

**Il concetto di integrazione verticale
come strumento analitico
e come comportamento strategico:
alla ricerca di una continuità
tra analisi economica ed economia industriale**

Luca Lambertini*

Dipartimento di Scienze Economiche
Università degli Studi di Bologna

novembre 1992

Sommario

L'obiettivo di questo lavoro è quello di delineare il ruolo ricoperto dal concetto di integrazione verticale nell'analisi del processo di creazione della ricchezza sia sotto il profilo aggregato, che interessa l'analisi economica, sia sotto il profilo individuale, che interessa la teoria dell'impresa e la teoria dell'organizzazione industriale.

**Rivolgo il mio ringraziamento a Giuliana Campanelli, Roberto Cellini, Roberto Censolo, Paolo Onofri, Luisa Segnana ed in particolare a Gianpaolo Rossini e Roberto Scazzieri per i puntuali e proficui commenti espressi su di una precedente stesura. Ringrazio inoltre Michele Burattoni e Dario Sermasi per il loro insostituibile supporto informatico. Naturalmente, la responsabilità di quanto scritto resta esclusivamente mia.*

Introduzione

*We have one sap and one root -
Let there be commerce between us.*
Ezra Pound

Il concetto di integrazione verticale ricopre un proprio ruolo tanto all'interno dell'analisi economica, ovvero nella tradizione dell'Economia Politica, quanto all'interno della relativamente moderna letteratura che si occupa di organizzazione industriale.

L'intuizione da cui ha avuto origine il presente lavoro è quella di presentare in una prospettiva congiunta le due linee lungo le quali si è sviluppato tale concetto, cercando di identificare possibili punti di contatto tra esse ed in ultima analisi di isolare il momento in cui questo ha smesso i panni dello strumento per indossare quelli di oggetto specifico dell'analisi stessa. In altre parole, l'intento con cui ci siamo avvicinati al tema dell'integrazione verticale è quello di andare oltre la semplice giustapposizione delle due interpretazioni sviluppate dalla letteratura, per quanto interessante possa apparire, in sé, un'operazione di questo genere. Il nostro obiettivo è più ambizioso ma certamente più attraente, ed è quello di individuare il *turning point*, sia cronologico che analitico, che ha determinato il mutamento di prospettiva, inducendo gli economisti, o almeno una larga parte di essi, a riconsiderare il tema dell'integrazione verticale sotto una nuova luce, vale a dire come nodo cruciale nell'ambito della teoria dell'organizzazione industriale.

Da Smith a Keynes,

The notion of vertical integration is implicit in all discussions on the theory of value of the Classical economists. The same thing can be said of the marginalist economists. ... Very rarely, however, is the logical process of vertical integration explicitly discussed. Generally it is simply taken for granted. (Pasinetti, 1973, p.1n)

In altri termini, all'interno dell'analisi economica si è fatto ripetutamente ricorso al procedimento logico di integrazione verticale in contesti il cui oggetto specifico si identificava, di volta in volta, nella teoria dei prezzi, nella creazione di ricchezza da parte del sistema economico nel suo complesso e da parte dei singoli settori al suo interno, ed infine nell'analisi del progresso tecnico.

Sinteticamente, quindi, la nozione di integrazione verticale ha rappresentato, all'interno dell'analisi economica, uno strumento dedicato allo studio di comportamenti e di grandezze di natura aggregata, offrendo un nesso sintetico tra mezzi di produzione e creazione di surplus, tale da consentire all'economista interessato alla *performance* globale del sistema e alle sue capacità autoriproduttive di trascurare qualsiasi problema riguardante l'organizzazione industriale del sistema stesso.

Il secondo ruolo individuato per l'integrazione verticale è intimamente connesso con uno degli obiettivi più rilevanti della teoria economica moderna, segnatamente quello di interpretare l'allocazione delle risorse e quindi delle attività economiche tra istituzioni alternative, vale a dire imprese, mercati e soluzioni organizzative ibride.

In questo senso, la nuova rilevanza assunta dal concetto di integrazione verticale è da ricondursi al fatto che esso rappresenta un punto di intersezione tra la teoria dell'impresa, la teoria dei contratti e la teoria dei mercati.

Quindi, ai nostri fini, l'interrogativo cruciale a cui dobbiamo rispondere è il seguente: quando e perchè, all'interno della teoria economica, ha iniziato a prendere corpo una teoria dell'impresa non più come semplice polo produttivo ma come organizzazione alternativa al mercato? E quali sono le motivazioni che spingono gli operatori a darsi una struttura che sia in grado di surrogare il mercato?

Come è del resto intuibile, queste domande sono intimamente connesse al superamento della visione del funzionamento del sistema economico *à la* Arrow-Debreu. E non è un caso che lo stesso Arrow abbia più volte sottolineato questa necessità, con riferimenti espliciti al peso della disponibilità delle informazioni rilevanti nell'ambito dell'organizzazione reale dell'attività economica:

The desirability of creating organizations of a scope more limited than the markets as a whole is partially determined by the characteristics of the network of information flows. (1974, p.37)

L'organizzazione del lavoro è la seguente. La sezione 1 è dedicata all'inquadramento del procedimento logico di integrazione verticale così come è stato impiegato all'interno dell'analisi economica. Nella sezione 2 tenteremo di mettere in luce l'importanza del contributo di Smith come legame tra analisi economica ed economia industriale, sottolineando il movente tecnologico e monopolistico all'integrazione verticale intesa come organizzazione della produzione. Le sezioni 3 e 4 sono dedicate invece al tema dell'integrazione come organizzazione delle transazioni, cioè come scelta allocativa tra mercato ed impresa, all'interno del paradigma istituzionale di Coase e Williamson. Nella sezione 5 esamineremo i modelli di integrazione proposti in ambito neoistituzionalista ed all'interno della teoria dei contratti, in cui l'integrazione si configura come una soluzione al problema dell'allocazione dei diritti al controllo sulle decisioni dell'impresa. Nell'appendice forniremo una stilizzazione grafica del percorso seguito dal concetto di integrazione verticale in ambito industriale.

1. Il concetto di settore verticalmente integrato nell'analisi del processo di formazione della ricchezza

Esaminare il ruolo del processo di integrazione verticale all'interno della tradizione dell'economia politica significa, in prima istanza, identificare due prospettive fondamentali sotto le quali è stato analizzato il processo di formazione della ricchezza da parte di un sistema economico.

La prima fa riferimento all'idea di *scarsità*, ed al rapporto esistente tra domanda effettiva, distribuzione di una dotazione finita di *inputs* produttivi tra impieghi alternativi e determinazione degli *outputs* finali.

La seconda fa invece riferimento all'idea di *producibilità*, non solo dei beni finali ma anche dei mezzi di produzione necessari a produrli, e quindi, in ultima analisi, ad una visione **circolare** del sistema economico, in cui ogni singolo bene può assumere la duplice veste di consumo finale e di fattore di produzione impiegato all'interno uno o più processi.¹

Lo strumento dell'integrazione verticale assume rilevanza all'interno di una concezione del sistema economico basata sulla prima prospettiva esposta, in quanto l'accento posto sulla connessione diretta tra dotazione iniziale di mezzi di produzione non prodotti e beni finali conduce alla rappresentazione del sistema per settori verticalmente integrati.

1.1. I contributi classici

Di volta in volta, nell'ambito dell'analisi economica, il concetto di integrazione verticale è stato applicato sia ai singoli settori che al sistema economico nel suo complesso.

Il primo esempio rilevante di analisi condotta **per settori** è la *Ricchezza delle Nazioni* di Adam Smith (1776 [1973]), in cui il processo di formazione del surplus complessivo a disposizione del sistema, cioè la quantità di beni complessivamente disponibile ai fini del consumo finale, è ricondotto alla quantità di lavoro direttamente o indirettamente necessaria per produrlo: in altre parole, la ricchezza netta del sistema è determinabile sulla base della dotazione iniziale di lavoro a disposizione dello stesso, della sua ripartizione all'interno dei diversi settori, e dell'efficienza dei processi produttivi impiegati.

Quest'ultima è connessa alla domanda effettiva esistente per ogni prodotto (il sapore keynesiano di questa affermazione non è per nulla accidentale): l'estensione del mercato determina il grado di divisione del lavoro, che a propria volta influenza la produttività, attraverso un più efficiente impiego della dotazione di lavoro e l'adozione di nuove tecniche.

Questa rappresentazione del sistema come apparato di trasformazione *a senso unico* si ritrova nei contributi di Malthus (1820 [1951]), Mill (1848) Marx (1867)² e Marshall (1890 [1920]), fino a Young (1928), Pigou (1929a,b) e Keynes (1936 [1973]) e, più oltre, Kaldor (1967).

1. Per un'efficace ed esaustiva analisi del rapporto tra il concetto di *scarsità* e quello di *producibilità* nella storia dell'analisi economica, che va evidentemente oltre la portata del presente lavoro, si rimanda il lettore ai fondamentali contributi di Baranzini e Scazzieri (1990); Quadrio Curzio e Scazzieri (1982a, 1982b, 1985).

2. Marx, in realtà, sintetizza l'approccio per settori verticalmente integrati di Smith con quello per industrie di Ricardo, arrivando così a configurare il lavoro come merce (ri)prodotta all'interno del sistema economico.

L'analisi condotta in termini di integrazione verticale induce a glissare sui nessi di natura circolare che pure caratterizzano l'apparato produttivo, convogliando l'interesse dell'economista sulle grandezze che si trovano alle estremità opposte del processo, cioè lavoro e consumo finale.³ La reiterazione del procedimento logico di integrazione verticale per i beni che rientrano nel processo come mezzi di produzione prodotti porta, in ultima analisi, ad identificare nel fondo di lavoro di un sistema la fonte stessa della sua ricchezza, ed è alla base del meccanismo di formazione dei prezzi che, come nel caso di Smith, instaura una relazione diretta tra salario e *prezzo naturale* di una merce. Questo approccio consente inoltre di individuare le condizioni necessarie al conseguimento di obiettivi di natura macroeconomica/aggregata quali il livello di occupazione ed il reddito, sulla base della relazione intercorrente tra dotazioni iniziali, domanda effettiva e livello di attività: in quest'ambito, il pensiero corre immediatamente a Keynes, in cui il ricorso all'integrazione verticale è spinto ai suoi limiti estremi, fino a trascurare ogni distinzione settoriale ed a stabilire un'unica correlazione di carattere quanto mai globale tra domanda, occupazione e reddito *aggregati*, al fine di studiare le fluttuazioni rilevate nelle ultime due grandezze, sulla base dell'entità e dell'andamento della prima:

The given factors allow us to infer what level of national income measured in terms of the wage-unit will correspond to any given level of employment: so that ... the national income depends on the volume of employment ... in the sense that there is a unique correlation between the two. (Keynes, 1936 [1973], p.246)

Così facendo, tuttavia, Keynes si preclude la possibilità di esaminare gli effetti dello stimolo esercitato dalla domanda sulla dinamica intersettoriale, ovvero assume che tale stimolo sia completamente neutrale ai fini della ripartizione della produzione (e quindi delle risorse) sui diversi settori: il processo di diffusione delle innovazioni come dinamica interna all'apparato produttivo rimane fuori portata.

1.2. Hicks e l'approccio neo-austriaco

L'opportunità di analizzare il percorso seguito da un impulso o da una perturbazione all'interno della struttura settoriale del sistema economico, appare maggiormente sfruttata in Hicks (1932, 1973) ed in tutti i contributi che possono essere fatti rientrare nell'approccio neo-austriaco da lui proposto e, successivamente, riceverà una propria collocazione all'interno della formulazione analitica più comprensiva del problema dell'integrazione verticale dovuta a Pasinetti (1973), pur con le cautele del caso.

Nei termini più generali, analizzare un processo di aggiustamento significa considerare il ruolo ricoperto dal fattore-tempo:

3. A rigore, questa affermazione non vale nel caso di Pigou, la cui analisi si basa, per quanto in misura largamente implicita, sulla presenza di legami *orizzontali* all'interno di una struttura produttiva che è riconducibile ad un modello di relazioni *input-output* *à la Leontief*, e quindi sulla natura sequenziale del processo di aggiustamento che segue una qualsiasi perturbazione, mentre appare pienamente giustificata a proposito, ad esempio, di Keynes.

The vertical concept of economic structure may lead itself to the analysis of the time structure of a diffusion process in which an explicit distinction is introduced among the different stages of the various production processes, so that the stages and their duration are explicitly considered. In this way it is possible to follow through the sequence of repercussions starting from a given original impulse by taking into account the speed of reaction of the different parts of the given economic system. ... Again, consideration of the time structure leads to the analysis of features that remain in the background whenever structural specification is confined to the static classification and typology of economic activities, as is the case with the theories of Pigou and Keynes. (Baranzini e Scazzieri, 1990, p.292-3)

Il modello neo-austriaco elaborato da Hicks pone l'accento, da un lato, sul fatto che il processo produttivo *consuma tempo* (riprendendo, sotto questo profilo, un'idea che può essere fatta risalire, in ultima analisi, a David Ricardo), e dall'altro sulla possibilità di ridurre i prodotti finali di consumo ad una quantità di lavoro pura, in modo da caratterizzare il lavoro come unico *input* primitivo completamente trasferibile da un processo all'altro, al cui interno la sua prima mansione consiste nel produrre i beni capitali necessari, che quindi vengono caratterizzati come beni intermedi specifici e rigorosamente non trasferibili, cioè interni alle singole tecniche (Baldone, 1984); la loro specificità rappresenta così il risultato dell'incorporazione della singola *tecnica* associata a ciascun processo.

In questo senso, appellandosi alla teoria austriaca del capitale (Böhm-Bawerk, 1889; Wicksell, 1901; von Hayek, 1941), la formulazione proposta da Hicks produce un concetto di capitale esclusivamente circolante che non è più riconducibile ad una dotazione di beni fisici, bensì ad un fondo-salari costituito dalla quantità di beni finali con cui finanziare il lavoro necessario al sistema produttivo. Endogenizzando il tasso di attivazione dei processi,

the model becomes sequential and can be used to describe the 'traverse' from one stady-growth path to another, characterized by the use of machines embodying a *superior* technique. (Amendola e Gaffard, 1988, p.31; enfasi nell'originale)

In ultima analisi, quindi, il metodo adottato da Hicks assume la connotazione di un procedimento di *integrazione intertemporale* della produzione (Baranzini e Scazzieri, 1990, p.294-5), in cui scompare la relazione smithiana tra stadi diversi, nella forma di un flusso di materiali in trasformazione progressiva, per lasciare spazio alla distribuzione temporale del fondo di lavoro disponibile lungo la catena di attivazione del processo produttivo:

Hicks's contribution allows us to identify a remarkable difference between the two types of sequential analysis that are associated with the concept of vertically integrated structure. On the one hand, in the case of Smith, vertical integration leads to a

'forward-looking' description of the actual transformation process; whereas in the case considered by Hicks vertical integration leads to a 'backward-looking' description. (Baranzini e Scazzieri, 1990, p.295)

La conclusione di carattere più generale che possiamo trarre dalle considerazioni finora svolte è quindi che il processo di integrazione, nelle varie forme in cui esso è stato condotto, è volto all'analisi di un sistema economico **nel suo complesso**, condotta nei termini di un nucleo circoscritto di variabili fondamentali in grado di sintetizzarne il comportamento: ciò comporta, tuttavia, la perdita dell'insieme di relazioni esistenti all'interno dell'insieme di processi produttivi che lo compongono. Alla luce di questo, il contributo di Pasinetti (1973) si concretizza in un apparato analitico in grado, da un lato, di sintetizzare i diversi approcci che abbiamo incontrato e, dall'altro, di non sacrificare la dinamica intersettoriale sull'altare dell'analisi aggregata.

1.3. Pasinetti ed il concetto di settore verticalmente integrato

Il contributo di Pasinetti (1973) consiste nella costruzione di una struttura analitica per settori che sia contemporaneamente in grado di rendere conto della dinamica globale del sistema.

Partendo da un'esposizione estremamente generale in cui ogni merce è prodotta a mezzo di merci ed il sistema è descritto *visivamente* in termini di 'industrie', ognuna delle quali produce un unico bene, Pasinetti giunge ad una descrizione *logica* del sistema attraverso la sua riduzione analitica ad un insieme di settori verticalmente integrati.

Vediamo di chiarire. Ogni industria è definita in termini di un flusso di quantità di lavoro e di un insieme di quantità fisiche (eterogenee) di merci che Pasinetti definisce rispettivamente 'coefficiente di lavoro diretto' ed unità di capacità produttiva diretta', necessari a produrre un'unità fisica della merce relativa. Il concetto di industria fa quindi riferimento a grandezze aggregate per loro natura **osservabili**.

Il settore verticalmente integrato è invece definito sulla base della riduzione di tali grandezze alle quantità di lavoro e di merci direttamente o indirettamente necessarie al sistema al fine di produrre una unità di un determinato bene finale, ciò che Pasinetti chiama 'coefficiente di lavoro verticalmente integrato' e 'unità di capacità produttiva verticalmente integrata' (1973, p.6): vale a dire, grandezze **non direttamente osservabili** nelle quali risulta compattato ciascun sottosistema corrispondente ad un bene finale. L'insieme di tali coefficienti riassume così il sistema nel suo complesso.

L'operazione di integrazione verticale condotta esplicitamente da Pasinetti mette in evidenza quanto era generalmente implicito in tutte le analisi della creazione di surplus ed in ogni discussione sulla teoria del valore e della formazione dei prezzi.⁴

4. La ripetizione del procedimento di integrazione, fino ad ottenere settori verticalmente integrati di ordine superiore, conduce alla riduzione dei prezzi a sommatorie di quantità di lavoro ponderate per potenze del saggio di profitto, proponendo così una soluzione al *problema della trasformazione* di Marx e generalizzando il metodo sraffiano di riduzione dei prezzi a quantità datate di lavoro.

Non solo: l'adozione del concetto di settore verticalmente integrato è di particolare rilievo ai fini dell'analisi del progresso tecnico e della dinamica ad esso associata, in quanto le grandezze che definiscono il settore non perdono il proprio significato in presenza di mutamenti nelle tecniche, ciò che non si potrebbe affermare a proposito di un'analisi tradizionalmente condotta in termini di industrie.

La descrizione del sistema produttivo per settori verticalmente integrati elimina il problema della ridefinizione di ciascuna industria a seguito di progresso tecnico o anche, più semplicemente, di rimpiazzo del capitale, che interrompendo il legame tra capacità produttiva e reddito a disposizione del sistema compromettono qualsiasi tentativo di analisi dinamica. Il ricorso ad una nozione analitica sintetica quale è quella di settore verticalmente integrato garantisce la continuità di una relazione logica instauratasi tra le grandezze rilevanti, nella misura in cui la rende indipendente da qualsiasi loro misurazione fisica.

2. Integrazione verticale ed economia industriale: una valutazione interlocutoria

Il concetto di integrazione verticale di cui ci siamo occupati finora, in generale, non ha legami con il problema dell'organizzazione industriale, in quanto appare come un procedimento di riduzione logica delle variabili fondamentali di un sistema macroeconomico ad una loro descrizione *permanente*, che agevoli l'analisi delle relazioni fondamentali tra esse.

Si tratta allora di ricercare, all'interno di quella letteratura, un momento in cui emerga, a fianco del tema della struttura del sistema produttivo, quello dell'organizzazione della produzione, cioè un momento in cui la teoria economica inizi ad occuparsi della produzione non più esclusivamente sotto il profilo aggregato, ai fini della creazione di surplus, ma sotto quello disaggregato, come agglomerato di comportamenti e di attività individuali.

Il momento in cui questo avviene è riconoscibile nella *Ricchezza delle Nazioni* di Adam Smith (1776).

Smith, ponendo in relazione la dimensione del mercato (cioè la domanda effettiva) e l'organizzazione della produzione, è il primo e per lungo tempo l'unico autore a porre l'accento sulla possibile formazione di imprese (intese per il momento esclusivamente come centri di produzione) di dimensioni più che puntiformi.

L'entità-impresa riemerge nella letteratura successiva in ambito marginalista neoclassico come agente della *mano invisibile* nell'ambito di una visione decentralizzata del sistema economico volta al raggiungimento dell'equilibrio generale walrasiano. L'impresa che qui compare è la funzione di produzione neoclassica, intesa come soluzione di un problema di scelta tecnologica efficiente. Alla base di questo risultato vi è tuttavia un'ipotesi forte, vale a dire quella relativa alla quasi-concavità della funzione di produzione, che esclude la possibilità di osservare rendimenti *ovunque* crescenti. Questo perchè il caso di rendimenti crescenti è incompatibile con una visione decentralizzata *di mercato* del sistema economico (e quindi anche con il risultato più notevole da essa prodotto, l'equilibrio concorrenziale); d'altra parte, è anche l'unico caso che renda concepibile l'esistenza di un'impresa-centro di produzione di

grandi dimensioni.⁵

Il problema diventa quindi analizzare il passaggio di consegne, *interno alla teoria*, dalla mano invisibile di Smith alla mano visibile di Chandler (1977), in una prospettiva in cui le imprese non siano più entità atomistiche immerse in un ambiente perfettamente efficiente e concorrenziale, ed in cui vi sia spazio per l'iniziativa imprenditoriale schumpeteriana.

Alla base del concetto di impresa, così come di quello di integrazione verticale e quindi, in sintesi, di autorità o gerarchia come scelta organizzativa, vi è l'intuizione che questa organizzazione sia in grado di produrre una *performance* superiore rispetto alla somma delle singole parti, sia sotto il profilo della gestione delle informazioni, sia sotto il profilo dei costi di produzione, sia, in ultima analisi, sotto il profilo del potere di mercato:

Assume that production of a final product can be split up into a series of technologically separable stages. Assume, furthermore, that each stage is organized as a simple hierarchy, each of which exhausts scale economies. The questions to be addressed... are (1) when will the production of components by such technologically separable units be exchanged within a firm rather than across an intermediate product market, and (2) how will such internal transactions be organized? Put differently, given that the final product is to be assembled from a series of separable components, which of these will be bought, which will be made, and how will the latter be organized? This is the vertical integration issue. (Williamson, 1975, p.82)

Questo ci conduce quindi a formulare alcuni interrogativi di fondo, ai quali la letteratura ha fornito risposte eterogenee. Perché esiste l'istituzione che chiamiamo impresa? E, posto che essa sorga in reazione ad imperfezioni del mercato, quali fattori ne delimitano le dimensioni, o, in altre parole, perché non si organizzano tutte le transazioni economiche all'interno di un'unica impresa grande quanto il mercato al quale si sostituisce?

A partire da Viner (1932), fino alla teoria dei mercati contendibili (Baumol, Panzar e Willig, 1982), passando attraverso Stigler (1951, 1968) e Bain (1956, 1959), la letteratura riguardante l'integrazione sia verticale che orizzontale, ed in generale la dimensione d'impresa, abbonda di spiegazioni di natura tecnologica, le quali in ultima analisi si traducono in considerazioni sull'andamento dei costi. Tutto questo appare perfettamente plausibile, ma, come fa osservare Williamson:

5. Si noti che la soluzione recentemente proposta dalla teoria della contendibilità (Baumol, Panzar e Willig (1982); Panzar e Willig (1981), tra gli altri) è solo parziale, valutando la (grande) dimensione dell'impresa in termini di efficiente allocazione delle risorse e, quindi, di economia del benessere.

Technology is fully determinative of economic organization only if (1) there is a single technology that is decisively superior to all others and (2) that technology implies a unique organization form. (1985, p.87)

Queste brevi considerazioni riassumono i punti salienti del dibattito che accomuna la teoria dell'impresa e la teoria dell'organizzazione industriale. Nelle sezioni che seguono tenteremo di fornire un quadro delle linee principali lungo le quali si è sviluppato.

2.1. Il teorema smithiano e la tradizione di Chicago

La prima visione dell'impresa che emerge dalla letteratura di radici neoclassiche e che si spinga oltre i confini della funzione di produzione governata dalla legge delle proporzioni variabili è quella di un'organizzazione delle attività di produzione e di scambio in grado, da un lato, di esaurire le economie di scala e di produzione multipla (*economies of scope*), dall'altro di aggirare gli apparati regolativi imposti al mercato dall'esterno ed i vincoli costituiti dalla formazione di cartelli al suo interno, con il fine ultimo di esercitare un potere di monopolio (e di discriminazione dei prezzi) su uno o più mercati contigui.

Questa visione scaturisce per via diretta dal teorema smithiano, *'the division of labour is limited by the extent of the market'*, le cui implicazioni sono state poi riprese ed approfondite da Young (1928) e da Stigler (1951). La tesi comune a questi contributi è che l'espansione del mercato dovrebbe produrre una forte tendenza alla disintegrazione (e viceversa), perchè una maggiore dimensione del mercato (cioè una maggiore domanda) favorisce l'assegnazione di ogni singola fase produttiva ad un'unica unità specializzata. Ciò che si dovrebbe osservare in un settore 'adulto' è quindi una sorta di integrazione (e di concentrazione) *orizzontale*. Si tratta, come si vede, di una tesi piuttosto peculiare, in quanto contrasta con l'idea molto diffusa secondo la quale ad un sistema industriale maturo si associa l'esistenza di imprese di grandi dimensioni, o comunque in crescita, il cui obiettivo (una volta che il proprio stadio primario abbia raggiunto una dimensione sufficiente) è quello di realizzare le economie offerte dagli stadi contigui. In questo senso, quindi, il modello Smith-Young-Stigler si focalizza sullo sfruttamento delle economie di scala nell'ambito di un singolo stadio, escludendo la possibilità di economie di produzione multipla (verticale) tra stadi contigui.

Tuttavia, ciò che rileva, in questa sede, è sottolineare come in Smith compaia per la prima volta, per quanto in modo comprensibilmente embrionale, il concetto di integrazione come nesso tra organizzazione della produzione e struttura del mercato, cioè come problema di organizzazione industriale.

La scuola di Chicago sviluppa l'intuizione smithiana in modo piuttosto eclettico. In Stigler (1951, 1968) e, successivamente, in Perry (1978), l'integrazione verticale assume l'ulteriore connotazione di uno strumento per ottenere profitti monopolistici e, nello stesso tempo, operare una discriminazione di prezzo sui mercati nei quali l'impresa è operante.

L'argomentazione monopolistica è sviluppata pienamente da Bain (1956, 1959), secondo il quale se, da un lato, l'integrazione è dovuta ad economie legate alla complementarità tecnologica dei processi impiegati, dall'altro, in assenza di tali moventi, la spinta all'integrazione sarà interpretabile sulla base dell'incremento dei poteri di monopolio,

in termini di barriere all'entrata, in quanto (a) l'integrazione a monte o a valle ottiene il risultato immediato di eliminare uno o più concorrenti potenziali per i mercati nei quali già si operi; (b) aumenta i requisiti in termini di fabbisogno di capitale iniziale per i potenziali entranti.⁶

Un'interpretazione affine, anche se non coincidente con questa, è quella che vede l'integrazione come una forma di comportamento strategico, con il fine ultimo di escludere dal mercato gli avversari non integrati, elevandone i costi (Salop e Scheffman, 1983).⁷

L'esame di questo ampio spettro di letteratura fa tuttavia sorgere il sospetto che dietro la panoplia dei moventi individuati si celi ancora qualcosa di non detto. E questo si manifesta nell'intuizione che il lato monopolistico dell'integrazione verticale, che finora ha sottolineato come l'impresa possa compromettere il funzionamento efficiente del mercato (e quindi appare come indizio per una sorta di incriminazione dell'impresa), non sia solo un obiettivo strategico a cui è mirato il suo comportamento, ma anche una soluzione dettata da elementi ancora celati dietro le quinte.

Un passo in questo senso viene compiuto già in ambito neoclassico, arricchendo la funzione di produzione del tema organizzativo. Il risultato è la **funzione di produzione di gruppo** (*team production function*) di Alchian e Demsetz (1972). L'integrazione delle funzioni produttive all'interno dell'impresa è esclusivamente motivata dalle caratteristiche di superaddittività e non separabilità della funzione di produzione.⁸ L'*output* complessivo è il risultato dei contributi forniti da più fattori produttivi, non valutabili singolarmente: questo fa sì che si verifichi un problema di *free riding*, in quanto i singoli fattori non sono adeguatamente incentivati. La soluzione proposta da Alchian e Demsetz consiste nell'introdurre un soggetto esterno (il *monitor*) che ha il preciso incarico di controllare i comportamenti individuali all'interno del *team*, ed al quale spettano i diritti residuali sulle entrate (al netto della retribuzione dei fattori. In questo senso, egli non è altro che il proprietario dell'impresa, in quanto la sua retribuzione coincide esattamente con i profitti.

E' tuttavia evidente che si tratta di una visione limitata del problema, in quanto operando in un mondo caratterizzato da certezza perfetta, il contributo dell'*input* organizzativo si esaurisce nella supervisione dei comportamenti dei singoli *inputs* produttivi. Il tema dell'integrazione come manifestazione del rapporto tra impresa e mercato può essere esaminato sotto un profilo alquanto diverso e di più ampia portata.

6. Aghion e Bolton (1987) e Hart e Tirole (1990) hanno sottolineato il ruolo dei contratti a lungo termine come barriera all'entrata. In particolare, Aghion e Bolton (1987, pp.392-6), mostrano come, in condizioni di informazione asimmetrica, la lunghezza dei contratti possa costituire un segnale della probabilità di entrata. Sul ruolo svolto dai contratti a lungo termine all'interno della teoria dell'impresa, rimandiamo alla sezione 5.

7. Questa interpretazione può essere fatta risalire fino ad Alfred Marshall, che osserva come il controllo unificato delle fasi di lavorazione relative ad un unico prodotto finale dia vita ad un monopolio molto più difficile da intaccare rispetto alle singole fasi gestite sotto proprietà separata (1920, p.495).

8. Una funzione di produzione $Y(x)$ è superadditiva (rispetto all'*output*) se $Y(x_i + x_j) > Y(x_i) + Y(x_j)$; è non separabile se $\frac{\delta^2 Y}{\delta x_i \delta x_j} \neq 0$.

3. La Natura dell'Impresa: un debito storico

L'intuizione originale di porre in discussione e rifondare le basi tecnologiche sulle quali si fondava l'impresa neoclassica appartiene a Ronald Coase (1937), che ha riformulato il problema dell'organizzazione del sistema economico in termini di comparazione istituzionale.

Nei confronti di questo lavoro sono debitrice le generazioni di contributi successivi che in tempi molto recenti ne hanno raccolto il testimone. In effetti, come ha avuto modo di sottolineare lo stesso Coase (1972), la sua teoria dell'impresa è sempre stata "molto citata e poco usata", senza peraltro che la scarsa sistematicità della stessa, come rileva Williamson (1975), possa apparire un motivo sufficiente.

L'obiettivo che Coase si proponeva consisteva nel colmare il vuoto che separava l'ipotesi per cui l'allocatione delle risorse viene effettuata tramite il meccanismo dei prezzi e quella, di sapore schumpeteriano, secondo la quale questa stessa allocatione si realizza attraverso l'attività coordinatrice di un imprenditore, che produce un'organizzazione 'altra' rispetto al mercato, l'impresa:

It can ... be assumed that the distinguishing mark of the firm is the supersession of the price mechanism. (Coase, 1937, p.389)

Si trattava quindi di raccogliere elementi sufficienti per poter giudicare se le due forme dovessero intendersi come alternative o piuttosto complementari, ovvero di identificare i motivi che portano al sorgere dell'organizzazione-impresa ed al suo progressivo sostituirsi al mercato.

La prima ragione, che costituisce il nucleo del contributo di Coase, è individuata nei **costi di transazione** che l'uso del meccanismo di mercato implica per gli operatori, costi che possono essere ampiamente ridotti facendo ricorso ad un'organizzazione alternativa, quale appunto l'impresa, che nasce come risposta al problema posto dalla necessità di negoziare un contratto separato per ciascuna transazione e dalla non improbabile impossibilità di farlo. Più precisamente,

A firm is likely .. to emerge in those cases where a very short term contract would be unsatisfactory. (Coase, 1937, p.392)

Quindi i contratti non vengono del tutto eliminati, ma vengono largamente ridotti: se un gruppo di fattori produttivi collabora in modo continuativo all'interno di un'impresa, può scegliere di stipulare un unico contratto a lungo termine destinato a sfociare nella forma-impresa, piuttosto che affidarsi al mercato con una serie di contratti a breve.

Il secondo punto cardine è il ruolo ricoperto dall'incertezza (e, come vedremo, dalla razionalità limitata), messo in evidenza prima da Knight (1921) e successivamente ripreso da Coase: la necessità di operare in un mondo incerto in cui l'informazione è scarsa e costosa, richiede la costituzione di un apparato organizzativo in grado di assicurare i flussi informativi in modo più efficiente rispetto al mercato:

With uncertainty entirely absent, every individual being in possession of perfect knowledge of the situation, there would be no occasion for anything of the nature of responsible management or control of productive activity. Even marketing operations in any realistic sense would not be found. The flow of raw materials and productive services through productive processes to the consumer would be entirely automatic. ... With uncertainty present, doing things, the actual execution of activity, becomes in a real sense a secondary part of life; the primary problem or function is deciding what to do and how to do it. (Knight, 1921 [1971], pp.267-8)

Da queste prime considerazioni sorge un ulteriore interrogativo, cioè se sia possibile individuare i fattori che intervengono a determinare la dimensione di un'impresa e, in ultima analisi, se è vero che l'impresa nasce come risposta all'esistenza di costi di transazione, perchè si continui comunque a ricorrere al mercato; perchè, in altre parole, non si faccia ricorso ad un'unica grande impresa, capace di *surrogare* completamente il mercato.

Posto che

A firm ... consists of the system of relationships which comes into existence when the direction of resources is dependent on an entrepreneur. ... A firm becomes larger as additional transactions ... are organised by the entrepreneur and becomes smaller as he abandons the organization of such transactions. The question which arises is whether it is possible to study the forces which determine the size of the firm. Why does the entrepreneur not organise one less transaction or one more? (Coase, 1937, p.393-4)

La risposta che Coase fornisce poggia sui rendimenti decrescenti attribuibili all'attività manageriale (o genericamente organizzativa, tenendo conto del fatto che, per il momento, la distinzione tra proprietà e controllo è ancora al di fuori dell'analisi).

L'andamento decrescente è ascrivibile, da un lato, al costo marginale crescente associato a ciascuna transazione aggiuntiva compiuta all'interno dell'impresa, fino a raggiungere quel punto in cui esso è pari al costo della stessa transazione condotta sul mercato oppure in un'altra impresa; dall'altro, è da addebitarsi ad una diminuzione delle capacità dell'imprenditore di allocare in modo efficiente i fattori produttivi, al crescere del numero delle transazioni portate a termine. Il risultato di questo processo è analogo a quello precedente: la perdita dovuta ad una allocazione delle risorse sempre più inefficiente diviene tale da eguagliare e neutralizzare totalmente il vantaggio precedentemente acquisito trasferendo la transazione dal mercato all'impresa.

Ciò significa che l'impresa surrogherà il mercato solo nella misura in cui quest'ultimo non si dimostrerà efficiente, e non oltre, lasciando comunque aperta la possibilità di un ritorno ad esso nel caso opposto.⁹

4. La Nuova Economia Istituzionale: Williamson

Oliver Williamson (1964, 1975, 1985, 1986 e 1989) riprende l'intuizione di Ronald Coase (1937), sviluppandone le tesi germinali fino a fondare quella che lui stesso ha battezzato *Nuova Economia Istituzionale*. Williamson riconosce il proprio debito nei confronti del precursore in termini molto espliciti:

it is not uncertainty or small numbers, individually or together, that occasion market failure but it is rather the *joining* of these factors with bounded rationality on the one hand and opportunism on the other that gives rise to exchange difficulties.
(Williamson, 1975, p.7)

In questa breve affermazione, che possiede l'impatto espressivo di una sentenza, è sintetizzato il senso complessivo dell'approccio istituzionale, attraverso il quale Williamson identifica i fattori umani ed ambientali che giustificano il sorgere dell'impresa.

Riconsideriamo i termini del problema. L'efficienza del mercato deve fare i conti con i costi di transazione, che potremmo definire *costi di funzionamento del sistema economico* (Arrow, 1969, p.48; cit. in Williamson, 1985, p.8), equivalenti agli attriti nei sistemi fisici. Di conseguenza, si può scegliere di abbandonare il mercato a favore dell'impresa, cioè di un'organizzazione di tipo gerarchico che consenta di evitare i rischi connessi alla stipulazione di una serie di contratti contingenti ed alla necessità di farne rispettare i termini. Il progressivo fallimento del mercato conduce alla sua eliminazione.¹⁰ Questa era la conclusione a cui era già arrivato Coase. Williamson la riprende per darle una fisionomia più netta. Il suo approccio può essere sintetizzato nei seguenti punti:

- (a) il mercato e l'impresa sono strumenti alternativi per mezzo dei quali portare a termine un determinato insieme di transazioni;
- (b) la loro efficienza relativa rappresenterà il criterio in base al quale assegnare una transazione al primo o alla seconda;
- (c) l'ordine di grandezza dei costi legati alla stesura ed all'esecuzione di contratti complessi sul

9. Questa tesi riemergerà in Williamson (1975).

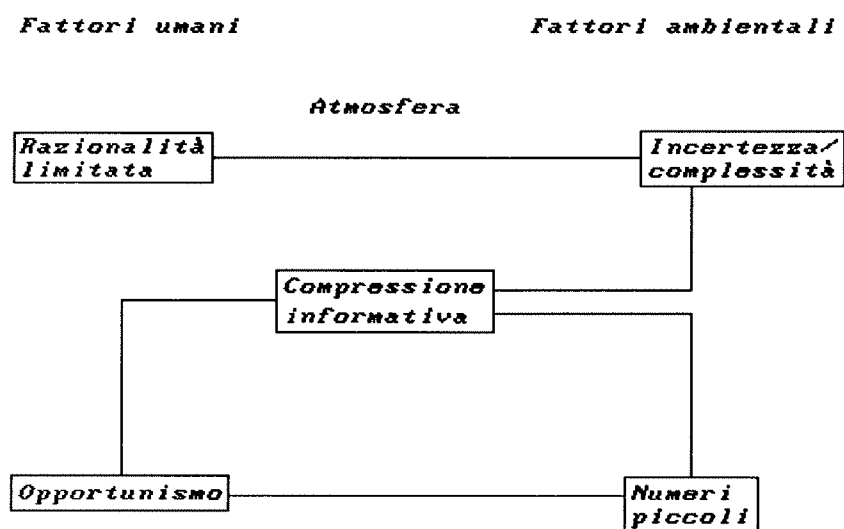
10. Williamson assume, sia per coerenza verso la storia, sia per chiarezza espositiva, che *in principio erano i mercati*, e specifica che il loro fallimento è presunto avvenire quando si passa all'organizzazione interna. Il tono abbastanza tautologico di un'affermazione come questa serve tuttavia a chiarire che si tratta appunto di un'ipotesi di partenza: le transazioni possono essere trasferite dal mercato all'impresa quando e perchè ne conseguono dei vantaggi (vedremo quali) ma resta valido anche il legame inverso, in quanto la presenza di transazioni che non vengono internalizzate implica il fallimento dell'organizzazione interna (Williamson 1975, pp.20-21).

mercato dipende dalle caratteristiche personali di coloro che sono responsabili delle decisioni, così come dalle caratteristiche del mercato in questione;

(d) Infine, gli stessi fattori, d'ordine sia umano che ambientale, che influenzano il funzionamento del mercato, agiscono anche all'interno dell'impresa, anche se in modo diverso.

I fattori ambientali a cui Williamson fa riferimento sono: (a) *incertezza* e (b) relazioni di scambio in condizioni di *numeri piccoli*. I fattori di carattere umano sono (a) *razionalità limitata* e (b) *opportunismo*.

A supporto dell'esposizione abbiamo di seguito riprodotto lo schema costruito da Williamson (1975, p.40) per evidenziare i legami esistenti tra i succitati fattori:



- Fig. 4.1

Per prima cosa, consideriamo la combinazione tra incertezza e razionalità limitata. Secondo Simon,

The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world. (Simon, 1957, p.198, enfasi nell'originale)

Se, tenendo conto di tali limiti, siano essi di natura neurofisiologica o di linguaggio, è impossibile o almeno estremamente costoso venire a conoscenza di tutte le contingenze future ed adeguarsi ad esse ex ante, allora un'organizzazione interna può sostituirsi ai contratti, soprattutto nei casi in cui si prevede che la relazione instaurata debba durare nel tempo. In tal modo, essa può proporsi come uno strumento per far fronte alla presenza contemporanea di incertezza e complessità attraverso un processo di intervento sequenziale e selettivo che non comporta i rischi che si correrebbero affidandosi esclusivamente al mercato. Il ricorso

all'impresa consente di economizzare in razionalità limitata in situazioni in cui i prezzi non sono più indicatori sufficientemente attendibili.

Il legame tra opportunismo e numeri piccoli è molto più intricato. In prima approssimazione, includere l'opportunismo nel quadro dell'analisi significa ampliare la portata dell'ipotesi neoclassica di massimizzazione di funzioni obiettivo individuali, in quanto dobbiamo ora tener conto di possibili comportamenti di tipo strategico fondamentalmente 'disonesti' o comunque poco trasparenti. D'altra parte, un'inclinazione di questo genere comporta rischi contenuti fintantochè si abbiano relazioni di scambio di tipo concorrenziale (caratterizzate cioè da *grandi numeri*); in altre parole, è dalla possibilità di comportamenti opportunistici che deriva l'importanza dei numeri piccoli.

Per comprendere esattamente cosa intenda Williamson con questa espressione, è necessario pensare al vantaggio 'della prima mossa' (*first-mover advantage*). Supponiamo che la condizione di partenza su di un certo mercato sia, con una ragionevole approssimazione, concorrenziale; supponiamo poi di metterci nella posizione di un produttore di un determinato bene finale: se, ad un ipotetico istante iniziale $t=0$, in cui si inaugurino le trattative, un qualunque soggetto ottiene un contratto esclusivo con un fornitore (a monte) e/o con un distributore (a valle), ne ricaverà un vantaggio in virtù della stessa natura esclusiva della relazione che si è creata, che intacca la competitività del mercato. Anche se inizialmente è coinvolto un grande numero di operatori, questo generalmente non sarà più vero in seguito, quando alla contrattazione ex novo si sostituisce un processo di esecuzione e rinnovo dei contratti esistenti, la cui 'metabolizzazione' conduce rapidamente ad una situazione di numeri piccoli.

Con questo, arriviamo al nocciolo della questione, rappresentato dalla nozione di **compressione informativa** (*information impactedness*), alla quale Williamson fa risalire (e riassume) entrambi i fenomeni di *adverse selection* e *moral hazard*.¹¹

Essa scaturisce dalla combinazione tra opportunismo, incertezza e numeri piccoli, e si verifica quando una delle parti coinvolte nello scambio possiede informazioni maggiori o più dettagliate rispetto all'altra, riguardo circostanze rilevanti ai fini dello scambio stesso, e la controparte non può entrarne in possesso se non sopportando costi eccessivamente alti. La categoria di informazioni di cui ci stiamo occupando può essere 'compressa' perchè (a) chi ne è in possesso non la divulga e la utilizza ai propri fini (ovvero, è un opportunist), oppure (b) anche se fosse seriamente intenzionato a renderla di dominio pubblico, non potrebbe farlo, perchè non è tecnicamente in grado di comunicarla in modo esauriente e preciso.

11. Tra i concetti di compressione informativa ed opportunismo adottati da Williamson e quelli di asimmetria informativa, *adverse selection* e *moral hazard* esiste una corrispondenza intuitiva ma per certi versi incompleta. Sinteticamente, possiamo dire che se, al momento della stesura di un contratto o della realizzazione di una transazione, esiste asimmetria ex ante, ci troviamo in condizioni di *adverse selection*; se invece vi è asimmetria ex post, ci troviamo in condizioni di *moral hazard*. A questo proposito, si rimanda ad Akerlof (1970) e Stiglitz (1985).

4.1. Dentro l'impresa: la gerarchia

Nel corso del capitolo precedente abbiamo più volte fatto riferimento all'impresa come ad una *gerarchia*. Come apparirà chiaramente col procedere dell'esposizione, l'approccio gerarchico è un modo per aprire la scatola nera che è l'impresa neoclassica ed esaminarne l'organizzazione interna (Holmström e Tirole, 1989).

La nozione di gerarchia è legata strettamente al nome di Oliver Williamson, ed è dalla sintesi del suo punto di vista al riguardo che prenderà le mosse questa breve panoramica.

La gerarchia può essere concepita come:

- 1) una forma organizzativa;
- 2) un sistema per la raccolta ed il trattamento dell'informazione;
- 3) una struttura d'autorità.

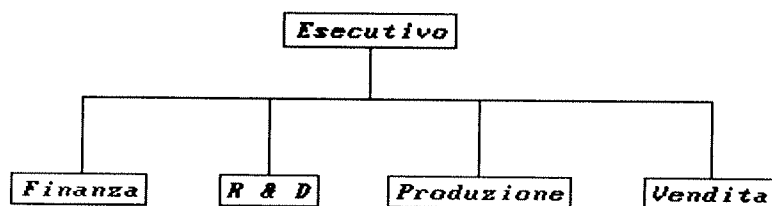
4.1.1. La gerarchia come forma organizzativa

In questo paragrafo ci occuperemo del mutamento strutturale che ha luogo all'interno della grande impresa nel momento in cui passa da un'organizzazione in forma-U ad una in forma-M. Quest'idea è stata concepita (e documentata) da Williamson (1967, 1970, 1975, 1985) e, prima di lui, da Chandler (1966).

La forma-U (unitaria) è quella che, generalmente, contraddistingue l'impresa giovane e quindi, di solito, di piccole dimensioni. Essa è concepita in modo da riunire le attività secondo le loro funzioni all'interno dell'impresa, sotto un'unico *staff* direttivo. Non è altro che una trasposizione del problema dell'integrazione verticale: è il modo naturale per organizzare e condurre attività multifunzionali, in quanto la specializzazione per funzioni consente di realizzare sia le economie di scala sia un'efficiente divisione del lavoro (Williamson, 1975, p.133). La motivazione alla base della sua scelta è la possibilità di evitare la duplicazione dei costi associati ad ogni funzione.

Ciò non toglie che la forma-U possa implicare le complicazioni tipiche della produzione *in team*, (cfr. sez.2.1) dal momento che le caratteristiche ultime ed il successo del prodotto finale sono decretati dal concorso di molteplici contributi difficili da valutare e supervisionare singolarmente da parte della direzione.

In Figura 4.2 è illustrata la struttura di un'impresa in forma-U:



- Fig. 4.2 - La forma-U

Secondo Chandler (1962), l'espansione radiale dell'impresa in forma-U è sottoposta ad effetti cumulativi di **perdita di controllo** (per la nozione di perdita di controllo - *control loss* - cfr. infra, par.3) che ne incrinano progressivamente l'efficienza e favoriscono il perseguimento di obiettivi distinti dal profitto, in quanto alterano la natura dei processi

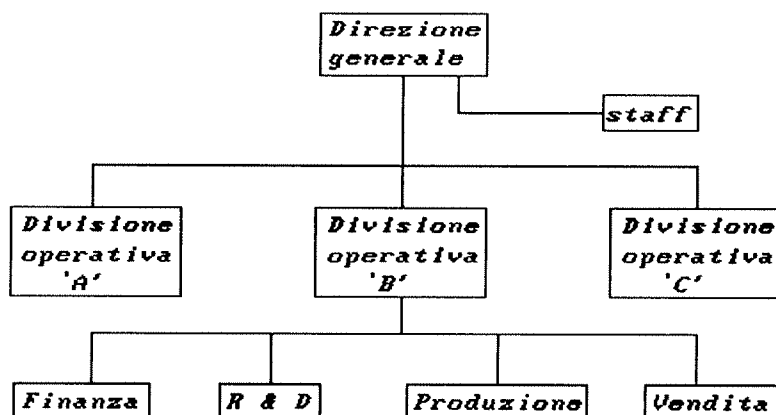
decisionali.

Operando in condizioni di sovraccarico, la forma-U e, soprattutto, il suo limitato *staff* dirigenziale, incontrano dei limiti invalicabili dovuti alla propria razionalità limitata: l'incapacità di identificare e tanto più di realizzare obiettivi organizzativi globali, combinata con l'insorgenza di atteggiamenti opportunistici, fa sì che si perseguano invece obiettivi parziali, specifici delle singole parti funzionali.

Una volta che la sua stessa espansione decreta la morte della forma-U, essa viene sostituita dalla forma-M (multidivisionale), rappresentata in Figura 4.3.

La forma-M può essere approssimativamente considerata come un grappolo di forme-U in scala ridotta, che ne costituiscono le singole divisioni, la cui *performance* può essere ora misurata in modo affidabile grazie alla presenza, in cima al 'grappolo', di un ufficio direttivo dotato di capacità ed incarichi di carattere generale, tra i quali la pianificazione strategica, l'allocazione efficiente delle risorse a lungo termine, la creazione e la gestione di un apparato di controllo esteso su tutte le divisioni.

[The] new structure left the broad strategic decisions as to the allocation of existing resources and the acquisition of new ones in the hands of a top team of generalists. Relieved of operating duties and tactical decisions, a general executive was less likely to reflect the position of just one part of the whole. (Chandler, 1962 [1966], p.383)



- Fig. 4.3 - La forma-M

In sintesi, quindi, il passaggio alla forma multidivisionale e la creazione di uno *staff* direttivo di vertice implica che le responsabilità di quest'ultimo si estendono sul seguente spettro di compiti: (a) identificazione di attività separabili all'interno dell'impresa; (b) conferimento ad ognuna di esse di una posizione quasi-autonoma (un'espressione equivalente per dire che ogni divisione è una quasi-impresa ed è responsabile delle decisioni operative); (c) controllo della *performance* di ognuna; (d) incentivazione; (e) allocazione dei flussi liquidi; (f) pianificazione strategica (diversificazione, investimenti, acquisizioni).

La *corporation* in forma-M può ricondursi ad un mercato finanziario in miniatura:

infatti, le disponibilità liquide di varia provenienza entrano in una competizione interna che ne seleziona la destinazione di investimento, invece di essere immediatamente reinvestite 'alla fonte'.¹²

Tuttavia, Williamson (1985) individua un *tradeoff* tra ampiezza e profondità, dal momento che la diversificazione può rivelarsi eccessiva, se la capacità di allocazione delle risorse viene spinta al proprio limite od oltre, provocando, al margine, un'allocazione inefficiente e fenomeni di opportunismo.

Per quanto riguarda il nesso che si può ricercare tra la forma organizzativa e gli assunti teorici sul comportamento dell'impresa, Williamson sostiene che

The organization and operation of the large enterprise along the lines of the M-form favors goal pursuit and least-cost behavior more nearly associated with the neoclassical profit maximization hypothesis than does the U-form organizational alternative.
(Williamson, 1975, p.150, enfasi nell'originale)

Ciò che però rende effettivo questo legame non è la divisionalizzazione in sé, quanto la separazione delle decisioni di carattere strategico dall'insieme delle altre. L'efficacia di questa ripartizione, a sua volta, si fonda sullo sviluppo di un apparato di controllo nelle mani dell'esecutivo di vertice.

La sua estraneità alle questioni 'tattiche' rende possibile l'insorgere di un impegno psicologico (*psychological commitment*) nei confronti della *performance* complessiva dell'organizzazione. Non solo: questo tipo di struttura combina l'esigenza di razionalità con l'opportunità di realizzare un sinergismo. Il tutto è più grande (in questo contesto, 'grande' sarà da interpretarsi come 'efficiente') della somma delle sue parti.

4.1.2. La gerarchia come sistema informativo

La disponibilità e la trasmissione di informazioni all'interno di una struttura organizzativa sono importanti in quanto agevolano i processi decisionali e migliorano il controllo sul comportamento dei suoi membri.

Nella letteratura relativa alla *team theory*,¹³ si è cercato di isolare l'aspetto decisionale, tramite l'ipotesi secondo la quale tutti i membri del *team* condividono lo stesso obiettivo (Marschak e Radner, 1972): il problema è definire come avvenga la raccolta delle informazioni e, soprattutto, attraverso quali canali vengano comunicate, ed è sviluppato in due fasi successive.

Il primo stadio consiste nell'individuare lo schema decisionale ottimale per

12. Il mercato finanziario può essere interpretato come uno strumento di controllo nei confronti dello *staff* dirigenziale (questa tesi è espressa, tra gli altri, da Marris (1964)) ma, secondo Williamson, essendo uno strumento esterno, è di efficacia limitata.

13. In questo contesto, il termine *team* assume quindi una connotazione diversa da quella associata alla *team production function* (cfr. sez.2.1).

l'organizzazione, in base al determinato bagaglio informativo che si è ottenuto. Il secondo stadio consiste invece nella valutazione dei sistemi informativi alternativi. Tuttavia, è in genere troppo complesso individuare la migliore struttura informativa, dato un certo livello di costi d'acquisizione e trasmissione dei dati, cosicché la letteratura ha preferito concentrarsi sulla valutazione dei benefici associati ad ogni sistema in modo 'discreto', lasciando che l'aspetto dei costi fosse analizzato in altra sede.¹⁴

Ciò non significa che questa teoria non abbia raggiunto risultati importanti. Il primo di questi è che, all'interno della regola decisionale ottimale, lo schema decisionale di ogni singolo membro è ottimale, dati gli schemi di tutti gli altri; ovvero, per usare il linguaggio della teoria dei giochi, si assiste alla realizzazione di un equilibrio di Bayes-Nash.

Inoltre, un risultato di validità generale che emerge dal secondo stadio, cioè la fase comparativa, è che, in completa consonanza con Blackwell, (1951), un sistema informativo, che chiameremo α , riceverà una valutazione più favorevole rispetto ad un altro, che chiameremo β , se è meno 'disturbato', o 'rumoroso', se cioè il segnale fornito da β si può interpretare in termini di α , più un fattore casuale che costituisce il 'rumore': in sintesi, α è una statistica sufficiente per β .

4.1.3. La gerarchia come struttura d'autorità

I 'vuoti' inevitabili nell'insieme dei contratti, la loro incompletezza, rendono necessario il ricorso alla negoziazione, che può essere governata designando un arbitro, al quale si riconosce l'autorità di risolvere la controversia: la delega d'autorità, e l'arbitrato interno che essa crea, sostituiscono il sistema giudiziario.

Come fa osservare Williamson (1985, pp.135-38) e come ribadiscono Holmström e Tirole (1989, p.106 e ssgg.), l'aspetto essenziale dell'autorità all'interno di un'organizzazione è la sua struttura verticale: la maggior parte dei conflitti che insorgono tra divisioni o tra singoli membri sono risolti da autorità superiori, tramite quello che Williamson definisce **intervento selettivo**. Da questo punto di vista, il ricorso alla gerarchia (in questo caso anche al posto del mercato 'giudiziario') si traduce in un processo sequenziale che elimina la maggior parte dei rischi connessi sia alla stesura dei contratti contingenti, sia al ricorso al giudizio di una corte.

La visione della gerarchia come struttura d'autorità presuppone un'attività di supervisione e controllo, a sua volta strettamente connessa alla trasmissione delle informazioni degli incentivi lungo la gerarchia stessa.

L'interesse di Williamson per il fenomeno della **perdita di controllo** (*control loss*) nasce dalla questione della dimensione ottimale dell'impresa. In contrasto con un'opinione molto frequentata, secondo la quale solo in un contesto dinamico sorge un'esigenza di

14. Per una eccezione notevole, si veda Geanakoplos e Milgrom (1984), i quali includono i costi nel proprio modello, attribuendoli a ritardi nel processo decisionale. Ne risulta una dimensione ottimale finita per la gerarchia, in quanto il beneficio marginale derivante da un'espansione del nucleo dirigenziale è decrescente. E' una conclusione molto vicina a quella a cui era giunta, per altra via, la Penrose (1959).

coordinamento, che a sua volta si riflette sul tasso di crescita dell'impresa,¹⁵ Williamson afferma che la crescita non rappresenta una *conditio sine qua non* per quella funzione di coordinamento che è componente tipica dell'attività manageriale. Sostiene piuttosto che il fatto di operare in condizioni di razionalità limitata costituisce un vincolo (quasi)-statico alla dimensione d'impresa, attraverso il meccanismo della perdita di controllo: la necessità organizzativa prende le mosse da questo, e le considerazioni riguardanti la crescita non fanno altro che sottolinearla.

Intuitivamente, di pari passo con l'espandersi di un'organizzazione, si indebolisce il controllo esercitato su di essa da parte del vertice:¹⁶ di conseguenza, si assiste costantemente ad un *tradeoff* tra qualità e quantità. In altre parole, il vertice esecutivo è costretto ad un compromesso tra ampiezza e profondità delle informazioni in suo possesso, ogni volta che si realizzi una espansione. Il deterioramento qualitativo subito dal flusso informativo diretto al vertice, così come dal flusso di istruzioni diretto verso la base, costituisce quella che Williamson chiama perdita di controllo.¹⁷

5. L'impresa come insieme di contratti a lungo termine

All'interno della letteratura neoistituzionalista e di tutti i contributi indebitati in qualche misura con l'intuizione di Coase, l'integrazione e, quindi, la dimensione, vengono interpretate sulla base dei problemi connessi alla stesura ed alla durata dei contratti: inaccessibilità delle informazioni, complessità ambientale, razionalità limitata e, in ultima analisi, fenomeni di *moral hazard*.

Alchian e Demsetz (1972) asserivano che l'essenza dell'organizzazione-impresa non è da ricercarsi nei contratti a lungo termine tra l'imprenditore ed i proprietari dei fattori produttivi, in quanto non esisterebbe necessariamente un vincolo contrattuale che obblighi né l'uno né gli altri a proseguire la propria collaborazione, se non intendono farlo. Per contro, l'approccio neoistituzionalista asserisce che il motivo di fondo per l'esistenza di quella particolare struttura contrattuale che è l'impresa risiede nella specificità di capitale, che produce *appropriable quasi-rents*, rappresentate dalla differenza tra il valore attribuibile al capitale nell'impiego per cui è stato concepito e quello che avrebbe nel migliore impiego alternativo (Klein, Crawford e Alchian, 1978).

Quindi, le relazioni a lungo termine e le caratteristiche strutturali delle imprese e dei

15. Secondo Kaldor (1934), in condizioni statiche i problemi di coordinamento sono assenti, mentre si manifestano in una prospettiva dinamica, nel cui ambito rappresenterebbero un effettivo limite alla crescita dell'impresa. Tuttavia, questo approccio appare eccessivamente ottimista, ed adottarlo significa 'entrare nel Nirvana dell'analisi economica' (Robinson, 1934). Ciononostante, è vero che, in un'accezione più misurata, questa argomentazione non sia priva di peso: l'espansione dipende dalle capacità di pianificazione e controllo, che a propria volta dipendono dall'esperienza; essendo questa un bene scarso e costoso, il tasso di crescita sarà limitato.

16. Questa è l'esatta formulazione della *legge del controllo decrescente*, nata all'interno delle teorie del comportamento burocratico (Downs, 1967).

17. Per una trattazione formale del concetto di perdita di controllo, si vedano Williamson (1967, 1986) e Calvo e Wellisz (1978).

mercati che da esse scaturiscono appaiono spiegabili in virtù dell'esistenza di transazioni riguardanti beni *specifici* alla relazione stessa, intendendo con questo una categoria molto ampia, dai beni capitali in senso stretto ad *inputs* di lavoro altamente specializzati.

La situazione tipica è quella di un investimento irreversibile in un bene capitale durevole il cui valore all'interno di quella particolare relazione sia superiore a quello che avrebbe al di fuori di essa, ed in virtù del quale le parti sperimentano quello che Williamson ha opportunamente battezzato *lock-in effect*. Questo in ultima analisi sfocia nella proprietà unificata, tramite l'integrazione verticale. Di conseguenza, si ha una condizione di numeri piccoli, ad evidente svantaggio di chi si affidi ancora alla libera contrattazione sul mercato.

Si consideri un bene che può essere prodotto utilizzando due diverse tecnologie, una 'generica' e l'altra 'specifica', nel senso che quest'ultima richiede ampi investimenti in beni capitali specifici e di conseguenza è più efficiente (genericamente parlando, migliore).

Se contrassegnamo con k la misura della specificità di capitale, e con s le misure prese a salvaguardia della transazione prevista dal contratto, è possibile ricostruire una mappa contrattuale come quella in Figura 5.1 (Williamson, 1985, pp.32 e ssg.).

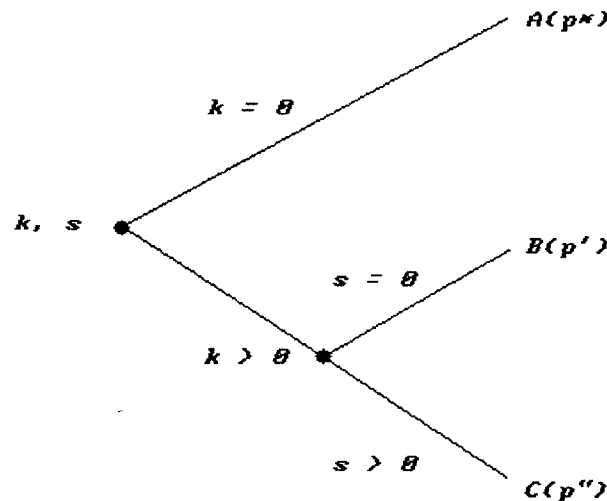


Fig. 5.1 - Mappa contrattuale

Le transazioni che fanno uso della tecnologia generica sono contraddistinte da $k=0$, mentre per quelle che ricorrono alla tecnologia specifica il valore di k sarà positivo. Per quanto riguarda le prime, il meccanismo di mercato è sufficiente, ed $s=0$; la natura delle seconde, per contro, è tale da incentivare le parti a prendere provvedimenti tali da garantirne la realizzazione secondo i termini contrattuali: quindi, potrà risultarne $s>0$.

L'interpretazione della Figura 5.1 è immediata: ad ogni schema contrattuale corrisponde un prezzo pagato dal compratore al produttore; resta da sottolineare la differenza tra le due situazioni che possono verificarsi quando vi sia specificità di capitale, a seconda che s sia nullo o positivo: alle soluzioni B e C sono associati due prezzi diversi, rispettivamente p' e p'' , tali che $p' > p''$. La ragione di questa differenza di prezzo è attribuibile ai meccanismi assicurativi introdotti nel contratto, che possono assumere tre forme principali. Il primo consiste in un

riallineamento degli incentivi tramite una clausola che preveda il pagamento di una penale in seguito ad interruzione prematura del rapporto. Il secondo consiste nella creazione e nell'uso di un arbitrato specializzato, cioè di una struttura di controllo ed intervento che eviti il ricorso al tribunale. Il terzo è rappresentato dall'introduzione di metodi e prassi che diano riscontro dell'intenzione delle parti a continuare il proprio rapporto.

Le transazioni localizzate nel punto B sono intrinsecamente instabili, e verosimilmente tenderanno a riaggiustarsi secondo le configurazioni previste dal punto A o dal punto C:

Inasmuch as price and governance are linked, parties to a contract should not expect to have their cake (low price) and eat it too (no safeguard). (Williamson, 1985, p.35)

Si prospetta quindi un *tradeoff* tra i vantaggi, in termini di efficienza e perciò di prezzo, e gli svantaggi, consistenti nel rischio strategico associato alle transazioni concernenti beni capitali specifici. Ciò che dobbiamo analizzare è il legame tra il livello di specificità del capitale, l'andamento del *tradeoff* individuato e la configurazione istituzionale adottata.

In caso di investimento in beni capitali specifici, l'identificazione reciproca delle parti ha un peso non trascurabile e la continuità e la natura del rapporto vanno, di conseguenza, tutelate. Non solo: se non è possibile specificare a priori la spartizione del surplus, cioè se non possono essere stipulati contratti completi, la redistribuzione dipenderà dalle rispettive posizioni contrattuali a posteriori. Queste, a propria volta, dipenderanno dalla configurazione istituzionale assunta dal rapporto, la quale dipenderà ampiamente dal livello di specificità dei beni capitali coinvolti. In presenza di capitale specifico, le parti non si affideranno al mercato perchè temono di non potersi appropriare del rendimento derivante dai propri investimenti, se in futuro dovessero trovarsi in un ambiente non più concorrenziale: il trasferimento delle transazioni all'interno di un'unica impresa è una garanzia nei confronti dei rischi e delle conseguenze di comportamenti opportunistici.

In altri termini,

transaction cost economics holds that a condition of large numbers bidding at the outset does not necessarily imply that a large numbers bidding condition will prevail thereafter. Whether *ex post* competition is fully efficacious or not depends on whether the good or service in question is supported by durable investments in transaction-specific human or physical assets. (Williamson, 1985, p.61, enfasi nell'originale)

Questa è stata battezzata da Williamson la *trasformazione fondamentale*: in corrispondenza di investimenti specifici, chi ottiene un contratto 'al primo *round*', se così si può dire, ne ricava un vantaggio (*il vantaggio della prima mossa*) che si ripercuote nel futuro, perchè le condizioni di parità concorrenziale osservabili all'istante iniziale risultano sconvolte, trasformandosi in una situazione di monopolio bilaterale.

Quello redistributivo non è l'unico motivo valido per stipulare un contratto a lungo

termine. Se le transazioni sono di tipo ripetitivo, un tale contratto riduce i costi di transazione connessi con una rinegoziazione pressochè continua; inoltre, una serie di contratti a breve potrebbe facilmente risentire dell'insorgere di fenomeni di opportunismo, se si avessero asimmetrie informative nel corso del rapporto.¹⁸

Fino a questo momento, ci siamo occupati dei vantaggi che un contratto a lungo termine può procurare: ma cosa si può dire in quanto ai costi? E' essenziale afferrarne la natura e le dimensioni, anche se rientrano in una categoria in sè piuttosto vaga come quella dei costi di transazione, per valutare se una grande impresa può o non può fare tutto quanto viene fatto da una serie di imprese più piccole.

La tipologia dei costi è sostanzialmente la stessa riscontrabile per i contratti a breve, e cioè (a) il costo della specificazione ex ante di ogni contingenza rilevante ai fini dell'esecuzione del contratto; (b) il costo associato alla specificazione del comportamento che ciascuna parte dovrà tenere in corrispondenza di tali contingenze; (c) il costo generato dalla necessità di eliminare il più possibile dalla stesura del contratto ogni fonte di ambiguità, in modo che esso possa essere fatto applicare anche da un soggetto esterno; e (d) il costo legale del ricorso agli organi giudiziari.

Ciò che cambia, rispetto ai contratti a breve, è evidentemente la magnitudo di queste categorie di costo, che in questo caso sarà notevolmente maggiore, soprattutto per quanto riguarda i punti (a) e (c).

A causa di questi costi, i contratti, quelli a lungo termine in particolare, saranno notevolmente incompleti:

The parties will quite rationally leave out many contingencies, taking the point of view that it is better to wait and see what happens than to try to cover a large number of individually unlikely eventualities. Less rationally, the parties will leave out other contingencies that they simply do not anticipate. (Hart e Holmström, 1985, p.132)

La soluzione adottata sarà un contratto a medio termine rinegoziabile una volta giunto a scadenza, a surrogare in modo più o meno imperfetto un contratto a lungo termine vero e proprio.

L'incompletezza che affligge i contratti a lungo termine deriva dall'impossibilità di descrivere compiutamente (a) lo stato del mondo in corrispondenza del quale avviene la transazione prevista; (b) le caratteristiche dei beni o dei servizi oggetto della transazione; (c) le azioni intraprese dai soggetti coinvolti.

Queste fonti di incompletezza saranno analizzate nel paragrafo 5.2, in un modello in cui questo problema fa premio anche sull'asimmetria informativa, in quanto entrambe le parti

¹⁸. A questo proposito, si veda Fudenberg e Tirole (1983). Per ulteriori estensioni, oltre ai citati lavori di Williamson, si rimanda ai contributi di Klein (1980); Klein e Leffler (1981); Klein, Crawford e Alchian (1978).

sono in grado di osservare lo stato del mondo e gli altri dati rilevanti, ma non sono in condizione di comunicare tali conoscenze a soggetti esterni, quali gli organi giudiziari. Esiste effettivamente un'asimmetria, ma affligge i rapporti tra *insiders* ed *outsiders*:

To use the jargon, incompleteness arises because states of the world, quality, and actions, are *observable* (to the contractual parties) but not *verifiable* (to outsiders). (Hart e Holmström, 1985, p.134)

All'interno di questo quadro, riacquista un ruolo di primaria importanza l'assegnazione dei diritti residuali al controllo: infatti,

only those with residual decision rights can establish a reputation. The other parties will simply follow prescribed conditions in the contract, which signal nothing about future intentions. (Holmström e Tirole, 1989, p.77)

Posto che sia eccessivamente costoso definire anticipatamente l'impiego o le caratteristiche di un determinato bene, si potrà scegliere di affidarne il controllo ad uno specifico soggetto, intendendo per controllo il diritto residuale alla decisione.

5.1. Modelli di integrazione: Williamson

Il punto di partenza scelto da Williamson (1985, pp.90-95, e 1989, pp.150-9) è l'opzione tra impresa e mercato, sulla base del grado di specificità del capitale. Siano $\beta(k)$ ed $M(k)$, rispettivamente, il costo associato al controllo gerarchico e quello associato al meccanismo di mercato. Si può assumere che $\beta(0) > M(0)$, dal momento che l'organizzazione interna sarà generalmente incapace di preservare gli incentivi *ad alto potenziale* tipici del mercato. L'operazione di integrazione avrebbe cioè degli effetti collaterali negativi, tali da impedire alla gerarchia di incentivare i propri membri in modo che si comportino come se fossero indifferenti al rischio.

Il risultato è quindi che la struttura istituzionale-organizzativa prescelta influenza la *performance* finale, perchè influenza in modo rilevante l'insieme delle informazioni disponibili, le quali a propria volta entrano sia nei meccanismi decisionali che nella costruzione degli schemi di incentivazione.

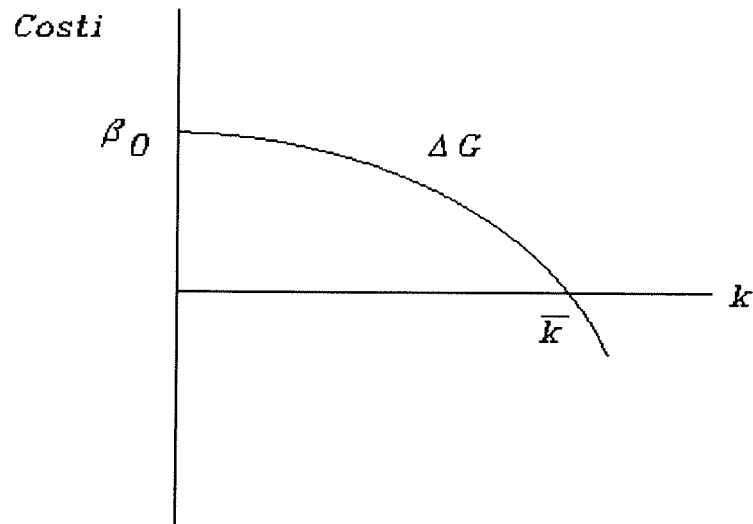


Fig. 5.2 - Costi di controllo comparati

In Figura 5.2 è rappresentato l'andamento di:

$$\Delta G = \beta(k) - M(k); \quad M' > \beta', \quad \forall k;$$

quindi ΔG misura lo scarto tra i costi di controllo di impresa e mercato (G sta per *governance*).¹⁹ Il mercato è più efficiente fintantoche la specificità di capitale è bassa, mentre da un certo valore di k in poi la soluzione gerarchica è preferibile, perchè è più adatta per fronteggiare una situazione di forte interdipendenza.

L'intersezione della curva con l'asse orizzontale individua il valore di transito per k , \bar{k} , in corrispondenza del quale si può optare indifferentemente per impresa o mercato.

Nella scelta entrano naturalmente anche altre considerazioni, relative all'andamento delle curve di costo, cioè alle economie di scala o di produzione multipla. I costi comparati di produzione (calcolati mantenendo costante il livello di produzione) sono individuati in Figura 5.3 da ΔC , funzione decrescente di k :

$$\Delta C = \alpha(k) - m(k).$$

ΔC è quindi la differenza di *steady state* (cioè con produzione costante) tra produrre internamente il proprio fabbisogno ed acquistarlo sul mercato.

Quando il mercato è in grado di realizzare tutte le economie possibili (verosimilmente, per transazioni estremamente standardizzate e per valori di k molto bassi), gli svantaggi in termini di costo di produzione legati alla gerarchizzazione sono alti. Aumentando progressivamente la specificità, diviene sempre più forte la spinta all'integrazione, perchè le

19. Williamson non esplicita il segno delle derivate, ma l'intuizione suggerisce che, per ottenere un andamento decrescente come quello rappresentato, è necessario che, al crescere di k , β diminuisca più rapidamente di quanto aumenti M ; da cui, arguendo, $M' > \beta'$.

economie 'aggregative' del mercato raggiungono il proprio limite asintotico.

La retta $[\Delta C + \Delta G]$ esprime l'andamento dei costi complessivi, ed è la funzione obiettivo da minimizzare. L'introduzione di economie dimensionali favorisce il ricorso al mercato per un più ampio spettro di specificità: infatti, il valore di transito, \hat{k} , è maggiore di quello registrato per i soli costi di controllo. Il valore ottimale k^* dipenderà dalla forma organizzativa prescelta.

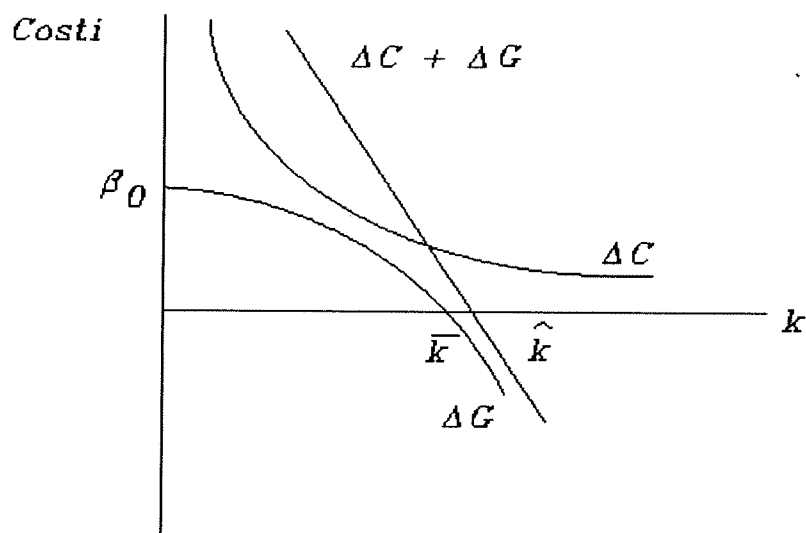


Fig. 5.3 - Costi comparati di controllo e di produzione

L'analisi della Figura 5.3 consente di derivare alcune conclusioni di rilievo: (a) quando $k^* \ll \hat{k}$, il mercato possiede un vantaggio significativo nei confronti dell'integrazione, per entrambe le categorie di costo; (b) il contrario vale quando $k^* \gg \hat{k}$, caso in cui è molto rischioso affidarsi al mercato; infine (c) quando k^* è prossimo a \hat{k} , si avrà una casistica mista.

L'implicazione nuova è che, data questa configurazione, non si procederà all'integrazione sulla sola base dei costi di produzione, perchè, avendosi $\Delta C > 0$ ovunque, sotto questo profilo il mercato è sempre in vantaggio. Questo però nell'ipotesi che la dimensione dell'impresa rimanga fissa; se la si abbandona, la conclusione è completamente diversa, ed è che la grande impresa mostrerà un più alto grado di integrazione, proprio perchè realizzerà un crescente risparmio nei costi di produzione al crescere della quantità di *input* necessaria alle sue linee di produzione, realizzando delle economie di scala all'aumentare della sua dimensione relativamente a quella del mercato sul quale opera.

In questo contesto è possibile chiarire ulteriormente il *tradeoff* interno all'organizzazione gerarchica tra la capacità di intervenire selettivamente per risolvere i problemi di distorsione degli incentivi e la sua relativa incapacità di ricreare gli incentivi ad alto potenziale posseduti dal mercato. La chiave di volta è rappresentata dal passaggio dalla forma-U alla forma-M (cfr. sez.4.1): a parità di altri fattori, la forma-M è in grado di disporre di una quantità di informazioni maggiore e di un dispositivo di controllo più efficace rispetto

alla forma-U, e quindi risulterà maggiormente integrata.²⁰

Ciononostante, l'integrazione non procederà all'infinito perchè l'intervento selettivo, a partire da un certo punto, non è più fattibile.

But there is more to it. If the use of high-powered incentives in firms is inherently subject to corruption, then the notion that the integrated firm can do everything that the nonintegrated parts could accomplish is a fiction. (Williamson, 1985, p.140)

La letteratura recente ha prodotto una ristretta gamma di modelli²¹ che formalizzano sia il problema dell'incompletezza dei contratti, sia quello dell'incentivo ad effettuare investimenti specifici ex ante, entrambi, come abbiamo visto, già affrontati da Williamson su di un terreno più intuitivo.

L'idea alla base di questa modellistica è che l'integrazione verticale sia una soluzione che consente di ovviare all'impossibilità di specificare all'interno del contratto tutte le circostanze rilevanti, in quanto nell'ambito dell'impresa integrata è possibile adottare meccanismi decisionali di tipo adattivo e sequenziale. Questo senza consentire alla parte che ha proceduto all'integrazione di accedere a nessuna particolare informazione in possesso della parte assorbita: l'integrazione si tradurrà allora nella necessità/possibilità di definire uno schema di incentivazione tale da indurre quello che ora è un agente interno all'impresa a rivelare le proprie informazioni private o, quanto meno, ad adottare un comportamento allineato agli obiettivi del principale.²²

La sezione che segue affronta questa letteratura prendendo in esame il modello proposto da Grossman e Hart (1986).

5.2. Modelli di integrazione: Grossman e Hart

Questi autori indagano sull'incompletezza dei contratti dovuta sia alla difficoltà di descrivere compiutamente gli stati del mondo, sia alla difficoltà di specificare adeguatamente (a) le caratteristiche di ciò che viene scambiato e (b) il comportamento che le parti devono tenere in corrispondenza di ogni evento rilevante ai fini dell'esecuzione del contratto.

Si assume una struttura della proprietà che riprende quella tipica della teoria dei diritti di proprietà,²³ nell'ipotesi che le differenze tra le diverse strutture organizzative siano riconducibili in ultima analisi all'allocazione dei diritti residuali al controllo. Tali diritti

20. Questo ribadisce la specifica tesi di Williamson, secondo la quale l'insieme delle informazioni disponibili non è definibile esogenamente ma, al contrario, si determina endogenamente in rapporto alla struttura organizzativa.

21. Grossman e Hart (1986), Tirole (1986) e Hart e Moore (1988, 1990).

22. Sul tema delle relazioni d'agenzia, si rimanda all'esauriente rassegna di Hart e Holmström (1987) e, per approfondimenti ulteriori, ai contributi fondamentali di Grossman e Hart (1983) e di Holmström (1979, 1982).

23. Per un inquadramento della *property rights literature* si rimanda il lettore ad Alchian e Demsetz (1972, 1973).

residuali sono proprio quelli che non sono stati preventivamente inclusi nel contratto in termini espliciti, e che determineranno le soluzioni adottate in tutta quella gamma di circostanze in relazione alle quali il contratto si riveli fortemente incompleto:

contractual rights may be of two types: specific rights and residual rights. When it is too costly for one party to specify a long list of the particular rights it desires over another party's assets, it may be optimal for that party to purchase all the rights except those specifically mentioned in the contract. Ownership is the purchase of these residual rights of control. (Grossman e Hart, 1986, p.692)

Non si fa alcuna distinzione tra proprietà del capitale e funzione di controllo; si definisce anzi la proprietà come la facoltà di esercitare il controllo sulle attività dell'impresa.

Quindi, a differenza di Williamson (1975, 1985) e di Arrow (1975), gli autori ritengono che l'integrazione non possa in se stessa rendere rilevabile alcuna nuova variabile. Di conseguenza, l'integrazione non facilita la stesura di una qualsiasi clausola contrattuale, non diminuendone il costo: il suo vantaggio risiede essenzialmente nel semplificare il controllo di tutto ciò che non è possibile includere nel contratto iniziale.

A scopo illustrativo, ricorriamo ad una semplice formalizzazione dovuta ad Hart e Holmström (1987, pp.134-41). Si consideri il caso di un fornitore ($S = seller$) e di un acquirente ($B = buyer$) che stipulino un contratto per la fornitura di una unità di un determinato *input*.

Il modello è biperiodale. Nel primo periodo (fase ex ante) entrambi compiono investimenti in capitali specifici, che chiameremo I_B ed I_S e che potranno assumere solo due valori: H (alto) o L (basso). Questi investimenti sono osservabili ma non contrattabili, o perchè tecnicamente troppo complessi per essere descritti in modo accettabile, o perchè consistono in un *input* manageriale in sè non verificabile. Nel secondo periodo (fase ex post) avviene lo scambio, ed eventualmente viene modificato il contratto. Analogamente, si assume che le caratteristiche (o la *qualità*) di ciò che viene scambiato siano osservabili dalle parti, e tuttavia troppo complesse da specificare nel contratto iniziale: delle conseguenze che questa non contrattabilità comporta, e che già ora sono intuibili, ci occuperemo più avanti. Per il momento è sufficiente sottolineare che questo basta a giustificare l'assegnazione dei diritti residuali al controllo, per stabilire quale delle due parti abbia la facoltà di ridefinire i termini contrattuali allo scoccare del secondo periodo. Al momento della transazione (data 1), S può fornire un *input* 'soddisfacente' o 'insoddisfacente', quest'ultimo definito come quell'*input* che ha costi e benefici nulli: il che è come dire che non viene nemmeno prodotto. Nella Tabella 5.1 figurano i *payoffs* per B ed S , in termini rispettivamente di benefici e costi, per un *input* giudicato soddisfacente, e secondo le combinazioni d'investimento:

	$I_B = H$	$I_B = L$
$I_S = H$	$(9, 5)$	$(8, 6)$
$I_S = L$	$(8, 6)$	$(5, 9)$

Tab. 5.1 - Payoffs

In ogni coppia (x, y) , il primo valore si riferisce al beneficio lordo v per B, il secondo al costo lordo c sopportato da S. Ad entrambi vanno sottratti i costi di investimento, che si assumono pari ad 1.8 se $I_{B,S} = H$, e a zero se $I_{B,S} = L$. E' immediato che il livello di investimento scelto da S influenza sia i benefici di B che i suoi propri costi, e lo stesso vale, specularmente, per l'investimento effettuato da B. Un investimento più alto può migliorare la qualità (se effettuato da S) oppure può diminuire i costi di produzione (se effettuato da B) di un *input* soddisfacente. Per chiarire, si può pensare a B come ad una centrale per la produzione di elettricità, e a S come ad una miniera di carbone.²⁴ $I_B = H$ potrà rappresentare un impianto migliore (caldaie e turbine più efficienti), capace cioè di operare anche con un carbone di scarsa qualità, ed $I_S = H$, analogamente, potrà essere un impianto di estrazione capace di fornire un minerale di carbone con una percentuale di impurità inferiore, in grado di essere usato con un buon rendimento anche da una centrale termoelettrica che impieghi una tecnologia non efficiente. L'idea che ne scaturisce è che, se B possedesse S, potrebbe esercitare i propri diritti residuali in modo da migliorare la qualità del carbone estratto, e viceversa che, se S possedesse B, potrebbe esercitare i propri diritti residuali per far sì che B usi impianti più efficienti, in grado di accettare carbone di qualità scadente. Queste osservazioni si riveleranno di grande utilità tra breve.

Dal momento che in questo modello l'incertezza è assente, il contratto ottimale è quello che massimizza il surplus totale. Nella configurazione di *first best*, $I_B = I_S = H$, con un beneficio netto complessivo pari a $(9 - 5 - 3.6) = 0.4$: ma questa configurazione si potrebbe ottenere solo se il livello d'investimento o la qualità dell'*input* fossero specificabili nel contratto. Infatti, poniamo che lo sia I : allora, il contratto ottimale fisserà $I_{B,S} = H$, attribuendo a B la facoltà di accettare in data 1 l'*input* al prezzo p' , o di rifiutarlo al prezzo p_0 , siccome la qualità è osservabile per entrambi. Se $5 \leq dp \leq 9$, S sarà disposto a fornire un *input* soddisfacente, che B accetterà. Per contro, se la qualità è contrattabile ex ante (ed è quindi specificato nel contratto che dovrà essere alta, ovvio) ciascuna parte è incentivata a porre I al livello che massimizza il proprio beneficio e quindi quello collettivo, cioè H , che è il valore

24. Non è una scelta casuale. Joskow (1985, 1987), ha analizzato la struttura dei contratti in questo settore, ricavandone risultati in consonanza con le previsioni della teoria: nello specifico settore, fornitori ed acquirenti sembrano preferire *longer commitments* ad una rinegoziazione frequente, in presenza di capitale specifico.

socialmente ottimale.

Se entrambe le variabili rilevanti (qualità e livello d'investimento) non sono includibili dettagliatamente nel contratto, la condizione di *first best* risulta irrealizzabile. Ma qual'è la soluzione contrattuale migliore? Ovvero: di quanto possiamo avvicinarci al *first best*?

Analizzeremo per primo il caso in cui non sia possibile rinegoziare il contratto. Un contratto di *second best*, posta la non verificabilità di investimento e qualità, dovrà basarsi su qualsiasi altra variabile che lo sia: senza dubbio potrà specificare sia p' che p_0 , dal momento che B ha la facoltà di rifiutare l'*input*. Perchè questo sia soddisfacente, occorre che:

$$dp \geq 6.8 \quad (= \pi_S); \quad (1)$$

per quanto riguarda B, il suo *payoff* netto dovrà soddisfare la seguente condizione:

$$dp \leq 7.2 \quad (= \pi_B); \quad (2)$$

in sintesi:

$$\pi_S \leq dp \leq \pi_B; \quad (3)$$

Ma, se S si aspetta che B scelga H , preferirà porre $I_S = L$, supponendo che, essendo la (3) soddisfatta, la transazione si concluderà ugualmente. Infatti, in virtù di questa deviazione, il suo *payoff* netto risulterà superiore, pari a $(p' - 6)$: quindi, ci sarà un forte incentivo a rompere l'equilibrio $I_{B,S} = H$ (il ragionamento simmetrico vale per B). L'inefficienza, quindi, sorge anche quando i soggetti hanno accesso alla medesime informazioni: in particolare, deriva dall'interesse di S (o di B) a contrarre I , in quanto questa mossa aumenta i suoi profitti, anche se va a scapito del benessere collettivo. Tuttavia, è possibile dimostrare che, posto che la rinegoziazione sia impossibile, il *first best* non solo è raggiungibile, ma può essere mantenuto: è sufficiente modificare appropriatamente i valori che compaiono sulla diagonale secondaria della Tabella 1, per verificare che esistono valori di dp tali che, sotto l'ipotesi che la rinegoziazione sia esclusa, il *first best* risulta sostenibile. Il punto è che, forse sorprendentemente, ricorrendo alla stessa argomentazione, si mostra che il *first best* può non essere sostenibile sotto l'ipotesi opposta, cosicché le parti possono ottenere un risultato migliore se si impegnano preventivamente a non rivedere il contratto (si suppone che i guadagni dalla rinegoziazione vengano spartiti equamente, *à la* Nash).

Questa gamma di situazioni mostra che l'integrazione può rappresentare la soluzione più desiderabile, cioè **può** incrementare l'efficienza complessiva. Ad esempio, se B si integra a monte, assorbendo S, acquisisce un controllo che gli consente di diminuire le possibilità di un comportamento opportunistico da parte di S. Nel caso che S ponga $I_S = L$, B può intraprendere un'azione 'di rappresaglia', tale da incrementare i costi di S, che l'input prodotto sia soddisfacente o meno, e B sarà maggiormente incentivato a farlo se questo aumenta il suo surplus. Sotto queste condizioni, si mostra che il *first best* è realizzabile e, per un appropriato valore di p' , la combinazione $I_B = I_S = H$ è un equilibrio di Nash: infatti, se S devia si espone

alla rappresentazione automatica di B, mentre se è quest'ultimo a deviare, l'*input* prodotto da S sarà insoddisfacente. È fondamentale sottolineare che l'integrazione presenta dei vantaggi sostanziali rispetto alla transazione di mercato se e solo se la rappresentazione non è includibile nel contratto iniziale: essa, quindi, costituisce un esempio di quei diritti residuali al controllo che, non essendo specificati nel contratto, devono essere assegnati tramite l'integrazione.

La questione può porsi in termini maggiormente intuitivi, ricordando le considerazioni riguardo le opportunità che si offrono alla parte che procede all'acquisizione, e facendo riferimento ora ad una situazione continua, non più dicotomica. Assumendo una posizione critica rispetto alle tesi di Williamson, Grossman e Hart intendono porre in evidenza il fatto che, se è vero che l'impossibilità di scrivere un contratto completo tra B ed S apre la strada a comportamenti opportunistici, è anche vero che la teoria dei costi di transazione, pur spiegando i motivi che conducono all'integrazione, non spiega a sufficienza cosa accada di quei comportamenti una volta compiuta l'integrazione. Si faccia l'ipotesi che le caratteristiche dell'*input* scambiato non siano specificabili ex ante, ma lo siano ex post, una volta che B ed S conoscano lo stato del mondo e gli aspetti rilevanti della produzione. Con le parti in possesso dello stesso patrimonio informativo, la ricontrattazione (non costosa) sfocerà sempre in una allocazione efficiente ex post dei diritti residuali, qualsiasi fosse quella di partenza. Ne risentirà tuttavia la ripartizione del surplus totale: e per questo tramite, l'allocazione dei diritti residuali produrrà una sorta di *feedback* sugli investimenti ex ante. Vale a dire che una qualsiasi struttura della proprietà si dimostrerà ex post efficiente, ma anche che ogni diversa struttura mostrerà effetti distorsivi caratteristici. Allora il nuovo interrogativo è: quale è la dimensione di queste distorsioni? Ovvero: quale è la struttura capace di ridurre al minimo questi effetti?

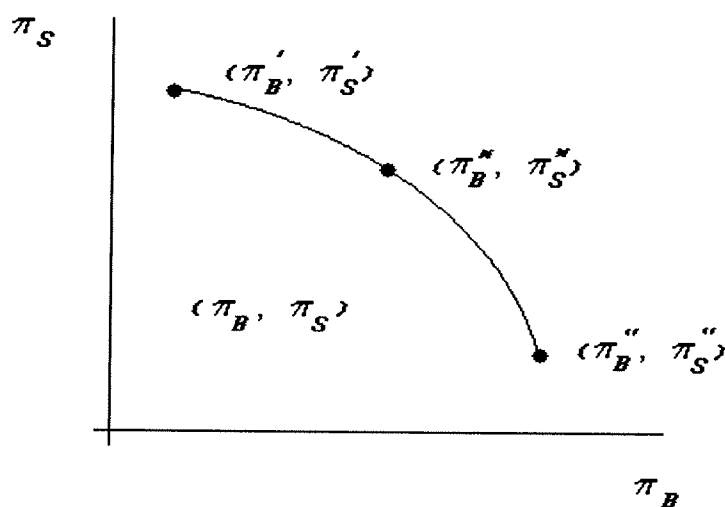


Fig. 5.4 - La frontiera del benessere

La curva in Figura 5.4 rappresenta la *frontiera del benessere*, cioè il confine entro il quale si troverà la combinazione dei benefici per B ed S (π_B, π_S). Se uno dei due controlla l'altro, l'equilibrio si collocherà in qualche punto lungo la curva, perché una parte ha il controllo sulle decisioni di entrambe.

In particolare, le due soluzioni estreme (π' e π'') indicano che l'assorbimento (di B da parte di S o viceversa) può fornire un risultato efficiente ma anche molto lontano da quello ottimale, caratterizzato cioè dai livelli di investimento ottimali per entrambi. La 'lontananza' del risultato effettivo dell'integrazione (che potrà essere in una posizione di confine) dal *first best* dipenderà dalla distanza che separa i rispettivi investimenti dai loro valori ottimali:

$$I_{B,S} > I_{B,S}^* \Leftrightarrow \pi_{B,S} > \pi_{B,S}^*$$

e viceversa.

Per chiarire: (a) il risultato 'non integrato', essendo le decisioni non coordinate, non solo non sarà ottimale, ma sarà ampiamente inefficiente, e cioè si troverà in qualche punto della zona delimitata dalla curva e dagli assi, a sud-ovest del punto di ottimo; (b) il risultato 'integrato' dovrà collocarsi sulla frontiera efficiente, ma in generale si avrà $\pi_{B,S} \neq \pi_{B,S}^*$, con il segno della disequaglianza determinato dalla loro posizione relativa.

L'argomentazione completa è più complessa, tuttavia queste note appaiono sufficienti. Si rammenti l'ipotesi secondo la quale l'obiettivo comune a B ed S è la massimizzazione del surplus totale. Ora, se una delle due imprese ha il controllo di entrambe, tenderà a sovrainvestire, relativamente al *first best*, mentre l'altra tenderà a sottoinvestire, perchè la prima potrà usare i propri diritti residuali per appropriarsi della fetta più grande del surplus ex post.

Il problema è che, in condizioni di non integrazione, entrambe tenderanno a sottoinvestire, perchè entrambe conservano il proprio potere decisionale, mentre per lo stesso motivo la divisione del surplus sarà più equa. Il controllo da parte dell'una o dell'altra sarà perciò desiderabile a seconda dell'importanza relativa attribuita ai rispettivi investimenti: in tal modo il sottoinvestimento da parte dell'impresa acquisita non risulterà troppo dannoso, ed il sovrainvestimento da parte dell'impresa che l'ha assorbita sarà a propria volta meno grave del sottoinvestimento altrimenti compiuto nello status quo ante.

L'analisi che precede mostra innanzitutto che, se gli investimenti iniziali sfuggono alla contrattazione, e se, in generale, questa è incompleta, l'integrazione può creare più problemi di quanti ne risolva, o comunque crea problemi nuovi: dal momento che l'integrazione è il risultato di un contratto incompleto ma che specifica l'assegnazione di specifici diritti decisionali, le scelte effettuate dalla parte che ha proceduto all'integrazione hanno la connotazione di un *esito di minaccia*. Di conseguenza, essa non sarà generalmente in grado di appropriarsi di tutti i benefici derivanti dalle proprie decisioni.

Un secondo punto che merita di essere sottolineato è che, a differenza della letteratura sui costi di transazione, qui

the relevant comparison is not between the nonintegrated outcome and the complete contract outcome but instead between a contract that allocates residual rights to one party and a contract that allocates them to another. (Grossman e Hart, 1986, p.716)

Non solo: dal momento che non permette di accedere alle informazioni private in possesso della parte assorbita, l'integrazione non rimuove gli incentivi all'opportunismo: piuttosto li modifica e li trasferisce in qualche punto all'interno della nuova organizzazione, andando a toccare la distribuzione del surplus ex post, e tramite questa la struttura degli investimenti ex ante.

Il problema informativo non si riduce a questo. La presenza di asimmetria informativa si tradurrà verosimilmente nel fatto che le due parti non avranno la medesima *percezione* della soluzione integrata e delle possibilità di profitto ad essa legate. Questo significa che sarà la parte globalmente in possesso delle maggiori informazioni a procedere all'integrazione, ovvero che i diritti residuali al controllo saranno appannaggio della parte che dispone di un particolare capitale specifico, identificabile come quel patrimonio di informazione imprenditoriale che consente di coordinare attività in origine disgiunte.

Conclusioni

Il concetto di integrazione verticale compare, all'interno della teoria economica, in due contesti diversi ed apparentemente autonomi.

In un caso, all'interno dell'analisi economica, si configura come uno strumento che consente di ottenere una rappresentazione sintetica del sistema economico, al fine di studiarne il comportamento aggregato, riducendone l'apparato produttivo ad un apparato di trasformazione di un fondo di risorse in un insieme di beni finali che rappresenta la ricchezza netta del sistema.

Nell'altro, in quanto comportamento strategico attuato da singoli agenti, costituisce uno dei poli attorno ai quali gravita l'attenzione degli economisti, nell'ambito della teoria dell'organizzazione industriale.

La dicotomia analitica e l'asincronicità che separano le due correnti di pensiero fanno tuttavia sorgere il seguente interrogativo: esiste un punto di contatto tra esse, sia storico che analitico? Ovvero, in quale fase dello sviluppo dell'Economia Politica entra in scena il tema dell'integrazione verticale nella sua seconda e più moderna accezione?

Questo equivale a chiedersi in quale momento la teoria economica abbia preso in esame, oltrechè la *performance* globale del sistema, anche il versante **individuale** della produzione di ricchezza e, quindi, il rapporto tra comportamenti individuali e struttura del sistema economico.

La risposta che si pare di poter fornire è che il primo punto di contatto emerge in Adam Smith (1776), e che tuttavia l'intuizione di Smith è solamente parziale, e le sue implicazioni verranno sviluppate molto più tardi, grazie a contributi che metteranno in luce aspetti impliciti (se non del tutto trascurati) di quello che è noto come *teorema smithiano*.

L'analisi del rapporto tra dotazione di lavoro e ricchezza fa sì che Smith adotti una rappresentazione dell'economia per settori verticalmente integrati; d'altro canto, la sua attenzione per la producibilità lo porta ad esaminare il problema dell'organizzazione efficiente dell'attività produttiva, attraverso la divisione del lavoro. In questo modo emerge il legame tra organizzazione della produzione e struttura e dimensione del mercato, che viene ripreso e sviluppato da Stigler (1951), il quale si chiede perchè non si dovrebbe sfruttare l'opportunità presentata da un'attività caratterizzata da costi decrescenti (almeno in un certo intervallo), concentrando (integrando) in un'unica unità tutta la produzione necessaria per il bene in questione, sia esso intermedio o finale.

L'intuizione di Smith induce quindi a pensare all'integrazione come ad un fenomeno essenzialmente *orizzontale*, manifestazione di una spinta monopolistica dettata essenzialmente da motivazioni di natura tecnologica.

In realtà, questa interpretazione appare fortemente incompleta, in quanto priva del supporto di un'adeguata teoria dell'impresa, che non sia semplicemente un centro di produzione, in altre parole quella 'scatola nera' che l'economia industriale ha ereditato dalla teoria marginalista neoclassica (il cui unico problema, in un mondo caratterizzato da certezza perfetta, è una scelta tecnologica efficiente), ma che si comporti invece come un agente economico reale. Questo significa che, implicita nel concetto di integrazione vi è l'eliminazione almeno parziale degli scambi effettuati su base contrattuale o di mercato: la mano invisibile di Smith viene soppiantata da quella visibile dell'impresa, che internalizza le

transazioni (Chandler, 1977).

Le prime risposte fornite dalla letteratura a questo proposito sono quindi di sapore monopolistico: l'integrazione viene interpretata come uno strumento per esercitare una discriminazione di prezzo o per ostacolare l'entrata. Affine a queste argomentazioni è l'interpretazione dell'integrazione come comportamento strategico, volto ad assicurarsi la fornitura di fattori (a monte) e l'assorbimento del prodotto (a valle).

Le potenzialità insite in questa tesi vengono sviluppate da Williamson (1975, 1985), che riprende l'intuizione iniziale di Coase (1937), in un contesto di comparazione istituzionale, in condizioni di incertezza ed informazione asimmetrica. Il ricorso all'integrazione (e quindi il sorgere dell'istituzione-impresa) si giustifica sulla base delle imperfezioni di mercati *non-clearing*: l'impresa si propone qui come risposta al fallimento del mercato.

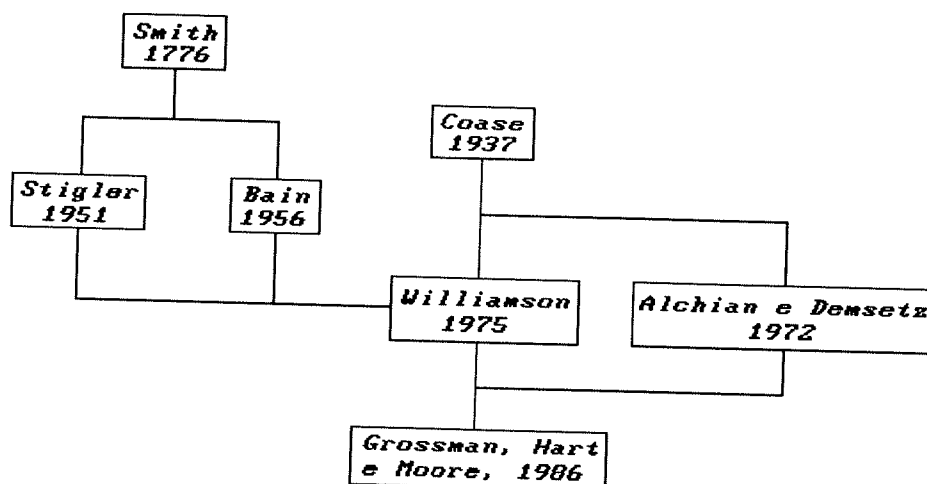
Alla luce di questo, l'integrazione non è da interpretarsi come un problema di relazioni produttive, bensì di scambio: l'esistenza di costi di transazione, pregiudicando il funzionamento efficiente dei mercati, conduce a surrogare questi ultimi con strutture gerarchiche basate su di un insieme di contratti che assegnino i diritti al controllo su decisioni di importanza strategica, relative ad una gamma di situazioni non compiutamente specificabili *ex ante*. La teoria dei contratti incompleti, sotto questo profilo, rappresenta il corrispondente formale delle argomentazioni sviluppate in ambito neoistituzionalista.

In un ambiente caratterizzato da forte incertezza, l'integrazione si prospetta quindi come una sorta di 'assicurazione' nei confronti dei rischi associati al ricorso al mercato e, parallelamente, come una soluzione che consente di accedere ad un patrimonio di informazioni altrimenti inaccessibile (Arrow, 1975; Holmström e Tirole, 1989; Williamson, 1975 e 1985).

Su quest'ultimo punto la letteratura recente, in particolare quella che si occupa dei diritti di proprietà, non è unanime. In effetti, la maggiore accessibilità delle informazioni non è un risultato *necessario* dell'integrazione (Grossman e Hart, 1986), perchè nulla garantisce che gli interessi individuali vengano intaccati in qualsiasi misura dal fatto che i soggetti che ne sono portatori operino all'interno della medesima struttura piuttosto che su di un libero mercato. Il reale vantaggio dell'integrazione sembrerà perciò scaturire dall'idea secondo la quale una proprietà unificata implichi un minor grado di incertezza; in altre parole, dal fatto che l'integrazione sia riconducibile ad un contratto che assegni ad una delle parti coinvolte (plausibilmente, si tratterà della parte globalmente più informata, o comunque di quella in possesso delle informazioni più difficili da ottenere tramite meccanismi di incentivazione interna) i diritti residuali al controllo in circostanze la cui ambiguità metterebbe altrimenti in discussione il rapporto stesso. Questo è particolarmente vero nel caso di relazioni in cui siano coinvolti capitali specifici che, in presenza di attriti informativi, generano comportamenti opportunistici, cioè un problema di 'appropriazione indebita' del surplus generato grazie al loro impiego.

Tutto questo induce facilmente il sospetto che, se non esistono concreti ostacoli tecnologici, aggirabili replicando i singoli impianti dimensionati in corrispondenza della scala efficiente, i vantaggi legati all'integrazione portino alla formazione di imprese di dimensioni paragonabili a quelle del mercato. Se ciò non avviene, la causa risiede in un *tradeoff* tra integrazione e controllo: il trasferimento delle asimmetrie informative all'interno dell'impresa,

che crea una catena di relazioni d'agenzia, e la perdita di controllo che si registra all'estendersi della gerarchia pongono un limite endogeno all'integrazione e quindi anche alla dimensione dell'impresa.



Una rappresentazione topologica della propagazione del concetto di integrazione nell'economia industriale

Se quindi, da un lato, è possibile identificare una cerniera storica tra le caratterizzazioni del concetto di integrazione verticale come strumento dell'analisi economica e come comportamento strategico in Adam Smith e, successivamente, in Young, dall'altro rimane aperta la questione, senz'altro più rilevante, di comprendere se esiste, e di quale natura sia, una cerniera più rigorosamente analitica tra i due approcci.

Si tratta, in altre parole, di chiedersi se sia non solo possibile ma anche lecito, e sotto quale specifico profilo, tentare di conciliare tra loro un approccio di analisi macroeconomico-strutturale dei flussi interindustriali *à la* Pasinetti ed un approccio di teoria dell'organizzazione industriale di natura esplicitamente microeconomico-strategica.

Tale interrogativo si pone, in primo luogo, in quanto quella che potremmo definire come la *terza via* all'integrazione, che è quella tracciata da Smith e seguita da Young, che pure evidenzia, pur senza svilupparle fino in fondo, talune potenzialità analitiche sul versante della dinamica strutturale, di fatto non coincide con alcuna delle altre due visioni; e, in secondo luogo, in quanto la prospettiva propria dell'analisi economica propone il tema dell'integrazione verticale come premessa all'esame delle conseguenze prodotte dal mutamento tecnologico sull'aggregato, mentre la prospettiva propria della teoria dell'organizzazione industriale lo propone nei termini, più circoscritti, dell'interazione tra la struttura di un singolo mercato (o, più correttamente, di più mercati sovrapposti) da un lato, ed il comportamento e l'organizzazione interna delle imprese in esso operanti.

Non è possibile, in questa sede, e sulla base delle nostre attuali conoscenze, fornire una risposta sia pur preliminare a queste domande; tuttavia, se questo studio ha consentito di individuare alcune linee lungo le quali indirizzare la ricerca futura, allora potremo dire che esso ha raggiunto il proprio scopo.