorse iniziali ω del sistema da questo vettore, si ottiene un vettore che misura la differenza fra domanda e offerta in tutti i mercati dell'economia. Tale vettore viene detto eccesso di domanda: $z=x-y-\omega$.
Se z=O e quindi $x-y=\omega$, per ogni bene la domanda netta del mercato eguaglia le risorse iniziali totali; in questo caso lo stato (x, y) viene detto equilibrio di mercato.
Il contesto in cui viene di solito esaminato il problema dell'equilibrio è quello delle cosiddette economie con proprietà privata delle risorse. In esse il consumatore i possiede non soltanto ωi, ma anche una quota dei profitti dei produttori. La frazione posseduta da i del profitto πi di j è un numero compreso fra 0 e 1 che denoteremo con Θij (ovviamente formula). In corrispondenza del vettore dei prezzi p, la ricchezza di i è data dal valore delle

sue risorse iniziali e dalla sua quota di profitto, ovvero da

formula.

In un'economia con proprietà privata delle risorse, un equilibrio di mercato è una (r+s+1)-pla di punti di IRn,$(x^*, y^*, p^*)=(x^{(1)*}, x^{(2)*}, ..., x^{(r)*}, y^{(1)*}, ..., y^{(s)*}, p^*)$, tale che:
a) nell'insieme dei panieri x(i) il cui valore non supera il vincolo di bilancio, cioè tali che

formula

è massimale rispetto a $\precsim_i$;
b) y(j)* massimizza il profitto rispetto al prezzo p*;
c) $x^*-y^*=\omega.p^*$ sarà detto vettore di prezzi di equilibrio.
Dato il vettore dei prezzi p, i sceglie xi nell'insieme di domanda

formula.

In un'economia con proprietà privata delle risorse ξ_i dipende soltanto da p e quindi possiamo scrivere $\xi_i(p)$. La somma delle domande individuali fornisce la domanda totale

formula,

che definisce la corrispondenza di domanda totale

formula.

Analogamente si definisce la corrispondenza di produzione totale

formula,

tale che

formula.

Consideriamo ora l'insieme dei vettori ottenuti sottraendo l'offerta dalla domanda, ovvero $\xi(p)-\eta(p)- +\omega=\zeta(p)$. L'eccesso di domanda $z=x-y-\omega$ è un vettore di questo insieme $\zeta(p)$. Resta così definita la corrispondenza (in generale multivoca) $\zeta:\rightarrow\zeta(\pi)= \xi(p)-\eta(p)-\omega$, che viene detta corrispondenza di eccesso di domanda dell'economia.
Risolvere il problema dell'esistenza dell'equilibrio significa dimostrare che esiste almeno un vettore di prezzi di equilibrio p*, cioè un vettore p* tale che $\zeta(p^*)$ contenga il vettore nullo. In formule il problema è: esiste un p* tale che $0\{\zeta(p^*)$? Se esiste un siffatto p* cercheremo successivamente di vedere se esso è unico.
Il problema è assai più semplice se ζ è una funzione univoca, poiché allora $\zeta(p)$ si riduce a un solo elemento e si tratta di vedere se esiste un vettore di prezzi p* tale che $\zeta(p^*)=0$, ovvero se l'equazione $\zeta(p)=0$ ammette una soluzione, e se tale soluzione è unica.
Accenniamo ora al modo in cui si formalizza il problema della stabilità globale o dinamica dell'equilibrio, limitandoci a tradurre matematicamente il cosiddetto processo di tâtonnement. Si suppone che esista un agente esterno al mercato chiamato 'banditore' che ha le funzioni di informazione e di coordinamento. La prima funzione consiste nell'informare gli agenti economici dei prezzi delle merci, ovvero nell''annunciare' un vettore di prezzi $p^{(1)}=(p^{(1)}_1 ,..., p^{(1)}n)$. Questa informazione è la base per la determinazione dei piani di consumo e di produzione e quindi delle funzioni $\xi(p^{(1)})$ e $\eta(p^{(1)})$. La seconda funzione del banditore consiste nel calcolo dell'eccesso di domanda $\zeta(p^{(1)})$. Se $p^{(1)}$ è un vettore di prezzi di equilibrio, gli scambi e la produzione sono possibili; altrimenti il banditore 'griderà' un nuovo vettore di prezzi $p^{(2)}$. Il processo iterativo continuerà finché non sia stato raggiunto un vettore di prezzi di equilibrio. Si noti che il vettore di prezzi p(i) viene cambiato dal banditore secondo una legge che dovrà essere ben definita e che gli agenti economici potranno operare soltanto se p(i) è un vettore di prezzi di equilibrio.
La legge di aggiustamento dei prezzi scelta da Walras riflette la cosiddetta 'legge della domanda e dell'offerta': se l'eccesso di domanda di una merce k è positivo, il prezzo di k cresce, se è negativo, il prezzo di k decresce. Se ora consideriamo il vettore dei prezzi p come funzione $p(t)=(p_1(t), ..., pn(t))$ del tempo t, l'ipotesi che pk(t) sia crescente si esprime matematicamente dicendo che la derivata di pk(t) rispetto al tempo è positiva. In simboli:

formula.

Analogamente, se pk(t) è decrescente,

formula.

Il fatto che una crescita (o una decrescita) di pk(t) corrisponde alla positività (o alla negatività) dell'eccesso di domanda si può esprimere asserendo che p·k(t) è proporzionale a ζi(p), cioè: p·k(t)=α**x**ζk(p), dove ak è un numero reale positivo. Otteniamo così il sistema di equazioni differenziali che descrive il tâtonnement:

formula.

Le soluzioni di questo sistema sono date da un sistema di n funzioni: $p_1=p_1(t)$, $p_2=p_2(t)$, ..., pn=pn(t). I vettori di prezzi di equilibrio sono le soluzioni costanti.
Il problema della stabilità globale ha pertanto la seguente traduzione formale: dato un processo di tâtonnement, descritto dalle funzioni $p_1=p_1(t)$, ..., pn=pn(t), che non parta da uno stato di equilibrio, è vero oppure no che esso tende verso un vettore di prezzi di equilibrio? Se ciò è vero quale che sia il vettore di prezzi da cui ha inizio il tâtonnement, allora vuol dire che le forze operanti nel mercato sono capaci di guidare il sistema verso uno stato di compatibilità delle azioni indipendenti.

4. Una valutazione critica del modello di Arrow e Debreu: struttura, risultati analitici, interpretazioni, sviluppi recenti

Se si vuole dare una valutazione critica del modello di Arrow e Debreu e, più in generale, del programma di ricerca sull'equilibrio economico generale, occorre partire dalla considerazione che vi è un'unità sostanziale tra le questioni storiografiche, i problemi e i risultati matematici. Questa unità è stata troppo spesso trascurata o dimenticata sia nella storiografia del pensiero economico, sia nel dibattito teorico sull'equilibrio economico generale.
La storia della teoria illumina in modo penetrante le finalità che hanno guidato la sua costruzione ed è quindi un elemento essenziale per capire perché essa sia giunta a porsi

determinati problemi formali. La riflessione sui problemi matematici offre una chiave importante per valutare i successi e gli insuccessi della teoria, anche sulla base dei criteri interni di scientificità che il programma di ricerca sull'equilibrio economico generale si è dato nell'arco della sua lunga evoluzione.

Abbiamo sottolineato come il modello scientifico dei teorici dell'equilibrio generale sia stato quello della fisica matematica classica; questa osservazione apre molti problemi nuovi, sia nella storia che nella valutazione dello stato della teoria, i quali possono essere riassunti in una serie di quesiti che hanno trovato prime significative risposte in diverse opere citate nella bibliografia. Quali sono stati i paradigmi di riferimento presi a prestito dalle scienze fisico-matematiche e perché essi hanno avuto un ruolo così importante nella costruzione della struttura formale della teoria? Quali problemi si sono posti nel trasferimento di quei paradigmi dal contesto delle scienze fisico-matematiche a quello dei fenomeni di mercato e come si è tentato di affrontarli e risolverli? Quali sono stati i canali di comunicazione tra il mondo scientifico e gli studiosi impegnati sulla frontiera della matematizzazione della scienza sociale? (È fuor di dubbio, infatti, che la teoria dell'equilibrio economico generale sia stata, fino a un'epoca recentissima, la frontiera della matematizzazione della scienza sociale). In che modo si sono sviluppati questi canali di comunicazione e che ruolo ha avuto il dialogo con i fisici matematici nella costruzione della teoria? Infine in quale misura e in quali periodi i fisici matematici si sono rifiutati di dare un contributo al progetto di matematizzazione perseguito dai teorici dell'equilibrio economico generale?

Si tratta di quesiti che portano a riflettere sulle ambizioni e le conquiste, reali o presunte, della teoria dell'equilibrio economico generale e impongono di verificare se i risultati ottenuti nella lunga evoluzione della teoria siano stati effettivamente all'altezza delle ambizioni di 'scientificità' che ne hanno costantemente ispirato la costruzione. L'assiomatizzazione della teoria completata negli anni cinquanta ha svolto un ruolo fondamentale per quest'opera di chiarificazione, forse più di quanto non sia stato riconosciuto finora, e ciò conferma il ruolo centrale che ha avuto sul piano analitico il modello di Arrow e Debreu.

Il problema che è al cuore della valutazione della teoria può essere formulato come segue. Nell'opera di Walras e di Pareto prende forma un nucleo di problemi analitici che si conserva come nucleo centrale di tutte le successive versioni della teoria. Si tratta dei tre problemi dell'esistenza, dell'unicità e della stabilità globale dell'equilibrio, in un mercato di concorrenza perfetta dove gli agenti razionali siano price-takers e la compatibilità delle

decisioni indipendenti degli agenti sia assicurata da un sistema di prezzi intesi come segnali. La dimostrazione matematica dell'esistenza dell'equilibrio è la risposta formale al problema della coerenza e della compatibilità dei fini, dati i vincoli della tecnologia e delle risorse. La dimostrazione matematica dell'unicità dell'equilibrio è la condizione perché abbia significato il confronto tra diversi stati di equilibrio dell'economia al variare dei vincoli o delle preferenze (statica comparata). Infine la dimostrazione della stabilità globale ha come scopo quello di mostrare come, sia pure in condizioni ideali e astratte, il meccanismo di mercato guidato dalla 'mano invisibile' dei prezzi converga verso lo stato di equilibrio.

Anche al di là dei problemi interpretativi e di 'realismo' della teoria, assai dibattuti nel corso della sua storia, l'assiomatizzazione ha portato alla luce le gravissime difficoltà interne allo stesso nucleo centrale della teoria: i due problemi dell'unicità dell'equilibrio e della stabilità globale del tâtonnement non soltanto sono rimasti irrisolti, ma appaiono insolubili entro la struttura formale della teoria. L'assiomatizzazione ha avuto l'importante funzione di evidenziare questa difficoltà analitica. Restano da valutare le conseguenze di questo fallimento sull'intero programma di ricerca sull'equilibrio economico generale.

Non è possibile, entro i limiti di questo articolo, presentare una rassegna esauriente dei risultati ottenuti sul piano matematico: ci limiteremo ad alcuni cenni rinviando a Ingrao e Israel (v., 1987) e a Weintraub (v., 1986) per maggiori dettagli.

Come si è già detto, il teorema di esistenza dell'equilibrio è senz'altro il risultato più rilevante e significativo della teoria. L'esistenza dell'equilibrio economico fu infatti dimostrata da Arrow e Debreu (nel 1954) sotto ipotesi assai poco restrittive: per l'eccesso di domanda $\zeta(p)$ si richiede soltanto che sia una corrispondenza (e quindi non necessariamente univoca) semicontinua superiormente. Il teorema di Arrow e Debreu costituisce il culmine di una serie di ricerche iniziate da von Neumann e pone al centro l'uso dei teoremi di punto fisso. Numerose generalizzazioni di questo teorema furono ottenute in seguito (fra le quali va menzionata quella dovuta ad A. Mas Colell).

Quanto all'unicità dell'equilibrio, fu quasi subito evidente che era facile produrre esempi di economie con più di un equilibrio. Restava aperto il problema di determinare le condizioni sotto le quali l'unicità si verificava. La ricerca svolta a partire dagli anni cinquanta ha dimostrato che non si riesce a dimostrare l'unicità dell'equilibrio, così come non si riesce a dimostrare la convergenza del tâtonnement, se non in circostanze molto particolari come quella in cui tutte le merci sono 'sostituti'. In una serie di lavori iniziati nel 1976 Debreu

analizzò a fondo la questione dell'unicità, mostrando non soltanto che essa in generale non è verificata, ma anche che gli equilibri possono essere in numero infinito. L'unica limitazione significativa (ed è questo il principale risultato di Debreu al riguardo) è che l'insieme delle economie che hanno un numero discreto di equilibri è 'generico', ovvero che le economie sono 'quasi tutte' di questo tipo.

La questione della convergenza del tâtonnement fu affrontata agli inizi degli anni cinquanta con alterni risultati, fino a che, nel 1960, Herbert Scarf non produsse esempi di economie globalmente instabili. Questa scoperta arrestò le ricerche per un lungo periodo fino a che, negli anni settanta, Stephen Smale tentò di superare la difficoltà proponendo un processo di aggiustamento dei prezzi diverso dal tâtonnement, che costituisce una generalizzazione 'globale' del classico metodo di Newton per il calcolo degli zeri: tale processo è convergente, al prezzo tuttavia di alcune ipotesi ineliminabili (come una forte restrizione degli stati iniziali ammissibili del processo medesimo) e di una difficoltà di interpretazione economica. Analisi successive hanno mostrato che il processo di Smale non può essere migliorato e che esso resta soprattutto un efficace e interessante strumento di carattere puramente matematico per la determinazione numerica di equilibri di un campo vettoriale.

Le difficoltà inerenti alle questioni dell'unicità e della stabilità sono legate a un difetto di specificazione delle ipotesi di base del modello e della forma delle funzioni eccesso di domanda aggregate. In termini più precisi il problema centrale è che la procedura di formazione dell'eccesso di domanda attraverso l'aggregazione delle domande individuali (a loro volta derivanti dalla struttura delle relazioni di preferenza) non lascia traccia della struttura di tali funzioni (le domande individuali) nella forma della funzione eccesso di domanda aggregata, la quale si presenta come una funzione quasi del tutto generica. Ciò è stato mostrato da Hugo Sonnenschein nel 1974 con un teorema dal quale risulta che ogni funzione polinominale può essere pensata come funzione eccesso di domanda $\zeta(p)$ di un'economia e quindi (a norma di un classico teorema di approssimazione di Weierstrass) che ogni funzione continua è una funzione eccesso di domanda. Il teorema di Sonnenschein illustra meglio di qualsiasi altro risultato il vizio di fondo della teoria: quanto sia illusoria la speranza di conseguire risultati così stringenti come l'unicità e la stabilità nel caso di una funzione eccesso di domanda praticamente arbitraria.Il giudizio di sintesi è che la teoria dell'equilibrio economico generale giunge a una incoerenza tra le proprie finalità e i risultati analitici ottenuti: sulla base delle premesse del modello assiomatizzato neowalrasiano di

Arrow e Debreu non è possibile dedurre in modo rigoroso quei risultati (stabilità globale e unicità dell'equilibrio) che pure sono un elemento centrale del nucleo paradigmatico della teoria.

È certamente legittimo chiedersi se la teoria dell'equilibrio economico generale possa fare a meno dei risultati di stabilità globale e unicità e, in particolare, se il tâtonnement sia davvero così importante per essa o sia soltanto un espediente di Walras, di cui è possibile sbarazzarsi senza troppe angustie. È tuttavia evidente come il fatto che i casi in cui l'equilibrio non è unico siano la quasi totalità sia di per sé un ostacolo grave a ogni tentativo applicativo e, più in generale, a ogni tentativo di ricavare dalla teoria dei significativi teoremi di statica comparata. È inoltre agevole constatare che non è affatto facile espungere la tematica della stabilità globale dal quadro walrasiano e neppure dal quadro assiomatizzato neowalrasiano. Espungere questa tematica significherebbe riproporre una lettura totalmente normativa della teoria dell'equilibrio economico generale: ciò è stato di fatto tentato e il filone algoritmico inaugurato da Scarf, che studia il calcolo degli equilibri in modelli di equilibrio applicati, elaborati per condurre valutazioni di politica economica (v. Scarf, 1973), si è mosso proprio in questa direzione. Questo nuovo indirizzo della ricerca è una riscoperta dell'anima normativa della teoria, che lascia tuttavia irrisolta la conciliazione con le finalità descrittive della teoria. La grande difficoltà è il paradosso di una teoria normativa dell'economia di mercato, che rinuncia alla dimostrazione dell'efficacia spontanea del libero mercato e giunge così a una equivalenza con la teoria della pianificazione. Anche una teoria normativa dell'equilibrio, se non vuol rinunciare a essere una teoria del mercato oltre che una teoria dell'equilibrio, non può prescindere dalla decentralizzazione delle decisioni e delle scelte e quindi da un meccanismo spontaneo che nello scambio guidi le decisioni indipendenti a convergere verso la compatibilità nello stato di equilibrio. È certamente legittimo criticare l'inadeguatezza della descrizione dei processi di aggiustamento dei prezzi mediante il tâtonnement; ma, pur prescindendo dal fatto che la ricerca di altri tipi di processi aventi caratteristiche di convergenza ha dato risultati assai modesti e pochissime speranze di conseguire miglioramenti, è paradossale pensare che si possa costruire una teoria normativa del libero mercato senza libero mercato e cioè sul solo presupposto di un algoritmo di calcolo scelto per motivi di convenienza matematica dagli specialisti del settore. In conclusione la convivenza precaria tra l'anima normativa e l'anima descrittiva attraversa tutta la storia della teoria dell'equilibrio economico generale, fino ai problemi aperti sulla frontiera della ricerca

attuale.
Le considerazioni precedenti conducono al quesito se il quadro assiomatico classico della teoria (e in particolare la sua espressione più compiuta, ovvero il modello di Arrow e Debreu) debba ancora essere considerato centrale oppure si possa configurare un suo abbandono, nella speranza di non incontrare, in nuovi contesti, analoghi problemi di coerenza tra premesse e risultati analitici, o addirittura di risolvere i problemi di indeterminatezza e di povertà di risultati che sorgono entro il quadro classico. Tuttavia, quando si esce dal quadro delle ipotesi alla Arrow-Debreu, si apre un vaso di Pandora: i problemi delle aspettative, dell'incertezza e del tempo, liberati dalla costrizione forzosa entro lo schema artificioso della perfetta preveggenza e dell'economia di mercati futuri completi, generano scompiglio, distruggono le pretese di imitazione dei modelli fisici e obbligano a rimettere in discussione radicalmente la metodologia stessa dell'equilibrio.
La letteratura recente ha sottoposto a revisione l'una o l'altra delle ipotesi nel tentativo di dilatare la struttura di base del modello assiomatico, abbandonando i tratti più artificiosi e irrealistici della teoria. Tre indirizzi di ricerca hanno perseguito questo obiettivo, accantonando i requisiti della stabilità e dell'unicità dell'equilibrio, come se tali requisiti non fossero essenziali al significato e alla validità della teoria, e rinunciando persino alla pretesa di pervenire alla dimostrazione di teoremi di esistenza sotto condizioni generali.Il primo indirizzo di ricerca ha prodotto modelli di equilibrio generale abbandonando l'ipotesi della concorrenza perfetta nell'ambito della produzione, vale a dire l'ipotesi classica che i produttori assumano come dati i prezzi di mercato. Come abbiamo visto, tale ipotesi comporta l'artificiosa presenza di un 'banditore', poiché nessuna delle imprese presenti sul mercato ha la capacità di definire i prezzi di vendita. I modelli di equilibrio con concorrenza monopolistica impostano nel contesto dell'equilibrio generale il problema della concorrenza monopolistica definito dall'economista statunitense Edward Hastings Chamberlin (v., 1933). La concorrenza monopolistica è quella forma di mercato che si ha quando numerose imprese di piccole dimensioni, in condizioni di libertà di entrata sul mercato, offrono prodotti differenziati che sono sostituti nelle funzioni di domanda dei consumatori.
Molte difficoltà sorgono nel trasferire questo impianto entro un modello alla Arrow-Debreu. Come definire la libertà di entrata, se imprese e beni sono in numero dato? Come inserire allora la differenziazione dei prodotti e la loro sostituibilità? Come far intervenire le curve di domanda dei prodotti differenziati nelle scelte di ciascuna impresa?Nei modelli di equilibrio

generale con concorrenza monopolistica elaborati nella letteratura recente le imprese sono in numero dato, ma hanno il potere di fissare il prezzo del proprio prodotto, tenendo conto della curva di domanda (la relazione tra prezzo e quantità domandata) per ciascuna impresa. In una classe di modelli le imprese effettuano le proprie scelte congetturando la curva di domanda che limita il loro operare sul mercato, sotto la condizione aggiuntiva, tuttavia, che le loro congetture siano corrette almeno in corrispondenza della soluzione di equilibrio, vale a dire che le curve di domanda soggettivamente percepite come vincoli dalle imprese passino per il punto di equilibrio quale risulta dalla soluzione del modello.
Questa impostazione è stata proposta da Negishi (v., 1961) e ripresa dallo stesso Arrow (v., 1971) in un celebre articolo sul ruolo dell'impresa nei modelli dell'equilibrio economico generale. È possibile dimostrare teoremi di esistenza dell'equilibrio anche per questi modelli, ma sotto ipotesi più restrittive di quelle del modello dell'equilibrio generale di concorrenza perfetta.
Le curve di domanda così inserite nel modello sono arbitrarie, perché rappresentano le aspettative soggettive delle singole imprese, senza una teoria che si proponga di darne conto. Sulla spiegazione teorica prevale allora la convenienza analitica: così, ad esempio, alle funzioni che rappresentano le aspettative delle imprese circa la relazione tra prezzi e quantità domandate sul mercato del loro prodotto si attribuisce a priori una certa forma per facilitare la ricerca di soluzioni. Per rimuovere tale arbitrarietà, in altre famiglie di modelli si è seguita la via di costruire curve di domanda oggettive, che riflettono non ipotesi arbitrarie formulate dalle imprese, ma le relazioni tra prezzi e quantità sui mercati descritti dalla teoria. Si immagina allora che ciascuna impresa compia le scelte di produzione (con l'obiettivo del massimo profitto) assumendo come date le scelte delle altre imprese e prevedendo esattamente gli effetti sul vettore dei prezzi di mercato di tutte le scelte di offerta. In questo contesto la dimostrazione dell'esistenza dell'equilibrio è però problematica e sono stati presentati controesempi di modelli che non ammettono soluzioni (v. Roberts e Sonnenschein, 1977).
Nei modelli di cui si è detto, come nell'originario modello di Chamberlin, non c'è interazione strategica tra le imprese: le imprese non scontano le reazioni delle imprese concorrenti alle decisioni assunte su prezzi o quantità prodotte, ma si limitano a prendere atto delle scelte effettuate dalle altre imprese per quel tanto che tali scelte si rivelano nei dati aggregati del mercato. In un'altra classe di modelli, al contrario, si esamina il problema della concorrenza

imperfetta a partire dalla teoria dei giochi, strumento teorico per rappresentare le situazioni di interazione strategica, nelle quali ciascun soggetto che assume decisioni è consapevole della capacità degli 'avversari' di rispondere con mosse razionali alle sue strategie di gioco.
Un secondo indirizzo di ricerca che si muove nella scia della definizione di concorrenza perfetta elaborata da Cournot mira a dimostrare che l'equilibrio di una economia di molti produttori tende al limite all'equilibrio walrasiano (v. Hildenbrand, 1973).
In sintesi, l'abbandono dell'ipotesi della concorrenza perfetta apre la difficile alternativa tra modelli che precipitano in rappresentazioni arbitrarie delle aspettative e modelli che, costruiti su premesse di coerenza e oggettività più solide, non possono però raggiungere risultati analitici forti per la complessità delle interazioni possibili tra gli agenti. Della stessa natura, ma molto maggiori, sono le difficoltà che si incontrano quando si tenta di adattare lo schema dell'equilibrio generale alla Arrow-Debreu per rimuovere le fantasiose ipotesi sulla struttura degli scambi nel tempo. I tentativi in questa direzione, fondati sulla ripresa del concetto di equilibrio temporaneo, sono approdati a sofisticati modelli di economie sequenziali con previsione perfetta.
Questo terzo filone di ricerca è stato inaugurato negli anni settanta soprattutto da J.M. Grandmont e poi è stato sviluppato dallo stesso Grandmont e da numerosi altri studiosi, fra cui vanno segnalati Y. Younès, F. Fuchs, G. Laroque, J.P. Benassy (per un'ampia rassegna, v. Grandmont, 1988).
I modelli di equilibrio temporaneo più esplorati nella letteratura sono modelli a generazioni sovrapposte: non si considera più una data finale degli scambi, un orizzonte finito dei piani e delle transazioni progettate e decise nell'istante iniziale, ma un orizzonte infinito, perché nei mercati si succedono, sovrapponendosi, generazioni di agenti economici, ciascuna delle quali ha un orizzonte di vita finito. Poiché gli scambi continuano in un futuro indefinito (ogni generazione muore, ma lascia spazio alla generazione divenuta adulta, che a sua volta convive con la generazione più giovane), i mercati hanno un numero infinito di merci e di agenti. Infinite sono le date alle quali potranno essere disponibili i diversi beni.
Una delle motivazioni principali di queste ricerche è legata alla critica dell'ipotesi che gli agenti economici siano illimitatamente preveggenti e del carattere statico della 'dinamica' che essa comporta. Per superare questo inconveniente si postula che i mercati si succedano nel tempo e che gli agenti economici apprendano soltanto in modo graduale le leggi dinamiche dell'economia in cui operano. Per dare un'idea di tale approccio, consideriamo un caso

elementare. In coerenza con quanto detto, supporremo che l'equilibrio al tempo t sia completamente determinato dalle previsioni fatte, allo stesso tempo t, dagli m agenti economici, circa lo stato in cui l'economia si troverà nell'immediato futuro t+1: indichiamo con $\chi i,e\ t+_1$ tali previsioni per l'agente i-esimo. Allora l'equilibrio sarà dato dall'equazione funzionale dove f descrive il risultato del processo di aggiustamento verso l'equilibrio del mercato per un dato sistema di aspettative.
L'equilibrio intertemporale è allora definito come una sequenza di stati ⟨xt> e di previsioni ⟨xi, $et+_1$ > tali che sia soddisfatta l'equazione dell'equilibrio (*) e si abbia xi, $et+_1=xt+_1$ per ogni i e ogni t.
L'informazione di ogni agente al tempo t è descritta dalla sequenza (xt, $xt-_1$, ...). Quindi le relazioni di previsione sono date da: dove ψi sono le funzioni di aspettativa. Le equazioni dell'equilibrio di mercato si scriveranno allora: La dinamica di un'economia è così descritta, date le f e le ψi, come una sequenza di equilibri temporanei che hanno luogo ricorsivamente nel tempo.
L'approccio degli equilibri temporanei è stato applicato anche al di fuori della cornice classica, abbandonando l'ipotesi di piena occupazione, e ha permesso di toccare questioni come quella della struttura dei modelli macroeconomici keynesiani della disoccupazione o quella dei fondamenti microeconomici della teoria monetaria. In alcuni modelli di equilibrio temporaneo si è tentato di inserire la moneta e di analizzarne il ruolo come mezzo di scambio. Inoltre la teoria si è sviluppata sia nel quadro deterministico che in quello stocastico.
Problemi centrali di questa teoria, sul piano analitico, sono la dimostrazione dell'esistenza dell'equilibrio nel breve periodo e della stabilità del medesimo sul lungo periodo. Occorre tuttavia osservare che, probabilmente a causa della sua notevole complessità matematica, questa teoria si presenta con caratteristiche molto più frammentate di quella classica ed è assai meno accentrata attorno a risultati generali. Lo stesso Grandmont ha recentemente osservato (v. Grandmont, 1988) che molto si deve ancora fare per ottenere in questo contesto formulazioni e risultati generali e compatti. Particolarmente critico è ancora il livello dei risultati ottenuti attorno al problema centrale della stabilità dell'equilibrio di lungo periodo. Al di là dei risultati formali conseguiti e ancora aperti, non appare risolto e neppure affrontato in termini davvero innovativi il problema di costruire dei modelli che contengano una descrizione effettivamente dinamica dei processi di mercato, come gli stessi autori

ammettono. In conclusione, ancora una volta, il problema che ha ispirato questo sviluppo si ripresenta in forma sostanzialmente irrisolta.In tema di dinamica, un dibattito si è sviluppato sul modello walrasiano con accumulazione di capitale, quel modello che secondo gli intenti di Walras doveva estendere la teoria dell'equilibrio generale al caso di mercati dove gli agenti risparmiano e le imprese offrono nuovi beni capitali, cioè a una rappresentazione schematica del processo di accumulazione e crescita in condizioni di concorrenza perfetta. Questo modello walrasiano, che ipotizza un saggio di rendimento uniforme per tutti i beni capitali di nuova produzione, ammette soluzioni solo sotto l'ipotesi che in equilibrio la produzione di alcuni beni capitali possa essere nulla. Questo risultato formale ha suscitato controverse interpretazioni. Secondo alcuni studiosi, ciò comporta una radicale inadeguatezza del modello walrasiano con accumulazione di capitale, che non potrebbe dar conto in modo coerente di un equilibrio con accumulazione di capitale nel lungo periodo, perché non garantirebbe simultaneamente la condizione di uniformità del saggio di profitto e la riproduzione di tutti i beni capitali. Secondo altri interpreti, tuttavia, non era intento di Walras comprimere il processo di accumulazione entro lo schema di un equilibrio di lungo periodo. Questi interpreti, più attenti alla lettura del testo walrasiano, sottolineano piuttosto come, pur nelle residue ambiguità del modello, la struttura proposta da Walras anticipi un modello di equilibrio temporaneo, una successione di equilibri nel tempo con ipotesi molto semplici sulla formazione delle aspettative. La difficoltà maggiore della originaria teoria walrasiana in tema di dinamica resta invece l'impossibilità di dar conto del mutamento dei dati (preferenze, tecnologia, risorse) e di conciliare, in un'unica struttura teorica, questi mutamenti di struttura con i processi di aggiustamento dei prezzi.

5. Conclusioni

La ricca e tormentata storia della teoria dell'equilibrio economico generale conduce ad alcune significative riflessioni.La prima riflessione riguarda l'enorme ambizione e forse l'ingenuità con cui è stato affrontato il problema dell'equilibrio generale del mercato. In un carteggio con Pietro Verri, nel 1773, Condorcet metteva in guardia Verri dal sottovalutare le difficoltà della formalizzazione nella scienza sociale, osservando che le maggiori menti matematiche dell'epoca non erano state in grado di risolvere il 'problema dei tre corpi', e commentava: "Essi suppongono che questi corpi siano delle masse senza estensione [...] e

questo problema, benché limitato da cento condizioni che lo facilitano, li ha occupati da venticinque anni e li occupa ancora. L'effetto delle forze che agiscono nella testa del commerciante, anche il più stupido, è ben più difficile da calcolare". Una problematica analoga, anche se affrontata in un'ottica diversa, emerse nel carteggio tra Walras e Poincaré: Poincaré sosteneva che fosse lecito assumere come punto di partenza una grandezza non misurabile (l'utilità espressa con una funzione arbitraria), ma non certo che una teoria costituita su queste basi fosse soddisfacente. Pur senza approfondirlo, Poincaré aveva toccato con grande acume un problema cruciale della teoria: il difetto di specificazione dovuto all'elevato numero di funzioni arbitrarie incluse nelle sue premesse.Tutte queste difficoltà acquistano un pieno significato se sono viste nel contesto del problema generale delle difficoltà che insorgono, in forme ormai ben note, quando si tenta di matematizzare dei campi di fenomeni ben più complessi di quelli fisici, quali sono i fenomeni biologici e sociali.

Un secondo ordine di riflessioni investe l'ipotesi che è a fondamento della descrizione del comportamento degli agenti economici nei modelli di equilibrio: l'idea che gli agenti economici sono razionali in un senso molto particolare, cioè sono capaci di scegliere in modo coerente secondo procedure di ottimizzazione, dati i segnali che ricevono dal mercato. Questa visione della razionalità individuale del consumatore è controversa e criticabile, poiché attribuisce agli individui capacità di calcolo esagerate, mentre omette di rappresentare quella più autentica razionalità, essenziale anche nell'attività economica, che è la capacità inventiva e più in generale la capacità di interagire in modo creativo con l'ambiente. Gli agenti dei modelli di equilibrio economico generale sono degli "idioti razionali", come li ha spiritosamente definiti l'economista Amartya Kumar Sen. I risultati della storia della teoria dell'equilibrio economico generale mostrano che l'ipotesi della razionalità individuale del consumatore, già discutibile dal punto di vista del realismo, è anche povera: è povera di implicazioni e produce scarsi risultati in termini di ipotesi falsificabili e di proposizioni determinate.

Infine una riflessione è stata avviata e dovrà essere approfondita sulla 'visione' del mercato che la teoria dell'equilibrio economico generale ci trasmette. Walras scelse il mercato d'asta come tipo reale da cui astrarre il tipo ideale della concorrenza perfetta. Questa scelta è stata densa di implicazioni: ha abituato l'economista teorico a pensare il mercato tout court come un mercato d'asta e, peraltro, come un mercato d'asta improprio e fittizio, perché perfetto dal

punto di vista dell'informazione e collocato in un tempo algoritmico astratto che non è quello delle decisioni reali prese sul mercato; ha abituato altresì l'economista teorico a pensare il meccanismo di mercato come riducibile a un meccanismo dei prezzi elementare e trasparente. Nei mercati reali, fuori dalle finzioni della teoria walrasiana, l'armonizzazione e la compatibilità dei fini individuali sul mercato sono raggiunte anche grazie all'operare del meccanismo dei prezzi, ma entro un sistema di istituzioni, di regole, di relazioni di fiducia, di forme e attività organizzative, che non assomiglia affatto al banditore walrasiano.
Il 'libero mercato' è un insieme di istituzioni in cui funzionano meccanismi allocativi e distributivi complessi, assai più ricchi e articolati, indipendentemente dalla loro maggiore o minore efficacia, di quelli descritti dalla teoria dell'equilibrio economico generale.
Ciò che la teoria dell'equilibrio generale ha eliminato dall'analisi della forma ideale del mercato di concorrenza, in quanto 'accidente' irrilevante o in quanto ostacolo al paradigma della razionalità, è proprio ciò che consente di dare un contenuto positivo di conoscenza alla ricerca economica. Gli 'attriti', le specificità storiche e istituzionali dei mercati, le innovazioni tecnologiche, la storia monetaria o industriale, i processi di mutamento sono il materiale indispensabile che fornisce contenuto e determinatezza alla disciplina e le dà una ragione d'essere a fianco di altre forme di riflessione sulle regole della vita associata.

Cicli economici
di Terenzio Cozzi

sommario: 1. Introduzione. 2. Misurazione e caratteristiche dei cicli: a) gli indicatori; b) la separazione di ciclo e trend; c) cicli di diversa lunghezza; d) caratteristiche dei cicli. 3. Le prime spiegazioni teoriche del ciclo. 4. Le teorie monetarie del ciclo: a) le formulazioni del periodo tra le due guerre; b) il monetarismo attuale. 5. Interazione tra moltiplicatore e acceleratore: a) il modello originario; b) le teorie non lineari: soffitti e pavimenti; c) i modelli econometrici disaggregati. 6. Teorie dello sviluppo ciclico: a) la teoria di Schumpeter; b) le teorie di derivazione marxista; c) una digressione sul ciclo politico; d) i modelli preda-predatore. □ Bibliografia. Introduzione

"Analizzare i cicli economici vuol dire né più né meno che analizzare il processo economico nell'era capitalistica". Così inizia la prefazione a Business cycles, la monumentale opera di J.A. Schumpeter. Vi sono sinteticamente espresse due idee fondamentali: la prima, che le teorie del ciclo riguardano i sistemi economici capitalistici; la seconda, che lo sviluppo di questi sistemi non procede in modo uniforme, ma è intrinsecamente caratterizzato da oscillazioni.

Alti e bassi dell'attività produttiva ci sono sempre stati. Basta pensare ai sette anni di vacche grasse e sette di vacche magre di cui parla la Bibbia. Ma le oscillazioni dei sistemi precapitalistici trovavano la loro causa in fattori esterni alla sfera economica, quali la variabilità dei raccolti, le conseguenze di guerre o di epidemie, ecc. Non si parlava allora di ciclo, ma di crisi e, più precisamente, delle cause esogene delle crisi.

L'espansione dell'area del mercato capitalistico e i cambiamenti socioeconomici e istituzionali che ne sono derivati hanno avuto l'effetto di modificare significativamente le caratteristiche delle oscillazioni, il loro modo di propagarsi, i comportamenti degli operatori, ecc. L'evoluzione ciclica è risultata determinata in misura sempre più accentuata da fattori endogeni e da meccanismi di propagazione peculiari del sistema capitalistico. Si è modificata in modo essenziale la natura dei processi ciclici, che diventano così un fenomeno tipico del capitalismo, del suo modo di operare, del funzionamento delle sue istituzioni, specialmente quelle finanziarie. Schumpeter, ad esempio, ritiene che l'impronta di tipo nuovo al fenomeno delle oscillazioni venga data in larga misura proprio dallo sviluppo del sistema bancario e dalla sua capacità di creare credito e, di conseguenza, considera l'affermarsi di questi aspetti come l'atto di nascita del capitalismo.

La seconda idea, quella che per i sistemi capitalistici l'andamento ciclico rappresenti una caratteristica non accidentale, ma distintiva e sostanziale, può essere intesa in due modi. Il primo, più comune e meno impegnativo, è che normalmente lo sviluppo di questi sistemi avviene in modo oscillante. Il secondo, molto più pregnante, è che l'andamento ciclico costituisce l'essenza stessa del processo di sviluppo capitalistico, ne è elemento costitutivo e ineliminabile.

Con la prima interpretazione si apre la strada a una netta separazione tra teoria del ciclo e teoria dello sviluppo. Ne deriva, come si vedrà, una semplificazione molto rilevante della trattazione analitica, che dà ragione del perché la grande maggioranza degli studiosi dei cicli si sia mossa su tale terreno.

Per la seconda interpretazione sono le stesse cause a determinare ciclo e sviluppo. Di conseguenza, ogni separazione concettuale o analitica non può che impoverire e distorcere la comprensione dell'unico fenomeno che si manifesta realmente: lo sviluppo ciclico. A Marx e Schumpeter va il merito di essere stati i pionieri di questa linea di pensiero estremamente ricca e affascinante, ma anche molto impegnativa sul terreno analitico.La trattazione di modelli di sviluppo ciclico richiede l'uso di complesse tecniche matematiche che permettano di analizzare le modificazioni di natura strutturale in stretta integrazione con quelle di natura congiunturale. Le prime, quelle che caratterizzano i processi di sviluppo, si manifestano spesso (non però sempre) in modo lento ma continuo; producono variazioni singolarmente piccole, che vanno per lo più nella stessa direzione e che, cumulandosi, diventano sempre più rilevanti. Le seconde, quelle tipiche degli andamenti ciclici, derivano da movimenti molto più veloci e sono quindi in grado di esercitare una robusta influenza sulla dinamica di breve periodo, ma possono con la stessa rapidità cambiare direzione.

L'analisi integrata di questi movimenti difformi risulta ulteriormente complicata dalla necessità di tenere in considerazione non solo gli andamenti delle variabili strettamente economiche, ma anche quelli delle variabili sociali, politiche e istituzionali. Le interrelazioni che si stabiliscono tra tutte queste variabili non possono infatti essere trascurate quando si voglia superare l'orizzonte del breve periodo. Lo studio dello sviluppo ciclico si amplia così fino a comprendere quello dell'intera evoluzione del sistema e in definitiva, come avviene infatti nelle trattazioni di Marx e Schumpeter, quello del destino futuro del capitalismo.
Misurazione e caratteristiche dei cicli

Gli indicatoriFinora si è parlato di cicli come alti e bassi dell'attività produttiva, come periodi di espansione e di depressione, ecc. Per rendere un po' più preciso il discorso, è necessario dire qualche cosa sugli indicatori dello stato di maggiore o minore prosperità del sistema economico. L'indicatore più usato è dato dall'andamento del prodotto interno lordo (PIL) in

termini reali, che misura nel modo più sintetico le fluttuazioni del complesso dell'attività produttiva alle diverse date. Ci sono anche molti altri indicatori di fenomeni parziali, ma di importanza anche notevole, quali l'andamento delle produzioni dei diversi settori, dei livelli occupazionali, degli investimenti, degli aggregati monetari e creditizi, delle domande di particolari prodotti, dei consumi di materie prime, dei prezzi, dei profitti, ecc.

Gli andamenti dei diversi indicatori non sono ovviamente identici, ma, in molti casi, sono abbastanza simili. Quanto meno, c'è una notevole concordanza di diversi indicatori nel segnalare - però con diversa tempestività - i periodi di accelerazione dell'attività produttiva e quelli di rallentamento.

Si deve anche rilevare che, considerando solo il PIL, si rischia di non porre in sufficiente evidenza le modificazioni di natura strutturale, che non sempre riescono a risaltare nell'andamento di un grande aggregato. Lo studio di queste modificazioni può però avere importanza cruciale per spiegare l'interrelazione tra ciclo e sviluppo.

Dal punto di vista empirico ci sono anche altre ragioni che consigliano l'utilizzazione di più serie storiche, anzi di un buon numero di esse. Il National Bureau of Economic Research (NBER) fin dagli anni venti ha sviluppato un metodo di analisi dei cicli che tiene simultaneamente in considerazione numerose serie. In ogni mese o trimestre un certo numero di queste segnalano espansione, altre contrazione. Si può avere perciò una, sia pur rozza, misura dell'intensità dell'espansione o della contrazione calcolando la percentuale delle serie che mostrano l'uno o l'altro andamento. Inoltre, è possibile notare che alcune serie tendono normalmente ad anticipare l'andamento di molte altre. Dalla loro osservazione si possono perciò trarre informazioni utili a scopo previsivo.

La separazione di ciclo e trendL'andamento delle serie storiche è normalmente molto accidentato; non è facile riconoscervi regolarità cicliche. Occorre un'opera di pulizia dei dati grezzi, innanzitutto per eliminare quelle componenti cicliche che hanno una spiegazione di tipo puramente stagionale (ad esempio la flessione della produzione industriale in agosto) o accidentale (ad esempio variazioni delle produzioni mensili causate dal diverso numero di giorni lavorativi), e poi anche per eliminare andamenti particolari che necessitano di una

spiegazione ad hoc, in quanto sono determinati da eventi importanti, ma unici, come gli effetti di una guerra, di una calamità naturale, ecc.

Una volta eseguite queste operazioni, rimane da affrontare il problema principale: quello del rapporto tra ciclo e crescita, due fenomeni che coesistono nella rappresentazione che ci forniscono i dati empirici. Come già detto, la grande maggioranza degli economisti ha ritenuto di risolvere il problema effettuando una netta separazione dei due fenomeni.

I metodi statistici utilizzati allo scopo differiscono nei particolari, ma nella sostanza consistono tutti nella valutazione di un andamento tendenziale (il trend) che dovrebbe dar conto della crescita, e nel calcolo degli scostamenti che la serie storica considerata mostra nei confronti del trend. Un esempio, con trend lineare, è riportato nella fig. 1. Come si vede, un andamento sempre sostanzialmente crescente viene scomposto in due componenti: una che mostra una crescita uniforme per un lungo arco di tempo e l'altra che invece oscilla assumendo valori positivi o negativi a seconda che la variabile considerata sorpassi o meno l'andamento tendenziale di lungo periodo.

Si è così aperta la strada per una spiegazione separata dell'andamento delle due componenti che rappresentano rispettivamente la crescita e il ciclo. Ma questo è proprio ciò che Schumpeter e Marx avevano negato fosse lecito fare. Si può per di più osservare che, con riferimento ai due autori, si era parlato di sviluppo e non di crescita, e che i due termini non sono perfetti sinonimi. Il significato di crescita si esaurisce infatti in quello di espansione quantitativa. Il significato di sviluppo è invece più ampio, riguarda anche modificazioni di tipo qualitativo e cambiamenti di natura strutturale, che possono trascendere gli aspetti puramente economici.

La separazione del trend dal ciclo dà luogo anche ad altre perplessità. Innanzitutto, bisognerebbe giustificarla teoricamente. Quasi sempre questo avviene solo in modo implicito, assumendo cioè che il trend rappresenti l'andamento di equilibrio del sistema e che gli scostamenti ciclici siano dovuti a squilibri. Ma, per accettare questa impostazione, è necessaria la preventiva dimostrazione che la crescita in equilibrio è possibile e che il meccanismo che la determina non viene a modificarsi quando il sistema si trova in condizioni

di squilibrio. In assenza di una dimostrazione di questo tipo, la determinazione del trend non può non contenere elementi di arbitrarietà.

Inoltre, il calcolo del trend viene di necessità fatto sulla base dei dati del passato. Ma il presente e il futuro possono avere caratteristiche molto diverse da quelle del passato, specialmente se si sono verificate modificazioni di carattere strutturale, il che è tanto più probabile quanto più lungo è il periodo di osservazione dei dati. Il trend può essersi modificato in tempi recenti, ma ciò può non apparire dal calcolo che risente di molti dati relativi a periodi più lontani. D'altra parte, se si considerano solo quelli più recenti, non è possibile calcolare un trend che abbia significatività statistica.

A fronte delle perplessità indicate, rimane tuttavia il fatto che la separazione di ciclo e trend offre un vantaggio di semplificazione e uno di misurazione. Esaminando il profilo degli scostamenti diventa infatti facile individuare le fasi con cui tradizionalmente si descrive il ciclo: ripresa, boom, recessione, depressione, successiva ripresa, ecc. E diventa anche facile calcolare la durata del ciclo come lunghezza del periodo che intercorre tra i punti più alti (i picchi) di due boom successivi (t_0-t_1) oppure tra i punti più bassi (le valli) di due depressioni successive (τ_0 τ_1), come illustra la fig. 2.

Cicli di diversa lunghezzaLa parola ciclo evoca, oltre al concetto di oscillazione che si ripete, anche quello di regolarità in un duplice senso: quello di intensità pressoché uguale dei vari cicli e quello di una loro durata non molto diversa. Gli studi empirici hanno messo in discussione entrambi i modi di intendere il concetto di regolarità. L'ampiezza delle fasi espansive e depressive dei cicli storici non ha avuto sempre dimensioni comparabili. E anche la loro durata è risultata molto variabile. Di conseguenza, alcuni teorici moderni hanno cominciato a dubitare dell'esistenza stessa di fenomeni che possano essere propriamente chiamati ciclici.

Tuttavia, non andando troppo per il sottile nella definizione delle regolarità cicliche, si è diffusa l'opinione - suffragata dall'autorità di Schumpeter - che sia possibile individuare su base empirica almeno tre tipi di cicli di durata ben diversa. Il primo è conosciuto come ciclo di Juglar, dal nome del francese Clément Juglar che lo individuò nel 1860. Egli ne aveva

fissato la durata in 10 anni per le tre fasi di "prosperità, crisi e liquidazione". Indagini successive hanno più approssimativamente fissato la durata tra 7 e 11 anni e hanno ripartito l'intero ciclo nelle 4 fasi indicate sopra.Sono stati poi individuati dei cicli di durata molto più breve che venivano a modificare, con alti e bassi, il profilo dei cicli di Juglar. Questi cicli brevi sono stati individuati dall'americano Joseph Kitchin nel 1923 e sono conosciuti con il suo nome. Hanno una durata non superiore a 40 mesi e, poiché sembrano mostrare un andamento sostanzialmente simile a quello delle scorte di prodotti finiti tenute dagli operatori, sono talvolta chiamati anche cicli delle scorte.

Un altro tipo di ciclo è quello delle 'onde lunghe' individuato nel 1922 dal russo Nikolaj D. Kondrat´ev. Il ciclo ha una durata di circa 50 anni ed è costituito da una fase ascendente e una discendente di lunghezza pressoché uguale. A queste onde lunghe ha prestato particolare attenzione Schumpeter, che le ha collegate alle conseguenze di innovazioni molto importanti, quelle che fanno epoca, quali la filatura meccanica e la macchina a vapore, e poi le ferrovie, l'elettricità, l'automobile, ecc. Secondo la cronologia schumpeteriana il primo ciclo lungo va dal 1787 al 1842, il secondo dal 1843 al 1897, il terzo dal 1898 al 1938. In tempi recenti la tesi è stata ripresa per suggerire che una possibile spiegazione della debole dinamica dell'economia mondiale a partire dagli anni settanta possa essere dovuta al suo ingresso nella fase discendente del quarto ciclo di Kondrat´ev.

La possibile compresenza di diversi cicli (ne sono stati individuati molti altri) mostra la complessità del fenomeno. Infatti, anche nell'ipotesi assurda che i singoli cicli abbiano regolarità perfetta e solo ampiezza e durate diverse, il loro operare simultaneo e sovrapposto è inevitabilmente destinato a provocare un andamento del PIL che può apparire del tutto irregolare per periodi anche molto lunghi. Poiché le serie empiriche mostrano andamenti del genere, l'ipotesi della compresenza di molti cicli può aver senso. Ma allora, calcolando il trend, si corre il rischio di individuarlo sulla base di ciò che potrebbe invece essere l'andamento tipico di una fase (ascendente o discendente) di un ciclo più lungo.

Caratteristiche dei cicliOgni ciclo storico ha ovviamente caratteristiche sue proprie in quanto si svolge in situazioni socioeconomiche diverse e mutevoli ed è influenzato in misura differente da fattori endogeni (economici) ed esogeni (extraeconomici). Ma l'unicità storica di

ogni singolo ciclo in ciascun paese non impedisce che essi abbiano anche alcune caratteristiche comuni.

Quali che siano le forze che danno inizio alla ripresa - che le diverse teorie metteranno in luce - esse prenderanno via via maggior vigore. Quali prima quali dopo, i vari indicatori incominciano a muoversi con maggior velocità; la ripresa si diffonde e si trasforma in boom. Crescono le produzioni, le vendite, i livelli occupazionali, il reddito reale, le capacità produttive utilizzate, le domande di materie prime e di prodotti finiti, la quantità di moneta, la sua velocità di circolazione, l'ammontare del credito e dei finanziamenti all'economia. Crescono anche i prezzi e i salari monetari: le fasi di boom hanno spesso dato origine ad acute tensioni inflazionistiche. Il buon andamento degli affari e dei profitti stimola gli investimenti reali; la borsa esalta questi fenomeni con movimenti speculativi che fanno salire rapidamente le quotazioni.

La fase di boom non dura però indefinitamente; l'attività economica prima incomincia a decelerare e poi a contrarsi. L'espansione troppo rapida porta con sé le radici della successiva contrazione, e ciò per un insieme di cause che hanno operato con intensità diversa nei diversi periodi.Fino alla prima guerra mondiale i cicli appaiono abbastanza regolari sia nella durata (7-10 anni) sia nelle loro caratteristiche. Di norma, la fase espansiva terminava perché l'eccessiva dilatazione del credito costringeva la banca centrale ad attuare misure fortemente restrittive. Ne derivavano crolli e panico in borsa, crisi finanziarie anche di vaste proporzioni, fallimenti di banche e imprese. Cadevano prezzi e salari e aumentava la disoccupazione. Nelle due più gravi depressioni del secolo scorso, quelle del 1856-1858 e del 1873-1879, in Inghilterra la disoccupazione aumentò rispettivamente dal 4,7% all'11,9 e dall'1,2% all'11,4.

La grande depressione del 1929-1932 appare per molti aspetti un fatto eccezionale. Innanzitutto per la sua maggior gravità, che Schumpeter attribuisce alla coincidenza delle fasi depressive di tutti i tre cicli di cui sopra. Poi, perché il precedente boom delle produzioni non era stato accompagnato da rilevanti aumenti di salari e prezzi. Infine, perché ha scardinato in modo inusitato le strutture finanziarie e creditizie e il sistema degli scambi e dei pagamenti internazionali. Il panico seguito al crollo di Wall Street, gli errori di politica economica, i fallimenti a catena delle banche, il ritiro dei capitali stranieri contribuirono a rendere la crisi

negli Stati Uniti molto più grave che altrove. La riduzione dei prezzi rese insopportabile il peso dell'elevato indebitamento delle imprese che in gran numero furono costrette a fallire. Tuttavia nell'industria i prezzi e i salari diminuirono meno che durante la crisi del 1873-1879, mentre la produzione si ridusse molto di più. Paolo Sylos Labini (v., 1970, pp. 94-96) spiega questo fatto con l'allontanamento da situazioni di concorrenza che si era verificato a seguito di massicci processi di concentrazione industriale. La disoccupazione salì a cifre impressionanti: oltre il 25% negli Stati Uniti, poco meno in Germania, oltre il 20% nel Regno Unito dove la crisi, che pur ebbe minor gravità, era sopraggiunta in una fase di disoccupazione già elevata. In Italia si arrivò al 15%, anche se l'isolamento internazionale l'aveva tenuta al riparo dai più gravi contraccolpi.

Nel secondo dopoguerra e fino al 1973 alcune caratteristiche del ciclo si modificano. Gli aspetti monetari appaiono di minor rilievo. Contano di più quelli reali. Il rallentamento produttivo che prelude alla fine del boom avviene perché, in un numero sempre maggiore di settori, le capacità produttive risultano sovrautilizzate e perciò non più in grado di far fronte a ulteriori crescite della domanda. La riduzione della disoccupazione, specie per la manodopera con particolari qualificazioni, pone altri ostacoli alla crescita delle produzioni. Nelle fasi finali del boom aumentano i salari e gli altri costi di produzione, mentre i profitti segnano il passo o incominciano a ridursi. Aumentano inoltre sia il costo del denaro sia le difficoltà di ottenere credito, e gli investimenti incominciano quindi a ridursi. La conseguente flessione della domanda si ripercuote dai settori inizialmente interessati verso gli altri con effetti sempre più diffusi.Il ciclo ha durata più breve e intensità molto minore di prima. Non c'è più caduta delle produzioni durante la crisi, c'è solo una crescita minore. Si ha un ciclo dei tassi di crescita: più alti nei boom e più bassi nelle depressioni. Inoltre, nei periodi di crisi, non si verifica più la riduzione di prezzi e salari.

Dopo il 1973 cambia ancora qualche cosa. Innanzitutto, compare sulla scena il vincolo energetico. Inoltre, con l'abbandono dei cambi fissi, la globalizzazione dei mercati finanziari e le innovazioni che sono state ivi introdotte riassumono un'importanza primaria gli aspetti finanziari. Il trend di crescita si riduce fortemente e i livelli di produzione tornano a diminuire durante le crisi. Lo stesso non avviene per i prezzi. Se si escludono quelli di alcune materie prime che si sono ridotti in concomitanza con le fasi discendenti del ciclo mondiale,

gli altri prezzi e i salari monetari non sono diminuiti. Sono però aumentati molto meno che nelle fasi di espansione, specialmente i salari, che in molti casi non sono neppure riusciti a seguire l'andamento dei prezzi e sono quindi diminuiti in termini reali.Nel corso dei cicli le aspettative hanno sempre un ruolo di grande rilievo: nelle fasi di espansione si diffonde un clima di ottimismo, in quelle di contrazione di pessimismo. Le decisioni imprenditoriali relative agli investimenti ne risultano particolarmente influenzate. Per questa ragione, oltre che per quelle che vedremo parlando della teoria dell'acceleratore, le fluttuazioni della domanda e della produzione di beni di investimento risultano più accentuate di quelle dei beni di consumo e dei servizi.

Occorre infine dire qualche cosa sugli aspetti internazionali del ciclo. Spesso c'è stato sincronismo tra gli andamenti dei diversi paesi. Le fluttuazioni si sono trasmesse dai paesi economicamente più importanti verso gli altri per il tramite della fitta rete di relazioni commerciali e finanziarie, queste ultime provocando forti effetti di esaltazione. Per comprendere le ragioni della trasmissione internazionale del ciclo basta osservare che una rapida espansione negli Stati Uniti provoca un aumento cospicuo delle importazioni americane. Per il resto del mondo vengono quindi a crescere le esportazioni con effetti positivi sui livelli di produzione. Il processo, una volta avviato, tende ad autoalimentarsi perché l'espansione di ciascun paese provoca aumenti della domanda negli altri. Con espressione colorita si dice che il paese che si muove per primo - gli Stati Uniti nel nostro esempio - svolge la funzione di 'locomotiva' che mette in moto l'espansione mondiale, la quale poi si autoalimenta. Ma quando la locomotiva si arresta o inverte la marcia, i contraccolpi vengono sentiti a livello mondiale e prende perciò avvio il processo depressivo, a meno che non si mettano in moto altre locomotive (ad esempio il Giappone, l'Europa) che abbiano una forza paragonabile a quella degli Stati Uniti e riescano perciò a sostituirsi a essi.
Le prime spiegazioni teoriche del ciclo

Si è visto come l'andamento del sistema risulti influenzato in tutti i suoi aspetti economici, sociali e politici dal fenomeno ciclico. Non può quindi sorprendere che sia difficile pensare a una sua spiegazione teorica in termini di un'unica causa. E infatti le teorie più accreditate non negano affatto la pluralità di cause; si limitano ad attribuire a una (o a poche) di esse peso (molto) maggiore rispetto alle altre. Lo stesso discorso vale per i meccanismi attraverso cui

cause e fenomeni interagiscono tra loro.

La riflessione sulle cause e sui meccanismi che generano i movimenti ciclici ha una storia ormai lunga. Molte delle teorie del ciclo che dominano la scena ancora oggi hanno i loro antecedenti in formulazioni apparse nel secolo scorso o agli inizi del nostro. Le trattazioni successive hanno beneficiato di un apparato analitico più rigoroso e di una mole molto più ampia di conoscenze empiriche. Ma non sempre sono risultate superiori a quelle precedenti per profondità di visione.

Le prime spiegazioni delle crisi commerciali erano fatte risalire alle conseguenze di fattori esogeni: dalla variabilità dei raccolti alle guerre, dallo sfruttamento di nuovi giacimenti auriferi all'apertura o chiusura di mercati stranieri. Ma, aspetto questo particolarmente importante, anche alle innovazioni tecniche (lord Overstone nel 1837 e Thomas Tooke nel 1838). Tuttavia ben presto ci si rese conto che una teoria puramente esogena del ciclo era del tutto insufficiente e che bisognava integrarla con spiegazioni di tipo endogeno, che dessero conto sia dei meccanismi con cui il sistema reagiva alle perturbazioni esogene, sia del perché le reazioni imprimessero al sistema un movimento ciclico.

Un altro aspetto che è stato sottolineato molto presto è quello dell'importanza della moneta e del credito. Ciò anche perché nel secolo scorso le crisi assumevano spesso la caratteristica di crisi finanziarie. Già Overstone e Tooke, che pur vedevano nei fattori reali la causa iniziale del ciclo, sottolineavano come le fluttuazioni del credito contassero molto nell'esaltare le fasi di espansione e nell'aggravare quelle di contrazione. Più tardi, John Stuart Mill sottolineava l'influenza dei fenomeni speculativi e degli errori di previsione degli operatori che provocavano fluttuazioni irrazionali nella domanda di credito, le quali, a loro volta, avevano l'effetto di esaltare i fenomeni ciclici.

Agli inizi del secolo scorso si trovano anche le radici delle diverse teorie che spiegano le crisi in termini di sottoconsumo e di eccesso di risparmio. Basta ricordare lord Lauderdale, Thomas R. Malthus e Simonde de Sismondi. Le critiche successive hanno messo in luce le difficoltà analitiche di queste teorie, che però hanno avuto una certa influenza su quelle ben più rigorose di stampo keynesiano. Esse continuano inoltre ad avere un certo rilievo nella

letteratura che si riallaccia a Marx e alla sua individuazione delle cause delle crisi, oltre che nella caduta tendenziale del saggio di profitto, nelle difficoltà di 'realizzo' provocate dall'insufficiente potere di acquisto delle masse lavoratrici e nell''anarchia del capitalismo', che produce 'sproporzioni' nella composizione delle domande e delle offerte a causa dell'impossibilità di coordinare e programmare le decisioni dei capitalisti.

La teoria delle sproporzioni, tra le differenti branche produttive e tra la capacità di produzione e quella di consumo, è stata sostenuta con molta lucidità da Michail Tugan-Baranovskij. L'aspetto più rilevante della sua analisi riguarda però il ruolo dell'investimento in capitale fisso. Egli ha messo in luce che le fluttuazioni più violente nel corso del ciclo avvengono nei settori che producono beni di investimento e si diffondono all'intero sistema economico a causa delle strette relazioni di interdipendenza che legano tra loro i diversi settori, l'espansione di uno creando domanda per gli altri. Sulla scia di Tugan si è poi mosso Arthur Spiethoff che ha approfondito la spiegazione del perché gli investimenti sono soggetti ad ampie fluttuazioni, sostenendo che la fase di prosperità ha inizio quando, essendo disponibile capitale finanziario in cerca di investimenti, si manifestano allettanti prospettive di profitto in seguito all'apertura al commercio di nuovi paesi o alla scoperta di nuove tecnologie. Non siamo ancora alla teoria delle innovazioni di Schumpeter, ma la strada è ormai aperta.

Va infine ricordato che la teoria dell'acceleratore, che assumerà un ruolo dominante nelle spiegazioni del ciclo che utilizzano modelli dinamici, è stata presentata originariamente da Albert Aftalion nel 1909. Nella sua trattazione si trova inoltre una buona analisi dell'importanza che hanno i ritardi nel determinare le oscillazioni cicliche. Tra il momento in cui si decide l'attuazione di investimenti e quello in cui le nuove attrezzature incominceranno a produrre passa un tempo piuttosto lungo. All'inizio di una fase di espansione, sotto la spinta di favorevoli condizioni dal lato della domanda, si intraprendono progetti di investimento che non danno subito luogo ad aumenti di produzione. Lo stimolo della domanda perciò rimane inalterato e nuovi investimenti continuano a essere messi in cantiere. Quando però arriva sul mercato l'aumentata produzione, essa risulta in eccesso rispetto alla domanda. I prezzi tendono perciò a cadere e molte imprese hanno perdite. Ne derivano una brusca riduzione degli investimenti e la crisi. Le teorie monetarie del ciclo

Le formulazioni del periodo tra le due guerreGli aspetti monetari risultano particolarmente appariscenti nel corso del ciclo, ragion per cui tutte le teorie ne hanno dovuto tener conto in un modo o in un altro. Le teorie monetarie hanno però una peculiarità: la moneta e il credito vi hanno un ruolo di gran lunga dominante. Il ciclo è descritto come "un fenomeno puramente monetario", secondo l'espressione di Ralph G. Hawtrey, il quale afferma anche che i flussi monetari sono la causa unica e sufficiente degli alti e bassi dell'attività economica. Se fosse possibile stabilizzare l'andamento dei flussi monetari, il ciclo scomparirebbe. Ma, secondo Hawtrey, questo è un compito estremamente difficile perché il moderno sistema creditizio è inerentemente instabile.

Sono le banche a dar avvio a un processo di ripresa per mezzo della riduzione dei tassi di interesse e delle facilitazioni nella concessione di credito. Hawtrey ritiene che i commercianti, specialmente i grossisti, reagiscano con particolare intensità a variazioni, anche piccole, nelle condizioni creditizie. Essi sono quindi indotti ad aumentare sensibilmente le proprie scorte di prodotti, passando gli ordinativi ai produttori. Si ha così un aumento delle produzioni, dell'occupazione e dei redditi. I consumatori incominciano a spendere di più; l'aumento della domanda provoca l'assottigliamento delle scorte presso i dettaglianti e presso i grossisti. Si ha allora un nuovo aumento degli ordinativi, dei redditi, della domanda di beni e quindi un nuovo assottigliamento delle scorte. Il processo diventa cumulativo ed è alimentato da una continua espansione del credito.

L'attività produttiva non può però crescere senza limiti: ci sono sempre quelli fissati dalle capacità produttive disponibili. Quando questi vengono raggiunti, i produttori incominciano ad alzare i prezzi, ma questo fatto fornisce un nuovo stimolo ai grossisti e ai dettaglianti: è infatti conveniente aumentare le scorte quando i prezzi stanno crescendo e si ritiene che cresceranno ancora. Perciò, fino a quando il sistema bancario continua a espandere il credito, il boom continua, ma con inflazione crescente. La banca centrale però non può accettare che l'espansione della circolazione monetaria metta a repentaglio il mantenimento della convertibilità (aurea) della moneta. Decide allora di attuare una stretta creditizia, alzando il tasso ufficiale di sconto e riducendo il rifinanziamento dell'economia.

Ha così inizio la fase recessiva che si evolve anch'essa secondo un processo cumulativo. Si riduce la produzione, aumenta la disoccupazione, diminuiscono le vendite e i prezzi. I grossisti cercano di liquidare le scorte ma trovano difficoltà a farlo a causa della caduta della domanda. La (relativa) rigidità dei salari monetari fa sì che essi si riducano con un certo ritardo rispetto ai prezzi delle merci e risultino perciò troppo elevati, date le peggiorate condizioni economiche. Si ha così una caduta dell'occupazione molto maggiore di quanto potrebbe altrimenti succedere.

Durante la depressione si ha il ripagamento graduale dei prestiti e il ritorno del circolante presso le banche. Le riserve, che prima si erano ridotte notevolmente rispetto alla circolazione, si riportano ai livelli normali e poi li superano ampiamente. Le banche decidono allora di ridurre i tassi sui prestiti e di facilitare il credito. Queste decisioni potrebbero dare l'avvio a una nuova ripresa nel modo che si è appena descritto. La posizione iniziale di Hawtrey era che effettivamente la ripresa sarebbe iniziata, ma l'esperienza della grande crisi degli anni trenta gli ha fatto sorgere più di un dubbio sulla questione se una politica di tassi di interesse molto bassi sia realmente in grado di provocare la ripresa di un sistema caduto in profonda depressione, almeno fino a quando non vengano rovesciate le aspettative di prezzi calanti.

La teoria di Hawtrey può essere considerata la capostipite delle teorie monetarie del ciclo. È certamente la più 'pura' e perciò la più unilaterale. Ha superato in notorietà quella più articolata di Ludwig von Mises, che riprendeva l'analisi del processo cumulativo di Knut Wicksell. Tale analisi si fondava sulla possibile discrepanza del tasso di interesse monetario, cioè quello sui prestiti fissato dalle banche, dal tasso naturale, cioè quello di equilibrio che eguaglia la domanda di prestiti per l'attuazione di investimenti all'offerta di risparmio. Se le banche praticano un tasso inferiore a quello naturale, la domanda di credito supera l'offerta e la differenza viene coperta dall'espansione monetaria, con effetti inflazionistici che continueranno a manifestarsi fino a quando le banche non decideranno di alzare il tasso di mercato per portarlo all'uguaglianza con quello naturale. Al contrario, se è quest'ultimo a essere più basso, ha origine una tendenza cumulativa alla riduzione dei prezzi.

Mises applica l'analisi wickselliana alla teoria del ciclo e spiega l'avvio della ripresa con la

decisione delle banche, che escono dalla depressione in condizioni di ampia liquidità, di ridurre il tasso monetario al di sotto di quello naturale. Una prima differenza con Hawtrey è che il livello allettante del tasso di mercato influenza le decisioni di investimento in capitale fisso direttamente e non in via mediata, cioè per il tramite delle decisioni dei grossisti. Una seconda differenza, che sarà poi analizzata in profondità da Friedrich A. von Hayek, è che il livello troppo basso del tasso sui prestiti provoca un eccesso di investimento che, a sua volta, determina un allungamento del periodo di produzione, cioè l'adozione di tecniche a più elevata intensità di capitale. Ma l'allungamento ha natura artificiale in quanto non è determinato da un aumento del risparmio volontario della collettività e dalla conseguente riduzione del tasso naturale. Ciò che infatti viene ridotto artificialmente, per decisione delle banche, è il tasso di mercato. L'eccesso di investimento che ne deriva deve trovare una controparte reale nel contenimento dei consumi. L'aumento dei prezzi ha proprio questo effetto e dà origine al fenomeno del 'risparmio forzato'. In altri termini, l'inflazione si incarica di estorcere dai consumatori le risorse necessarie a far aumentare il capitale reale.

Questo processo non può però andare avanti all'infinito: la distorsione deve essere corretta. Da un lato, il risparmio forzato tenderà a ridursi, perché i lavoratori incominceranno a ottenere aumenti dei salari monetari superiori alla crescita dei prezzi e potranno quindi consumare di più. Dall'altro lato, il sistema bancario sarà costretto a bloccare l'espansione creditizia. La fase depressiva che ne seguirà sarà tanto più acuta quanto più robusto era stato il precedente boom le cui distorsioni debbono essere curate.

Come si vede, la teoria di Hayek è molto più raffinata di quella di Hawtrey e tira in ballo anche aspetti non esclusivamente monetari. Tra l'altro, viene sottolineata l'influenza che possono avere i cambiamenti tecnologici nel favorire l'avvio della ripresa. Ma è sempre l'instabilità del sistema bancario a provocare quella del sistema economico. Per evitare le depressioni bisognerebbe evitare di stimolare artificialmente il sistema. Bisognerebbe cioè attuare una politica monetaria 'neutrale', che non provochi ampliamenti o restrizioni arbitrarie del credito. Ma lo stesso Hayek sottolinea la difficoltà del compito quando rileva come, per il singolo banchiere, sia impossibile distinguere in un aumento dei depositi ciò che è dovuto a risparmio volontario da ciò che è dovuto a risparmio forzato.

Il monetarismo attualeLe teorie monetarie del ciclo che avevano dominato la scena negli anni venti e trenta sono cadute pressoché nell'oblio con il trionfo dell'impostazione keynesiana. Il loro prestigio continuava però a rimanere alto presso l'Università di Chicago sotto l'influenza di Milton Friedman. È al suo lungo lavoro teorico ed empirico che si deve il ritorno sulla scena delle teorie monetarie nella riformulazione che oggi va sotto il nome di 'monetarismo'.

La tesi di Friedman è che esiste una stretta relazione tra i movimenti ciclici della quantità di moneta e quelli del reddito reale. Lo stesso invece non avviene per i movimenti secolari della moneta che hanno principalmente influenza sul livello dei prezzi. La causazione va dalla moneta al reddito e ai prezzi e non viceversa. Sono le variazioni autonome della quantità di moneta, gli shocks monetari non quelli reali, i responsabili delle fluttuazioni. La crescita reale di lungo periodo è invece indipendente dalla moneta, derivando soltanto dalle decisioni di risparmio - e quindi di investimento - e dall'evoluzione della tecnologia. C'è la netta separazione di ciclo e trend: la moneta è responsabile del primo, i fattori reali del secondo. Nel lungo periodo le variazioni della quantità di moneta hanno effetto solo sul livello dei prezzi.In assenza di disturbi monetari l'economia di mercato sarebbe stabile; i prezzi relativi dei beni e dei servizi produttivi si muoverebbero in modo da assicurare una crescita ordinata in condizioni di sostanziale piena occupazione e in assenza di inflazione. La trattazione di Friedman, molto più raffinata sia sotto l'aspetto teorico sia sotto quello dell'analisi empirica, ha quindi riportato in auge le teorie tradizionali e la connessa convinzione del liberismo secondo cui gli interventi volti a regolare l'andamento del sistema economico non sono utili; sono anzi dannosi. La prescrizione monetarista è perciò quella di non utilizzare a scopi anticongiunturali la politica monetaria e anche quella della spesa pubblica, se essa può provocare variazioni arbitrarie della quantità di moneta. Questa deve invece aumentare a un tasso costante che consenta solo il finanziamento della crescita reale, cioè non inflazionistica, di lungo periodo. A tal fine sono necessarie modifiche istituzionali che vincolino il comportamento delle banche centrali.

Le tesi monetariste hanno avuto notevole diffusione negli anni settanta-ottanta, soprattutto a causa della dimostrata incapacità delle politiche macroeconomiche di far uscire il sistema mondiale da condizioni di stagflazione (ristagno più inflazione) in cui continuava a rimanere invischiato. Negli stessi anni Robert E. Lucas jr. ha proposto una nuova versione del

monetarismo, ancora più radicale di quella di Friedman e ben più raffinata sotto il profilo analitico. Lucas si è proposto di integrare la visione monetarista del ciclo nell'ambito della teoria dell'equilibrio generale, ciò che sembrava quasi una contraddizione in termini: il ciclo era infatti tradizionalmente visto come un fenomeno di squilibrio. Egli assume che tutti i mercati siano sempre in equilibrio e che gli operatori si comportino razionalmente prendendo le proprie decisioni sulla scorta di tutte le informazioni disponibili e delle previsioni che il modello che descrive 'correttamente' il funzionamento del sistema può da esse far derivare.

I movimenti ciclici si manifestano, in equilibrio, perché gli operatori interpretano variazioni del livello assoluto dei prezzi come variazioni transitorie della struttura dei prezzi relativi. Ciò accade perché gli operatori dispongono soltanto di informazioni limitate. Così ad esempio capita che i lavoratori interpretino un aumento dei salari monetari, almeno parzialmente, come un aumento transitorio dei salari reali e quindi decidano di lavorare temporaneamente di più, riservandosi di lavorare di meno quando i salari scenderanno al di sotto del livello normale. In questo modo Lucas riesce a dare una spiegazione delle fluttuazioni dei livelli di attività, in concomitanza con quelle dei prezzi e dei salari monetari, senza tirare in ballo il concetto di disoccupazione involontaria. Questa, per definizione, non può esserci in condizioni di equilibrio e anzi - egli ritiene - non può neppure essere definita in modo rigoroso, in quanto in ogni tipo di disoccupazione c'è sempre un elemento di volontarietà (se uno accettasse salari molto più bassi di quelli prevalenti sul mercato, potrebbe facilmente trovare lavoro).

La causa dell'errata percezione dell'effettiva situazione da parte degli agenti è data dall'incertezza provocata dall'erraticità della politica monetaria. Risulta perciò inevitabile la conclusione di tutte le teorie monetarie del ciclo: per ridurre l'instabilità del sistema reale bisogna eliminare il potere discrezionale delle autorità monetarie e costringerle a seguire regole fisse, indipendentemente dallo stato della congiuntura. Interazione tra moltiplicatore e acceleratore

Il modello originarioLa pubblicazione della Teoria generale di Keynes nel 1936 ha provocato una svolta importante nella teoria del ciclo. L'instabilità del sistema reale non appare più come conseguenza dell'instabilità monetaria, anche se questa può contribuire a esaltare il

fenomeno ciclico, ma come conseguenza dell'instabilità degli investimenti e quindi della domanda globale.

Keynes non ha formulato alcuna teoria dettagliata del ciclo economico, ma ha sottolineato come gli investimenti siano soggetti a fluttuazioni ampie e improvvise. Essi sono infatti decisi dagli imprenditori sulla base delle aspettative circa la loro profittabilità futura. Ma l'incertezza che domina la scena economica rende le aspettative molto precarie. Se queste cambiano improvvisamente, cambiano anche, e in misura molto ampia, le decisioni di investimento.

Una variazione degli investimenti fa variare la domanda globale, il reddito e perciò i consumi che da esso dipendono. Si ha quindi una nuova variazione della domanda, del reddito, dei consumi e così via. Il moltiplicatore misura l'effetto complessivo sul reddito di una variazione degli investimenti. La spiegazione della dinamica di questi è fornita dall'acceleratore, che postula che gli imprenditori intendono avere a disposizione uno stock di capitale proporzionale al livello della domanda. Se questa è aumentata nel recente passato, essi decidono nuovi investimenti; se è diminuita, cercano di disinvestire. Si ha perciò una relazione di proporzionalità tra il livello degli investimenti (netti) e le variazioni della domanda.

Un modello che faccia interagire moltiplicatore e acceleratore può dar luogo a soluzioni cicliche. Il più semplice, e più famoso, è quello presentato da Paul A. Samuelson nel 1939, che, in termini matematici, dà luogo a un'equazione alle differenze finite, lineare, del secondo ordine. L'andamento temporale del reddito, che si ottiene come soluzione, dipende dai valori dei parametri, vale a dire dal moltiplicatore e dal coefficiente di accelerazione che collega il livello degli investimenti alle variazioni della domanda. Per certi valori l'andamento non è ciclico ma è di tipo esponenziale: il reddito tende a portarsi verso un valore costante (di equilibrio) oppure tende a crescere continuamente a un tasso costante. Per altri valori dei parametri si ha invece un andamento oscillatorio. Ma le oscillazioni si smorzano con il passar del tempo oppure, al contrario, diventano sempre più ampie. Solo in un caso limite, estremamente improbabile, esse rimangono di ampiezza costante. Nel caso di oscillazioni smorzate, il ciclo è destinato a scomparire. Nel caso contrario, esso permane ma aumenta

sempre di intensità, in contrasto con ciò che suggerisce l'osservazione empirica.

Prima di esaminare quali ipotesi siano state formulate per sopperire a queste difficoltà, è opportuno fare alcune osservazioni generali. In tutti i modelli dinamici il possibile andamento oscillatorio deriva dalla presenza di ritardi nelle relazioni che legano le diverse variabili (ad esempio consumo e reddito, investimenti e variazioni del reddito, ecc.). È proprio la struttura dei ritardi a far sì che il sistema reagisca in modo oscillatorio a ogni impulso (shock) che lo sposti dalla posizione di equilibrio.Supponiamo ad esempio che aumenti la spesa pubblica. Il moltiplicatore farà aumentare il reddito e i consumi. Ne seguirà, per effetto dell'acceleratore, un aumento degli investimenti (anche in scorte). Scatterà allora nuovamente il moltiplicatore e si avrà un nuovo incremento del reddito. Se questo incremento è superiore al precedente, continuerà la crescita. Se invece è inferiore, per effetto dell'acceleratore, il livello assoluto degli investimenti risulterà inferiore a prima e si avrà quindi l'inizio di un processo di riduzione del reddito e dei consumi. Il contrario accade quando il rallentamento nella discesa del reddito riduce il disinvestimento: la domanda non cade più e anzi ricomincia a crescere.

Come si vede, gli shocks (ad esempio aumento della spesa pubblica) sono esogeni, ma il movimento ciclico che ne risulta ha natura endogena, derivando dalle reazioni ritardate delle diverse variabili. Ragnar Frisch già nel 1933 aveva mostrato come una successione di shocks esogeni di tipo puramente casuale potesse dar origine a movimenti ciclici non destinati a venir meno con il passar del tempo, anche nel caso in cui i parametri dell'equazione alle differenze avessero valori tali da provocare una soluzione di oscillazioni smorzate e perciò non permanenti. Essendo gli shocks casuali, non sono essi a provocare l'andamento ciclico, anche se sono proprio essi a tenerlo in vita. Si è, a questo proposito, indicata l'analogia con il movimento di una sedia a dondolo sollecitata da spinte irregolari.

Si può infine osservare che un modello che si riduce in definitiva a un'equazione alle differenze (o differenziale), lineare, del secondo ordine, non può spiegare contemporaneamente il ciclo e la crescita. O spiega l'uno, nel caso di soluzione oscillatoria, o spiega l'altra, nel caso di soluzione di crescita esponenziale. Su alcuni tentativi di superare questa grave limitazione si dirà più avanti.

Le teorie non lineari: soffitti e pavimentiSi è visto che, nel caso di oscillazioni smorzate, la permanenza del ciclo nel tempo può essere assicurata da una successione di shocks irregolari, di natura casuale. Nessuna teoria basata sull'interazione tra moltiplicatore e acceleratore trascura questo aspetto, ma quasi tutte non gli attribuiscono importanza esclusiva. D'altra parte, le indagini empiriche sembrano suggerire che i valori dei parametri del modello siano tali da dar luogo a una soluzione di tipo 'esplosivo', cioè a oscillazioni amplificate oppure a crescita esponenziale. Le teorie più note hanno preso atto di ciò ma hanno introdotto dei limiti, superiori (soffitti) e inferiori (pavimenti), che bloccano gli andamenti eccessivamente divergenti del sistema facendolo rimbalzare in direzione opposta.

Andamenti esplosivi si hanno in corrispondenza di valori elevati dell'acceleratore. A un aumento della domanda (esogena), segue allora un rilevante aumento degli investimenti e perciò una crescita rapida. Il boom si autoalimenta e il sistema viene spinto verso il limite superiore fissato dalla piena utilizzazione delle capacità produttive (specie nei settori che producono beni di investimento) oppure dalla piena occupazione della manodopera (con le caratteristiche richieste).

Per i paesi, come l'Italia, molto aperti al commercio internazionale il soffitto può essere incontrato anche molto prima, cioè quando il disavanzo della bilancia dei pagamenti diventa difficilmente sostenibile. Possono allora intervenire provvedimenti fiscali o monetari (come nel caso delle teorie monetarie) volti a raffreddare l'economia.

Quale che sia il primo soffitto nelle cui vicinanze ci si viene a trovare, la crescita reale del sistema deve rallentare. Ma allora interviene l'acceleratore a far crollare gli investimenti e a far avvitare il sistema nella depressione. Anche questa non può però continuare indefinitamente. C'è un limite determinato dal fatto che le imprese non possono disinvestire di più di quanto lo consente la possibilità di non rimpiazzare gli impianti giunti al termine della propria vita economica. In altri termini, un pavimento è fissato al livello zero per gli investimenti lordi.

Ma prima di arrivare a questo limite estremo, il sistema può incontrare altri pavimenti. Uno è

rappresentato dalla possibilità che certi investimenti siano 'autonomi', decisi cioè sulla base di considerazioni diverse da quelle di cui tiene conto la teoria dell'acceleratore. Si pensi alle opere pubbliche, all'edilizia, agli investimenti di tipo innovativo o ai programmi pluriennali incominciati in precedenza, ecc. Un altro limite inferiore può essere fissato dai cosiddetti 'stabilizzatori automatici' della domanda, cioè dai vari meccanismi di sostegno dei redditi e dei consumi - cassa integrazione, sussidi di disoccupazione, prestazioni di sicurezza sociale, ecc. - di cui i paesi moderni si sono dotati, oltre che dal sistema dell'impostazione fiscale progressiva, che comporta minori prelievi sui redditi calanti.

Quando la domanda cessa di ridursi e incomincia eventualmente a crescere, quando cioè si incontra un pavimento, l'acceleratore torna a operare in senso positivo. Il sistema allora rimbalza e ha inizio la fase di ripresa.

L'introduzione di limiti superiori e inferiori ha consentito un maggior realismo della teoria e ne ha eliminato alcune importanti deficienze. Ne rimangono però altre. Innanzitutto i pavimenti sembrano posti a un livello troppo basso, anche se considerazioni di tipo istituzionale - quali quelle a cui si è accennato - possono contribuire a farli alzare. In quasi tutti i casi reali le depressioni non sono apparse così gravi da far ritenere che sia stato raggiunto uno dei pavimenti sopra indicati. Inoltre, non è realistico pensare che il rimbalzo verso l'alto avvenga appena il sistema ha raggiunto il pavimento. Perché la ripresa prenda avvio è infatti necessario che si verifichi un rovesciamento delle aspettative che si sono notevolmente deteriorate nel corso della crisi. Ma non c'è niente nel modello che possa garantire un evento del genere. La spiegazione del punto di svolta inferiore sembra quindi abbastanza insoddisfacente. Questa critica non vale però solo per la teoria di cui si sta parlando. Anche le altre trovano infatti difficoltà a spiegare l'avvio della ripresa. A vantaggio di questa rimane comunque il fatto di aver individuato chiaramente le ragioni del perché il boom debba prima o poi finire e di aver anche suggerito a Richard M. Goodwin, che per primo aveva introdotto soffitti e pavimenti, l'ipotesi che si debba cercare di formulare una teoria del ciclo che tenga in considerazione l'esistenza del solo limite superiore.

C'è infine la questione della crescita ciclica cui sopra è stato accennato. Per rendere possibile la crescita, John R. Hicks ha ipotizzato che sia il pavimento ad alzarsi con il passar del

tempo. Più precisamente, ha introdotto un trend esogeno di crescita per gli investimenti autonomi. Per l'innalzamento del soffitto non sorgono nuovi problemi, in quanto si può ipotizzare un aumento esogeno del numero dei lavoratori e della loro produttività.Per alcuni aspetti simile alla trattazione di Hicks è quella di Marco Fanno, che è analiticamente meno elaborata ma, nella versione iniziale, è precedente a quella dell'autore inglese. Rispetto a questa presenta il vantaggio di abbozzare una spiegazione in termini economici della crescita degli investimenti autonomi nel tempo.La soluzione di crescita ciclica che Hicks e Fanno riescono in tal modo a ottenere non soddisfa però pienamente, perché la crescita non è affatto spiegata; è solo ipotizzata in modo esogeno. C'è inoltre un aspetto tecnico che lascia perplessi. L'introduzione di limiti superiori e inferiori ha infatti reso non lineare il modello originario: l'acceleratore non funziona più come prima quando si raggiungono i limiti. È proprio questo il pregio del modello in quanto, se si rimane in ambito lineare, il ciclo o scompare o diventa esplosivo. Ma in un modello non lineare non è matematicamente corretta la separazione del trend dal ciclo, salvo che nel caso del tutto particolare di un trend costante e identico per il soffitto e il pavimento. Non vale infatti il 'principio di sovrapposizione' che permette, nei casi lineari, di studiare un tipo di movimento per volta e metterli poi assieme. Se si ritiene che le non linearità siano essenziali, come avviene per le teorie sotto esame, la scappatoia di aggiungere un trend a un movimento ciclico non può essere considerata soddisfacente.

I modelli econometrici disaggregatiSecondo il messaggio keynesiano le fluttuazioni sono provocate principalmente dall'instabilità della domanda globale. Un attento controllo di questa può perciò dare un notevole contributo alla stabilizzazione dell'economia. Allo scopo possono essere utilizzate la manovra della spesa pubblica, la politica fiscale e quella monetaria, quest'ultima però con maggior cautela per evitare effetti indesiderati sugli investimenti (cioè sulla variabile che essa più è in grado di influenzare).Regolare efficacemente un sistema economico concreto è però un problema complesso per la cui soluzione non basta guardare soltanto agli aspetti aggregati. Bisogna prestare particolare attenzione a molte caratteristiche, istituzionali o di altro genere, che possono influenzare le modalità di comportamento del sistema. In particolare, occorre badare alle strette interrelazioni che si stabiliscono tra i diversi settori produttivi, confrontare l'andamento delle domande e delle capacità produttive settoriali, considerare gli effetti di cambiamenti nella

composizione di variabili aggregate quali reddito, spesa privata e pubblica, gettito fiscale, ecc. Inoltre, occorre anche una più attenta considerazione delle relazioni commerciali con l'estero, dei movimenti di capitale in entrata e in uscita, del fenomeno della trasmissione internazionale del ciclo.

La più accurata rappresentazione del sistema economico ha perciò comportato l'introduzione di molte nuove variabili e un'accentuata disaggregazione di altre. Di conseguenza, i modelli attualmente utilizzati per l'analisi del ciclo comprendono un numero molto elevato di variabili e di relazioni stimate per via econometrica. La loro formulazione matematica è quindi particolarmente complessa anche per la presenza di complicate strutture di ritardi. Quasi tutti i modelli contengono relazioni che possono essere fatte risalire ai concetti originari di moltiplicatore e acceleratore nelle loro diverse varianti. Possono perciò essere idealmente collegati all'impostazione keynesiana.

L'utilizzazione dei modelli econometrici a scopi previsivi ha avuto esiti alterni: talvolta le previsioni si sono dimostrate sufficientemente accurate, altre volte - e soprattutto in tempi recenti - hanno lasciato molto a desiderare. Lo stesso discorso può essere ripetuto per l'utilizzazione dei modelli allo scopo di valutare gli effetti di particolari misure di politica economica. Si ricorre in tal caso a procedimenti di simulazione. Si parte cioè da una soluzione del modello per dati valori dei parametri e delle variabili esogene; si cambia uno o più di questi valori (ad esempio coefficienti di imposizione fiscale, spesa per investimenti pubblici in infrastrutture, ecc.); si risolve il modello con i nuovi dati e si vede come cambiano gli andamenti delle variabili endogene che interessano (ad esempio livello di occupazione, dei prezzi, ecc.).

Il relativo insuccesso che i modelli econometrici hanno registrato, sia sul terreno previsivo sia su quello della valutazione delle politiche economiche, può essere facilmente spiegato. In tempi recenti si sono infatti verificate, a livello mondiale, rilevanti modificazioni di tipo strutturale. Basta pensare agli shocks petroliferi e alla diffusione del fenomeno della stagflazione. Se il passato non appare più una valida guida per il presente e per il futuro, non può stupire che lo stesso valga per i modelli econometrici che sono stimati su dati del passato. Inoltre, e soprattutto per quel che riguarda le valutazioni degli effetti delle misure di

politica economica, occorre osservare che se l'esperienza ha insegnato che quando il sistema va in un certo modo vengono presi certi provvedimenti, gli operatori hanno imparato la lezione e hanno corrispondentemente previsto il comportamento delle autorità. Ciò può averli indotti a prendere decisioni diverse da quelle che prendevano prima. Il modello, che non contemplava questa possibilità, non ha potuto fornire valutazioni corrette.

Pur non negando l'importanza di queste limitazioni dei modelli econometrici, non si deve concludere che essi non possono in ogni caso fornire indicazioni utili per una politica di controllo del ciclo. Una tale conclusione, che trova ad esempio sostenitori tra i più accesi monetaristi, sembra andare oltre il segno e derivare dalla convinzione liberista dell'inopportunità di politiche di regolazione dell'economia più che dalle effettive limitazioni dei modelli econometrici. Questi devono essere migliorati e aggiornati. Deve essere aumentata la loro capacità di individuare tempestivamente i cambiamenti strutturali e di valutarne le conseguenze. Devono essere migliorate le procedure di stima econometrica, anche al fine di poter prevedere meglio il comportamento degli operatori. In altre parole, i modelli sono strumenti che devono e possono essere perfezionati, alcuni anche radicalmente modificati. Ma ciò non significa affatto negare, in via di principio, la loro utilità come guida per una più efficace politica di controllo del ciclo. Teorie dello sviluppo ciclico

La teoria di SchumpeterLa tesi principale di Schumpeter è che le fluttuazioni rappresentano "la forma che lo sviluppo assume nell'era del capitalismo". Sono cioè la conseguenza necessaria della rottura dell'equilibrio stazionario che deve verificarsi perché si abbia lo sviluppo.Il motore del processo è l'imprenditore innovatore, che introduce nuovi prodotti o nuovi processi produttivi, che si inserisce su nuovi mercati o trova nuove fonti di approvvigionamento, che migliora l'organizzazione produttiva o commerciale, ecc. Per realizzare le innovazioni, di regola, l'imprenditore ha bisogno di sottrarre le risorse reali dalle combinazioni produttive esistenti, di rompere cioè le precedenti condizioni di equilibrio. Le risorse finanziarie gli sono fornite dal sistema creditizio con creazione di moneta. Una volta introdotta l'innovazione, l'imprenditore si trova in condizioni di vantaggio rispetto ai concorrenti. Realizza quindi maggiori profitti. Ma questi hanno natura temporanea perché altri imprenditori (imitatori) introdurranno l'innovazione. Si acuirà allora la concorrenza e i profitti tenderanno a scomparire. Il sistema tornerà così a portarsi a una nuova situazione di

equilibrio, caratterizzata da combinazioni produttive più efficienti.

Le fluttuazioni si manifestano perché, di norma, le innovazioni non si distribuiscono nel tempo in modo uniforme. Avvengono invece 'a gruppi' o 'a sciami', perché la comparsa di uno o più innovatori facilita quella di altri in numero crescente. Le ragioni sono molteplici: ci sono i processi di imitazione; c'è il fatto che un'innovazione ne sollecita altre nello stesso o in altri settori; c'è l'espansione che si autoalimenta e migliora le aspettative imprenditoriali, ecc. La comparsa più o meno simultanea di un gran numero di innovatori dà origine al boom che viene esaltato da comportamenti speculativi. Dopo un certo tempo però, con l'arrivo sui mercati dei frutti del processo innovativo, ci si accorge che l'offerta di beni è cresciuta in modo eccessivo rispetto alle possibilità di assorbimento dei mercati. Alcune imprese, specie quelle che si sono mosse in ritardo, vengono a trovarsi in difficoltà. Il sistema bancario, che prima aveva espanso troppo il credito, chiede il rimborso dei prestiti. Incomincia quindi la depressione che è la conseguenza inevitabile del precedente boom.ale la pena di sottolineare che lo sviluppo ciclico non è la conseguenza delle nuove scoperte rese disponibili dal progresso delle scienze che, tra l'altro, può avvenire in modo continuo. Sono le innovazioni, cioè l'applicazione a scopi produttivi di vecchi o nuovi ritrovati tecnici o organizzativi, che hanno andamento discontinuo e procedono per ondate. Il fenomeno è tipicamente economico e richiede quindi una spiegazione in termini economici. Schumpeter ha posto le basi per una teoria endogena del progresso tecnico. Ma la riflessione successiva non ne ha ancora dato una formulazione del tutto soddisfacente.

Le teorie di derivazione marxistaSul terreno tradizionale delle teorie delle sproporzioni e del sottoconsumo la riflessione recente non è andata molto al di là della riesposizione, con qualche maggior affinamento analitico, delle tesi di Marx (spesso riprese nella versione di Tugan-Baranovskij). Ci si sarebbe aspettati di più, specie se si pensa che l'esperienza degli anni settanta-ottanta è stata caratterizzata da rilevanti squilibri tra domande e offerte settoriali e da notevoli cambiamenti nei prezzi relativi e nella distribuzione del reddito.

Sviluppi più interessanti si sono invece avuti sul terreno dell'integrazione e del confronto fra tradizione marxista e analisi keynesiana. Il punto di partenza è stato la riaffermazione del ruolo dominante che hanno le decisioni di investimento dei capitalisti. Si è mostrato come le

difficoltà di funzionamento del capitalismo trovino spiegazione proprio nel ruolo duplice, e in un certo senso contraddittorio, degli investimenti: da un lato, creatori di domanda e quindi causa della prosperità ma, dall'altro lato, anche creatori di capacità produttive aggiuntive e quindi ostacolo a ulteriori ampliamenti. Secondo la felice espressione di Michal Kalecki, "la tragedia dell'investimento è che esso provoca le crisi perché è utile. Senza dubbio molta gente considererà questa teoria come paradossale. Ma non è la teoria che è paradossale, ma il suo oggetto, l'economia capitalistica" (v. Kalecki, 1939; tr. it., p. 176).

In coerenza con questa posizione, assume enorme importanza l'analisi delle motivazioni degli investimenti. Le funzioni degli investimenti che Kalecki ha proposto derivano, in senso lato, dalla teoria dell'acceleratore. Se ne differenziano, oltre che per l'introduzione di diversi tipi di ritardi, per la considerazione esplicita dell'influenza che i profitti, realizzati e attesi, hanno sulle decisioni di investire. I profitti realizzati sono però funzione diretta degli investimenti per effetto del meccanismo del moltiplicatore che Kalecki deriva, in modo originale, introducendo l'ipotesi che i consumi dipendano dalla distribuzione del reddito tra lavoratori e capitalisti. Ne risulta un'interazione affatto particolare tra moltiplicatore e acceleratore, in cui svolgono un ruolo dominante i profitti realizzati e attesi, questi ultimi dipendenti anche dal divario tra stock di capitale desiderato e stock effettivo.

I modelli di Kalecki hanno soluzioni cicliche qualitativamente non dissimili da quelle discusse in precedenza. In particolare non risulta spiegata endogenamente la crescita, anche se appare interessante il suggerimento che essa possa derivare come conseguenza di variazioni lente, ma esogene, dei diversi parametri del modello. La conclusione più rilevante è però quella che la crescita può essere spiegata solo considerando gli effetti delle innovazioni sulle decisioni di investimento. Questa conclusione, che si rifà a Schumpeter, è stata in tempi recenti almeno parzialmente condivisa anche da autori che, come Joseph Steindl e altri, si sono mossi lungo linee in parte simili a quelle di Kalecki, però con l'obiettivo di fornire una spiegazione del ristagno degli anni settanta più che una teoria dello sviluppo ciclico.

Una digressione sul ciclo politicoKalecki è considerato anche il progenitore delle cosiddette teorie del ciclo politico-elettorale. L'idea originaria (1943) era che il sistema capitalistico non

potesse rimanere per troppo tempo in condizioni di crescita e piena occupazione, perché altrimenti i lavoratori acquisirebbero troppa forza contrattuale e tenderebbero a "sfuggire al controllo", per cui "i capitani d'industria sarebbero ansiosi di dar loro una lezione". La presenza di tensioni inflazionistiche e le pressioni delle grandi imprese sui governi porterebbero inevitabilmente all'adozione di politiche restrittive. Ne seguirebbe la depressione, nel corso della quale si verrebbe a ristabilire la "disciplina del lavoro". Ma l'aumentata disoccupazione provocherebbe una forte pressione delle masse per obbligare i governi a intraprendere politiche espansive. Cambiamenti nel tipo di pressioni politiche darebbero così origine a oscillazioni attorno a un trend di crescita vicino alla piena occupazione.

Le teorie più recenti prendono dall'impostazione originaria l'idea che cause politiche possano provocare fluttuazioni, ma la sviluppano con logica diversa. I politici vogliono vincere le elezioni; gli elettori sono influenzati, al momento del voto, dalle condizioni economiche del momento; perciò, in prossimità di elezioni, i politici attueranno politiche espansive in modo da andare alle urne in condizioni di elevati livelli dell'occupazione e dei redditi, senza che ancora si siano realizzati rilevanti aumenti dei prezzi. Passate le elezioni, ci sarà la necessità di misure restrittive, ma i politici non se ne preoccupano più di tanto perché contano sulla memoria corta dell'elettorato.

Dal punto di vista formale l'analisi del ciclo elettorale potrebbe non richiedere alcuna nuova strumentazione. Basterebbe valutare le conseguenze di variazioni nelle variabili esogene (ad esempio nella spesa pubblica). Dal punto di vista sostanziale si ritrova invece l'idea di Marx e Schumpeter che l'analisi economica del ciclo non può essere avulsa dall'analisi dell'evoluzione sociopolitica del sistema capitalistico.

I modelli preda-predatoreL'idea che il capitalismo generi contraddizioni al suo interno, perché è dominato da una tendenza all'espansione eccessiva e perché ha bisogno di periodi di disoccupazione che servono a ripristinare la profittabilità degli investimenti e la disciplina nel mondo del lavoro, ha avuto da parte di R.M. Goodwin una formulazione molto interessante, che ha aperto la strada a un'ampia messe di contributi, non pochi dei quali di autori italiani.

Il capitalismo è caratterizzato da un accentuato conflitto tra lavoratori e capitalisti per la distribuzione del reddito. La dinamica della distribuzione influenza le decisioni di investimento e quindi la crescita e questa, a sua volta, influenza la distribuzione. Quando i salari aumentano e i profitti calano, i capitalisti riducono gli investimenti; aumenta quindi la disoccupazione e si ha un contenimento dei salari. Il contrario accade quando i profitti aumentano. Per descrivere formalmente questo tipo di dinamica, Goodwin fa ricorso alla teoria matematica della lotta per la vita sviluppata da Vito Volterra con riferimento alle specie biologiche. Se in un dato territorio convivono prede (pesci piccoli) e predatori (pesci grandi), si possono originare movimenti ciclici perché, quando i predatori sono numerosi, le prede diventano sempre più rare cosicché i predatori, per mancanza di cibo, dovranno ridursi. Ma allora le prede, non più troppo cacciate, tornano ad aumentare. Migliora perciò la situazione per i predatori che possono espandersi nuovamente, e così via. Basta porre l'analogia tra profitti e predatori e tra salari e prede per comprendere la logica fondamentale dell'originario (1967) modello di Goodwin.

Nelle fasi di espansione l'interazione tra moltiplicatore e acceleratore produce una crescita troppo rapida, che assorbe la disoccupazione, fa aumentare i salari più della produttività e determina quindi una riduzione della quota dei profitti. Si hanno allora un deterioramento delle aspettative, un rallentamento e quindi una caduta degli investimenti. Ne deriva un aumento della disoccupazione che ha l'effetto di ridurre il potere contrattuale dei lavoratori e di bloccare la crescita salariale. Dopo qualche tempo, per effetto del progresso tecnico, vengono a ristabilirsi condizioni di profittabilità. Il sistema ricomincia allora a espandersi con velocità crescente.È importante sottolineare come, secondo Goodwin, la struttura del capitalismo sia rappresentata da un meccanismo omeodinamico attivato da variazioni nelle quote distributive. Ne risulta un andamento ciclico di breve periodo attorno a un trend di crescita esponenziale. Le quote distributive oscillano attorno a un livello costante di lungo periodo e lo stesso succede per il margine di disoccupazione.

Nel modello originario il ciclo permane e non esplode perché il sistema di equazioni differenziali che lo compone rappresenta un oscillatore (unilaterale, perché c'è solo il limite della piena occupazione) non lineare, di tipo conservativo. Il modello è però strutturalmente instabile, in quanto piccole variazioni nei suoi parametri possono dar luogo a soluzioni

qualitativamente diverse da quella di oscillazioni costanti. È stato però possibile superare questa difficoltà a seguito di opportune riformulazioni delle ipotesi e a prezzo dell'utilizzazione di strumentazioni matematiche più complesse (v. ad esempio Medio, 1979).

Goodwin ha volutamente presentato un'analisi molto stilizzata e semplificata del funzionamento del capitalismo, confidando che in seguito si sarebbe potuto sviluppare il modello su basi più realistiche. Qualche cosa in questo senso è stato fatto nelle trattazioni più recenti e si sono ottenuti risultati interessanti (v. Vercelli, 1982), più però sul piano formale che su quello sostanziale. Un aspetto in particolare richiede ulteriori approfondimenti: quello, consueto, dell'avvio della ripresa. La spiegazione che danno i modelli preda-predatore appare ancora troppo meccanicistica. La ricostituzione di condizioni di profittabilità dopo una depressione e il miglioramento delle aspettative avvengono perché il tempo passa e il progresso tecnico fa aumentare esogenamente la produttività più dei salari. È forse troppo poco per spiegare l'avvio di una ripresa destinata, in breve volger di tempo, a prendere velocità sostenuta e ad assumere carattere autopropulsivo. Occorre probabilmente una spinta di intensità maggiore, quale potrebbe provenire da un massiccio sforzo di investimenti collegato all'introduzione di innovazioni di grande portata, di quelle che 'fanno epoca' e hanno perciò l'effetto di modificare in modo significativo le modalità di funzionamento del sistema.

Si ritrova così la visione del capitalismo presentata da Marx e Schumpeter con finalità diverse: per mostrarne gli aspetti negativi il primo e quelli positivi il secondo. Il capitalismo è un sistema strutturalmente instabile che cambia continuamente, talvolta in modo decisamente brusco e perciò discontinuo. Per tener conto di questo aspetto, alcuni studiosi stanno ora considerando la possibilità di utilizzare la teoria matematica delle catastrofi, sviluppata da René Thom per studiare sistemi che si evolvono in modo continuo e che all'improvviso, quando appunto si ha la catastrofe, cambiano radicalmente tipo di evoluzione dando origine a nuove forme di vita (morfogenesi, in biologia) o, più in generale, a nuove forme di comportamento. I risultati ottenuti su questo promettente terreno di ricerca sono però ancora troppo esigui per giustificarne una trattazione più ampia.

Crisi economica e finanziaria

di Pierluigi Ciocca e Gian Enrico Rusconi

Crisi economica e finanziaria
di Pierluigi Ciocca

sommario: 1. Il concetto. 2. Le crisi nella storia: a) la mappa delle crisi; b) crisi mancate o sventate. 3. Le cause delle crisi. Tre scenari di teoria: a) il filone Thornton-Bagehot; b) speculazione e ciclo creditizio; c) la dinamica del capitalismo e le crisi. 4. La probabilità delle crisi: crescente o decrescente? □ Bibliografia.

1. Il concetto

Le crisi sono qui intese come fasi in cui, nelle economie di mercato, una contrazione profonda e non breve delle attività produttive e d'investimento si unisce a rapide, forti flessioni nei valori di cespiti patrimoniali, 'reali' o finanziari. Fra le manifestazioni di mal funzionamento, non ci si riferisce quindi a quelle indicate con termini generici che, sia pure in vario grado, s'applicano a ogni sistema economico-sociale: difficoltà, instabilità, trasformazione, decadenza. Ma non ci si riferisce neppure, fra le manifestazioni più propriamente tipiche del capitalismo industriale dell'età contemporanea, teorizzate dalla scienza economica, a quelle ricomprese in altri concetti e termini specifici: ciclo, recessione, depressione, ristagno, stagflazione. I fenomeni d'instabilità che queste diverse categorie logiche rappresentano sono legati da molteplici nessi; solo i principali, concernenti in modo più diretto le crisi, potranno essere richiamati in questa voce. Le ripercussioni sociali e politiche delle crisi economiche possono essere estremamente gravi. Nascono dallo spreco, irreparabile, delle risorse disponibili ma non utilizzate; dal rischio che l'intensità della contrazione rallenti la crescita dello stesso potenziale produttivo attraverso cui soddisfare, nel lungo periodo, i bisogni della società civile; dal trauma della perdita del lavoro per chi lo aveva e dalla prospettiva della inoccupazione per i giovani; dalle spinte disgregatrici che il calo delle attività economiche e le mutazioni di prezzo dei cespiti patrimoniali imprimono attraverso una redistribuzione di reddito e di ricchezza arbitraria, incontrollata.
Il crinale della crisi si situa fra la dimensione 'reale' e la dimensione monetaria dell'economia. Nei sistemi di mercato quel crinale è reso strettissimo dalle aspettative dei

produttori-investitori, da un lato, e dei risparmiatori, dall'altro. Sebbene in questi casi sia meno probabile, la crisi può scoppiare anche se l'economia non è entrata nella fase discendente di un ciclo, o non versa in uno stato di depressione, o non subisce il prevalere di tendenze al ristagno. La mutevolezza delle aspettative e la centralità che esse assumono nel processo economico possono provocare cadute di produzione e forti scompensi nei mercati monetari e finanziari, autonomamente ovvero amplificando spinte in atto. Per la natura stessa di queste forze, nelle economie capitalistiche le crisi, sempre possibili, non sono affatto periodiche o ineluttabili. Dai primi tentativi di inizio Ottocento sino alle più recenti teorie statistico-matematiche la scienza economica è stata costantemente impegnata a identificare i fattori capaci di provocare le crisi, la politica economica a prevenirle e a circoscriverne le conseguenze.In un'economia di mercato sviluppata è diffusa la detenzione, e sono intensi gli scambi, di beni patrimoniali (terreni, fabbricati, mezzi di produzione) e di valori mobiliari. L'una e gli altri trovano alimento nel credito, bancario e non bancario. Allorché prevale l'attesa che cespiti reali e titoli diminuiscano di prezzo, vi sarà una corsa a venderli e ad acquisire moneta (oro, argento, biglietti di Stato o degli istituti di emissione, depositi presso le banche più solide), per detenerla o per rimborsare i debiti prima che le passività eccedano le attività, provocando l'insolvenza. Gli investimenti produttivi possono risentire fortemente delle peggiorate previsioni di profitto, del più elevato costo del capitale, della convenienza ad acquistare i beni strumentali che già esistono, meno cari di quelli di nuova produzione; il calo della ricchezza e del reddito, il deterioramento delle prospettive di lavoro contraggono i consumi. Il ridursi della domanda globale provoca la flessione del prodotto, degli scambi con l'estero, dell'occupazione, dello stesso livello generale dei prezzi. Le spinte recessive interagiscono sinché forze di segno opposto, spontaneamente espresse dal sistema o sollecitate dall'intervento della politica economica, non arrestano e invertono il movimento. La sequenza attraverso cui il processo si svolge e la misura in cui l'economia viene colpita nei suoi diversi aspetti variano in relazione a un'ampia gamma di elementi. In questo senso ogni crisi presenta una propria specificità di forme e si propone pertanto quale oggetto d'indagine all'analisi storica, prima ancora che a quella teorica. Una sequenza stilizzata viene riprodotta nella tabella. Compilata (v. Fisher, 1933) sulla base dell'esperienza accumulata sino agli anni trenta, essa dà l'idea della complessità delle forze in gioco, sebbene l'accento posto sull'indebitamento e sulla flessibilità dei prezzi verso il basso (deflazione dei debiti) sia apparso a più d'uno studioso, in seguito, eccessivo.Va sottolineata la natura intrinsecamente

monetaria delle economie capitalistiche. In tali economie il circuito produttivo è messo in moto ed è diretto dagli imprenditori. Essi hanno la proprietà, o il controllo, degli impianti, dei macchinari, dell'organizzazione della produzione. Hanno, inoltre, la disponibilità di ampi mezzi finanziari: possiedono moneta o possono procurarsela a credito, attraverso le banche, gli altri intermediari finanziari, la borsa. Con questi mezzi liquidi è possibile pagare i salari ai lavoratori prima ancora di aver prodotto e venduto la merce. La vendita dei prodotti e il denaro così incassato devono consentire poi il rimborso dei prestiti e il pagamento degli interessi ai creditori, l'ottenimento di un profitto netto - scopo ultimo della produzione, realizzato in forma monetaria attraverso l'accumulo di una ricchezza fungibile - e l'inizio di un nuovo ciclo su scala più larga.La moneta e il credito sono essenziali allo svolgersi di questo ciclo produttivo, a cui ne corrisponde uno monetario e creditizio. Tendono quindi ad assumere le caratteristiche tecnico-organizzative richieste da un mercato improntato alle esigenze delle imprese e condizionato dalle loro scelte. E tuttavia, da formidabile leva di sviluppo, moneta e credito possono trasformarsi in fattore di acuta instabilità dell'intera economia. Entrano, direttamente o indirettamente, nella produzione di tutte le altre merci. Ma il tesoreggiamento può essere preferito all'investimento produttivo; il credito a chi produce può così rapidamente venir meno. Ancorché tutelata da vincoli contrattuali e da norme di legge, la fitta rete di obbligazioni finanziarie si fonda da ultimo sulla fiducia, riferita sia all'intero mondo degli affari, sia alla specifica attività e al particolare debitore. Lo scemare della fiducia riduce la propensione agli investimenti, o a finanziare a lunga scadenza e a costi accettabili chi li effettua; accresce la propensione a detenere mezzi liquidi, quale precauzione di fronte all'incertezza e al rischio; può provocare effetti di contrazione a catena, tanto più intensi ed estesi quanto più brusco è il calo di fiducia e minore è la misura in cui la domanda straordinaria di mezzi liquidi viene soddisfatta.
Uno scompenso monetario e finanziario è quindi parte integrante della crisi, sia quale manifestazione sia quale concausa. Storicamente, alle recessioni produttive profonde si sono associati dissesti e forte instabilità nel sistema bancario, nei mercati monetari e finanziari. Ciò spiega anche perché i banchieri siano stati fra i primi a comprendere la natura e la gravità del problema, e le banche centrali fra le prime istituzioni a ricercare, attraverso il 'credito di ultima istanza' accordato alle banche in difficoltà, una risposta di politica economica.

2. Le crisi nella storia

La consapevolezza dell'esposizione del sistema al rischio di crisi è documentabile presso i 'pratici', operatori dell'industria e della finanza, già fra lo scorcio del Settecento e gli inizi dell'Ottocento. Quella consapevolezza si fondava sulle esperienze d'instabilità del mondo degli affari nella fase dell'ascesa e dell'affermarsi del capitalismo industriale, segnatamente nell'Inghilterra del XVIII secolo. Essa si estese poi nel secolo successivo ai governanti e agli studiosi, di fronte al dato, ovvio ma illuminante, che le crisi avvenivano: sperimentate più spesso da singoli paesi, talvolta, e con sincronismo tendenzialmente crescente, dal complesso delle economie di mercato.Una visione d'assieme è stata a lungo ostacolata da carenze statistiche. Mentre le cronache coeve e gli stessi resoconti analitici successivi delle vicende monetarie e finanziarie abbondano, sono a lungo mancati dati quantitativi di sintesi, sia pure di prima approssimazione, sull'attività produttiva. Contributi quali quelli di Bairoch, Maddison, Mitchell, consentiti dalla costruzione di serie di contabilità nazionale relative a ciascuna economia, hanno colmato in par te la lacuna: per l'Europa a partire dal 1830, per gli Stati Uniti e pochi altri paesi a partire dal 1870. Al tempo stesso, per la storia d'Europa, ma con ampio riferimento agli Stati Uniti, Kindleberger ha censito le fasi di tensione finanziaria in senso stretto, definite come "il deterioramento - acuto, concentrato nel tempo, metaciclico - dell'insieme o di un sottoinsieme di indicatori finanziari" (v. Kindleberger, 1987, p. 339). Se a ognuno di questi casi di tensione finanziaria si accostano i dati ora disponibili sull'attività produttiva, diviene possibile tracciare una mappa delle crisi, 'reali' e finanziarie, nei secoli XIX e XX.

Emerge, quale primario elemento fattuale, che nelle attività economiche le fasi di caduta forte, non breve, non limitata all'accidente dei cattivi raccolti, racchiudono tensioni finanziarie, ovvero sono da queste di poco precedute o seguite. Forse meno scontata è una convinzione da sempre espressa dai banchieri centrali: più d'una crisi, che avrebbe potuto avvenire, sarebbe stata evitata attraverso misure di sostegno. Non di rado gli interventi sarebbero riusciti a prevenire gli stessi focolai dell'instabilità. Naturalmente, manca la controprova: la storia con i 'se', controfattuale, non è forse impossibile, ma è certo difficile. Nondimeno, in altri casi le tensioni finanziarie si sono palesate e hanno teso a diffondersi, ma non sono arrivate a intaccare l'attività produttiva, anche grazie all'azione di un prestatore di ultima istanza, o ad altri provvedimenti correttivi.

a) La mappa delle crisi

Se si prescinde dai periodi bellici e postbellici, le crisi su scala internazionale più gravi, che hanno segnato la storia economica dell'età contemporanea, si situano negli anni 1833-1842, negli anni settanta sempre dell'Ottocento, negli anni trenta del Novecento.Durante il primo di questi tre periodi, il prodotto lordo in termini reali diminuì in Europa del 9% nel 1833, del 4% nel 1841; in entrambe le occasioni l'attività economica restò per circa un biennio sui più bassi livelli toccati. Negli Stati Uniti la contrazione produttiva probabilmente si limitò al settore dei beni strumentali. Le tensioni finanziarie divennero acute alla fine del 1836 nel Regno Unito, nel giugno e nel settembre del 1837, rispettivamente, in Francia e negli Stati Uniti. Nel Regno Unito si arrestò la tendenza a creare nuove società bancarie; divennero frequenti i fallimenti di banche, specialmente di quelle più esposte con i settori cotoniero e ferroviario; la fiducia subì un crollo generale. Negli Stati Uniti il panico finanziario del 1837 si estese al mercato azionario, con una catastrofica flessione dei corsi, connessa anche con la caduta dei prezzi dei terreni, in particolare di quelli utilizzati per la produzione di cotone; fra il 1838 e il 1843 la quantità di moneta diminuì di un terzo, a seguito dei ritiri dei depositi bancari, il numero delle banche di un quarto. Attraverso una rete di relazioni commerciali e finanziarie che aveva già raggiunto dimensioni internazionali, la crisi inglese e quella americana ebbero ripercussioni sul sistema creditizio e sul mercato monetario della Francia, del Belgio, della Prussia.

Negli anni settanta del secolo scorso prese l'avvio una tendenza cedente dei prezzi che ha a lungo indotto a connotare l'intero periodo 1873-1896 come periodo di 'grande depressione' internazionale. Gli studi più recenti hanno invece negato che alla deflazione si sia unito, nell'intero ventennio, un generale ristagno economico. E tuttavia gli anni settanta furono segnati da fasi di caduta del prodotto lordo in Europa - non solo nel 1870, in connessione col conflitto franco-prussiano (-7%), ma anche nel 1875-1876 (-4% nel biennio) e nel 1879-1880 (-7% sempre nel biennio) - e dalla fase di crescita zero (il 1874-1875) di una giovane economia in forte, tendenziale sviluppo, come quella degli Stati Uniti. Nella sfera finanziaria le difficoltà più acute su scala mondiale presero l'avvio nel 1873. La tensione che allora si determinò nelle principali piazze del mondo fu probabilmente la più acuta ed estesa del secolo. Essa cominciò a Vienna, in maggio; toccò altri paesi europei, come l'Olanda, l'Italia, il Belgio; poi si estese agli Stati Uniti, in settembre, e infine al Regno Unito, alla Francia, alla Russia. Furono severamente colpiti gli intermediari bancari e finanziari esposti con i settori

delle ferrovie, dei prodotti primari, dell'edilizia. Ma il malessere finanziario, esploso nel 1873, proseguì negli anni successivi. Sul fronte bancario i dissesti furono numerosi ovunque. Negli Stati Uniti il numero complessivo delle banche commerciali - che era raddoppiato, da 1.500 a 3.000, nel periodo 1863-1873 - diminuì di 200 unità nel 1874 dopo la moratoria del 1873 e vide la sua crescita arrestarsi per circa un decennio. In Italia, fra il 1874 e il 1879, il numero delle società bancarie per azioni scese da 143 a 101, il loro capitale nominale complessivo da 792 a 269 milioni di lire dell'epoca. Profondità e durata dello squilibrio finanziario emergono soprattutto nella caduta dei corsi azionari. Dopo la chiusura di Wall Street, dal 20 al 30 settembre 1873, i mercati borsistici vissero ripetuti momenti di panico, dominati come furono da "vendite di titoli massicce e concentrate, con considerazione solo secondaria per il prezzo di realizzo" (v. Morgenstern, 1959, p. 543). Da allora i corsi restarono in flessione per diversi anni in quasi tutte le borse: la discesa fu del 40% negli Stati Uniti, sino al 1877, del 60% in Germania, sino al 1877-1878, del 35% nel Regno Unito, sino al 1879, del 30% in Italia, sino al 1877.

La crisi degli anni 1929-1932 prospetta un nesso ancor più evidente fra dimensione 'reale' e dimensione finanziaria, nel disastro che allora colpì il complesso delle economie di mercato, incluse quelle periferiche, e che in Europa concorse all'ascesa del nazismo. Se si considera il prodotto lordo complessivo di sedici paesi - i principali paesi dell'Europa occidentale, gli Stati Uniti, il Giappone, l'Australia, il Canada - le statistiche registrano una caduta continua per un triennio, pari, rispetto al livello del 1929, al 17% nel 1932; la produzione tornò ai livelli del 1929 solo nel 1936. La disoccupazione colpì la forza lavoro, anch'essa in diminuzione, in una misura compresa, nei diversi paesi, fra il 15 e il 30%. Il volume del commercio mondiale si contrasse più del reddito, di circa un quarto, a seguito delle barriere, tariffarie e non tariffarie, introdotte quasi ovunque. I movimenti internazionali dei capitali a medio termine furono prossimi ad annullarsi. La grande depressione degli anni trenta passò alla storia come 'crisi del 1929', in ragione del fatto che sul fronte finanziario essa prese le mosse dal crollo del mercato azionario di New York nell'ottobre di quell'anno (v. Galbraith, 1955). Dopo un calo del 5% il giorno 3 e del 6% il 23, i corsi precipitarono il 28 e il 29: la flessione del 23% in questi ultimi due giorni è rimasta la più acuta registrata in cento anni negli Stati Uniti, sino alle due giornate operative 16 e 19 ottobre 1987, allorché l'indice accusò una discesa del 26%. Ma anche in Europa la fase acuta della recessione coincise con i dissesti della Creditanstalt, la prima banca d'Austria, nel maggio del 1931, e della tedesca

Danat-Bank, in giugno, oltre che con forti pressioni sia sulla sterlina, ampiamente convertita in oro dai non residenti, sia sul mercato bancario e monetario interno inglese, in luglio. L'eccesso d'offerta dei titoli si estese a tutti i mercati azionari: rispetto al 1929, nel 1932 i corsi erano diminuiti del 90% negli Stati Uniti, del 60% in Germania, del 70% in Francia, del 50% nel Regno Unito. Difficoltà, fallimenti, fusioni, dure razionalizzazioni non risparmiarono le banche di nessun paese. Nell'economia più colpita, quella degli Stati Uniti, nel 1930-1932 lo stock di moneta, prevalentemente costituito da depositi bancari, diminuì del 27%, sebbene la base monetaria non avesse subito contrazioni. La flessione del livello medio dei prezzi fu del 31%. Delle 24.000 banche commerciali esistenti 5.000 fallirono; nel complesso, anche attraverso autoliquidazioni e fusioni, il loro numero si ridusse del 40%; i dissesti si estesero alle casse di risparmio e alle compagnie di assicurazione; fu necessario riedificare la struttura bancaria e finanziaria dalle fondamenta, a partire dalla generale chiusura delle banche proclamata dal presidente Roosevelt il 9 marzo del 1933. Ma il travaglio del sistema bancario non fu minore, ad esempio, in Italia, dove il numero delle aziende di credito diminuì da 4.657 a 2.042 fra il 1926 (l'anno d'inizio delle difficoltà per l'economia italiana, con l'apprezzamento della lira verso 'quota 90' rispetto alla sterlina) e il 1936 (l'anno della nuova 'legge bancaria', tuttora vigente).

Questi tre periodi storici si staccano dagli altri per entità di manifestazioni e vastità di ripercussioni. Ma altre crisi generali ebbero luogo, ancorché estese a un numero più limitato di paesi e di più breve durata. Un esempio lontano nel tempo è quello dell'Europa nel 1847-1848. Il fatidico anno 1848 vide il prodotto lordo diminuire del 7%. Nel biennio 1847-1848 decine di ditte industriali e di case bancarie fallirono nel Regno Unito e in Francia, altre in Olanda, in Prussia, in Italia. La tensione monetaria indusse a sospendere il Bank act in Inghilterra nell'ottobre 1847; richiese un sostegno di riserve auree alla Banca di Francia da parte della Banca d'Inghilterra. Ma la crisi, oltre a essere relativamente breve, venne acuita da fattori speciali: la speculazione sui cereali indotta da raccolti eccezionalmente cattivi nel 1845-1846 e molto migliori nel 1847, le forti oscillazioni nei prezzi delle scorte, le convulsioni politiche nel continente, il timore di una rivolta cartista in Inghilterra.

Un esempio meno lontano è quello della crisi nell'area dell'OCSE, innescata nel 1973-1975 dal brusco rialzo del prezzo del petrolio e dalle fluttuazioni del dollaro. La crescita si arrestò nel 1974, e nel 1975 si trasformò in una flessione del volume degli investimenti privati non

residenziali dell'8% e in una, più lieve (0,2%), del prodotto lordo. I sistemi bancari del Regno Unito (insolvenza delle fringe banks londinesi), della Repubblica Federale Tedesca (fallimento della Herstatt) e degli Stati Uniti (fallimenti della Franklin Bank e della Security National Bank, instabilità estesa a grandi gruppi bancari) attraversarono momenti difficili, che vennero tuttavia superati, sia pure grazie anche a massicci interventi di sostegno da parte delle autorità monetarie e di supervisione bancaria.Qualora si facesse riferimento a singoli paesi, piuttosto che all'insieme delle economie di mercato, la mappa delle crisi, 'reali' e finanziarie, si amplierebbe notevolmente. Fra gli episodi più noti relativi alle principali economie vi sono quelli degli Stati Uniti nel 1893-1894 e nel 1908, della Francia nel 1883-1885, del Regno Unito nel 1892-1893, dell'Italia nel 1889-1894, gli "anni più neri" dell'economia italiana (v. Luzzatto, 1968).

b) Crisi mancate o sventate

Sono quindi rari, forse inesistenti, i casi in cui una contrazione produttiva non sia stata accompagnata da tensioni bancarie e finanziarie. Meno rari sono quelli in cui gli scompensi manifestatisi nel mondo della banca e della finanza non si sono risolti in una contrazione estesa al reddito reale, agli investimenti produttivi, all'occupazione. Nella maggior parte di questi ultimi casi la crisi venne tuttavia circoscritta da avvenimenti, più spesso da provvedimenti, che ristabilirono in tempo la fiducia. Gli espedienti più comuni per arrestare il panico finanziario, precedenti e poi talora complementari a una politica monetaria incentrata sul credito di ultima istanza, furono la sospensione della convertibilità dei biglietti in metallo prezioso, la moratoria dei pagamenti, la chiusura temporanea di banche e borse, l'offerta di garanzie ad hoc da parte dello Stato o da parte di associazioni di banche, l'assicurazione dei depositi.

Intercorrono circa due secoli fra l'esempio di Londra nel 1793 e quello di Wall Street nel 1987, ma l'intervento esterno fu in entrambe le occasioni risolutivo. Nel 1793 il panico si diffuse, nella City e nel paese, a seguito del fallimento di banche operanti fuori Londra; la domanda dei biglietti della Banca d'Inghilterra crebbe a dismisura, per ragioni precauzionali; la stretta monetaria si fece rapidamente durissima. La catena dei dissesti venne però spezzata, e non si estese al mondo della produzione: bastò allo scopo il mero annuncio della decisione del Parlamento di mettere a disposizione dei mercanti solvibili titoli a breve dello Scacchiere, facilmente liquidabili. Nell'ottobre del 1987 la 'guerra' dei tassi d'interesse fra Stati Uniti,

Giappone e Germania Federale, a sostegno dei cambi delle rispettive monete, sospinse il tasso d'interesse a lungo termine molto al di sopra del rendimento dei titoli azionari, sopravvalutati; al crollo da primato storico dei corsi azionari a Wall Street fecero seguito quelli delle altre principali borse. Ma il cedimento venne contrastato dalla rapida correzione di segno della politica monetaria e dalla dichiarata determinazione delle banche centrali a impedire, con ampia immissione di liquidità nei mercati, che l'attività produttiva venisse pregiudicata.
Se gli episodi locali di tensione finanziaria rientrata, per accidente o per intervento esterno, sono stati non infrequenti (di nuovo l'Inghilterra nel 1797, 1810, 1825, 1866, 1890; la Francia nel 1818; l'Italia nel 1907), sono solo sporadici quelli in cui la tensione bancaria e finanziaria ebbe pieno corso, con manifestazioni acute e diffuse fra gli intermediari, senza tuttavia associarsi a una contrazione delle attività produttive. Almeno un episodio va richiamato. Il 1857 fu un anno di crisi finanziaria di notevole gravità, di vasta portata internazionale. Essa prese l'avvio in agosto dal tracollo della speculazione sui titoli ferroviari negli Stati Uniti; in poche settimane coinvolse le ditte inglesi più impegnate negli scambi con gli Stati Uniti e quelle di Amburgo. In Inghilterra venne elevato sino al 12% il tasso di sconto e fu sospeso il Bank act. Ciò creò il panico nella Borsa di Parigi e tutti i centri finanziari dell'Europa centro-settentrionale ne risultarono scossi. Nonostante tutto nel periodo 18561858 il prodotto lordo dell'Europa si accrebbe considerevolmente.

3. Le cause delle crisi. Tre scenari di teoria

Il laboratorio della storia offre, dunque, una casistica variegata: prova che le crisi sono un'eventualità più che possibile; mette in luce il nesso fra l'acutezza delle crisi più gravi e la loro duplice dimensione, a un tempo 'reale' e finanziaria; esclude che una tensione finanziaria, pur forte, si risolva inevitabilmente in una contrazione profonda dell'attività produttiva; suscita la questione delle politiche economiche anticrisi. Sollecitata dai fatti della storia, l'analisi economica ha ricondotto a tre fondamentali famiglie di teorie l'identificazione dei fattori e delle condizioni che tendono a provocare la crisi e dei modi per prevenirla o superarla.

a) Il filone Thornton-Bagehot

Il primo filone è quello che nell'Ottocento, con Thornton e con Bagehot, diede nobiltà teorica e veste istituzionale alla pratica del credito di ultima istanza. Secondo questa linea d'analisi, in un'economia monetaria, ancorché non soggetta a fluttuazioni endogene, la fiducia può venir meno. Se non viene ripristinata, può aver luogo una crisi: finanziaria prima, a seguito della stretta monetaria risultante dal tesoreggiamento e dalla minor velocità di circolazione della moneta, 'reale' poi, a seguito del diffondersi del pessimismo e del più elevato costo del credito per i produttori. Un prestatore di ultima istanza può ristabilire la fiducia: con la sua disponibilità a far credito, prima ancora che con l'effettiva immissione di mezzi monetari nel mercato.

Lo schema è di una chiarezza cristallina, nella sua semplicità. Implica un concetto d'instabilità analogo a quello che il senso comune applica di fronte all'equilibrista che percorre il filo teso fra i due campanili di una piazza. L'uomo può cadere per mille motivi. È opportuno quindi usare una rete di protezione, che, dandogli sicurezza, ridurrà le stesse probabilità di caduta.

Nella versione originaria di Henry Thornton, "le cause che possono far variare la rapidità di circolazione delle banconote sono molteplici. In generale, una situazione di elevata fiducia favorisce un aumento della velocità di circolazione, [...] contribuirà a far sì che occorrano minori riserve per far fronte agli imprevisti. In una situazione siffatta, ciascuno confida che, se una richiesta di pagamento, oggi dubbia ed eventuale, dovesse effettivamente pervenirgli, sarebbe possibile farvi fronte al momento: si è riluttanti all'idea di incorrere nei costi connessi con la vendita di merci o con il presentare una cambiale allo sconto con lo scopo di costituirsi una riserva con molto anticipo rispetto a quando sarà necessaria. All'opposto, allorché sopraggiunge una stagione di sfiducia, la prudenza consiglia di non dar peso alla perdita di interesse risultante dal detenere banconote per qualche giorno in più. [...] [Nel 1793] a diffondere una sensazione di generale solvibilità bastò l'aspettativa di un'offerta di buoni dello Scacchiere, che quasi ogni commerciante avrebbe potuto ottenere e, come era risaputo, vendere in cambio di banconote, a loro volta convertibili in ghinee" (v. Thornton, 1802; tr. it., pp. 47-50).

Il sistema bancario, con riserve liquide che rappresentano una frazione di passività a vista e a breve, è particolarmente esposto a cali di fiducia. L'illiquidità e la temuta insolvenza di una banca possono provocare ritiri di depositi, panico finanziario, dissesti aziendali, decurtazione d'attività nell'intera economia. I banchieri hanno da sempre ravvisato nella natura

strettamente bilaterale del rapporto di fido, fra il cliente finanziato e la banca, e nella mobilità dei depositi che la banca si è impegnata a restituire le ragioni per cui le probabilità che la crisi di una unità scuota la credibilità generale sono più alte nel sistema bancario che in ogni altra industria. Nel linguaggio della teoria economica odierna (Arrow) i contratti di debito-credito, di necessità proiettati nel futuro, sono intrinsecamente contrassegnati dalla con-certezza del loro rispetto e da carenze d'informazione. La fiducia, da ultimo fondata sulla consuetudine di rapporti fra banca e cliente, integra e sostituisce le informazioni trasmesse attraverso i mercati, là dove esse sono insufficienti, asimmetriche, mancanti. La completezza dell'informazione sui debitori non v'è neanche da parte delle banche nei confronti dei clienti usuali. Ancor meno può esservi da parte del mercato in generale. Il mercato secondario, sviluppato nel caso dei titoli azionari e obbligazionari, è sostanzialmente inesistente nel caso dei fidi bancari.

Proprio perché non fondabile sulla pienezza dell'informazione, il 'credito' di cui le banche e i loro clienti godono può rapidamente e diffusamente dissolversi. La stabilità del sistema bancario e finanziario costituisce la ragion d'essere originaria, permanente, anche se non più la sola, delle banche centrali. Prevenire l'insolvenza di banche solo illiquide, contrastare il pericolo di dissesti a catena: queste sono state, e restano, le primarie responsabilità assunte dalle banche centrali, già nel corso dell'Ottocento.

Elasticità e discrezionalità sono a fondamento del central banking. Esse nascono storicamente nel campo d'attività detto 'credito di ultima istanza': quello dei fidi accordati dalla banca centrale, direttamente e su base bilaterale, a singoli intermediari finanziari. Lo strumento del credito di ultima istanza deve esistere per dare al sistema bancario la sicurezza che sarà arrestato l'effetto 'domino' dell'illiquidità, con crisi a catena; al tempo stesso, le singole banche non devono aver mai la certezza d'accedervi, che le indurrebbe a comportamenti imprudenti e lassisti.

La protezione dall'illiquidità non va data agli insolventi: qualità ed efficienza dell'attività bancaria e finanziaria sarebbero altrimenti minate alla radice. Se solvibile e riconosciuto tale, l'intermediario finanziario illiquido non incontrerà difficoltà, purché il mercato monetario sia ragionevolmente concorrenziale. Altrimenti, mancando queste due condizioni, occorrerà un intervento esterno. Dovrà effettuarlo un soggetto istituzionale dotato delle necessarie risorse, capace di sceverare, eventualmente anche attraverso gli strumenti della vigilanza sulle banche, i casi di illiquidità da quelli in cui essa è commista con l'insolvenza. Qualora vi sia

sospetto di insolvenza, la valutazione che la banca centrale è chiamata comunque a dare, spesso in tempi brevi e sotto la pressione degli eventi, deve estendersi alla probabilità che negare il sostegno estenda la tensione. Se questa probabilità viene giudicata alta, le conseguenze di una crisi economica devono essere poste a confronto con le inefficienze che nel più lungo periodo deriveranno dal fare eccezione alla regola di non rifinanziare intermediari insolventi. È evidente la delicatezza di una siffatta valutazione costi-benefici da parte della banca centrale, stretta fra i due estremi di non sostenere gli intermediari mal gestiti e di non consentire che l'instabilità dilaghi.

b) Speculazione e ciclo creditizio

Il tratto caratterizzante dell'indirizzo Thornton-Bagehot è di non inscrivere la crisi in una teoria delle fluttuazioni. Quanto all'impulso generatore della crisi, l'analisi resta aperta alla considerazione di un'ampia gamma di possibili fattori, interni e internazionali, economici e non, esogeni ed endogeni. L'economia è in ogni momento, qualunque sia la sua condizione congiunturale e strutturale, soggetta a turbative della fiducia. L'acutezza delle ripercussioni viene quindi a dipendere, oltre che dalla natura e dall'intensità dell'impulso destabilizzante, dallo stato dell'economia in quel particolare momento.

Questa apertura analitica, che pure per un verso costituiva il pregio dello schema Thornton-Bagehot, suscitò insoddisfazione in chi cercava di delineare un meccanismo di instabilità radicato nel modus operandi del sistema economico: un meccanismo capace di generare la crisi, o il rischio della crisi, in modo endogeno e ricorrente, ancorché non necessariamente con periodica regolarità. La frequenza delle fluttuazioni economiche e il loro più sistematico censimento nel corso dell'Ottocento contribuirono a stimolare questo indirizzo di ricerca:

"Nel XIX secolo l'andamento dell'industria e del commercio fu segnato da cicli chiaramente definiti. Prosperità, boom, crisi, recessione e ripresa si susseguirono con una regolarità che fece pensare all'inevitabilità: divenne naturale riguardare la crisi come portato del boom, la depressione come portato della crisi. Nel XVIII secolo i cicli ricorrenti avevano presentato un'evidenza molto minore: vi furono fasi di espansione e di contrazione delle attività economiche e una crisi poteva avvenire, durante tali fasi, in ogni momento. Gli accadimenti connessi con guerre o conflitti interni bastano a dar ragione del panico nel 1701, 1715, 1745, 1778 e forse nel 1797. A chi ne subì le ripercussioni, queste crisi erano apparse non come eventi nell'ordine delle cose, che persone d'esperienza avrebbero potuto attendersi, ma come

fulmini a ciel sereno. Furono (o sembrarono) causa, non conseguenza, del disordine economico" (v. Ashton, 1959, p. 136).

Dagli inizi dell'Ottocento la costante osservabile con maggiore immediatezza nelle fluttuazioni economiche fu rappresentata dalla correlazione degli alti e bassi speculativi con l'espansione e contrazione del credito. Non sorprende, quindi, che in questa coppia di elementi venisse in primo luogo individuata la ragione della frequenza, se non del ricorrere, delle fluttuazioni e, all'interno di queste, delle crisi.

Da un lato, nuove occasioni di rapido guadagno, all'esaurirsi di quelle già sfruttate, si prospettano quasi senza soluzione di continuità a chi è più pronto, stando alla 'specula', ad approfittarne (v. Kindleberger, 1978 e 1987). Gli investimenti speculativi trovano sostegno in un'offerta di credito entro certi limiti elastica. Stimolo agli affari ed espansione del credito si autoalimentano, sino a degenerare in diffusa euforia. La tensione giunge al culmine allorché la stessa elevatezza di prezzo dei beni oggetto della speculazione - prodotti, cespiti 'reali' o finanziari - ingenera il timore di un'inversione di tendenza, ovvero quando l'accesso al credito diviene più difficile e costoso. Il processo allora s'inverte, con un movimento che va dalle attività su cui la speculazione si è incentrata ai mezzi liquidi. Il moto inverso può essere così rapido da risolversi in un panico finanziario, tanto diffuso da provocare, attraverso le aspettative e le interrelazioni interne al sistema economico, una contrazione produttiva profonda.

Dall'altro lato, l'instabilità è intrinseca al settore creditizio, più pronunciata quando questo è basato su un sistema di banche commerciali capaci di 'moltiplicare' il credito (v. Hawtrey, 1932). Il processo di espansione-contrazione del credito può essere in ogni momento avviato da deviazioni rispetto allo stato iniziale, in grado di cumularsi e autoalimentarsi a lungo per poi mutare di segno. Le banche possono incoraggiare la domanda di credito attraverso una deliberata strategia d'espansione della loro attività, fondata su minori tassi d'interesse, o semplicemente tardando a innalzare i tassi d'interesse quando il futuro si configura per le imprese come più redditizio (v. Wicksell, 1934-1935). Viene così stimolata e sostenuta l'attività, speculativa o produttiva, dei soggetti più attenti alle condizioni di finanziamento. Con essa, lo sono l'attività economica in generale e l'ascesa dei prezzi, sino a quando l'offerta di credito non incontra un limite nella ridotta liquidità delle banche. La contrazione può essere non meno cumulativa e traumatica, in specie se prevalgono aspettative di tipo estrapolativo su prezzi e profitti.

Sia nel caso in cui il primum movens è la speculazione, poi assecondata dal credito, sia nel caso in cui è questo a sollecitarla, il rischio di una caduta produttiva e, insieme, di uno squilibrio finanziario - il rischio di una crisi - si accresce una volta superato il limite superiore di un andamento economico considerato come intrinsecamente oscillatorio. Immediato, in questi stessi schemi, è il tipo di risposta a cui la politica economica è chiamata: stabilizzare il processo e attenuare le fluttuazioni. Ciò significava, analogamente a quanto suggerito dal filone Thornton-Bagehot, sostenere il sistema bancario e, attraverso di esso, l'intera economia, nella fase recessiva. Ma significava anche, forse soprattutto, agire per tempo al fine di graduare la spinta espansiva, prevenendone gli 'eccessi'.
La funzione di creditore di ultima istanza assegnata alla banca centrale veniva quasi naturalmente a estendersi sino a trasformarsi in politica monetaria stabilizzatrice, o anticiclica. Attraverso la manovra dello sconto - a cui si unirono poi le operazioni su titoli nel mercato aperto e l'influenza esercitata sulle riserve detenute dal sistema creditizio presso la 'banca delle banche' - la domanda di fondi andava mitigata e l'offerta frenata prima della fase acuta dell'espansione, fermo restando che, di fronte a un panico finanziario e a una contrazione produttiva incipiente o in atto, ogni sostegno dovesse venire assicurato. La moneta doveva essere, di necessità, manovrata, rimuovendo ogni automatismo o limite rigido alla sua temporanea espansione, quale, ad esempio, il gold standard. Non essendo le oscillazioni, per la natura stessa delle forze in gioco, di ampiezza e durata definite o regolari, veniva demandato all''arte' del banchiere centrale di stabilire tempi, modi, dosaggio dell'azione stabilizzatrice.

c) La dinamica del capitalismo e le crisi

Il terzo indirizzo teorico è meno facilmente sintetizzabile, per l'ampiezza della letteratura e per le caratterizzazioni dei principali contributi. Il filo conduttore consiste nel riguardare le fluttuazioni, e le crisi all'interno di queste, come inerenti alla dinamica di fondo delle economie capitalistiche: del capitalismo visto come "sistema che si evolve [...], generatore intermittente ma incessante di mutamento morfologico" (v. Goodwin, 1986, p. 14). Per Marx e per Schumpeter la speciale capacità del capitalismo di sviluppare le forze produttive e di trasformare la società è insita nella spinta con cui il flusso delle innovazioni tecniche si traduce in più alta produttività, attraverso l'accumulazione di capitale motivata dal profitto. Sia Marx sia Schumpeter erano al tempo stesso convinti che il processo fosse caratterizzato

da un'instabilità connaturata, estesa al livello generale dell'attività economica. Il contributo decisivo di Keynes, incentrato sul concetto di domanda effettiva e sull'analisi delle sue determinanti, ha consentito di chiarire come alla base delle fluttuazioni vi siano le decisioni d'investimento degli imprenditori, soggettive, esposte all'incertezza, non coordinabili.Le scelte d'investimento determinano, in quanto spesa, con effetto moltiplicativo, il livello del reddito e il grado di utilizzo degli impianti e del lavoro (v. Kahn, 1931). A capacità pienamente utilizzata, lo sviluppo entra in una regione d'instabilità (v. Harrod, 1973); vi è allora un'alta probabilità che il potenziamento degli impianti, con cui i singoli imprenditori si studiano di corrispondere agli incrementi di domanda attesi globalmente, ecceda quello sostenibile dal risparmio di cui l'economia è capace, ovvero non raggiunga i valori necessari a mobilitare pienamente le risorse non consumate. Il moto dell'economia si fa, ciclicamente ancorché non periodicamente, espansivo e recessivo, entro il limite superiore fissato dalle risorse più scarse e il limite inferiore fissato dagli ammortamenti, da componenti esogene della domanda e soprattutto dalla spinta del progresso tecnico: "Il punto di svolta superiore non è difficile da spiegare: nel mezzo dell'espansione si attiva l'acceleratore e il sistema diviene instabile. I ritmi della crescita si fanno insostenibili. Il limite è nelle risorse e negli alti prezzi relativi del lavoro e delle materie prime. Allora la crescita rallenta e il costo reale del produrre sale, provocando l'inflazione come immediata risposta dei produttori. L'inflazione maschera, ma non può rimuovere, il vero problema: la decelerazione. L'investimento viene tagliato, le aspettative subiscono un crollo drammatico, l'economia diviene stabile al più basso livello dettato dalle spese non sistematiche, come quella governativa o quella proveniente dal commercio estero. Su questo livello inferiore l'economia si attesta, finché un'innovazione adeguata non la risolleva. Si spiegano così le peculiarità storiche delle fluttuazioni. Questi movimenti non sono a stretto rigore dei cicli. Appaiono simili perché la spinta capitalistica al profitto implica un'incessante caccia alle innovazioni capaci di economie nei costi. La crescita, presto o tardi, non può non riavviarsi, ma acquista rapidità eccessiva e torna infine a interrompersi" (v. Goodwin, 1986, p. 21).
È questa l'estrema sintesi di uno schema eclettico Marx-Keynes-Schumpeter, pensatori per altri aspetti diversissimi. L"apporto' di ciascuno di loro allo schema, sebbene di complessa e per taluni aspetti di dubbia individuazione (v. Sylos Labini, 1970), va ricercato nella sottolineatura dei motivi dell'accumulazione capitalistica fatta da Marx; nella centralità che, per Schumpeter, le innovazioni assumono nella dinamica dell'economia; nell'analisi

imperniata su domanda effettiva e incertezza, proposta da Keynes. L'accostamento del loro pensiero in una sintesi richiede almeno due ordini di qualificazioni. La prima consiste nel sottolineare ancor più il carattere monetario e finanziario - da 'Wall Street' - dell'economia. La scissione dell'investimento dal risparmio della singola impresa e il ricorso alla finanza esterna ampliano le oscillazioni degli investimenti. Allorché le "aspettative subiscono un crollo drammatico", l'entità e la diffusione delle attività finanziarie e la fuga verso quelle più liquide e meno rischiose acuiscono la recessione, sino alla crisi. In secondo luogo, le fonti d'instabilità dell'intera economia non si limitano alle forze che interessano direttamente la domanda aggregata e l'offerta aggregata. L'instabilità può essere provocata anche da squilibri all'origine microeconomici: se i meccanismi di allocazione delle risorse non operano al meglio, se vi sono asimmetrie nei movimenti dei prezzi e nei comportamenti dei soggetti, se le difficoltà di alcuni settori dell'economia si generalizzano attraverso le aspettative e le interrelazioni tessute dalla finanza (secondo meccanismi che si inscrivono nel filone marxiano delle 'crisi da sproporzioni'). La stessa capacità di crescita e di trasformazione delle economie capitalistiche acuisce il problema, centrale in ogni economia, di far corrispondere alla mutata composizione della domanda gli opportuni adeguamenti di struttura dell'offerta (v. Pasinetti, 1984).

È evidente come da questo terzo filone d'analisi, e segnatamente dall'apporto di Keynes, sia discesa l'indicazione di una più articolata politica economica volta a stabilizzare l'economia e a prevenirne le crisi, configurabili sempre più come un rischio connaturato al sistema, non rimuovibile. Il suggerimento, ampiamente recepito dai governi, soprattutto dopo la crisi del 1929 e sino agli anni settanta, è stato di unire al credito di ultima istanza e alla politica monetaria della banca centrale una serie di altre politiche: di bilancio e fiscali, valutarie, dei redditi, allocative, strutturali. Si è comunque affermato il convincimento che la recessione e l'eventuale crisi vadano contrastate sostenendo con ogni mezzo la domanda globale.

Le difficoltà degli anni settanta, che l'attivismo della politica economica non sventò, e poi il buon andamento delle economie di mercato industrializzate negli anni ottanta, detti di deregulation, hanno riproposto antichi motivi liberisti, di non intervento dello Stato sull'andamento ciclico dell'economia. La convinzione analitica secondo cui l'operare della mano invisibile che guida i mercati non vada comunque turbato si unì durante tutto l'Ottocento all'idea vagamente puritana secondo cui dagli eccessi della speculazione ci si poteva mondare solo lasciando che la crisi facesse il suo corso. Ancora nel Novecento questi

orientamenti hanno trovato più d'un conforto teorico (v. Hayek, 1931) sino alle elaborazioni che, pur non escludendo interventi nelle crisi più acute, considerano la politica economica attiva come fattore →d'instabilità, causa o concausa delle fluttuazioni e delle stesse contrazioni produttive (v. Friedman, 1959; v. Lucas, 1983). La querelle fra interventisti e astensionisti prosegue irrisolta sul piano analitico. Sul piano politico e istituzionale, di fatto, gli stessi governi più liberisti, sotto la pressione degli eventi e dell'opinione pubblica, sono di rado riusciti a trattenersi dall'agire per cercare di contrastare una grave crisi, in atto o soltanto temuta.All'altro estremo vi è l'indirizzo critico di matrice marxiana. Marx vide nelle crisi ricorrenti la manifestazione delle contraddizioni di fondo del sistema capitalistico, che egli riteneva insuperabili. Il tema del 'crollo' del capitalismo ha da ultimo perso rilievo presso gli stessi studiosi più vicini al pensiero originario di Marx. L'accento è stato variamente posto sui tre modelli di crisi, ricondotta da Marx alla caduta tendenziale del saggio del profitto, al sottoconsumo, alle sproporzioni, all'interno di una visione unitaria che configurava il rapporto fra denaro (D) e merci (M) nel capitalismo secondo la sequenza $D \rightarrow M \rightarrow D' > D$, contrapposta alla sequenza $M \rightarrow D \rightarrow M$, propria delle società precapitalistiche. L'intervento statale può solo procrastinare il progressivo acuirsi di tali contraddizioni, che comunque sollecitano, se non il crollo del sistema, la scelta ideologica e politica di una sua profonda trasformazione (v. Sweezy, 1942). Fra le molteplici versioni che tale indirizzo radicalmente critico ha assunto, l'interesse analitico maggiore, anche attraverso importanti apporti di studiosi non marxisti, è stato probabilmente rivolto alle teorie (v. Kalecki, 1971; v. Steindl, 1952) che legano le depressioni a tendenze al ristagno di lungo periodo espresse dai connotati oligopolistici del capitalismo industriale.

4. La probabilità delle crisi: crescente o decrescente?

La migliorata comprensione della natura dell'instabilità, delle sue manifestazioni; la concezione di una politica economica anticiclica, a cui molti continuano a riconoscere efficacia, e l'approntamento degli strumenti per attuarla; la diffusione di correttivi automatici, in una cornice istituzionale disegnata anche per la stabilità dell'economia; l'irrobustimento del settore bancario e finanziario, favorito dalle legislazioni e dalla vigilanza prudenziale; lo stesso mutarsi della struttura produttiva, da prevalentemente agricola a prevalentemente terziaria: questi e altri elementi hanno tendenzialmente agito e agiscono nella direzione di

rendere le moderne economie di mercato meno esposte alle crisi e più robuste.
Al tempo stesso altri elementi spingono nella direzione contraria, di accrescere l'instabilità. Fra essi, nella condizione economica e politica con cui si sono aperti gli anni novanta, si situa la mancanza di un paese leader, capace di fungere da prestatore di ultima istanza per l''economia-mondo', mentre il coordinamento fra i tre poli principali - Stati Uniti, Giappone, Europa - stenta a progredire. In assenza di politiche di bilancio e dei redditi coordinate, l'integrazione finanziaria, la dimensione dei flussi dei capitali e la loro velocità di movimento rispetto al flusso delle merci, l'inflazione latente sovraccaricano di compiti le politiche monetarie. L'incertezza è accresciuta, sono sospinti all'insù i tassi dell'interesse. Per entrambe le vie possono determinarsi crolli di borsa, insostenibilità del debito estero dei paesi in via di sviluppo, cadute nella propensione a investire: non può escludersi quindi che la lunga espansione degli anni ottanta venga interrotta da recessioni.
Fra queste forze di segno opposto sembrano prevalenti, non solo agli ottimisti e agli apologeti, ma alla maggioranza degli studiosi, quelle che tendono a ridurre i rischi di instabilità nel lungo periodo. Almeno due fondamentali dati di fatto confortano questa generalissima valutazione. La variabilità annuale del prodotto interno lordo dell'Europa presenta una chiara tendenza di lungo periodo alla riduzione dal 1830 a oggi; la tendenza è ancora più continua se si prescinde, considerandola un caso speciale, dalla crisi del 1929. Una crisi della vastità e dell'acutezza assunte da quella del 1929 non ha più avuto luogo in oltre mezzo secolo, pur essendo stata - negli anni settanta e nel 1987 - da molti paventata o addirittura prevista.
La natura stessa delle crisi, tuttavia, esclude che possa mai darsi la certezza di un loro venir meno. Da ultimo riconducibili all'ingovernabilità delle aspettative e delle decisioni d'investimento in economie intrinsecamente monetarie, le crisi possono riproporsi, sebbene le probabilità che ciò accada siano forse oggi minori di ieri. La risposta all'onesto quesito "può il 1929 ripetersi?" (v. Minsky, 1982) deve quindi essere affermativa. Qualora l'evento si materializzasse, l'efficacia della reazione, allo stato attuale delle nostre conoscenze, resta comunque affidata alla messa in atto dei rimedi estremi suggeriti da Thornton e da Keynes: fornire al sistema tutta la liquidità necessaria e sostenere con ogni mezzo la domanda globale. (V. anche Capitalismo; Cicli economici; Politica economica e finanziaria; Sviluppo economico).

bibliografia

Arrow, K. J., Limited knowledge and economic analysis, in "American economic review", 1974, LXIV, 1, pp. 1-10.
Ashton, T. S., Economic fluctuations in England, 1700-1800, Oxford 1959.
Bagehot, W., Lombard Street: a description of the money market, London 1873 (tr. it.: Lombard Street, Torino 1986).
Bairoch, P., Europe's gross national product: 1800-1975, in "Journal of European economic history", 1976, V, pp. 273-340.
Balke, N. S., Gordon, R. J., The estimation of prewar gross national product: methodology and new evidence, in "Journal of political economy", 1989, n. 1, pp. 38-92.
Buiter, W. H., Macroeconomic theory and stabilization policy, Manchester 1989.
Ciocca, P., L'instabilità dell'economia, Torino 1987.
Fisher, I., The debt-deflation theory of great depressions, in "Econometrica", 1933, I, pp. 337-357.
Friedman, M., A program for monetary stability, New York 1959.
Galbraith, J. K., The great crash: 1929, London 1955 (tr. it.: Il grande crollo, Torino 1972).
Goodwin, R. M., The M-K-S system: the functioning and evolution of capitalism, in The economic law of motion of modern society: a Marx-Keynes-Schumpeter centennial (a cura di H. J. Wagener e J. W. Drukker), Cambridge 1986, pp. 14-21.
Harrod, R. F., Economic dynamics, London 1973.
Hawtrey, R. G., The art of central banking, London 1932.
Hayek, F. A. von, Prices and production, London 1931 (tr. it.: Prezzi e produzione. Il dibattito sulla moneta, Napoli 1988).
Kahn, R. F., The relation of home investment to unemployment, in "Economic journal", 1931, n. 162, pp. 173-198.
Kalecki, M., Selected essays on the dynamics of the capitalist economy, 1933-1970, Cambridge 1971.
Keynes, J. M., The general theory of employment, interest, and money, London 1936 (tr. it.: Teoria generale dell'occupazione, dell'interesse e della moneta, Torino 1978²).
Kindleberger, C. P., The world in depression, 1929-1939, London 1973 (tr. it.: La grande depressione nel mondo, 1929-1939, Milano 1982).

Kindleberger, C. P., Manias, panics, and crashes, New York 1978 (tr. it.: Euforia e panico. Storia delle crisi finanziarie, Roma-Bari 1981).
Kindleberger, C. P., Financial crisis, in The new Palgrave. A dictionary of economics (a cura di J. Eatwell, M. Milgate, P. Newman), London 1987, vol. II, pp. 339-340.
Lewis, W. A., Growth and fluctuations, 1870-1913, London 1978.
Lucas, R. E. jr., Studies in business-cycle theory, Cambridge, Mass., 1983.
Luzzatto, G., L'economia italiana dal 1861 al 1894, Torino 1968.
Maddison, A., Phases of capitalist development, Oxford 1982 (tr. it.: Le fasi di sviluppo del capitalismo, Milano 1987).
Marx, K., Das Kapital, 3 voll., Hamburg 1867-1894 (tr. it.: Il capitale, 3 voll., Roma 1970).
Matthews, R. C. O., A study in trade-cycle history, Cambridge 1954.
Milward, A. S., Saul, S. B., The development of the economics of continental Europe, 1850-1914, Cambridge, Mass., 1977 (tr. it.: Storia economica dell'Europa occidentale, 1850-1914, Bologna 1979).
Minsky, H. P., Can 'it' happen again? Essays on instability and finance, Armonk, N. Y., 1982 (tr. it.: Potrebbe ripetersi? Instabilità e finanza dopo la crisi del 1929, Torino 1984).
Mitchell, B. R., European historical statistics, 1750-1975, London 1980.
Morgenstern, O., International financial transactions and business cycles, Princeton 1959.
Nardozzi, G., Crescita, in Dizionario di economia politica (a cura di G. Lunghini), vol. VII, Torino 1983, pp. 115-219.
Pasinetti, L. L., Dinamica strutturale e sviluppo economico. Un'indagine teorica sui mutamenti nella ricchezza delle nazioni, Torino 1984.
Schumpeter, J. A., The theory of economic development, Oxford 1961 (tr. it.: Teoria dello sviluppo economico, Firenze 1971).
Steindl, J., Maturity and stagnation in American capitalism, Oxford 1952 (tr. it.: Maturità e ristagno nel capitalismo americano, Torino 1960).
Svennilson, I., Growth and stagnation in the European economy, Geneva 1954.
Sweezy, P. M., The theory of capitalist development: principles of Marxian political economy, 2 voll., New York 1942 (tr. it.: La teoria dello sviluppo capitalistico, Torino 1951).
Sylos Labini, P., Problemi dello sviluppo economico, Bari 1970.
Temin, P., The Jacksonian economy, New York 1969.
Temin, P., Did monetary forces cause the Great Depression?, New York 1976.

Thornton, H., An enquiry into the nature and effects of the Paper Credit of Great Britain, London 1802 (tr. it.: Indagine sulla natura e sugli effetti del credito cartolare in Gran Bretagna, a cura di P. Ciocca e V. Sannucci, Torino 1990).
Vicarelli, F., Keynes. L'instabilità del capitalismo, Bologna 1989.
Wicksell, K., Lectures on political economy, 2 voll., London 1934-1935 (tr. it.: Lezioni di economia politica, Torino 1950).

Crisi sociopolitica

di Gian Enrico Rusconi

sommario: 1. Il problema di una definizione: a) linguaggio corrente e criteri scientifici; b) la crisi di sistema. 2. Una retrospettiva storica: a) le origini del concetto; b) una cultura della crisi. 3. La 'crisi del capitalismo' nella tradizione marxista. 4. La 'crisi di legittimazione del capitalismo maturo'. 5. La crisi nelle teorie struttural-funzionaliste e dell'azione collettiva: a) l'approccio struttural-funzionalista e le crisi della modernizzazione; b) conflitto sociale e violenza collettiva. 6. La 'crisi delle democrazie'. 7. Le crisi internazionali. 8. Conclusione. □ Bibliografia.

1. Il problema di una definizione

a) Linguaggio corrente e criteri scientifici

Nel linguaggio corrente la parola 'crisi' indica una situazione di grave difficoltà, di interruzione di normalità, di rottura d'equilibrio, per il cui superamento sono richiesti impegno e risorse in grado di ricreare un nuovo equilibrio e una nuova normalità. Il processo di crisi include pertanto anche le reazioni dei soggetti interessati, individuali e collettivi, e quindi le loro decisioni. Crisi è inoltre un termine usato per designare fenomeni di ampia portata storica (crisi dell'industrializzazione, crisi di secolarizzazione), fenomeni storicamente circoscritti ma non meno complessi (crisi delle democrazie), oppure anche processi delimitabili tecnicamente (crisi di governo nel quadro delle regole costituzionali, crisi economica descrivibile con indicatori precisi). Con quest'ampiezza di significati si ritrova nel lessico di tutte le discipline che studiano sistemi e comportamenti umani: dall'economia alla psicologia, dalla filosofia alla politologia, alla sociologia.

Ma come ogni termine largamente diffuso e intuitivo, anche quello di crisi si rivela refrattario a una definizione scientifica univoca. Alcuni studiosi, persino alcune enciclopedie, depennano addirittura la voce 'crisi' dal vocabolario delle scienze sociali, sostituendola con concetti ritenuti scientificamente più solidi, quali conflitto, anomia, disfunzione, rivoluzione, catastrofe, ecc. In realtà, ciascuno di questi concetti rischia di creare problemi analoghi, non appena si allontana da un'accezione rigorosamente e convenzionalmente circoscritta. Da qui l'opportunità di mantenere il concetto di crisi, articolandolo nel modo più preciso possibile. I criteri scientifici di chiarezza, coerenza, capacità di spiegazione, cui deve sottostare il concetto di crisi (come qualunque altro concetto delle scienze sociali), non devono far dimenticare che la polivalenza e in qualche caso persino l'ambiguità del linguaggio corrente rispondono alla necessità di collegare aspetti e dimensioni della realtà altrimenti inconfrontabili. Anche questo è un dato con cui le scienze sociali devono fare i conti. Il fatto che il termine crisi sia un codice o un topos retorico del nostro tempo rende soltanto più impegnativo lo sforzo di definirlo.
Per questa operazione procederemo in più direzioni. Dopo una breve retrospettiva storica, prenderemo atto dell'uso molteplice del concetto nei diversi orientamenti scientifici di ieri e di oggi. Pur tenendo conto della ricchezza di motivi di questa ricognizione, ai fini della operatività scientifica dovremo ricondurre il concetto di crisi ad alcuni parametri sociologici e politologici legati al concetto di sistema. Crisi nel suo significato sociopolitologico forte non può essere che sinonimo di crisi di sistema o di disintegrazione dell'intero sistema sociale. Gli altri significati vanno commisurati a questo.

b) La crisi di sistema

Il concetto di disintegrazione presuppone che un sistema sociale e politico in condizioni normali sia in equilibrio, sia integrato, ovvero 'stia insieme' nelle sue parti e in ciascuna di esse. Queste sono definite sottosistemi (economico-produttivo-distributivo, politico-istituzionale-regolativo, culturale-normativo). Si distingue tra integrazione sistemica, prodotta dal funzionamento dei sottosistemi economico-produttivi e dell'apparato politico-amministrativo secondo regole di efficienza, e integrazione sociale, prodotta da o riferita al consenso collettivo sulle norme, alle procedure di legittimazione e in generale ai processi culturali-normativi. Questi ultimi non si limitano a costituire una parte del sistema, ma svolgono la funzione insostituibile di dare identità al sistema stesso tramite la dominanza di

valori collettivi che storicamente possono variare o combinarsi diversamente (valori 'nazionali', 'democratici', 'socialisti', ecc.).
Si ha dunque disintegrazione o crisi di sistema in senso pieno non soltanto quando una o più parti sottosistemiche non funzionano più secondo le attese (depressione economica, paralisi delle istituzioni politiche, dissensi e conflitti generalizzati), ma quando a partire da queste disfunzioni o crisi parziali si innesca un processo a spirale che intacca l'integrazione sociale, l'identificazione dei membri con il sistema e quindi l'identità stessa del sistema. Si genera allora uno stato di anomia in senso proprio, che può variare per intensità (v. Anomia).
Per sviluppare un concetto di crisi così articolato sono necessari più approcci complementari. Innanzitutto occorre una definizione qualificata degli elementi sociali, politici, culturali costitutivi di un sistema e precedenti la sua fase critica. Non basta, ad esempio, riconoscere la natura capitalistica del sottosistema economico: occorre determinarla in modo più specifico nel suo stadio di sviluppo. Lo stesso vale per la caratterizzazione di un sottosistema politico formalmente definibile come democratico-parlamentare: dietro questa formula ci sono differenti equilibri politici, culture, aspettative sociali. Solo una precisa identificazione del sistema consente di risalire alle radici e ai motivi scatenanti della crisi, che possono essere di volta in volta di origine economica o di mobilitazione sociale non direttamente dipendente da cause economiche. Non meno essenziale, infine, è lo studio del decorso puntuale della crisi nelle componenti istituzionali, extra- e anti-istituzionali, compreso lo studio delle strategie e delle chances degli attori in gioco. Il rapporto tra le 'grandi cause' e la contingenza del comportamento effettivo degli attori è l'aspetto più difficile dell'analisi delle crisi sociopolitiche.
La crisi di sistema così delineata rappresenta un concetto-limite, che trova pieno riscontro soltanto nelle grandi rivoluzioni che segnano cesure storiche, oppure in eventi epocali quali la dissoluzione degli imperi antichi. Al di qua di questi eventi le crisi presentano tratti diversificati, anche nel tempo. Così le profonde crisi di trasformazione socioculturale, legate all'industrialismo e in generale alla modernizzazione, percorrono più epoche e regimi politici. Viceversa, episodi di drammatica crisi politico-istituzionale (pensiamo al passaggio dalla IV alla V Repubblica francese) possono incidere solo marginalmente sui caratteri di fondo di una società. Le crisi dei regimi democratici europei, in coincidenza con l'affermazione dei fascismi nella prima metà del XX secolo, sono verosimilmente gli esempi più vicini al tipo di crisi di sistema tracciato sopra.

2. Una retrospettiva storica

a) Le origini del concetto

Se è vero che la nostra epoca, più di altre, applica a se stessa l'idea di crisi, il termine e il suo primo uso storico-scientifico risalgono all'età classica greca: κρίσις deriva da κρίνειν, che vuol dire discernere, discriminare, decidere. Sin dalle origini, però, questo significato è preso nella competizione tra due paradigmi. C'è innanzitutto il paradigma medico, ippocratico, applicato al decorso della malattia che crescendo di intensità entra in una fase decisiva e critica, appunto, che prelude a un mutamento radicale di stato - risanamento o decesso. Ma già Tucidide, che nella descrizione della peste di Atene del 430 a.C. usa questo schema, introduce nella sua Guerra del Peloponneso altri significati del termine. Oltre che come sinonimo di risoluzione decisiva della guerra tra Persiani e Greci, la crisi è intesa come procedimento, dibattito, esame di dispute all'interno di un paradigma giudiziario. Al di là di questi riferimenti puntuali, si può dire che Tucidide presenti la crisi peloponnesiaca combinando i due paradigmi, clinico e giudiziario, e offrendo il primo esempio di un'analisi storico-politologica che prescinde da ogni visione mitica, ciclica, organica. La crisi è identificata non soltanto nelle sue radici economiche e politiche, ma anche nella contingenza del comportamento degli attori storici (v. Starn, 1971).

Questa concezione tuttavia rimane senza seguito nella successiva letteratura storica e filosofica antica e medievale: la parola crisi viene usata marginalmente e nella sua accezione medica, organologica. Ma a partire da questo significato, nell'età moderna, precisamente nel Seicento, l'analogia tra crisi del corpo fisico e crisi del corpo sociale e politico mette in moto prospettive nuove. Il passo definitivo per l'acquisizione di nuovi contenuti etico-politici al concetto di crisi avviene nel secolo dei lumi, con l'associazione tra crisi e rivoluzione. Rousseau, nell'Émile, saluta e annuncia il secolo delle crisi e delle rivoluzioni nella prospettiva del rivolgimento delle istituzioni che riporterà alla natura. La crisi è portatrice di un verdetto storico e si erge quindi a tribunale morale. Non diversamente, in America, Thomas Paine saluta e segue la Rivoluzione con un giornale intitolato "The crisis", che dà alle vicende della guerra civile i connotati dell'esecuzione di un giudizio morale.Presto però questa euforia per una crisi che attraverso il mutamento politico vuol essere rinnovamento etico e culturale urta contro i propri limiti. "Il concetto di crisi, con il significato di anarchia

politica che Rousseau vi ha collegato, la crisi come dissolvimento di qualsiasi ordine, come crollo di tutti i rapporti di proprietà, collegato a convulsioni e imprevedibili disordini, la crisi come crisi politica dell'intero Stato non fu affatto il significato centrale del concetto in cui si sarebbe condensata la coscienza borghese della crisi" (v. Koselleck, 1959; tr. it., p. 211). Il radicalismo con cui la grande critica illuministica attacca l'ancien régime, e ne esalta la crisi, rivela presto dietro le sue ragioni universalistiche i motivi e le domande della classe borghese. E questa chiede presto una deradicalizzazione, anzi una restaurazione dell'ordine sociale.

Il concetto di crisi rispecchia questo mutamento di clima. Esso diventa sintomo dei mali morali e sociali che la Rivoluzione non ha corretto, anzi ha esasperato. La crisi non è più teleologicamente guidata verso un sicuro esito benefico, ma si connota di ambiguità e pessimismo. È segno di malessere e di contraddizione. Jacob Burckhardt e Karl Marx, da prospettive opposte, fissano ciascuno con il proprio apparato teorico questa accezione di crisi. Nella sua lezione sulle 'crisi nella storia', Burckhardt traccia una sorta di percorso costante del processo critico: protesta, rivolta, tensione utopica, assestamento e restaurazione di un nuovo ordine. Ma ciò che più lo affascina è la misteriosa vitalità e varietà che la crisi imprime alla storia con le sue rotture e innovazioni (v. Burckhardt, 1905).

b) Una cultura della crisi

Tra la fine del XIX e l'inizio del XX secolo si crea quella che si può chiamare una 'cultura della crisi'. Da un lato il marxismo fa della crisi inevitabile del capitalismo l'asse portante della sua analisi sociale e della sua proposta politica; dall'altro lato l'intera cultura borghese diventa ambivalente verso i processi della modernizzazione che ha messo in moto e che sembrano rivoltarlesi contro. Lo scoppio della prima guerra mondiale e i suoi effetti devastanti segnano lo spartiacque per il dilagare di un Kulturpessimismus, in particolare nell'area germanica. È impossibile riportare sotto un comune denominatore la vastissima letteratura della crisi che arriverà sino alla fine degli anni quaranta, fondendosi e stemperandosi nella cultura dell'esistenzialismo. Basti menzionare qui due titoli di libri di qualità assai diversa, eppure entrambi sintomatici di un'epoca: Il tramonto dell'Occidente (1918-1922) di Oswald Spengler e La crisi della civiltà (1935) di Johan Huizinga.

La crisi della civiltà è registrata anche nello Handwörterbuch der Soziologie (1931), a cura di Alfred Vierkandt, una delle opere più impegnative e significative delle scienze sociali tra le

due guerre. In essa lo 'stile di vita' capitalistico-borghese, di massa, è presentato come un modello negativo, segno della distruzione di ogni autentico valore, della sostituzione della quantità alla qualità, della brama di sensazioni forti ed epidermiche, del culto della violenza - in breve una somma di disvalori, vista in contrapposizione a un idealizzato mondo precapitalistico. Ma quasi a completamento della coeva diagnosi marxista, centrata sulle cause e sulle manifestazioni economiche della crisi, la fenomenologia qui registrata è quella della disintegrazione sociale, della perdita di senso e di identità, della frustrazione e alienazione individuale e sociale, dell'eclettismo e del contrasto di valori, una secolarizzazione accompagnata, peraltro, da forme di riscoperta del sacro. È una diagnosi che con poche varianti di sostanza si riproporrà sino ai giorni nostri. Negli anni venti e trenta essa può essere al servizio di posizioni etico-politiche indifferentemente progressiste e liberal-democratiche o nichiliste, irrazionalistiche.

Concludiamo questo succinto quadro ricordando due lavori filosofici che, pur costruiti su basi teoretiche inconfrontabili, convergono nel riportare la crisi alle radici del razionalismo occidentale. Edmund Husserl ne La crisi delle scienze europee e la fenomenologia trascendentale scrive: "Per penetrare il groviglio della 'crisi' attuale era indispensabile elaborare il concetto 'Europa' in quanto teleologia storica di fini razionali infiniti; era indispensabile mostrare come il mondo europeo sia nato da idee razionali, cioè dallo spirito della filosofia. La crisi poté considerarsi un apparente fallimento del razionalismo" (v. Husserl, 1954; tr. it., p. 341). Da posizione opposta Max Horkheimer e Theodor W. Adorno, nella Dialettica dell'illuminismo, sostengono che "il concetto stesso di illuminismo, non meno delle forme storiche concrete, delle istituzioni sociali cui è strettamente legato, implica già il germe di quella regressione che oggi si verifica dovunque [...]. La maledizione del progresso incessante è l'incessante regressione" (v. Horkheimer e Adorno, 1947; tr. it., pp. 5 e 44). 3. La 'crisi del capitalismo' nella tradizione marxista.

La centralità dell'idea di crisi nell'analisi marxista è bene espressa dallo stesso Marx nel Poscritto (1873) alla seconda edizione del Capitale: "La cosa che più incisivamente fa sentire al borghese, uomo pratico, il movimento contraddittorio della società capitalistica sono le alterne vicende del ciclo periodico percorso dall'industria moderna e il punto culminante di quelle vicende: la crisi generale. Essa è di nuovo in marcia, benché sia ancora agli stadi preliminari" (tr. it., p. 45). Sono tre le dimensioni del processo evocato: il movimento contraddittorio, che inerisce al modo di produzione capitalistico, il ciclo periodico delle crisi

e la crisi generale. La natura economica del processo critico non deve far dimenticare che esso tocca le radici del capitalismo come formazione sociale. La contraddizione del modo di produzione capitalistico, decifrabile economicamente come caduta tendenziale del saggio di profitto, investe infatti il sistema nella sua globalità perché mette a nudo la contraddizione tra il carattere sociale della produzione e il carattere privato della sua appropriazione. L'analisi marxiana della forma economica della crisi si è rivelata tutt'altro che convincente e conclusiva, come mostra il lunghissimo dibattito che ne è seguito (v. Colletti e Napoleoni, 1970). Nodo centrale della controversia è la Zusammenbruchstheorie o teoria del crollo del capitalismo, determinabile con indicatori economici oggettivi. Se è fuor di dubbio che Marx non coltiva attese meccanicistiche della fine del capitalismo, non è men vero che l'intero suo sforzo scientifico ha di mira la dimostrazione dell'insuperabilità delle contraddizioni capitalistiche. Probabilmente gli equivoci della teoria del crollo e del connesso atteggiamento 'attendista' di certa cultura marxista consistono nell'attribuzione di pretese predittive di breve e medio termine a quelle che intendevano essere soltanto analisi di struttura. L'equivoco si fa inevitabile quando l'analisi di Marx, verso la fine del XIX secolo, diventa l'ideologia ufficiale dei grandi movimenti operai europei, innanzitutto della socialdemocrazia tedesca. La teoria del crollo, nelle sue semplificazioni subculturali, adempie la tipica funzione di rassicurazione di un'identità collettiva.Il dibattito revisionista iniziato da Eduard Bernstein investe anche la questione del crollo, provocando un intenso lavoro di riflessione. Tutti i principali teorici della socialdemocrazia intervengono nel dibattito con argomenti nei quali è impossibile tenere separata l'analisi economica dalla polemica politica. Rosa Luxemburg ne L'accumulazione del capitale (1913) ribadisce con energia la prospettiva della crisi intrinseca al capitalismo, collegandola alle tesi sulle condizioni di accumulazione e sviluppo del capitalismo che esigono il sistematico sfruttamento del mondo non capitalistico. In questa maniera il capitalismo, ampliandosi, annienta i presupposti stessi della propria esistenza: quando l'ultima area 'esterna' non capitalistica sarà incorporata nel sistema capitalistico, il sistema stesso sarà destinato al tracollo.

Su posizioni diverse si trova il teorico socialdemocratico più ascoltato in tema di analisi economica, Rudolf Hilferding. Nel suo Il capitale finanziario (1910) un'intera sezione è dedicata alle crisi cicliche, imputate sostanzialmente a squilibri e sproporzioni interne al capitale, da non confondere con tendenze al crollo. Rimane viva invece la prospettiva di una crisi sociale e politica del capitalismo, da guidare con mezzi politici verso una soluzione

socialista.Karl Kautsky, l'interprete ufficiale dell'ortodossia socialdemocratica, non esita a riconoscere che il problema più difficile lasciato in eredità da Marx è quello delle crisi periodiche. Dopo aver ribadito il postulato del nesso necessario tra capitalismo e crisi, ne ripropone una definizione generica, parlando di esplosioni delle contraddizioni che il modo di produzione capitalistico porta in sé, che lo rendono vieppiù intollerabile e spingono strati sempre più vasti della popolazione a lottare contro di esso (v. Kautsky, 1910).In definitiva, nella dottrina marxista la crisi si presenta con una componente economica che ha uno svolgimento oggettivo, autonomo, e una componente soggettiva, sociopolitica, dettata dall'intollerabilità della condizione di vita imposta dal capitalismo. Alla dialettica è affidato il compito di tenere insieme, sul piano analitico e su quello pratico-operativo, i due fattori altrimenti scoordinati. Questo schema non muta sostanzialmente per decenni al di là delle diverse collocazioni e divisioni: revisionisti contro ortodossi, riformisti contro rivoluzionari, socialisti contro comunisti. L'ala revisionista-riformista studia i meccanismi economici e politici che consentono al capitalismo non soltanto di sopravvivere ma di prosperare; l'ala ortodossa-rivoluzionaria insiste sui costi permanenti della sopravvivenza del capitalismo, che lo allontanano ma non lo sottraggono al suo destino finale di crisi. In entrambe le prospettive si riconosce l'importanza crescente dello Stato come regolatore dell'economia e delle sue crisi, sia nella linea dell'interpretazione di Hilferding del 'capitalismo organizzato' sia in quella di Lenin del 'capitalismo monopolistico di Stato'. Sono elementi relativamente nuovi rispetto all'originaria analisi marxiana, destinati a diventare preponderanti nell'impianto del neomarxismo contemporaneo.

Se passiamo a considerare la crisi nei suoi tratti politici, dobbiamo tener presenti due posizioni, esemplari per la diversità dell'esperienza politica che riflettono. Lenin delinea una 'legge fondamentale' della crisi rivoluzionaria che - a dispetto del suo intento militante e ideologico - non si discosta dagli schemi politologici odierni: formazione di un'élite antagonistica al sistema esistente e forte di un'ampia base sociale (gli operai coscienti); incapacità dell'élite al potere di contenere la pressione di massa, con la conseguente crisi di governo; mobilitazione e politicizzazione generalizzata estesa alle masse apatiche; tracollo del regime (v. Lenin, 1920). Più complessa è l'analisi della crisi di sistema fatta da Antonio Gramsci, che rimanda a rapporti di forza internazionali, a rapporti sociali oggettivi dettati dalle forze produttive, a rapporti di forza politica e partitica, a rapporti di forza militare. Dopo aver distinto tra "movimenti organici" e "fenomeni di congiuntura", Gramsci esclude

che le crisi storiche fondamentali siano determinate immediatamente da crisi economiche. Viene così in primo piano il fattore soggettivo, politico, organizzativo. Il contrasto, nel regime liberale, tra rappresentati e rappresentanti porta a una vera e propria "crisi di egemonia" della vecchia classe dirigente, incapace di svolgere il proprio ruolo nel confronto di masse alienate o mobilitate da rivendicazioni che nel loro complesso disorganico costituiscono una rivoluzione. Si parla di "crisi di autorità" intendendo crisi di egemonia o crisi dello Stato nel suo complesso (v. Gramsci, 1974, p. 75). Avendo davanti agli occhi l'esperienza italiana della crisi dello Stato liberale di fronte al fascismo, Gramsci registra la vittoria di un partito totalitario, reazionario, guidato da un capo carismatico, contro "l'immaturità delle forze progressiste".

4. La 'crisi di legittimazione del capitalismo maturo'

Il capitalismo contemporaneo ha mutato profondamente alcuni suoi tratti, non soltanto rispetto ai tempi di Marx, ma anche rispetto agli anni trenta che ne hanno registrato la crisi più acuta. Il mutamento è avvenuto anche grazie alle strategie di superamento di quella crisi, agli interventi statali anticiclici, alle politiche sociali di larga scala, alle pratiche di compromesso sociale o di scambio politico. Tutto ciò ha inciso sul concetto di crisi del o nel capitalismo, accompagnandone le trasformazioni.Questo processo di riflessione si estende per alcuni decenni. L'esperienza della depressione economica degli anni trenta, rimasta nella memoria storica come la 'grande crisi', ha messo in moto gli studiosi sociali più vigili alla ricerca dei fattori che hanno consentito il contenimento di quella che, per molti, si annunciava come la crisi catastrofica del capitalismo. Essi hanno studiato la profonda trasformazione del sistema capitalistico: il ruolo determinante dello Stato e delle sue istanze pianificatrici, la politicizzazione dell'economia. L'Istituto per la ricerca sociale di Francoforte (e poi nella sede americana di New York: v. Jay, 1973) è il luogo dove si producono le analisi più acute, rivolte non direttamente al concetto di crisi o alla sua fenomenologia, bensì alla natura e alla dinamica del capitalismo di Stato, nelle sue forme totalitarie e non (v. Pollock, 1941).
In realtà, solo lentamente questa riflessione produce i suoi effetti su quel tipo di analisi sociale - di ispirazione neomarxista - che è parte della 'teoria critica della società'. Occorre attendere gli anni settanta per registrare un'esplicita riconcettualizzazione del tema della crisi.Jürgen Habermas, ad esempio, ritiene che il concetto di crisi in senso forte, Systemkrise,

sia valido solo per la formazione capitalistica liberale, mentre per il capitalismo maturo o avanzato la crisi prende la forma di un deficit o "crisi di legittimazione". Nel capitalismo liberale infatti la crisi è direttamente immanente al principio organizzatore del sistema, al mercato, al rapporto tra lavoro salariato e capitale, ancorato giuridicamente nella struttura di classe e di dominio, per quanto reso anonimo dai meccanismi della rappresentanza universalistica. La crisi economica, prodotta da imperativi contraddittori, innesca immediatamente la crisi sociale e politica. Questo esito è evitato nel capitalismo avanzato con l'attivazione sistematica dell'intervento statale in economia, con l'espansione degli apparati di assistenza sociale, con la ricomposizione quasi politica del conflitto salariale. In questo modo il sistema come tale è immunizzato sia rispetto alle contraddizioni economiche che agli effetti dirompenti del conflitto di classe, i cui costi sono dirottati in modo anonimo sulla società intera (inflazione cronica, crisi fiscale permanente dello Stato, squilibri salariali a svantaggio dei gruppi più deboli). Dissolta l'identità delle classi, frammentata la loro coscienza politica, anche in forza della spoliticizzazione indotta dalle democrazie di massa, la crisi di sistema viene evitata anche se non ne sono estirpate le radici. L'unica base di legittimazione del sistema diventa il 'risarcimento' a gruppi e a classi, in cambio della passività nei processi di formazione della volontà politica. Si producono in questo modo continue crisi di razionalità, motivazione, legittimazione. In particolare "una crisi di legittimazione si produce non appena le pretese di risarcimento conformi al sistema aumentano più rapidamente della massa dei beni disponibili o quando ingenerano aspettative impossibili da soddisfare con risarcimenti conformi al sistema" (v. Habermas, 1973; tr. it., p. 35).

Al centro di questo processo contraddittorio per cui da una parte viene bloccata la crisi di sistema, mentre dall'altra si ricreano in continuazione motivi di crisi, c'è lo Stato. Esso si confronta con il compito impossibile di correggere le conseguenze sociali disfunzionali di una produzione orientata al profitto privato nel momento stesso in cui deve garantirne le condizioni ottimali. Questo rimane "l'errore di costruzione" del capitalismo (v. Offe, 1972). Ma una teoria della crisi oggi deve spostare la sua ottica dalla dinamica puramente economica alle patologie della vita quotidiana, in particolare alla "colonizzazione del mondo vitale per opera di un sistema amministrativo autonomizzato" (v. Habermas, 1976).

La rilevanza dell'approccio marxista e neomarxista per la problematica della crisi non discende dalla quantità e qualità di analisi da esso prodotte, spesso con il recupero di

approcci di altra matrice (sistemica, funzionalista o di psicologia sociale). La ragione è un'altra: il marxismo nelle sue varianti presume di determinare la crisi con la categoria della contraddizione, che nella sua valenza logica, forte, consente una lettura unitaria e coerente della complessa fenomenologia della crisi. La critica e il rifiuto della categoria della contraddizione caratterizzano invece le teorie non marxiste della crisi, ancora prima della contestazione degli indicatori empirici. L'abbandono di tale categoria nella spiegazione della crisi o il suo uso meramente metaforico equivalgono all'abbandono del paradigma marxista (questa osservazione vale anche per chi usa il concetto di contraddizione in modo indeterminato, interscambiabile con 'dilemma' o 'paradosso'). Tra i tentativi di conservare l'ottica macrostorica della crisi va segnalato l'approccio della world-economy (v. Wallerstein, 1983). Qui la crisi appare come "tensione strutturale" che porta un sistema-mondo verso una graduale disintegrazione, che è nel contempo transizione verso stati non facilmente decifrabili in anticipo. Questa concezione della crisi - per la natura degli indicatori dell'economia-mondo e per l'ampiezza della durata dei periodi considerati - risulta essere utile per la comprensione critica dei grandi trends del passato, piuttosto che per la diagnosi di crisi in atto o virtuali a breve scadenza. 5. La crisi nelle teorie struttural-funzionaliste e dell'azione collettiva.

a) L'approccio struttural-funzionalista e le crisi della modernizzazione

Nell'approccio struttural-funzionalista tutto ciò che mette in pericolo la stabilità di un sistema, postulato e requisito del suo buon funzionamento, può definirsi crisi. Ma non necessariamente viene usato e tematizzato questo termine, anzi di solito gli si preferisce quello di 'disfunzione', inteso - secondo la lezione di Talcott Parsons - come processo o insieme di condizioni che intaccano l'integrità e l'efficacia di un sistema. A sua volta disfunzione può significare situazioni diffuse di anomia e alienazione rispetto ai valori dominanti, oppure circoscritti atti di devianza e ribellione. In generale l'indicatore più evidente e misurabile di disfunzione è il conflitto che si genera in un sistema sociale, o meglio, in uno o più dei suoi sottosistemi. Ma è importante sottolineare che il conflitto come tale non è necessariamente distruttivo rispetto al sistema cui inerisce: può spingerlo verso autocorrezioni, ponendo così le premesse per una migliore, nuova stabilità.

Questo vale soprattutto per le crisi studiate dalla letteratura sociopolitologica della modernizzazione, che vi vede strozzature inevitabili ma alla lunga benefiche per le società

sulla via della modernità. L'attenzione è posta in particolare sulla crisi come disincronia tra modernizzazione sociale e modernizzazione politica. "Se i ritmi di mobilità sociale e di diffusione della partecipazione politica sono molto elevati, mentre i ritmi di organizzazione politica e di istituzionalizzazione sono molto bassi, si ha come risultato instabilità politica e disordine" (v. Huntington, 1968; tr. it., p. 17). In quest'ottica la crisi nasce dal ritardo delle istituzioni politiche rispetto ai mutamenti sociali, dal divario tra aspettative e obiettivi raggiunti, dal contrasto tra vecchi e nuovi valori che genera frustrazioni e alienazioni in larghi strati sociali.

Per coprire l'intero arco della modernizzazione è stata ipotizzata una rete di crisi intese come nodi ineludibili: crisi di identità, crisi di legittimità, crisi di penetrazione, crisi di partecipazione, crisi di distribuzione (v. Binder e altri, 1971). In verità i processi qui descritti come 'crisi' sono le difficoltà che incontra la gestione politico-istituzionale di tutte le società in via di modernizzazione, intendendo quest'ultima secondo gli standard di valore e di funzionamento dei sistemi liberal-democratici di capitalismo avanzato. Le difficoltà riguardano la capacità e la misura di controllo del governo (crisi di penetrazione); il grado e le modalità di partecipazione al processo decisionale (crisi di partecipazione); le forme e la qualità del consenso (crisi di legittimità); i problemi connessi con la distribuzione dei beni materiali e immateriali e in genere la questione della disuguaglianza socioeconomica (crisi di distribuzione); da ultimo le difficoltà di identificazione dei cittadini con le decisioni del governo (crisi di identità).

In questo genere di studi la crisi è vista sostanzialmente come somma di problemi di governabilità di un sistema, come sfida all'élite politica perché innovi nel suo comportamento, pena il suo violento ricambio. In questo modo la tematica della crisi diventa strettamente affine e complementare a quella del processo rivoluzionario. In effetti in molta letteratura politologica non è facile distinguere tra i due approcci. Chalmers Johnson, ad esempio, sulla base di alcuni assunti struttural-funzionalisti descrive la crisi rivoluzionaria nel modo seguente: la prima fase è la 'deflazione del potere', che obbliga i detentori dell'autorità a un uso sempre più esplicito e massiccio della forza per far rispettare le norme, sino ad arrivare a uno 'Stato di polizia'; se anche con queste misure i leaders non riescono a contenere il processo di crisi, si ha una vera e propria 'perdita d'autorità' con la conseguente loro delegittimazione; a questo punto intervengono fattori di 'accelerazione', imprevedibili, che esaltano le chances delle élites emergenti, alternative (v. Johnson, 1964).

Schemi analoghi, più o meno sofisticati, ritornano in tutta la produzione politologica contemporanea; ma è bene ricordare che essi si ritrovano già nella storiografia dedicata alle crisi di regime e alle rivoluzioni. Non a caso l'importante lavoro di Karl Dietrich Bracher dedicato alla dissoluzione della Repubblica di Weimar scandisce le fasi della crisi in: perdita di potere, vuoto di potere e presa di potere da parte delle nuove élites antagoniste (v. Bracher, 1955). Modelli più complessi sono offerti dagli studi che mettono a confronto molti casi storici sul lungo periodo. Charles Tilly, esaminando il "secolo delle ribellioni 1830-1930" e i processi di formazione degli Stati nazionali, propone una sequenza così articolata: creazione di domande e di pretese inaccettabili per l'establishment politico; mobilitazione e rapida diffusione di tali domande presso strati sempre più vasti della popolazione; incapacità del potere costituito di soddisfare, respingere o reprimere le domande sociali; creazione di una nuova coalizione di forze d'opposizione in grado di sfidare efficacemente il potere esistente; influenza di tale coalizione su alcuni settori dell'establishment e su alcune risorse del potere; lotta tra le forze antagoniste, che può concludersi con la netta vittoria della nuova coalizione emergente o con una serie di compromessi; ricostituzione di un nuovo ordine (v. Tilly, 1978).

b) Conflitto sociale e violenza collettiva

Dall'esame condotto sin qui risulta evidente il ruolo che nella crisi e nelle teorie sulla crisi hanno l'azione collettiva, la violenza collettiva, le mobilitazioni, i movimenti di rivolta - in una parola, le svariate forme di conflitto. È evidente il nesso tra crisi e conflitto, definito come "una lotta per valori o per la rivendicazione di status, potere e risorse scarse; in questo scontro l'intenzione dei gruppi in conflitto non è soltanto quella di raggiungere i valori desiderati, ma anche di neutralizzare, colpire o eliminare l'avversario" (v. Coser, 1967, p. 232).

Di fatto gli studi sull'azione collettiva, specificamente quelli sulla violenza collettiva, riprendono e ricodificano gli schemi del processo di crisi visti sopra, spostando l'ottica dalla dinamica del sistema al comportamento degli attori. Attenzione speciale è prestata alle motivazioni e alla fenomenologia delle aspettative (aspettative crescenti, aspettative prodotte dall'alternarsi di soddisfacimento e diniego, frustrazioni da mobilità sociale discendente, senso di deprivazione relativa a confronto con altri gruppi sociali, incongruenza di status). Ciascuna di queste sindromi di frustrazione diventa virtuale disponibilità alla mobilitazione conflittuale.

L'apporto più interessante di queste analisi sta nel riconoscimento della razionalità di risorse di identità e di solidarietà, in atteggiamenti individuali e collettivi che sono facilmente fraintesi come irrazionali e distruttivi. È un'ottica complementare e correttiva rispetto a quella che vede la crisi esclusivamente come esplosione di distruttività e irrazionalismo. Detto questo, occorre dire che gli studiosi del conflitto e dell'azione collettiva non sono sempre in grado di determinare se i comportamenti conflittuali siano generatori di crisi o semplici manifestazioni di una crisi già presente. Analogamente, non è sempre facile distinguere un'azione conflittuale anti-istituzionale, che innova positivamente nel sistema, da un'azione distruttiva del sistema stesso, a meno che questa non prenda la forma esplicita di una 'guerra interna' (v. Eckstein, 1964). Ribellioni, conflitti sociali, violenze collettive non segnalano automaticamente una crisi di sistema. La violenza infatti è soltanto una delle variabili che accompagnano le 'sfide decisionali' che davvero qualificano questa crisi (v. Gurr, 1970).

Un gruppo di studiosi di Stanford, sulla base di alcuni casi storici comparati, ha introdotto il concetto di systemic crisis per integrare mobilitazioni sociali, formazione di coalizioni e leadership entro un quadro analitico che ricalca la già nota dinamica: domanda/sfida → struttura/istituzione politica → risposta/risarcimento (v. Almond e altri, 1973). Si ha crisi sistemica quando si interrompe "il flusso sincronico" tra i passaggi. Ma non ogni desincronizzazione è già crisi sistemica. Questa ha luogo soltanto quando si spezza o si blocca la struttura o costituzione politica che definisce l'insieme delle regole per "allocare autorità e risarcimenti". Questo modello, se da un lato parla di crisi sistemica in riferimento alla radicalità della messa in discussione dell'apparato politico-istituzionale, dall'altro non può qualificare ancora come "strutturale" questo tipo di crisi ("strutturale" è il qualificativo che gli autori citati riservano a processi di maggiore profondità storica).

Alla fine di questa breve rassegna di modelli ci troviamo dinanzi a due grosse difficoltà. Innanzitutto non è chiaro il ruolo esatto da assegnare alla crisi politica - visibile e analizzabile politologicamente - nel quadro e rispetto a una crisi di sistema nel senso pieno che abbiamo postulato all'inizio delle nostre considerazioni. Ci si avvicina alla soluzione di questa difficoltà se si individuano i fondamenti sociali del potere politicamente formalizzato, e quindi le regole del potere sociale tout court che caratterizzano il sistema nel suo complesso.In secondo luogo le generalizzazioni politologiche, del tipo sopra riportato, appaiono spesso scarsamente significative agli storici e agli specialisti di singoli casi

storicamente rilevanti. Al di là della tradizionale controversia sul rapporto corretto tra generalizzazione e specificità storica, il punto in questione è la precisa identificazione del sistema. A esso si collega la maggiore o minore importanza da attribuire al sottosistema economico rispetto a quello della cultura, o al ruolo del sottosistema dei partiti rispetto a quello della conflittualità sociale. L'identità di un sistema è data dalla sintesi o dall'equilibrio che di volta in volta si stabilisce tra questi fattori ed elementi. La problematica della 'crisi delle democrazie' ci consente qualche esemplificazione in merito.

6. La 'crisi delle democrazie'

Quando si parla di crisi delle democrazie nel XX secolo in Europa si intendono tre tipi o casi: 1) crisi di regime o crollo del sistema parlamentare democratico; esempi tipici sono l'istaurazione del fascismo in Italia, la crisi della Repubblica di Weimar con l'assunzione del cancellierato da parte di Hitler, il tracollo della Repubblica spagnola a seguito dell'insurrezione militare franchista; 2) severa crisi istituzionale nel quadro del mantenimento delle regole democratiche; il caso più evidente è il passaggio in Francia dal regime parlamentare della IV Repubblica al regime presidenziale della V Repubblica; 3) crisi di funzionamento o di governabilità, imputata dalla pubblicistica politica (soprattutto degli anni settanta) in misura più o meno acuta a tutte le democrazie europee.In tutti e tre i tipi ricordati si possono registrare fenomeni di caduta di consenso e di legittimazione, alta conflittualità sociale con manifestazioni di violenza politica, inefficienza dell'apparato statale. Ma, a parte il variare della loro intensità, ciascuno dei fenomeni citati può presentarsi con indicatori assai diversi: la caduta di consenso può generare sia disaffezione e apatia che radicalizzazione estremistica; l'intensificazione della conflittualità sociale può prendere la forma di cicli di scioperi legali o selvaggi oppure quella di una guerra civile strisciante; la paralisi dell'apparato statale può non escludere complicità con le forze eversive. Insomma, gli indicatori di crisi, per esser significativi per un singolo caso, richiedono tali e tante precisazioni da diventare spesso non significativi per altri casi.
Alcuni modelli politologici hanno proposto una omologazione delle crisi in riferimento alla dislocazione spaziale delle forze partitiche. Si è inteso mostrare che un sistema politico a pluralismo polarizzato, bloccato da partiti antisistema, è il comune denominatore della crisi della Germania weimariana, della crisi della IV Repubblica francese e virtualmente della

dinamica di crisi di altri regimi odierni (a cominciare da quello italiano). In realtà questi schemi interpretativi, per quanto illuminanti, ripropongono l'interrogativo della rilevanza da assegnare al sottosistema partitico per l'individuazione e la spiegazione delle cause e delle tappe di una crisi di sistema. Infatti se questa ha il suo luogo manifesto nei meccanismi politico-partitici, un giudizio comparativo sul loro funzionamento non può prescindere dalla diversa loro integrazione con i valori socialmente condivisi nei diversi sistemi formalmente democratici.

Per evitare di entrare nel merito sostantivo e valutativo della 'democraticità' di un sistema, si è suggerito di tenere distinto il 'crollo della democrazia politica' dalla più generale 'crisi delle società democratiche'. Il crollo politico è individuato nel "trasferimento di legittimità da un complesso di istituzioni a un altro", per opera di "opposizioni sleali" al regime esistente (v. Linz e altri, 1978; tr. it., p. 55). In realtà questa ottica, anche se politologicamente operativa, per essere davvero efficace nella comparazione dei sistemi deve qualificare in qualche modo sostantivamente i contenuti dell'azione delegittimante e rilegittimante dei gruppi di potere (com'è implicito, del resto, nel concetto di 'slealtà' politica), deve fare i conti con i valori vissuti e le aspettative concrete di democraticità dei soggetti coinvolti.In breve, la necessità di concentrare l'analisi sul regime politico-partitico non può far dimenticare che una democrazia funziona grazie alla integrazione di tutte le sue componenti, in particolare grazie all'equilibrio - di volta in volta mutante - tra regole sociali ed economiche e procedure di rappresentanza politica.

La gravità di una crisi della o nella democrazia si misura dall'intensità con cui sono coinvolte tutte le sue parti costitutive. Da questo punto di vista i casi sopra riportati si collocano idealmente su una scala di maggiore o minor prossimità a una crisi di sistema intesa in senso pieno. Quello che più vi si avvicina è il caso della democrazia della Germania di Weimar (1918-1933). Qui la crisi più drastica colpisce l'ambito politico-istituzionale con l'istaurazione della dittatura nazionalsocialista e la distorsione coatta dei valori normativi dominanti in direzione del razzismo e dell'etnocentrismo nazionalistico estremo. Ma questo processo non intacca, ad esempio, la struttura sociale esistente e il potere economico, esercitato in forma capitalistica, che esce addirittura rafforzato. Questo elemento di continuità rispetto al regime precedente segna, peraltro, una profonda alterazione e cesura dell'equilibrio sociale che costituiva l'identità del sistema weimariano. La paralisi del parlamento e la fine della costituzione democratica (inspiegabili, del resto, senza l'emergenza di movimenti

collettivi e senza il ruolo esercitato da élites del potere sleali) seguono la rottura delle regole sociali che sino ad allora avevano garantito l'esistenza di una "democrazia contrattata", dove le regole del gioco politico erano tutt'uno con le regole del compromesso sociale (v. Rusconi, 1977).

L'esperienza della democrazia weimariana e della sua crisi è istruttiva per un'altra ragione. Essa offre l'esempio più rilevante di quel processo socioeconomico in fieri sul continente europeo che si sarebbe chiamato 'pluralismo corporatista'. Esso caratterizza l'inserimento della classe operaia organizzata in un sistema stabile di contrattazione e di garanzie, sotto la supervisione dello Stato, e lo stretto intreccio tra economia privata e intervento statale nell'economia (v. Maier, 1975). Si anticipa così l'odierno assetto neocorporativo o corporatista degli interessi economici e sociali, basato sulla collaborazione tra lavoro e capitale e garantito da una politica governativa orientata alla costruzione dello Stato sociale di diritto.

Ebbene, questa esperienza viene espressamente negata in coincidenza con la crisi della democrazia weimariana, ma viene ripresa, sviluppata e perfezionata dopo la catastrofe nazista e la guerra - in sintonia con analoghi processi in atto in tutte le democrazie avanzate. Si conferma un trend irreversibile. In quest'ottica l'intreccio di crisi economiche, sociali, politiche che è associato emblematicamente all'anno 1929 (chiamato nella pubblicistica l'anno della 'grande crisi') si rivela retrospettivamente un passaggio decisivo per la messa in opera di quella 'architettura della stabilità' socioeconomica che, nel secondo dopoguerra, avrebbe caratterizzato le democrazie europee avanzate.

Questa affermazione non è in contrasto con il tema della crisi di governabilità delle democrazie contemporanee, che alimenta il dibattito sociopolitologico. Infatti, nonostante i toni allarmati, se prescindiamo dalle diagnosi di matrice marxista, la problematica della crisi è circoscritta di fatto alla gestione delle risorse istituzionali, economiche, morali di cui dispone (o dovrebbe disporre) un governo di sistemi altamente complessi. All'ordine del giorno ci sono le difficoltà del crisis management o 'controllo della crisi' piuttosto che l'eventualità di una crisi di sistema. Questo non vuol dire sottovalutare la serietà dei problemi di ordine economico (deficit statale crescente, crisi fiscale dello Stato, disoccupazione endemica) o di ordine sociale (emergenza di una conflittualità selvaggia, saturazione delle pratiche neocorporative, disfunzionalità del Welfare State), per tacere dell'acuirsi dei problemi ecologici. Ma questi fenomeni, pur segnalando difetti di funzionamento del sistema,

non portano generalmente a diagnosi o a prese di posizione politiche radicali, rivoluzionarie (salvo che nelle declamazioni ideologiche o accademiche).
Posto centrale in questa fenomenologia della crisi occupa lo Stato sociale o Welfare State, cioè quel complesso di norme legislative e di pratiche amministrative che hanno esteso e rafforzato i diritti sociali di tutti i membri delle democrazie avanzate. Non è sempre chiaro (ed è comunque motivo di controversia) se la crisi abbia le sue cause fuori dai meccanismi istituzionali del Welfare o non ne sia il prodotto endogeno. Si ha verosimilmente un processo circolare di causa-effetto, che trapassa da un settore all'altro.Cominciamo dal settore economico-finanziario: qui gli squilibri del deficit statale da un lato derivano da una spesa sociale sfuggita a ogni controllo, ma dall'altro non paiono consentire ristrutturazioni che non peggiorino le condizioni dei ceti meno privilegiati. Molti dei servizi erogati sono qualitativamente scadenti, affidati a procedure inefficienti, con scarse possibilità di controllo soprattutto da parte dei cittadini-utenti. Tutto ciò ha un immediato contraccolpo sull'apparato politico-istituzionale, ma di segno contrario. Mentre i responsabili del governo sono sollecitati a prendere iniziative di razionalizzazione e ristrutturazione nella direzione dell'efficienza, della trasparenza, del controllo, le forze partitiche non sono in grado di sottrarsi alle pressioni del loro elettorato che portano praticamente in direzione opposta. Il risultato è che - pur nell'oggettivo miglioramento dello standard medio di vita - non sono diminuite le disuguaglianze sociali, anche per l'ineguale distribuzione degli accessi ai vantaggi del Welfare e più in generale al godimento di beni pubblici. L'effetto più vistoso è la ricomparsa di nuove forme di povertà e di emarginazione sociale.
Queste considerazioni giustificano che si parli di una 'crisi del Welfare' come momento caratterizzante le democrazie contemporanee. Ma questa espressione segnala disfunzioni e difficoltà di governo piuttosto che latenza o imminenza di una crisi di sistema.

7. Le crisi internazionali

Accenniamo brevemente alla tematica della crisi nei rapporti internazionali, più precisamente alle situazioni di potenziale conflitto armato tra gli Stati. La crisi, cioè, è qui intesa come tensione prossima alla guerra. Dal punto di vista analitico la problematica è semplificata dall'assunto che i protagonisti della crisi sono i governi, considerati come attori unitari. L'esame della dinamica decisionale intragovernativa è demandato ad altri

approcci.Generalizzando dalla letteratura scientifica consolidata (v. Herman, 1972; v. Brecher, 1980), possiamo tracciare un modello di crisi prossima alla guerra con tre indicatori: a) la percezione da parte di uno o più governi di una grave minaccia all'integrità dei valori e degli interessi da esso rappresentati; questo indicatore può essere trattato in modo più o meno critico, a seconda che siano in pericolo effettivi interessi nazionali oppure interessi considerati tali soltanto dal governo in carica, oppure sia minacciata semplicemente l'autoimmagine dei governanti; b) la percezione dell'inesistenza di spazi per una manovra diplomatica e negoziale alternativa all'azione di guerra; c) la percezione di limiti ristretti di tempo per una decisione; sotto questo indicatore è riportato anche l'effetto 'sorpresa' che può accompagnare la crisi.
Lo schema ora tracciato è costruito sulla 'percezione' della situazione critica da parte di uno o più soggetti. Non esclude affatto l'esistenza di cause oggettive e remote, anche se la letteratura scientifica tratta la crisi prevalentemente nell'ottica della sfida decisionale. Quali che siano i dati storici e materiali dell'origine, la crisi è studiata innanzitutto come punto di svolta per il decisore che vede messa in gioco in modo acuto, ineludibile e carico di emozioni la sua scala dei valori (v. Deutsch, 1982).
Questo approccio dà molto spazio alla psicologia decisionale e alle tecniche del crisis management, anche se in proposito non mancano difficoltà e incertezze. Talvolta infatti crisis management è sinonimo di tecnica per fermare la spirale verso la guerra, considerata evento comunque negativo, da evitare con la ricerca di compromessi sempre possibili; talaltra è sinonimo di tecnica di apprendimento delle intenzioni effettive e delle chances dell'avversario per imporgli la propria posizione prima di ricorrere al conflitto aperto. Nell'assunzione dell'uno o dell'altro significato giocano, spesso in maniera inconsapevole, postulati di valore dello studioso o dell'osservatore.
La tematica del crisis management non esclude, anzi implicitamente presuppone, l'esistenza di una crisi sistemica, intendendo per sistema un complesso di elementi che 'stanno insieme', creando l'equilibrio dello status quo ante. La determinazione più precisa di questo equilibrio richiede un'analisi politologica che identifichi di volta in volta la situazione che va in crisi. Essa può essere il sistema multipolare degli Stati nazionali o l''equilibrio delle potenze' della prima metà del XX secolo oppure la struttura bipolare di due superpotenze, senza dimenticare sistemi ed equilibri regionali che interferiscono, con vari effetti di disturbo e di crisi. Soltanto sullo sfondo di queste più complesse dinamiche sistemiche acquista cogenza la

problematica del 'controllo della crisi'.Lo studio delle crisi internazionali ha recentemente adottato una tecnica di analisi nota come 'teoria dei giochi' (v. Schelling, 1960; v. Snyder e Diesing, 1977; v. Brams, 1985; v. Zagare, 1987), che rappresenta una branca autonoma rispetto alla problematica sviluppata sotto questo nome nelle discipline economiche e matematiche. Le crisi internazionali sono lette in chiave di teoria dei giochi a livelli analitici molto diversi: si va da semplici visualizzazioni schematiche delle preferenze degli avversari, ricondotte a due posizioni-base (cooperazione o cedimento versus non cooperazione o scontro), a sofisticati calcoli di probabilità dei possibili esiti del gioco, che tengono conto di variabili quali il potere della minaccia o l'efficacia di azioni di inganno o di reciproco fraintendimento.

La tecnica dei giochi assolve una duplice funzione. Innanzitutto ricostruisce crisi storiche (preferibilmente quelle politico-militari del secondo dopoguerra: crisi di Berlino del 1948, crisi dei missili a Cuba del 1962, crisi internazionali legate ai conflitti arabo-israeliani) attraverso tipologie che integrano le analisi storiche tradizionali. In secondo luogo traccia scenari di possibili crisi future, con particolare attenzione alle crisi delle due superpotenze nucleari, a partire dallo stato di reciproca deterrenza (v. Comunismo). Sebbene la tipologia dei giochi sia molto varia, ci sono due giochi standard, usati per identificare e analizzare le crisi internazionali: il 'dilemma del prigioniero' (prisoner's dilemma) e il 'gioco del pollo' (chicken game). Il primo mostra come gli attori, perseguendo razionalmente i loro interessi immediati, producano - in forza della dinamica di interdipendenza in cui si trovano - un esito irrazionale per entrambi. Si tratta di un esito irrazionale non soltanto perché non voluto, ma perché oggettivamente il peggiore tra gli esiti possibili. Questo esito non è direttamente imputabile a errori di calcolo o intenzionalità soggettive, ma alla struttura stessa della situazione. Esso è usato per spiegare crisi scatenate o portate al limite di rottura dalla spirale degli armamenti (soggettivamente intesi come sicurezza propria, non come minaccia ad altri) o da altri vincoli (di alleanza o clientela di Stati coinvolti in conflitti non voluti dall'attore in gioco). Il 'gioco del pollo', invece, definisce una situazione caratterizzata dalla sfida (unilaterale o reciproca) basata sulla presunzione che l'altro cederà (si comporterà da pauroso o da 'pollo'). Questo gioco, che può presentarsi come un vero e proprio esercizio di assunzione di rischio, per saggiare la resistenza dell'altro, è spesso usato per ipotizzare scenari di 'confrontazione' tra le superpotenze nucleari.

Le crisi internazionali studiate attraverso la tecnica dei giochi mostrano con nettezza

l'interdipendenza delle mosse degli attori, gettano luce sui meccanismi cognitivi che ineriscono alle percezioni reciproche degli avversari, offrono un quadro articolato delle probabilità degli esiti della situazione, che tende così a essere fortemente razionalizzata ma anche ipersemplificata. Di conseguenza, per la comprensione delle crisi internazionali e del loro decorso e soluzione, occorre introdurre ulteriori elementi di spiegazione che rimandino alla specificità e irripetibilità di ogni evento storico e al mondo dei valori degli attori coinvolti, da cui discendono le grandi scelte politiche che decidono le crisi (v. Rusconi, 1987).

8. Conclusione

Nel corso della nostra analisi abbiamo più volte rilevato come la realtà della crisi sia strettamente legata alla percezione e al modo stesso di esprimersi dei soggetti coinvolti. Il risultato è che il termine crisi viene applicato a una gamma vastissima di fenomeni, con un ampio margine di polivalenza e ambiguità di significati. Di fronte a questo labirinto di significati non sono mancati tentativi di forzare l'uscita sostituendo il concetto di crisi con quello di 'catastrofe' (nel senso elaborato da R. Thom: v., 1972). Per la verità, l'adozione di questo modello di analisi sociopolitologica introduce un certo rigore terminologico, ma a prezzo di spostare, alla fine, sul concetto di catastrofe le stesse difficoltà di determinare i fattori di quella che abbiamo chiamato crisi in senso forte o crisi di sistema. Rischia di essere una soluzione nominalistica. Non ci rimane quindi che mantenere la tematica della crisi nella sua complessità.

Se lasciamo sullo sfondo le accezioni più estese, riservate a fenomeni epocali come la 'crisi della modernizzazione' o anche a fenomeni più circoscritti ma controversi come la 'crisi del capitalismo' (su cui ci siamo a lungo soffermati), possiamo dire che nel lessico e nella cultura diffusa sono identificabili tre grandi tipi di eventi chiamati 'crisi'.

1. Eventi internazionali, ben definiti nello spazio e nel tempo, provocati o legati a conflitti strategico-militari, effettivi o soltanto virtuali. Le vigilie dei due grandi conflitti mondiali del XX secolo sono contrassegnate da questo tipo di crisi, anche se nella letteratura scientifica è rimasta emblematica soprattutto la crisi del luglio 1914. Nel secondo dopoguerra una serie di conflitti o sfide regionali e centrali (soprattutto le crisi di Berlino e di Cuba) ha arricchito la casistica di questo tipo, aumentando, peraltro, anche l'esperienza del crisis management.

2. Eventi di carattere fondamentalmente economico, con immediati riflessi sociali e politici a livello mondiale: esemplare è la crisi del 1929. Sotto questa tipologia sono classificabili altri episodi di assai minore gravità, come la cosiddetta 'crisi petrolifera' degli anni settanta e la catena degli squilibri economico-finanziari che accompagna, da un decennio, la vita economica mondiale.
3. Eventi di natura politico-istituzionale e sociale interni ai vari paesi e contrassegnati da un alto tasso di conflittualità. È a questa tipologia che abbiamo dedicato maggiore attenzione: dalle crisi dei regimi liberali in Europa negli anni venti e trenta ai temi contemporanei della 'crisi di legittimazione' o del Welfare State. Ciascuna nazione offre un quadro differenziato di questi sintomi di crisi che investono, in modo più o meno acuto, il sistema politico e le sue articolazioni sottosistemiche (crisi di rappresentanza, crisi dei partiti, crisi delle regole istituzionali, ecc.). Il concetto generale di crisi di sistema - così come l'abbiamo illustrato - fornisce gli indicatori grazie ai quali possono trovare posto le diverse fenomenologie di crisi. Si parla di crisi sociopolitica in senso forte soltanto se e quando le disfunzioni sistemiche sono accompagnate dalla disaffezione e dal venir meno della identificazione con il sistema da parte dei suoi membri. Un'analisi scientifica operativa deve mantenere vive e correlate due dimensioni: quella oggettiva delle disfunzioni e paralisi dei sottosistemi e quella dell'identità di appartenenza dei membri. Sono costoro in definitiva che, da un lato, 'subiscono' la crisi ma, dall'altro, 'mettono in crisi' il sistema stesso. Soltanto un approccio che operi con entrambe le dimensioni è in grado di decifrare, pronosticare ed eventualmente collaborare a risolvere una crisi sociopolitica.