Part I

The Actor

Introduction to Part I

A Definition of the Actor

A theory of the organization of economic action calls for a model of how human beings are believed to behave in economic actions. In fact, the concept of *actor* is widely used in organization. But why do we speak of actors and not more simply of individuals – of persons? There is a reason, and it relates to a basic methodological aspect of a discipline – the elementary unit of analysis.

Economics is very clear in this regard. The elementary unit of analysis is the individual. Individuals are in fact the legitimate owners of interests and rights. These are also deemed to be the not further divisible units in which preferences and subjective utilities are formed. It is the principle of methodological individualism: every socio-economic formation that is more complex than the individual must be explained and justified starting from individuals who are "free" to order alternatives and "states of the world" according to preference, on the basis of the consequences for their own interests (Schumpeter 1942; Arrow 1951).

From sociology there have arisen methodological conceptions that are opposed to methodological individualism. In a famous essay, Wrong (1961) contrasted the "undersocialized" view of actors in economics with an "oversocialized" conception prevailing in sociological studies: individuals are not free to act, their actions are in large measure determined by the social environment in which they find themselves, by groups (more or less large), by social norms and inherited culture, and by resource constraints. To that extent, the sociology of organization has often assumed aggregated units of analysis (the family, the firm, and the national society).

Psychological accounts of behavior, including economic behavior, usually interested in the most micro and individual level of analysis, have often found it necessary to introduce an infra-individual notion of self (Elster 1985). Contradictory beliefs and conflicting desires need not be considered as irrational as they may stem from the different, equally legitimate interests of the same physical person as projected through time, or across different contexts, or at different hierarchical levels of cognition (for example, the immediately sensible and the more immaterial and constructed, the agent and the controller).

Organizational analysis has been characterized by the assumption of a unit of analysis that is (1) variable according to the scope of the study and (2) often at the "meso" level (greater than the individual but less than the

system analyzed – Tosi 1992; Grandori 1995). A great number of organization theorists is in fact reluctant to adopt monolithic conceptions of systems composed of many people, sensitive to the problem of the potential differentiation of interests, and inclined to concede that people have the capability for choice and strategic behavior. On the other hand, organization theorists have a distinct awareness of the limits of cognition and individual rationality, of influence processes in small groups, and of the homogenization of perceptions and judgments produced from having access to the same information. To that extent, organization theory usually accepts the idea that the interpretation of one's own interest and the formation of preferences on the part of individuals happens often, even if not always, in groups. For example, for many analytical purposes, a group of workers employed in the same firm and in the same position in the productive process and in the organization structure can be considered to be an actor, characterized by homogenous perceptions, interests, and preferences. Groups of decision-makers responsible for different firm functions (production, marketing, and research, etc.) can be seen as single actors participating in an inter-functional project.

What matters for defining a unit of analysis, then, are analytical purposes. Depending on what problem is studied, some distinctions may become negligible – for example, intra-individual or small group differentiations if interfirm alliances are studied: if the formation of a joint venture is studied, firms can be defined and modeled as actors. Therefore in this text we define an actor as *a social entity in which no problem of inter-personal comparison of utility and of information transmission is considered to be relevant in relation to the problem being examined.* For example, the actors will typically be individuals when we are dealing with motivations and incentives or of job design; the actors will typically be interest groups or organization units when we are dealing with the analysis and design of the macrostructure of the firm; and the actors will typically be firms when we are dealing with networks of firms.

Chapter 1 is dedicated to developing a model of actor, focusing on its "endowment" of knowledge and interests. Chapter 2 presents a process model of decision and motivation, integrating the different "models of rationality" in a multiple rationality model.

If we look in more detail at the box of the configuration of actor in Figure I.1, we obtain Figure I.2, summarizing the structural and processual elements relevant for understanding the economic behavior of actors.

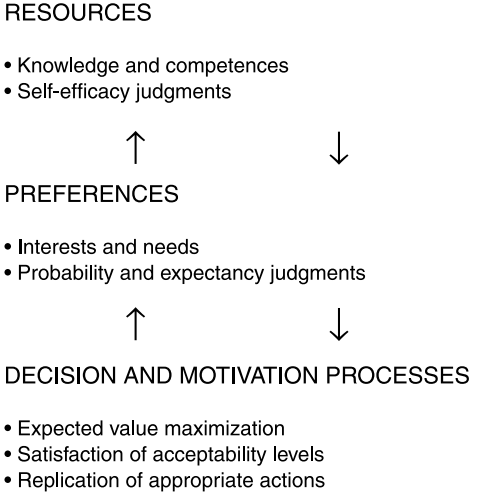


Figure I.2 Elements of the configuration of actors

Chapter 1

Knowledge and Preference

Two main structural features of actors can explain economic behavior – in so far as it depends on actors: what they know and are capable of doing (knowledge and competence) and what they want (their preferences and interests). These two elements substantiate the *identity* of an actor – the response to the questions: Who am I? What can I do? What do I wish to do? – whether the actor is individual or collective. In this perspective, an entity can be treated as an actor if it perceives itself as, or acts as if it were, on a certain matter, a unit of preference and knowledge.[1] Figure 1.1 outlines the conceptual structure of the chapter.

THE ECONOMIC ACTOR AND THE STRUCTURE OF KNOWLEDGE

Data, information, and knowledge

The "material" which is transformed in a decision-making process consists of information. This is the basic element that constitutes all the decision components that will be described. For example, in a decision-making process concerning the purchase of a car, all the following observations are information: the fact that the car we currently own no longer works well, the data relative to the performance of alternative vehicles, prices, the image of the car with respect to the image we have of ourselves, the level of aspiration concerning the type of car desired, and the experience and the evaluations formulated on the performance of that car once acquired.

First, there is a distinction to be made between the notions of "information" and that of "data." Data (news, facts, numbers, reports, etc.) must be perceived by a subject, interpreted, and stored in relation to others in order to become information. Second, different bits of information should be put in relation to one another in order to guide choice and learning: this is the network of relationships between cognitions that constitutes the "knowledge" of the subject. As Bateson (1972) evocatively illustrates: "Those of you who believe to see me, raise their hand. I see a lot of hands raised . . . I deduce that folly likes to stay in company. Naturally you do not 'really' see me: what you 'see' is a bunch of information about me, that you synthesize into a visual image of me. You construct that image."

Types of knowledge and competence

On the basis of contributions to the theory of knowledge in various fields (Simon 1960, 1990; Kuhn 1962; Polanyi 1967; Nelson and Winter 1982; Bandura 1986; Grant 1996),

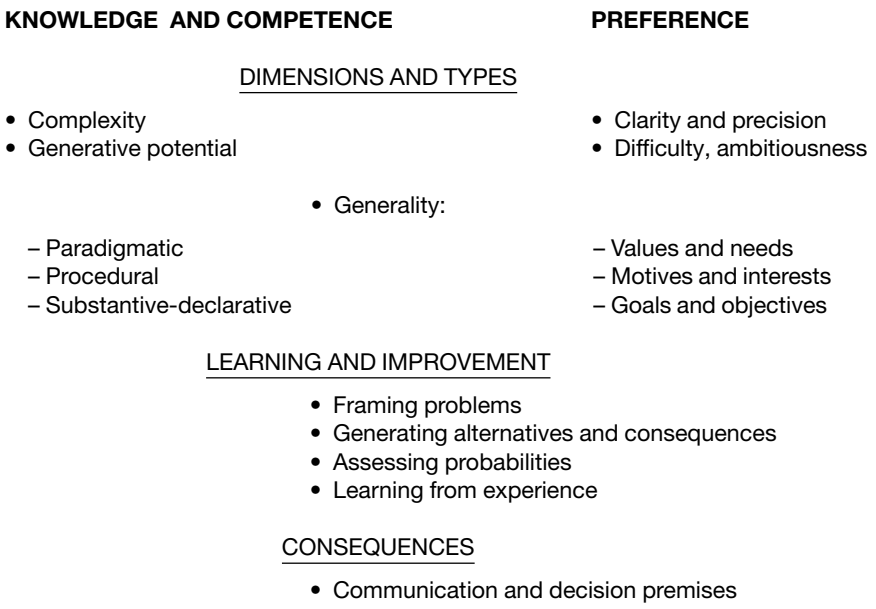


Figure 1.1 Knowledge and preference

knowledge relevant for economic action can be represented along some dimensions with important organizational implications. In this chapter we consider those dimensions that qualify the type of knowledge used by an actor in relation to an economic world: (1) the hierarchical organization of knowledge and competence; (2) the extent to which actors are aware and can transfer these resources (tacitness); (3) the quantitative and qualitative characteristics of information which make knowledge more or less *complex* and more or less *incomplete*, thereby generating a state of uncertainty in the subject; (4) the *combinative and generative potential* of knowledge and competence with respect to activities.

Those dimensions of knowledge and competence that characterize the *relations* among actors (in particular as deriving from the division of labor, as specialization and specificity) will be considered in the opening and founding chapter of Part III (Chapter 8), after having analyzed the problem of interdependence and coordination between multiple actors (Part II).

In addition, here the focus is on actors as subjects endowed with knowledge and competence resources. In Chapter 8, the relation between actors and the resources they "possess" – distinguished in human resources (human knowledge and competence) and technical resources (machine-embodied knowledge and competence) – will be taken as problematic, as a relation that can be organized in various ways, depending, among other things, on the nature of knowledge as described below.

THE HIERARCHICAL STRUCTURE OF KNOWLEDGE

"Paradigmatic" and "critical" knowledge

A first and framing component of actors' knowledge consists of underlying concepts and theories that enable them to make sense

of observed phenomena, hypothesize correlations among phenomena, and define economic problems; and of cognitive schemes which direct their attention, defining categories of subjects and objects and evaluating them positively or negatively.

Examples of basic assumptions about economic behavior are: Is the economic environment an external given, or can it be modified? Is work a costly activity or is it beneficial for the person who does it? Are other economic actors potential competitors or potential partners? (Schein 1985). Why is this type of knowledge present? Does it have any positive property? What are its organizational consequences? What are its limits?

Logicians and cognitive scientists have stressed the technical impossibility or the disproportionate cost of insisting on directly and knowingly checking all information and options on which human beings in general, and economic actors in particular, base their own conduct. It is sensible to trust teachers, to rely on past experience, to accept good "rules of thumb." It would be neither convenient nor feasible to start from scratch each time, rechecking all the information and assumptions on which inherited knowledge is based; and if it were attempted, it would entail enormous risks of error and inferior performance with respect to other actors (Bandura 1986). On this basis, a process of natural selection could even be expected to occur, in which unreceptive subjects would be eliminated in economic action while the number of "docile" ones in the population would increase (Simon 1990). In this way, a "species" of economic actors possessing the trait of "docility," or a capacity to accept a cumulated knowledge base, would even be privileged in the competition for survival.

However, each of us, as human beings, economic actors, and "informal scientists" in everyday life, has come up against the very difficult problem of drawing a boundary between the knowledge we are willing to accept as non-problematic, and the knowledge we wish to scrutinize critically and subject to conscious learning processes. Not by chance is this also a core dilemma in science (Kuhn 1962; Lakatos 1970). Indeed, the dilemma is typical of all processes involving knowledge acquisition and learning. On this basis, the nature and dynamics of economic and organizational knowledge can be and have been assimilated to the much more extensively studied nature of scientific knowledge: *a "core" of knowledge "assumed" to be outside discussion – called here "paradigmatic"*[2] – surrounded by a "belt" of hypotheses which is subject to critical examination (Argyris and Schön 1978; Duncan and Weiss 1978; Weick 1979b).

This schematization of knowledge can help to predict which economic actions will be more "inert" and which are more adaptive: economic actions that come into being as a result of "paradigmatic" knowledge will show a high level of inertia and will be more subject to natural selection processes than learning processes. Moreover, the "platform" of an actor's paradigmatic knowledge helps to predict the learning trajectories that can be followed by starting from that basis, or by working within that frame. An illustration of the dependence of learning paths and solutions found on paradigmatic and framing knowledge is given here in Box 1.1.

Beyond predicting and understanding behaviors, if knowledge is so structured, a *knowledge engineering* problem can be posed: Is there an "efficient boundary" of paradigmatic knowledge? How wide a set of shared and taken-for-granted beliefs should there be? Consider that this component of economic knowledge is scarcely modifiable

Box 1.1
Physicians, generals, and videotapes

How can the knowledge of a particular battle tactic help us to solve a medical problem? Gick and Holyoak conducted important experiments on the effectiveness of transferring knowledge and models to solve new problems. The following is a well-known experiment: suppose that you are a doctor and you are dealing with a patient who has been diagnosed with stomach cancer. It is impossible to operate; however, there is a kind of radiation that can destroy the tumor if administered at high intensity. Unfortunately the high intensity of the radiation would at the same time destroy the good organs and tissues affected by the radiation; at a lower intensity, good tissues would not be damaged but the radiation would not be enough to cure the patient. How do you cure the patient without damaging the good organs and tissues?

In their experiment, Gick and Holyoak posed this problem to groups of subjects. In order to provide these groups with source models, they were told different versions of a military incident. A general wants to capture a fort centrally located in a region. There are several roads that lead to the fort but they are all mined, so although small groups could go through with few risks, a larger contingent would certainly cause the mines to explode. However, the general needs all his troops in order to launch the winning attack against the fort. The different versions of the story offer different conclusions. In one version, for example, the general discovers an unguarded road that leads to the fort and sends all his troops along that route. In another version, the general divides his troops into several small groups and sends them simultaneously along all the routes so that they finally meet at the fort.

The subjects in the experiment turned out to be particularly sensitive to the source model that was presented to them in the form of a military incident. For example, about 75 percent of those who heard the version of the story where the general divides his troops came up with the correct solution to the medical problem – apply low-intensity radiation to the stomach from several directions, so as to pass only a limited amount of radiation through each healthy organ. On the other hand, only 10 percent of those people who were not told about the source model came up with the correct solution. It is interesting to note that all those who were presented with various versions of the story tended to develop different solutions. For example, those who read the version of the "unattended route" tended to suggest solutions based on the identification of an open passage – i.e. the esophagus – through which the high-intensity radiation could be given.

The Gick and Holyoak experiment enables us to understand the structure of competence transfer, as well as giving us a language to describe and analyze it. We can use these initial tools to reconstruct a richer and more complex example that might be faced by firms on a daily basis. The case of the development of the market for videotapes has gained a lot of attention, because of the disparity of competitive results among those firms that had first invested in product research, and the others that came into the market as technological followers. The video cassette recorder (VCR) case can be interpreted as a parable on the advantages of arriving late, and on the risks faced by technological pioneers. However, in light of the concept of knowledge transfer, the case can be interpreted in a very different way.

The two largest groups of firms that have created the VCR market highlight two different source models concerning innovation processes which led these organizations to focus their efforts very differently. On the one hand, because of their own previous experience in radio and television broadcasting, the first VCR pioneers (RCA and Ampex) initially concentrated their efforts on high-performance and high-cost recording tools aimed primarily at the market for technical instruments for television broadcasting. Later, however, they found themselves unable to reconvert such technologies to products for the large consumer market (primarily owing to manufacturing innovation problems.) On the other hand, the first followers in the development process (JVC, Sony, Matsushita) focused their efforts on developing a product aimed at the mass consumer market. As a result of these different objectives (reflecting the different routines, competences, and experiences of the firms), RCA and Ampex concentrated their efforts on sporadic research on important technological breakthroughs. JVC, Sony, and Matsushita, instead, continuously focused the process on developing a product that could be produced cheaply and on market research, owing to their experience as manufacturers of mass-produced electronics. As a result of these different technological strategies, RCA and Ampex found themselves excluded from the "window of opportunity" when it was effectively open during the second half of 1970.

The VCR case illustrates the sensitivity of strategic decisions of firms on their competences and accumulated experience (Prahalad and Hamler 1990.) Like the experimental subjects of Gick and Holyoak, firms have been exposed to different source models and have thus developed different representations of the problem to be solved and different behavioral strategies. The future of firms largely depends on their ability to use their own past intelligently.

Source: Warglien (1990).

because it is learned in an uncritical and unconscious way, or because it is believed to be "self-evident" and "outside discussion," or because it is made unmodifiable by convention. It will change very slowly or, rarely, through difficult "revolutions." Hence, *there is a trade-off between the informational efficiencies it brings about and the inertia it generates.* On one side, an actor should be able to exploit a basis of learned lessons and to act without verifying all assumptions on which action is based; on the other side, it should be able to question those assumptions and to explore radically new hypotheses of action, at least from time to time, for generating innovations (March 1992). Learning processes and organizational solutions can be shaped so as to achieve these capabilities. This issue is addressed in the next section on decision and learning processes, as well as in the second and third part of the book on coordination mechanisms and organization forms.

"Substantive" and "procedural" knowledge

Another distinction contributing to the hierarchical structuring of knowledge is that between "substantive" and "procedural" knowledge. "Substantive" knowledge consists of substantive propositions about what is observed and what to do, while "procedural" knowledge consists of rules on how to observe and how to find out what to do (Simon 1976; Anderson 1983). Hence

procedural knowledge is a "higher order" set of " mental programs" which governs the more operative notions of substantive knowledge.[3] For example, to possess knowledge about electronics and mechanics does not make an expert electromechanical technician; or to know that substituting a certain component of a broken machine usually fixes it is different from knowing how to search for other alternatives if the machine does not start again. The relative incidence of substantive or procedural knowledge in the endowment of a subject, therefore, is a determinant of the degree of discretion and expertise of the actor.

"Tacit" and "explicit" knowledge

People know more, and are able to do more, than they can explain. This component of knowledge is defined as the "tacit" knowledge of an actor (Polanyi 1972). The notion of tacit knowledge does not point at those components of know-how that have *not yet* been analyzed and made explicit, but could be codified without loss of information. For example, in the transformation of wood, skilled workers may apply sequences and procedures which have been learned through time and never been declared in an explicit procedure. However, suppose that the combinations of pace, motion, materials, tools, and remedies for problems encountered could be observed or reconstructed by the master worker so as to be communicated to others. In this case, the difference would be between unexpressed/informal knowledge and codified knowledge, and the organizational issue would be simply whether it pays off to structure information (for example, to allow production on a larger scale). By contrast, *knowledge is tacit if it is intrinsically difficult to identify which information and procedures are applied in successful processes.* Art, sport, and research and development are examples of activities largely based on tacit knowledge. An example of tacit economic knowledge is that used by a master artisan, as illustrated in Box 1.2.

The presence of tacit knowledge in many economic activities has far-reaching consequencies for their organization (illustrated in Parts III and IV). Knowledge of this type can be incorporated in human resources, in technical instruments, or in the organizational routines of a collective actor as a firm. By definition, however, it is difficult to diffuse or transfer it without common practice and reciprocal observation; therefore, economic action and interaction, where tacit knowledge is involved, cannot be effectively organized by mechanisms that do not allow this type of information exchange or learning.

Competences

The notion of competence further contributes to an understanding of the distinction between tacit and explicit knowledge, and the hierarchical organization of actor's resources that are relevant for organizing. The concept of competence is used in economic and organizational analyses as a wider construct, *including the notion of knowledge as well as those of skills and capabilities*: a *component of expertise that is "embodied"* in the whole body of the actor – the actor's force, senses, equilibrium, responsiveness – and not only in the actor's brain. Although the terms just used are typical of physical persons, it is common now to speak of "a firm's competences," precisely indicating not only its knowledge base, but also those capabilities that are embodied in its people, organization, and technologies (Richardson 1972).

Different types of competence can be

Box 1.2
The potter

Imagine observing a potter at work. On the shelves we can admire vases of various colors and shapes. To produce these pieces, the potter's job consists of various steps: s/he works the clay with the hands in order to obtain a lump, making sure that it is of the right consistency and has the appropriate degree of moisture; s/he then works the clay on the wheel to the desired shape, using the hands to correct small imperfections; in order to create spouts and handles, s/he uses special spatulas and tools; s/he then puts the vase into the oven for the first firing, taking care that the temperature is set correctly and that the vase is at the proper distance from the fire. If necessary, s/he fires the vase a second time. S/he then proceeds to decorate the vase, picking and choosing various colors. The potter evaluates the results obtained during each step of the work; if the piece is not satisfactory in terms of quality or shape, it is discarded. We label the potter as having a practical knowledge which allows him/her to "feel" the quality of the material that s/he uses, the consistency of the mixture of clay, the appropriateness of a decoration, and the effects that colors have on the clay. S/he knows that for different kinds of products and shapes, different kinds of clay are necessary, or alternatively that the mixture may change. The potter has a varied portfolio of models to choose from which can, however, be adapted to various needs upon request. When s/he works at the wheel, s/he knows what the pressure of his/her hands should be in shaping a piece and how to allow, with hands and fingers, for any possible imperfection; s/he also knows how best to synchronize and coordinate the various steps in such a way as to make the most efficient use of time. S/he is able to achieve the goal of producing a good piece based on certain design criteria, the characteristics of the material used, and the kinds of tools utilized. As in the case of the ski-instructor (see Box 1.3), the potter is an example of someone perfectly in synchronization with a continuous cycle of actions and steps, in which the material, the tools, the hand movements, the knowledge, and the actions constitute a single, consistent system.

Source: Lanzara (1992).

defined according to the extent and depth at which they are incorporated in actors.

Spencer and Spencer (1993) propose the image of an iceberg or of Chinese boxes to represent the idea that, in addition to visible layers, competence has "sunk" layers that are very difficult to distinguish from the identity of actors themselves and difficult to change. The traits and talents with which a person is gifted, the person's physical and cultural heritage, and personality, constitute the core, or deepest layer, of human competences; the knowledge and energy incorporated in technical assets can be seen as a complementary component of the "core competences" of a firm (Prahalad and Hamler 1990).

An intermediate component is seen as constituted by *skills or capabilities*. This layer of competence is seen not only as a set of resources but also as operating knowledge capable of dictating actions in response to the circumstances encountered, a set of "if–then"

rules that connects repertories of possible actions to types of problems and conditions (Nelson and Winter 1982). This layer of competences then parallels (and includes) that of procedural knowledge. For example, a person who is very knowledgeable in a given subject area may be very competent in terms of notions, but may not necessarily have the capability to act well on that knowledge. It is important that *competence as a stock of resources be coupled with application routines or action programs*. The "database" made up of substantive knowledge should be made operational by a procedural, action-oriented component. Habitualization and interiorization should transform calculated action into automatic action. Competence in sports is often cited as an example of this dynamic and procedural component of competence, as shown in Box 1.3.

The most explicit, codified, and manageable layer of competence is identified with explicit knowledge and codifiable expertise. For example, the ensemble of technical notions and symptom-diagnosis relations employed by a financial analyst or a physician to solve standard problems of investment or health care may be examples of this layer of codifiable competence, based on explicit knowledge.

Box 1.3
The ski-instructor

By observing a ski-instructor descending a slope we can evaluate – without hesitation – the instructor's level of ability, and thus we know that we are not dealing with an amateur. We appreciate the correctness of the instructor's position, the harmony and the fluidity of the movements, the smoothness with which the turns are coordinated, and the way the instructor controls and maneuvers skis and poles according to the characteristics and the difficulty of the slope (steepness, curves, etc.) We also notice the expert skier's ability to correct any mistakes or react to difficult moments. What impresses us is the elegance and the spontaneous way in which such experts ski, as well as their lightness, which makes it all seem effortless. The skier's practical knowledge – what is referred to as "knowing how to ski" – is in part codified within the sensory-motor skills, in such a way that the skier's muscle behavior responds instantly and almost automatically to the signs and the characteristics of the slope. Thus, one could almost say that the skier's behavior functionally "complements" the technical characteristics of the tools employed (skis, poles, booths) and the morphological peculiarities of the slope. However, the skier's competence also lies in the ability to anticipate what comes next, such as identifying what the toughest sections of the slope are, evaluating snow conditions, and choosing the easiest and least dangerous route to descend the hill. It is obvious that the skier who "really knows how to ski" must possess an explicit knowledge about a physical activity called "skiing," but the cognitive activity that distinguishes the competence of the skier is directed at the production of the skill during the action itself.

Source: Lanzara (1993).

Individual and collective competence and knowledge

The example of sport competence is also useful in highlighting how competences are more closely incorporated in actors and have a larger tacit component with respect to knowledge: therefore they are less easily modified and, in order to be transferred, may require more intense reciprocal interaction and observation between various subjects. However, it would be wrong to conclude that competence (let alone knowledge) is an exclusive or typical patrimony of the individual. On the contrary, if individual actors were to develop their competences only on the basis of direct experience, without taking advantage of the experience of others, learning processes would be extremely costly, slow, and boring. In addition to "inherited" knowledge and competence, many capabilities develop because the experience of other actors believed to be comparable is observed and these experiences are imitated or re-elaborated (Bandura 1986). This process of "social cognition" is fundamental in explaining the efficient development of sophisticated individual capabilities, which would be somewhat improbable without the support of collective capabilities, understood as reciprocal learning networks.

Second, not all types of competences require the same degree of interaction to be effectively transferred. For example, skiing capabilities can be taught fairly smoothly and even transferred on a competence market. This is not true for other competences that, however tacit or explicit, are nonetheless very personal, particular, and applicable only in a single environment – very "specific" to a use or a user (Williamson 1981a; Chapter 8).

Combinative and generative potential

Following a conceptual, primogenital path, the first important instance where the economic actor was modeled as a set of resources – and particularly of competences – is usually recognized in the work of Edith Penrose (1959). Looking for answers to classic economic questions on the nature of the firm and the reason for, the direction of, and the potential for its growth, Penrose identified a powerful explanatory factor in the capacity of both technical and human resources to deliver a wide range of potential services, and, in particular, a range that is often wider than the sum of their actual uses.

> Strictly speaking, it is never the resources in and of themselves that constitute the input of productive processes, but only the services that the resources can deliver. . . . The resources consist in a set of potential services and can, in large part, be defined independently of their use, in contrast to the services themselves that imply an activity or function. As we shall see, the source of the uniqueness of every firm lies, in large part, in this distinction.
>
> (Penrose 1959: 25)

At the individual actor level, its set of talents and competences is seen as capital that can be put to different uses, and hence as an input to the definition of new tasks and services rather than a consequence of task requirements (a resource-driven approach). At the level of collective actors, it is the combination of the uses that resources are put to that becomes of prime importance, since the variety of the possible combinations makes collective competence "distinctive," unique, and imitated only with difficulty to a much greater degree than individual competence, and, from an economic point of view, creates a possible additional source of uniqueness and "competitive advantage" (Penrose 1959; Barney 1990).

The very fact that competences are not easily divisible from the actors possessing them contributes to creating surpluses of competences with respect to the particular uses for which the resource was acquired. Competences, and more precisely the possible combinations and recombinations of competences, are a formidable base of value creation and task design (Chapter 8). Let us call this dimension *a competence "generative potential" – applicable to both human and technical resources*.

Complexity and incompleteness

Knowledge may often be in short supply – with respect to what would be necessary for complete coverage of all the relevant information for solving problems. Problems obviously can be more or less complex. As they grow larger – including many elements and relations among them – the knowledge required for solving them becomes "*computationally complex*," as in the prototypical example of the game of chess. As problems include difficulties in diagnosing what is observed, what actions cause what results, or even what is a positive or negative consquence, the knowledge required for solving them becomes "*epistemically complex*" – as in the prototypical example of the game of scientific discovery. This growth in knowledge complexity at a certain point will succumb to incompleteness, given the existence of limits to human rationality. The *lack of knowledge* about some elements in decision-making gives rise to a state of *uncertainty*. *This contingency is of core importance for understanding and designing which decision process to follow.*

For example, Thompson and Tuden (1959) in a seminal contribution hypothesized that different decision strategies are feasible contingent to the initial state of knowledge of two main decision inputs: objectives and cause–effect relationships. Developing this approach further, *knowledge relevant for decision-making* can be characterized as a system of conjectures regarding at least the following four fundamental classes of decision inputs:

- *objectives* (hypotheses about what is desired and perceived to be obtainable)
- *cause–effect relationships* (hypotheses about which actions or alternatives are related to which desired results)
- *probability judgments* on the likelihood of consequences
- *observational judgments* (estimates and measures about what events and "data" are observed).

The next section will examine how knowledge of these elements can be improved. The next chapter on decision and motivation processes addresses the problem of what decision strategy is applicable if uncertainty on some or all the elements cannot be further reduced.

IMPROVING JUDGMENT UNDER UNCERTAINTY

The processes of perception and judgment on the basis of which decisional inputs are defined are subjective and fallible. This consideration would, however, have few practical consequences if one were not able to identify some of the systematic weaknesses of human judgments and how to overcome them. A vast area of research in cognitive psychology has generated a mass of reliable and interesting results in this respect, especially on individual and group decision-making (Carroll and Payne 1976; Nisbett and Ross 1980; Einhorn and Hogarth 1981; Kahneman *et al.* 1982; Bandura 1986).

On the basis of these researches, it is

possible to make a sort of inventory of the principal systematic distortions of human judgment (obviously a limited inventory, not an exhaustive or absolute one). The term "biases" is used because it refers to undesired effects, optical illusions, and unconscious errors that, if only they were "seen" by decision-makers, would be willingly corrected.

These corrections cannot and should not be interpreted as the restoration of an impossible absolute or global rationality. They can and should be seen as a guide to an improved use of our inescapably fallible and bounded rationality, to the capacity for discerning between "good" and "bad" heuristics as a function of the decision situation, to the development of capabilities useful in making judgments under uncertainty, and to improving the quality of decision inputs.

"Heuristics"

The term "heuristics" refers to any mental rule or procedure capable of generating or finding something that is being sought. In other words, it is a method of search, which may be more or less well grounded in experience, more or less structured, and more or less "substantive" (prescribing an action) or "procedural" (prescribing a method), more or less "tacit" or "explicit." For example, all of the following rules indicating how to act or how to process information are heuristics: "Take an umbrella if you go out when it is cloudy," or "Hit the tennis ball at the highest point of the parabola after it bounces" (Russo and Schoemaker 1989); when searching, apply the principle of scanning "breadth first" and go "in depth" next (Newell and Simon 1976); "Look where information is readily available" (Tversky and Kahneman 1974); "Lower expectations if it is difficult to find an alternative that is satisfactory" (Simon 1955a); "Infer that A is the cause of B, if B is always observed after having observed A" (Nisbett and Ross 1980).

The use of heuristics is necessary in all problems where search is important. Still, there are important differences among heuristics. The difference between an expert and efficient decision-maker and a novice in a field depends to a large extent on the variety of heuristics they possess, on their degree of general validity, and on the awareness of possible distortions (Simon 1987, 1977; Kahneman, *et al.* 1982).

In this section we will present a selection of main distortions arising from the use of heuristics related to four main decision inputs and to their formulation processes:

- the definition of problems
- the search for information and alternatives
- the calibration of probability judgments
- the making of inferences based on observed experience.

Defining problems and interpreting observations: "framing" and "cognitive distance"

Many, if not all, decision processes begin with problems. However, no problem really exists in nature. It is always the fruit of a mental model, a series of perceptions and interpretations by an actor. For example, to state that "there are investment opportunites to be found in new technologies" defines a problem and a possible course of action, but implies several judgments and mental operations: allocating attention to technological innovations rather than to other aspects of reality; explicitly or implicitly defining desirability criteria, i.e. parameters for evaluating outcomes; diagnosing that the current state of affairs is not satisfactory or could be

improved, i.e. that there are "performance gaps" (Simon 1947, 1955a; March and Simon 1958).

Merton (1949) said that any way of seeing is also a way of not seeing. Defining problems is not only subjective but also selective in that it implies considering certain aspects of reality and ignoring others, stating what is "in" the problem and what is "outside" the problem. Real phenomena have infinite aspects and no decision-maker could consider them all. There is a considerable difference, however, between a decision-maker who considers one or a few aspects of a problem and one who considers many (Payne 1976), as illustrated by the study described in Box 1.4.

"Structured" and "unstructured" problems

The fact that the definition of a problem always implies building a model of the reality (Simon 1955a) does not imply that the nature of reality has no influence whatsoever on the complexity of the problem. The analogy between problem-solving in economic action and in scientific discovery (Simon *et al.* 1981) has shed light on the reasons why, in some areas, problems are typically "well structured" (well defined, with clear boundaries, a finite number of potential alternatives and a single best solution), while in other areas they are "ill structured" (Simon 1973). For example, physics problems lend themselves to a higher degree of structuring, on average, than those in medicine or sociology. Therefore, it is not true that the wider the definition of problems, the better – nor conversely. The ability to choose the least number of aspects that, at the same time, includes the largest number of relevant factors for an effective

Box 1.4
Framing investment decisions in information technology

In a research project about the decision processes related to the purchase and introduction of automated information systems, the decision-makers in various firms were found to define the problem in at least three ways, depending on the breadth of their perspective. For some, the problem concerned investing in technologies (equipment, software, and data elaboration techniques); for others, it included the issue of which informational sub-systems within the firm could use computerized information system services (order management, personnel management, general accounting, etc.) and how they could use those services. In the broadest formulation of the problem, issues related to information technology investments were considered to be problems connected to the evolution of the information systems of the firm as a whole; they therefore had an impact not only on all of the firm's activities, but also on the decisions and motivation of the people. The capacity to define problems in a wide rather than a narrow way was found to be related to the level of actors' experience in the decision area; this was consistent with the theoretical proposition that people learn the consequences of the action and their own preferences (judgments of desirability) on the basis of experimentation.

Source: Research described in C. De Vecchi and A. Grandori (1983) *I processi decisionali d'imprese*, Milan: Giuffrè.

solution, is a fundamental cognitive capability (in scientific activity as well as in decision-making). There is a trade-off between effort and accuracy, and between completeness and manageability, in the definition and structuring of problems.

Beyond the awareness of these trade-offs, problem-formulation may be improved by the awareness of some fundamental biases that can influence it. Two classes of biases have shown to be particularly important in framing problems: *cognitive dissonance* and *prospect effects*.

Consider the image drawn in Figure 1.2. What do you see?

The lines that make up the drawing can be interpreted in two different ways. For example, a curve can be interpreted as a chin or a nose depending on how it is related to other elements in reconstructing the meaning of the whole image – that could be an old and a young woman simultaneously. Typically, after one interpretation has been adopted, it becomes difficult to see other possibilities. An explanation of this phenomenon is that our mind is geared or "programed" to eliminate inconsistencies, contradictions, and incoherent information, discarding or even not seeing those elements which do not fit or have no significance according to one's own scheme of interpretation. Cognitive psychologists have called it a tendency to reduce cognitive dissonance (Festinger 1957). The implication of this cognitive tendency for problem-formulation is that once a point of view, an interpretation of observed "data," has been accepted, the ability to see alternative interpretations is inhibited and this can lead to rigidity and conflicts.

Figure 1.2 Gestaltic figures

Other important framing effects are generated by even more subtle shifts of meaning with respect to the choice of language and reference systems. A famous experiment designed to show these effects (Tversky and Kahneman 1981) is reported in Appendix 1. It should be played in a large enough group (such as a standard classroom); half of the participants solve the problem in Appendix 1.1A and the other half the problem in Appendix 1.1B.

On average, the choices made regarding the problem as formulated in the two ways are systematically different. In the 1.1A formulation, decision-makers tend to be risk prone: the majority select the uncertain alternative over the certain alternative, even though the expected value of the two alternatives is the same, in terms of expected saved lives. In the 1.1B formulation, decision-makers tend to be risk averse: the majority prefer the sure alternative over the uncertain alternative of

equal-expected value. Nonetheless, if the two formulations of the problem are compared, one realizes that they have identical characteristics with respect to the number of lives saved or lost as a result of the two restructuring plans. Tversky and Kahneman attributed the systematic difference in choices to a "prospect" effect. Decision-makers assess consequences with respect to some reference point – for example, the current situation, a "neutral" outcome, the best or the worst that can happen to them. They can then perceive or express the consequences as losses or gains with respect to that point of reference. Adopting a "positive frame" (see consequences as possible gains) or a "negative frame" (see consequences as possible losses) has a significant impact on the choices made: for example, other things being equal, positive frames make people more flexible in accepting solutions, less demanding, and less innovative; while negative frames encourage one to take risks and to flee from current states of affairs.

Frames are frequently adopted unconsciously, or by chance, or people are simply not aware they may exist. It is not difficult to imagine the potential for biases and conflicts that this may entail. For example, part of the systematic conflict that exists between technical functions and units – such as manufacturing – and commercial functions – such as marketing – is due to the different mental schemes that have been formed as a result of working in different information environments: production plants and raw materials, customers, new techniques, or scientific discoveries; and to the resulting inability to see other aspects of the shared reality within the firm (March and Simon 1958; Lawrence and Lorsch 1967; Chapter 11). In economic negotiations, the framing of consequences in terms of potential "concessions" (losses) or as potential gains with respect to not reaching an agreement has a paramount impact on which agreement will be reached, and on whether it will be reached (Chapter 6).

Cognitive differentiation and distance

The establishment and consolidation of actor-specific frames favor the differentiation between cognitive endowments and styles. In other words, the systematic differences in the type of information that actors take into consideration and the way of interconnecting them become fixed in their minds as a sort of mental software.[4] The consolidation and relative inertia of frames give rise to significant, predictable, and relatively stable "cognitive differentiation" (Lawrence and Lorsch 1967) or "cognitive distance" (Nooteboom 1999) among actors. The diversity of cognitive frames is sustained at the individual and collective level in economic action by the separation between units (groups, departments, or entire firms) that perform activities of different informational nature, which, in turn, attract and reinforce cognitive personalities whose traits are consistent with the nature of the activities and information (Tosi 1992). The specialization of knowledge and tasks, and its advantages, sustains it. Cognitive differentiation in itself must therefore be considered physiological. However, some of its consequences are undesirable such as the rigidity of frames and the inability to see new aspects of reality; or the misunderstandings and conflicts among actors with different mental outlooks.[5]

Antidotes

Communication difficulties do not imply and should not lead to total incommunicability.

As among people who speak different languages, the difficulties diminish as reciprocal exposition and communication intensity increase (provided there are no underlying conflicts of interest). Indeed, the investment in communication channels or "intermediaries" among actors using different mental schemes and styles is traditionally considered to be a good antidote to interfunctional and interpersonal conflicts (Lewin 1948; Lawrence and Lorsch 1967) and has been recently recommended to sustain knowledge transfers between business units (Grant 1996) or different firms in alliances (Lutz 1999).

A second type of antidote is multiplying the number of frames available to a decision-maker (Russo and Schoemaker 1989). An "open mentality" can be acquired and formed, in this perspective, by broadening individual cognitions and using groups. Individuals can cultivate additional areas of interest that differ from their principal fields of action. They can reduce their degree of specialization and focalization on particular tasks and information. All this helps to reduce the undesirable effects of framing and of the differentiation between cognitive styles. However, it may be difficult for all this to occur at individual level; and, if it does, it may remove the desirable effects of specialization and focalization. As a result, multiple competences' group problem-solving, or the use of third parties with an intermediate orientation, is often a more effective antidote than individual polyvalence, especially in unstructured and important activities (Lawrence and Lorsch 1967; Chapter 11).

"Local knowledge"

Research is a costly and difficult process. Therefore, actors might deliberately choose to limit it, as will be illustrated in the next chapter. Here, instead, the unconscious and spontaneous tendency of mind to orient and restrict attention in certain directions will be considered.

Tversky and Kahneman (1974), in their most influential and seminal work on the subject, identified three basic heuristics which tend systematically to restrict and bias the type of information we tend to consider, by catching us in a "local knowledge" trap: "availability," "representativeness," and "anchoring."

Availability

Consider the experiment described in Box 1.5. This judgment can be said to be "under uncertainty" because subjects do not usually possess all the relevant information for expressing an opinion. The usual process for generating information is then to generate examples and try to recall instances of the phenomenon. As a result, the judgment on which factor is more important turns out to be influenced (and biased) by the ease of retrieval of information. What are the characteristics of easily retrievable information? Not only the frequency with which one is exposed to it, but also its salience: its familiarity, vividness, and cognitive and emotional intensity.

In the example in Box 1.5, the vividness of information derived from journalistic sources and its frequency – biased in favor of impressive deaths – systematically bias respondents' judgments on the relative importance of different causes of death. The availability heuristics can therefore have important consequences on economic action. Insurance may or may not be bought depending on the salience of risks. Investments may or may not be made depending on how close and familiar the cases of failure and success are; staff

Box 1.5
An experiment on availability

Here are some possible causes of death, listed in sets:

Set 1	Lung cancer	vs	Car accidents
Set 2	Emphysema	vs	Murder
Set 3	Tuberculosis	vs	Fires

For each set, please select the item that it is believed will cause the most deaths in a year's time. The average percentages of the answers by the people interviewed by Russo and Schoemaker are shown in Appendix 1.2.

Source: Experiment described in Russo and Schoemaker (1989).

promotions may favor more visible rather than better performers; project costs and times are usually underestimated because it is more difficult to imagine what can go wrong rather than a sequence of normal activities.

Representativeness

Now try the question posed in the problem described in Box 1.6, before continuing with reading the rest of the chapter. Typically, in answering this question, people vastly overestimate the probability that Stefano is a librarian. Why?

On the basis of self-reports on mental operations by subjects, it appears that what governs assignments to categories is, above all, a judgment on the similarity between the qualitative description of the individual and the stereotype of that category, i.e. on how "representative" the individual case is of the category. Background information, such as the incidence of occupational categories in the population, or what is the presence of shy and introverted people in any profession, is scarcely taken into consideration, if at all (Tversky and Kahneman 1974).

More generally, the effects of chance and stochastic phenonema are poorly understood and little considered by decision-makers. This leads to believing in predictions and estimates with unjustified confidence. For example, evaluations of the probability of success of a person in a job, made on the basis of the correspondence between candidate profiles and job descriptions, are vulnerable to a representativeness bias .

Anchoring

No estimate or judgment is possible without reference points. Nonetheless, depending on the type of reference point employed, rather diverse estimates and different actions are produced. They should therefore be used knowingly.

For example, submit the following two estimating tasks to two different groups of people. What is the estimated product of $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$, in 5 seconds? And what is the estimated product of $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, in 5 seconds? Tversky and Kahneman (1974) report that the average estimate for the ascending sequence is 512, and for the descending sequence 2,250 (the correct answer is 40,320).

Another experiment, conducted by Russo and Schoemaker (1989), is described in Box

Box 1.6
An experiment on representativeness

This is a description of a person made by someone who knows him best: "Stefano is very shy and reserved, always available but showing little interest in people and the world in general. He is submissive and has a need for order and structure; he is very detail-oriented."

Question: what would you say the probability is that Stefano is:

- a farmer %
- a salesman %
- a librarian %
- a physician %

(Time allotted to answer: 2 minutes.)

The sum of the four probabilities does not necessarily have to add up to 100.

Source: Experiment described in Tversky and Kahneman (1974).

Box 1.7
An anchoring experiment

Russo and Schoemaker (1989) asked about 100 managers the following question: What do you think the prime rate will be in 3 months? The average answer (this happened in 1983 when prime rate was around 11 percent) was 10.9 percent.

Later, they put the following two questions, in sequence, to another group of managers: Do you think that in a 3 month period the prime rate will be more or less than 8 percent? What do you think it will be? The goal was to verify if the first question, aimed at anchoring the managers to 8 percent, would have lowered the estimates compared to the ones expressed by the anchored group. The average response was 10.5 percent.

The following question was then posed to a third group: Do you think that in a 3 month period the prime rate will be more or less than 14 percent, and what will it be? The average estimate was 11.2 percent.

Source: Experiment described in Russo and Schoemaker (1989).

1.7. People frequently make estimates starting from a known initial value and adjust them in the direction that they believe to be correct. For example, what will sales amount to in the next budget period? It is reasonable to begin with current sales and "adjust" in the right direction (is there economic expansion or contraction?). Empirical research has shown, in general, that adjustment is not sufficient. If estimates are anchored, they tend to be systematically biased toward the anchor and not sensitive enough to the corrective factors that should have been taken into consideration. If estimates involve many related variables

rather than just one, the effect is likely to be stronger. As Tversky and Kahneman note, anchoring contributes to explaining why people often badly underestimate, for example, the probability of failure of complex systems and actions – such as a nuclear plant, the human body, or the launch of a new product. In fact, people anchor their estimates to the failure probability of each single component, which may be low. If, however, the number of components is high, the estimate of the joint probability of failure is not sufficiently adjusted.

Antidotes

Many aspects of organizational systems and structures provide a response to these fundamental biases in judgments, and offer tools for improving the validity and reliability of the knowledge used in organizational decision-making. For example:

- The use of checklists on "factors to be considered" and *decision-support systems* is not a useless overhead invented by experts to squash "intuition." Intuitive decision-making in situations of uncertainty is subject to systematic error.
- *Explicit and codified personnel evaluation systems* are useful for reducing the above biases – and the unequities which are likely to go with them.
- "Zero-based budgeting" and other "zero-based" management techniques – "starting from zero" the analysis of activities and of resources needed – are corrections to the anchoring effect in the growth of activities and of the resources allocated to them, that in each period tend to be anchored to what was done in the previous period.
- When making important decisions it is useful to consider the apparently "unuseful" and boring output of information system services on "base-rate" information, and on the probability distributions of the variables at hand.
- If adequately varied in their composition and free in their dynamics, the use of groups can reduce the local search trap, generating richer and more varied information.

Overconfidence and underconfidence

All the three heuristics described so far may intervene, and intermingle, when a judgment on one's own knowledge and competence is involved. This judgment is usually called a "confidence" judgment: assessing how much one knows with respect to what one does not know on a matter.

Try to make an estimate of the size of a variable that, under normal information conditions, implies a judgment under uncertainty: for example, try to estimate the unemployment rate or the percentage of women in important positions in a foreign country, or the time it takes to manufacture a car. In order to control the experiment, choose examples of variables on which statistics are available for subsequent control of the quality of estimates. Then, attempt to estimate the probability or confidence that the estimate made was correct, using one of the following two methods:

- set an interval in which the value of the variable should fall, and express the probability that the estimates made will fall within the interval
- alternatively, set a level of confidence to reach, say 80 percent, and then estimate how large a value interval should be in order that, in eight out of ten cases, the estimate will be correct.

Both statistics can be constructed either by the same person on many estimates, or by several persons on even a single estimate.

In order to control how well "calibrated" the probability estimates are, one can compare the subjective probabilities of being right, or the subjects' degree of confidence, with the observed frequency of correct responses. The usual outcome of these experiments is that the average subjective confidence in one's own judgment is far greater than the frequency of correct answers (Lichtenstein *et al.* 1982). For example, in an experiment that we conducted with MBA students, they were asked to put in order three management schools, in terms of the overall number of course participants on an annual basis. The average subjective probability of having given the right order was about 60 percent, whereas the correct ordering was actually produced by only four students out of 120. The reason for such a wide discrepency can be traced to a combination of distortions, arising from qualitative information about schools being more available than quantitative, considering image and fame as "representative" of size, regardless of structural factors, such as market structure (for example, the number of comparable schools in the national market) and the school strategy (for example, the breadth of the course portfolio).

Underconfidence and self-efficacy

Confidence judgments are inescapably involved in the application of an actor's knowledge and competence in order to reach certain performances. Possessing competence does not necessarily mean using it. An intermediate variable of great importance for explaining performance differences, which plays a role even before expected benefits are considered, is an actor's perception of his/her own competence and of its expected link with performance ("self-efficacy" judgments – Bandura 1986). This perception may easily be uncalibrated. Experienced subjects are easily overconfident, because they assimilate new tasks to those in which they have already succeeded, or even develop a self-confidence that is independent from the task in hand; while the lack of experience and "noviceness" tends to produce underconfidence (Lichtenstein, *et al.* 1982; Bandura 1986).

Antidotes

The management of self-efficacy and the antidotes to uncalibrated confidence are a problem of sustaining and correcting learning processes. Different types of learning processes can contribute, depending on the possibility of direct experience and on the complexity of the task.

Direct experimentation

"Enactive attainments" – i.e. the direct experience of one's own success – "constitute the most influential source of information on effectiveness because they are based on authentic mastery experiences" (Bandura 1986: 399). In fact, it has been empirically shown that regular access to feedback on one's own performance is the most systematic factor in the calibration of probability and confidence judgments (Lichtenstein *et al.* 1982). Direct feedback is not, however, always so clear and objective as in the example of the long jump. In economic activity, the availability of feedback frequently depends on the potential and capacity of others – teachers, bosses, clients – to give it.

Vicarious learning

Actors adjust their own confidence judgments by observing how other actors who have comparable resources and competence succeed or fail. Observation and imitation are of course particularly important in activities that are new to the subject, or in activities where performance has few objective measurements and depends above all on comparison with others (Bandura 1986). However, learning through imitation requires similar and stable conditions, and is not likely to lead to discovering talents and competences of special quality and level.

Modeling

A more ambitious approach, more likely to enhance success chances and perceptions of chance in complex and new activities, involves the use of experiences for the construction of models of behavior correlated to success in the task, which can then be applied to the specific initial condition and starting competences available to the subject ("*modeling*"). This approach could be seen as an effort to construct causal explanations of performance in specific action fields. Box 1.8 describes an example of how to construct a model of the activities and behaviors that can lead to the completion of a challenging and complex task, and can help in building confidence.

Learning traps

As John Stuart Mill said: the logic of science is also the logic of economics and life. This thesis was more recently reinforced by similar opinions expressed by Karl Popper (1989), Herbert Simon (Simon *et al* 1981), and Karl Weick (1979b). To understand fully the difficulties of learning in economic and organized activities, however, some complications should be considered. While learning, economic actors pursue desired results. This can make learning from experience more difficult – for example, because there is interaction between the observer and the observed, or because actors face trade-offs between acquiring knowledge (on what actions are most fruitful for example) and short-term pay-offs. In this section, three fundamental types of obstacles typical to the learning processes in economic action will be illustrated.

The first and most important learning trap is generally valid for any process of research, while the other traps are more specific to economic learning. Consider the experiment described in Box 1.9.

The most frequent response by subjects in this experiment is to "test" series of numbers that conform or are deduced from the rule they have in mind. For example, if the rule is thought to be "a difference of two units," one would tend to try a series such as 10–12–14; or if the rule is thought to be a constant difference, one would try 10–15–20. The answer is always favorable, and no experiment is particularly informative. Confirming one's own hypothesis could go on forever. This would increase a psychological sense of confidence that the right rule has been found, whereas knowledge does not increase at all. For example, in Wason's experiment (1960), only 6 out of 29 participants had actually discovered the right rule the first time they thought they knew the answer. The most informative experiments are, on the contrary, "falsifying" experiments such as: supposing that the rule is "all numbers are even," try odd numbers; supposing that the law is that numbers should be ascending, try a descending series. The answers to these "falsifying experiments" are highly informative because, at least on structured tasks, they

Box 1.8
Tenure track modeling

Suppose you are the chairperson of an academic department and you want to help a new assistant professor to get tenure. Let us further assume that it is a publish-or-perish university and that the assistant professor is in the summer of her fourth year with a below-average, but not hopeless, record for scholarship. You sit down with her and go over her record. So far she has had two articles published in good journals, one in a mediocre journal, and has four papers under review. She also has six more projects in the pipeline, which can be submitted in the next year. The first task is to figure out what will be needed to make tenure. Let's say she will need about 10 papers, eight in good journals. Since tenure review will occur in two years, and since all projects do not work out, you would suggest that she has all six of the "in the pipeline" projects submitted by January of her last year (i.e. in the next 18 months.) This is to allow time to revise and resubmit before the September deadline. Furthermore, you advise that all "revise and resubmit" revisions on these and the "under review" manuscripts be done within 30 to 60 days. These are the goals. How do you get commitment? If the professor decides that she does not really want to be an academic in this institution, suggest that she look for work elsewhere. But if she does want to succeed, then the main issue is confidence-building. Express confidence based on the work to date. Suggest role models. Be supportive: ask her what you can do to help (for example, some extra assistants for data analysis; time off in the summer; reduced committee work). If previous rejections of papers have been demoralizing, suggest some alternative strategies (for example, reframe and submit to a different journal, combine two papers into one, etc.).

To insure careful tracking (feedback regarding progress), have her make a schedule indicating when each in-process manuscript will be submitted. Go over time-utilization issues (goal priorities) and strategies (for example, has she delegated as much of the busy work as possible? Is she going overboard on teaching? Is she working enough hours? Is she going to too many professional meetings? Is she spending too much time writing conference papers?).

To further help develop effective plans, have her consult with other junior and senior faculty members to see if they have any tips for her. Persuade her to let colleagues (expeditiously) review her papers before submission and also help to interpret letters from editors. Finally, tell her you want her to make it (if you do) and the reasons why.

Source: Locke (1996).

allow potential rules to be excluded or eliminated with certainty. (Wason's experiment is usually played with the coordinator applying a rule of "strictly ascending whole numbers.")

The indications of this experiment coincide with the single most important prescription of scientific research methodology which is taught to any professional researcher: no series of favorable cases can prove that a theory is true; there is more information in the confutation of hypotheses than there is in their confirmation; even though many hypotheses

Box 1.9
Wason's experiment

Suppose that a pre-defined rule is used to create triples of numbers. Among all possible triples, some obey the rule and some don't.

The following triple obeys the rule: 2, 4, 6.

Suppose you are a scientist, or a person interested in studying that rule, and you are trying to understand what is the law that links the numbers.

Based on triples you have created, you can ask the game leader if those triples obey the rule or not; you would get fair and honest answers.

Continue to test triples until you have defined a rule about which you feel reasonably certain.

Source: Experiment described in Wason (1960).

and conjectures are initially formulated by observing and recognizing empirical regularities, they can be improved or corroborated only by falsifying rival hypotheses, or parts of the initial hypothesis that are contradicted by the data and are substituted (Popper 1935; Hanson 1958; Simon 1977).

The tendency to search for confirmatory examples does not run the risk of being overestimated. For example, the fact that actors tend to judge the strength of a relationship between two variables, X and Y, based only on the frequency of favorable evidence is well documented (Einhorn and Hogarth, 1978). Decisions commonly involve hypotheses of this type. Consider any decision rule such as: if an alternative meets certain requirements, it is accepted; if not it is discarded, like rules used for hiring or promoting personnel, admitting students to schools, or granting credits. They are hypotheses that a relationship between (at least) two variables exists: if alternatives with sufficiently high values on an X property are accepted (people hired, credit granted), positive consequences will follow on certain result variables Y (for example, performance, credit repayment). For example, suppose that a new rule for promotion to sales management positions is adopted in a firm. Assume that 80 percent of those promoted are successful because, for example, they meet budgeted targets. What conclusions could be drawn on the validity of the promotion procedure, i.e. on the hypothesis that high values of X (the positively evaluated candidate characteristics) are correlated with high levels of Y (performance measurements)? Actually, no conclusion. To draw conclusions about that hypothesis, it would be necessary to consider what the success and failure rates would have been for candidates promoted according to alternative rules (an alternative hypothesis) or even randomly selected (a null hypothesis).

As Einhorn and Hogarth observed, decision-makers tend not to consider the entire four-cell quadrant of information (X acceptable/not acceptable; Y successful/not successful) even when this information is available at low cost. In addition, in economic activity, this full experimental information is often quite costly to obtain, because it may imply lower results (produced by the application of the hypotheses that turn out to be "wrong"), or because discarded alternatives may be no longer available, or because causal

attributions when actors' deliberate action is involved are particularly difficult, as illustrated below.

Therefore, distinctive complications of learning processes in economic action are: that often discarded alternatives disappear (partial feedback); that an actor's hypotheses influence observed behaviors (interaction effects); and that hypothesis-testing can imply a loss of income. All these factors accentuate the self-confirmation trap.

Partial feedback

Situations where feedback is partial are, for example, those in which it is not possible to control what would have happened if the discarded alternatives had been accepted: the people not hired, the trading offers rejected, the partners with whom a deal is not concluded.

Treatment effects and income losses

Treatment effects and income loss effects are well exemplified by the decision-making process of a waiter hoping for a tip (Einhorn and Hogarth 1978) (analogous examples could be a firm trying to learn which clients to target with higher discounts, or a tax office trying to choose which taxpayer to audit). The waiter has a theory regarding which characteristics of clients are good predictors of the size of the tip. However, since tipping is influenced by the quality of service provided and action is costly, the waiter has an incentive to provide a superior service to those clients with good prospects of giving a tip, and inferior service to those without (according to his theory about what tip predictors are). In this way, theories of an acting learner easily become self-fulfilling prophecies. How would it be possible for the waiter to discover whether his theory is valid? He should be willing to test different predictors, with the risk of obtaining a lower income, in case his initial theory might be right.

Attribution errors

Learning economic actors are often active and interested subjects, also as far as causal attribution processes and the ability to correct mistakes are concerned.

Nisbett and Ross (1980) identified a general tendency, when social phenomena are considered, to hypothesize that subjects, actors, and decision-makers – their will and action – are "causes" of events and results, rather than other more structural and more "exogenous" factors. Management decisions are believed to influence firm profits more than economic and sectoral trends. Managers are rewarded according to results, often regardless of the extent to which the results stem from their actions. The principal cause of the success of the Japanese economy is often sought in some attribute of Japanese mentality, for example.

This bias of "illusion of control" (Langer 1975) has received various psychological explanations, including the fact that people participate in social phenomena as actors and that information on actors' purposes and actions is more "available" than information on structural trends. Knowing that this bias is at work in causal attributions, one would be advised to force oneself to consider alternative, more structural, explanations of action outcomes, before attributing them directly to actors' will.

Self-serving biases

A complication in following this prescription may arise when considering that actors may assign positive or negative preferences to

Box 1.10
An experiment on commitment

Assume that you are a loan director at a bank. A customer comes in and asks for a $100,000 loan to start a business. After a thorough analysis of the request, you personally decide to grant the loan. Six months later, the customer comes back and says: "I have some good news and some bad news. The bad news is that the company is facing some problems. As a matter of fact, without any further help we will not be able to pay back the loan. The good news is that I am quite confident that if you lend us another $100,000, we could turn the situation around." Would you grant the loan?

Source: Experiment described in Bazerman (1986).

different causal attributions. Examine the problem posed in Box 1.10.

In this case, it is probable that the decision-maker will not consider the outcome as a personal failure and will renew the loan. For, if the results linked to the action are negative, then the decision-maker has an incentive to attribute the cause to other factors, and continue the action, however risky, rather than accepting certain losses (reputation, position, income) (Staw 1976b).

Antidotes

As appears particularly clear from the last examples considered, the possibility for *effective learning* is linked, first of all, to the presence of an organizational climate where *experimentation and errors are not immediately and readily punished* (for example, because of a mythicization of the principle that "only results count") (Chapter 9). This is a condition for people not washing their hands of mistakes but discussing them (Staw 1980; Popper 1989).

Second, effective learning is linked to the availability of many theories and hypotheses that can feed the processes of falsification and improvement of hypotheses. Examples of these contexts can be, both within and between firms, *multicompetence groups* (Chapter 5; Part III).

Third, it is important to remember to observe and evaluate action results, maintain systematic records on feedback, and not get rid of decisions as soon as they are "made." Obvious as it may seem, it is not rare to find actors who do not learn from experience simply because they have no available attention to pay to it. Firm *performance evaluation systems* do have this important memory function.

Fourth, it is important that actors be endowed with resources – time, attention, and material resources – that exceed those required in performing currently known actions (*slack resources*), in order to allocate some resource to the search for and test of new actions, and to finance the eventual diseconomies (expected income losses) of learning.

Fifth, if the idea that research and learning activities are an important – and ever more important – part of economic action is taken seriously, training in research and learning methods would not be out of place in economic education, along with the more traditional training in decision methods.

PREFERENCE

Human behavior, and economic behavior in particular, can and should be explained by taking into account that actors can be purposeful and can have "motives" (Elster 1985).

Early psychological approaches to motives emphasized a deep natural and emotional level of analysis and explanation of behavior, in terms of instincts. A psychoanalytical approach to motivation falls outside the scope of this book, and is anyway considered "antiquated" and "outdated" by psychologists as well, owing to various methodological limitations: for example, "the disturbing fact that the list of the instincts has continued to grow, reaching a number of approximately six thousand" and the difficulty in testing instincts empirically in a way that does not imply inferring them precisely from the behaviors that they should explain (Bandura 1986: 11).

Early economic theories of motivation suffered from the opposite problem: they had a very "narrow" view of preference, supposing that economic actors are motivated by (1) individual self-interest (rather than, for example, the interests of a community they identify with); and (2) wealth (rather than, for example, the access to a variety of resources) (Smith 1776; Taylor 1947).

Subsequently, in both fields there has been a gradual shift from content-based, substantive models of interests, to content-free and "procedural" notions of utility and purposeful action.

From substantive to procedural models of preference

Early social psychology models of motivation based on the content of "needs" opened up the "portfolio" of preferences, suggesting that people in organized settings assign positive preferences to resources that are not easily valued in monetary terms. On the other hand, they suffer from the same methodological limitation of early psychological perspectives: the unachievable goal of listing all the relevant content of interests. This is the case of the much criticized, but much used typology of needs developed by Maslow (1964). In contrast to an instinct, a "need" is a conscious drive to action, and consists in the perception of a "deficit," or of a "gap" "to be filled." It is supposed that it is the non-satisfaction of a need that drives toward an action that will re-establish equilibrium (Maslow 1953; March and Simon 1958). In other words, a need can be seen as a very particular type of goal: regarding its contents, it is defined over goods and services that by their nature or social convention are perceived to be necessary; regarding its logical form, it is an aspiration level, a target (Chapter 2) since subjects look for a quantity of such goods sufficient to fill the perceived deficit, not the maximum possible quantity.

A first part of Maslow's model arranges needs in five classes based on content:

- physiological needs (such as sleep or hunger)
- security needs (not being threatened and subject to risk)
- needs for belongingness (to be accepted and to have an identity in relation to a reference group)
- needs for esteem (to have confirmation of one's worth)
- self-actualization needs (to realize one's competence and preferences in intrinsically rewarding activities).

A second part of the model hypothesizes that needs are hierarchically ordered and that this order helps to predict behavior: "lower level"

needs must be sufficiently satisfied before "higher level" needs can be perceived in a way strong enough to drive action and orient behavior.

Both parts of the model can be and have been widely criticized. On the one hand, it appears intuitively attractive, and favorable examples can be found to confirm it, especially at levels of relative deprivation and dissatisfaction of primary needs (for example, if one has nothing to eat, all energy will be employed to satisfy that need; only when primary needs are settled, can attention be turned to other objectives). On the other hand, many objections and counter-examples to Maslow's hypothesis can be easily found and they actually led to revisions and modifications of needs theory. As a preparation for a better appreciation of these revised models, it may be useful to try to formulate a list of possible criticisms and counter-examples. An example of such a "list" of objections based on classroom discussion on this question is reported in Box 1.11.

Many subsequent revisions of Maslow's model have reacted to those possible criticisms – leveled against the content of classes as well as the hierarchical hypothesis – by trying to clarify the underlying cognitive processes.

Box 1.11
Counter-examples and objections to the needs hierarchy model

- Needs are culturally determined. It would be necessary to have different models for different cultures and countries and for different periods in history in order to be able to predict and explain behavior.
- Satisfaction levels for each category of needs are extremely different from person to person. It would be necessary to have different need hierarchies for different types of actor in order to explain and predict behavior.
- Rather than a scale, one often finds oneself facing a dilemma: more security, or greater development and self-realization? More esteem and socialization, or more consumption and more individual success? The actions required often conflict.
- When one cannot satisfy a better need – for example, self-realization – other needs are re-evaluated; for example, one invests in sociality.
- Altruistic behavior, "pride" in giving up even primary goods, even by "poor" people, would be inexplicable if the hierarchical hypothesis were true.
- Needs change over the course of one's working life so that a young person has different needs from a person at a later point in the life cycle.
- Needs depend on experience. A person who has had little, expects little. One who is used to receiving a lot, needs a lot.
- The same person, in the same time period, may have different needs when "active" in different activities or tasks.
- Why not the need for power, for fairness, or for novelty and change? Why not simply one need for freedom to choose what material and immaterial goods to look for, and in what measure?

Dual preference structures

One group of important studies tried to make the model compatible with the empirical observation that actors are normally able to perceive higher order needs, even though they do not feel lower order needs are satisfied (Lawler and Suttle 1972). In addition, these researchers were concerned with the potentially unfair consequences of the original formulation in terms of resource allocation (if primary needs are not satisfied, it is useless to offer people benefits that respond to higher level needs, i.e. the less one has, the less one receives). Herzberg and colleagues (Herzberg *et al.* 1959; Herzberg 1966) demonstrated how people are simultaneously sensitive to "lower" and "higher" needs, but interpret them in different ways. On the one hand, the needs for material resources, consumption, security, and affiliation are perceived as "deficits" and create dissatisfaction if left unanswered. On the other hand, the needs for growth, professional development, and self-actualization in work are perceived as "surpluses" and generate positive incentives to act so as to reach high levels of satisfaction, rather than to correct a negative situation. This result was reached through a special interview technique (called "critical incident"), in which people were asked to describe situations in their work life where they felt "particularly satisfied" and "particularly dissatisfied" and to describe in detail the causes and conditions that solicited those perceptions and judgments. The results showed that the factors mentioned as a cause of satisfaction were qualitatively different from those mentioned as a cause of dissatisfaction. Reaching objectives, the recognition of results attained, the content of work, the level of responsibility, and the possibility of promotion and professional advancement were predominantly mentioned as causes of satisfaction. They were therefore grouped in a single category and called *motivator factors*. Organizational procedures, the style of supervision, the quality of interpersonal relations, the physical working environment, compensation, physical conditions, and personal safety were predominantly mentioned as causes of dissatisfaction and were grouped as *hygiene factors*.

The hypothesis that preference structures are "dual," that they are compounded by a "positive" and a "negative" field which are treated differently, finds support in more recent cognitive research. Economic actors have been found to perceive utility in different ways depending on whether consequences are classified as possible losses or possible gains with respect to a situation subjectively believed to be neutral or balanced (Kahneman and Tversky 1979) (see above).

In the light of subsequent research, however, the perception of a particular content or resource as a hygiene or a motivator factor appears to be contingent rather than universal. For example, work satisfaction may vary as a function of age, instruction level, level of experience, and past results (Lawler 1973). Or again, the surprising placement of compensation among hygiene factors, in Herzberg's studies, may well have stemmed from having conducted a test on categories of employees (accountants and engineers) whose pay was predominantly fixed and relatively low; in all likelihood, different results would have been obtained had incentives contingent to performance been important, or had monetary rewards been very high and connected to status and social consideration.

Learned preferences

McClelland's theory of learned needs (1961, 1965, 1987) marked a clear departure from an innate concept of needs. Based on behavioral experiments mainly conducted in the laboratory, it showed that subjects acquire and learn certain needs rather than others, both by inheritance from societal culture and by direct experience. For example, people who realize they are able to develop high level task-oriented competence with ease will acquire a high need for achievement.

These results are broadly consistent with those obtained in cognitive research on preference learning and the perception of utility. Preferences should be explained rather than assumed, and they can be explained as outcomes of learning processes (March 1994). As for any knowledge, actors partly "receive" and partly construct, on the basis of experience, their knowledge of what they prefer. They accept as a matter of social inheritance a variety of basic values on what are "positive" and what are "negative" things (Schein 1985). But value-sharing and identification with given models contribute in defining only a deep and basic level of identity. Most economic actors, individual and collective, are capable of constructing an original identity, not only of identifying with a larger community or with given models. Actors evaluate their direct experience and the experiences of comparable others in order to learn what they can seek for and what they can praise (Bandura 1986).

In conclusion, it can be observed how the more valid and enduring parts of the so-called "content models" of motivation sought their foundations in cognitive learning process models underlying the formation of needs. As was argued for competence content classifications, the possibility of explaining and predicting behavior well on the basis of content typologies seems limited for various reasons including the following:

- the list of contents tends to expand indefinitely
- content typologies say something about what people prefer, and the procedure of "assuming" rather than surveying the content of preferences is of dubious legitimacy
- needs are primary and profound interests, and therefore are rather remote from action: they require to be operationalized in order to guide action and be predictive.

A logical hierarchy of preferences

In fact, more recent approaches to motivation structures favor the analysis of the logical structure of preferences conceived as layers of knowledge about oneself, rather than the listing of possible contents. If a hierarchy of needs exists, it may be the general logical hierarchy that organizes and "operationalizes" an actor's knowledge. According to more recent motivation theory (Locke 1991), needs and values are deep preferences that, even though partly learned in a task-specific way, are internalized as a characteristic of an actor as a whole. In the light of this, the affirmation that they are difficult to change may be justified. Moreover, it can be understood why they are only weakly correlated with specific behavior patterns. For many other passages are necessary in order to move from those primary options to actions, as shown in Figure 1.3. An intermediate judgment is that of attributing weights or values or degrees of utility to categories of desired (or shunned) outcomes, in order to understand what one's own *interests* in given situations are (for example, how one feels about a job that is more self-actualizing but offers

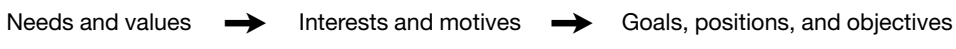


Figure 1.3 A logical hierarchy of preferences
Source: Adapted from Locke (1991).

relatively low monetary rewards). A second kind of judgment is to define specific *positions or goals* that indicate what operative objectives are being pursued (for example, a job position that offers at least a certain level of income, based in one's home town, and not overflowing into weekends).

Among the advantages of this logical or "content-free" framework there are the following:

- the content of utility may be quite different according to the problem at hand, the perception of external constraints and the nature of the game being played, the type of actor considered – individual or collective, acting as a person or in an organizational role
- the framework is compatible with economic ideas of motivation because in economic thinking there has also been an evolution from content-laden views toward content-free and processual views.

Utility as content-free preference

In fact, in economic thinking, the early notion of motivation based on self-interest as profit-seeking (Smith 1776) has been supplanted by notions of motivation based on subjective utility, in which not only the particular content of money and profit but even that of self-interest is not essential. In support of this contention, in a recent revisitation of the concept of rationality, Simon (1997) quotes Marshall's words that what is characteristic of the motivation of economic actions is "a *free choice* by each individual of that line of conduct that, after careful deliberation, seems to him the best suited *for attaining his ends, whether they are selfish or unselfish*."

Hence, in economic and cognitive research, economic actors are seen as utility-seekers, capable of translating into their subjective rankings of alternatives all the relevant outcomes they might value (Raiffa 1968; Simon 1996).

Only in particular cases, in highly structured and competitive situations, the relevant outcomes may be well represented by "profit" or net monetary rewards (Chapters 2 and 3). In most situations system efficiency does not coincide with profit or shareholder value maximization (Radner 1987). In addition, individual actors as well as firms can viably pursue a variety of other objectives, such as: growth (Marris 1964); "slack resources" (time, staff, equipment, money, or other resources that are in excess with respect to current productive needs and can be used at discretion (Williamson 1964, 1970); rents (from a variety of sources of monopolistic positioning) (Barney 1991); productive missions and the utility of final consumers under a constraint of economic viability of the producer (Masini 1978; Coda 1996).

Most economic contributions, in fact, employ the procedural notion of "utility maximization" for constructing models of optimal behavior, given some assumptions on the content of objectives that are relevant in solving a given problem. The content of utility can vary: if the problem is a technical investment, it may be the monetary return on investment; if the problem is the choice of a marketing technique, utility may be substantiated by customer satisfaction; if

the problem is product innovation, utility may be operationalized as the number of patents.

This variety in utility content is not without implications for the feasible processes of decision-making, however. "Utility maximization" is one possibility, which is feasible if utility and information are structured in a particular way. The next chapter discusses that model of decision in detail and a variety of legitimate alternatives to it.

SUMMARY

Chapter 1 presents a structural model of the economic actor, defined as a subject endowed with knowledge and preferences. In the first part of the chapter, some basic features of economic actors' knowledge with important organizational consequences were identified:

- *Knowledge is hierarchically structured*, whereby a core of basic principles and theories are articulated in operational and testable propositions, and a body of "substantive" notions is made operational by a body of "procedural" know-how. This dimension of knowledge governs the *degree of innovativeness* of economic behavior through the capacity and possibility to put into question and change "higher order" layers of knowledge; and to the extent behavior is constrained by detailed operational rules of action.
- *Knowledge is differentiated* into a "*tacit*" component (difficult to explain and communicate) and an *explicit* component (declarable, codifiable). This dimension governs the degree of transferability of the body of knowledge on which a certain class of actions is based, and influences the types of organizational mechanisms which can support the transfer.
- *Knowledge and competence* can possess higher or lower *combinative and generative potential.* This dimension influences the effective path of evolution and growth of knowledge and activities based on it
- *Knowledge* can be more or less complex. Complexity can be a matter of the quantity of inter-related information (*computational complexity*) or a matter of observational and causal ambiguity (*epistemic complexity*). The former component can be handled by investing in documental and information technology support, and by some organizational attributes (decentralization, formalization); the second component can be helped by direct experimentation and hypothesis testing, and by other organizational means (teaming, discretion).
- *Knowledge is always fallible*, and economic knowledge is no exception. In the second part of the chapter, the rich inventory of available tools that can *improve the validity and reliability of observations, judgments, estimates, and inferences under uncertainty* was examined. The principal cognitive biases in problem-formulation (for example, framing effects), in the selective perception of information on alternatives and probabilities (for example, "local knowledge" and "over-/underconfidence" effects), and in learning from experience (for example, self-confirmation and causal attribution errors) were considered and their possible individual and organizational antidotes examined (such as forming groups, using "management systems" which provide checklists of relevant information and alternatives, and avoiding incentive systems which punish "errors").

The third part of the chapter presents elements for the analysis of *preference*. Early

theories attempting to model the content of preference were mentioned, but it is argued that they have been overcome by logical and procedural (learned) conceptualizations of preferences. It is shown that the logical structure of utility judgments – as a form of knowledge of the self – exhibits the same hierarchical stucture of knowledge in general: there are *basic value judgments*, assessments of what the *interests* of an actor and the actor's *motives* for action are in a specified situation, and operational *goals* setting *objectives* to be reached. The more operational and precise utility judgments are, the more they constrain and orient behavior (see Chapters 2 and 7).

Exercise: The palimpsest decider

For some years, channel 1 of the Italian public television company, RAI, has decided to produce and broadcast the program *Carramba che Sorpresa* in the early evening. The new director of RAI1 must decide the palimpsest for the next season. He comes from an important magazine and he wonders what can have persuaded the previous management to confirm this program for three whole years.

Our young director seems to have understood: you choose the program that gives you the best audience. Easy enough – says an elderly director who has survived any amount of restructuring – because RAI is a "generalist" television station and, as such, must succeed in getting mass audiences: "The more people Fra [Francesco Sgobbi, the new director] can catch, the better."

If *Carramba* gets an audience, the problem of decision is thus very simple since the preference function of the decider (to maximize the audience) is circumscribed. Moreover, the result can easily be measured (audience data are provided at intervals of 15 minutes!). In the space of a few weeks at RAI they can tell whether a program is successful, or whether it's a classic flop and they must take the appropriate steps. However, before broadcasting a program, it must be planned, contracts must be drawn up and, often, the production and infrastructures must be set up. All this involves a lot of expense. Thus, as our director immediately learnt, it is necessary to limit as far as possible cancellation of programs as a result of lower than expected audience ratings. On the other hand, it is anything but easy to predict exactly what audience a certain program will get. In general, the deciders base their decision on the intervals: for example, say the experts, *Carramba* might oscillate between 28 percent and 34 percent of audience share. Anyway, at RAI as with other television companies, they are well aware that surprises are all too frequent.

Our new director thinks that you couldn't do better than *Carramba* for a Thursday evening. It's his opinion that "Over these last years *Carramba* has always done well [even with the old director] so why shouldn't it go well for me?" And then, seven million viewers means lots of advertising. This is very important for the RAI as advertising accounts for about half of its annual available revenue (the other half comes from TV licences). Our young director has no doubts: go for *Carramba*.

Convinced of the rightness of his decision, the director of the network is leafing through a newspaper and, in the show pages, he comes across the following headline: "Does RAI1 intend to serve up junk TV once again this year?" According to the journalist: "Aside from the audience ratings, RAI performs, or ought to perform, a public service." "People," he says, "pay the license fee and have the right to a proper service, not just to programs that gratify firms interested merely in advertising." The week before, too, that journal printed an article saying that RAI1 on Thursday evenings "resembled a South American TV station rather than a European one." And

yet, recalls the director, when RAI1 decided to broadcast a program on religions in peak time there was a slump in audience ratings and some newspapers talked of "public money used to finance programs watched by a mere handful of pseudo-intellectuals; when people work, and pay, they want to relax a bit in the evening." In addition, the advertisers were enraged and, brandishing their contracts, they demanded a refund of part of the money they had spent.

The preference functions seem to have become at least two in number and could even be in conflict with one another. The same problem has been rendered all the more critical by the fact that, alas, for a television program there is no clearcut criterion for calculating the level of a public service. Our director thinks: "Audience ratings, advertising, public service, orientation of the watchdog commission of the Chamber of Deputies, and we've yet to take account of the opinions of the media and the TV critics . . ."

Our decider is more and more confused, and he decides that the public service interest is best served if many people make use of the service; so, better *Carramba* than the philosophy lecture or the history program that almost nobody would watch. And yet, when he discusses the matter with the head of "educational" programs, he is told: "Look, if a philosophy lecture in peak evening time scores a resounding success, that means that you've stimulated a million ordinary Italians to do a bit of thinking." This sounds convincing enough, but the director who called him "a Fra," and who seemed to know his onions, had taught him that a generalist television must get as high ratings as possible, so even if the philosophy lecture attracts a million viewers, it's no good because the competition from a swimming contest or a comedian turns that million into ten!

What with all these pressures, plus his own confusion, our decider has opted to shoulder his own responsibilities and, thinking only of the audience, he states his decision to the palimpsest meeting to confirm *Carramba*. Unexpectedly, during this meeting all hell breaks out.

The director of RAI's channel 3 says: "Come on now, this is the only evening when I try with my program to reach the target of elderly and middle-aged viewers and you stick *Carramba* in my way. Very good, is this war then?" The director of RAI's channel 2 echoes him: "We're in the same firm, so don't toy with the idea of shifting *Carramba* to Fridays, that's our most important evening, otherwise that'll be the end of everything."

Our decider thinks to himself: "Why in heaven's name did I ever take on this job!?"

By Giuseppe Soda

Questions

- Who are the relevant actors in the matter?
- What type of knowledge can they act upon?
- How can they define their objectives and preferences, in content and form?
- What are the likely consequences, for the decision at hand, of different ways of framing the problem?

Appendix 1.1A

A matter of life and death

A dangerous disease is spreading throughout the region where you are the Assessor of Public Health. It is estimated that 600 people are risking their life in the course of the year.

You can choose between two possible plans for intervention:

- Under Plan A, 400 lives would be lost.
- Under Plan B, with a probability of 1 out of 3, no one would die; with a probability of 2 out of 3, all 600 would be lost.

Which plan would you pick?

Appendix 1.1B

A matter of life and death

A dangerous disease is spreading throughout the region where you are the Assessor of Public Health. It is estimated that 600 people are risking their lives in the course of the year.

You can choose between two possible plans for intervention:

- Under Plan A, 200 people of those exposed to the risk will surely be saved.
- Under Plan B, with a probability of 1 out of 3,600 people will be saved; with a probability of 2 out of 3, no one would be saved.

Which plan would you pick?

Appendix 1.2

Results of the experiment on "availability"

The table in Appendix 1.2 shows for each couple of causes of death the average response, the real frequency, and the frequency with which two typical daily newspapers published episodes regarding the various causes over a period of a year.

Cause of death	*Responses for each couple (%)*	*Total no. of cases in the USA (in thousands)*	*Episodes reported by newspapers per year*
Lung cancer	43	140	3
Car accidents	57	46	127
Emphysema	45	22	1
Homicides	55	19	264
Tuberculosis	23	4	0
Fires	77	7	24

Source: Russo and Schoemaker (1989).

Chapter 2

Decision and Motivation

Knowledge and interests are to be "processed" for leading to action. This can be done in different ways. Beyond all improvements in hypotheses, conjectures, and initial knowledge, problems differ as to the incompleteness of ex ante knowledge, or degree of "uncertainty" about relevant objectives, relevant alternatives, outcome probabilities, or even the very "data observed" and therefore different decision behaviors are comparatively more effective in solving them.

The first section of the chapter shows how decision-making research leads us to conceive an actor capable of adopting different decision strategies as a function of the state of knowledge on decision matters.

The second section of the chapter shows how the available models of motivation processes repropose substantially the same differentiation between modes of taking decisions about economic behaviors, with specific reference to work behaviors. This homology helps in comparing different motivation processes in terms of applicabilty and effectiveness conditions – a prescriptive and comparative approach currently undeveloped in motivation studies.

The third section uses justice and equity theories to show how decision and motivation about economic behaviors are, and need to be, driven not only by effectiveness and efficiency criteria but also by equity criteria. In all these three sections and fields, the three basic forms of value maximizing, heuristic, and non-calculative automatic rationality are found to represent the fundamental logical strategies available, outlined in Figure 2.1 along with the type of information required for their application, illustrated throughout the chapter.

DECISION STRATEGIES

Rationality can assume different configurations, that, in economic behavior, materialize in different decision processes or strategies. These decision strategies are formulated here in a way conducive to comparative evaluation according to three criteria:

- to what extent are they able to link actions – and the results those actions are expected to produce – to the preferences and objectives of decision-makers (effectiveness)
- to what extent do they economize on the scarce resource of cognitive capacity and effort (efficiency)
- to what extent are they able to resolve conflicts between different actors with different objectives using that strategy (conflict resolution capacity).

The main decision-making models that have

CONFIGURATION OF KNOWLEDGE AND PREFERENCE	DECISION LOGIC
• Structured problems • Complete and clear preference ordering • Unrepeated action	• Value maximizing
• Unstructured problems • Incomplete preference orderings • Unrepeated action	• Heuristic problem solving
• Unknown cause–effect relations • Unclear preferences • Repeated action	• Non-calculative appropriateness

Figure 2.1 Types of decision and motivation processes

been identified in economics, organization, and management can be described as particular, salient, and effective *combinations or configurations of rules and procedures for defining and modifying the fundamental decision inputs: procedures for defining objectives, for generating and evaluating alternatives, and for learning from experience*. In other words, one decision model or strategy differs from another if it is characterized by a different approach to any of these fundamental cognitive activities. The initial information conditions that make these diverse approaches or strategies applicable can be and will be specified.

On the basis of the vast research on economic decision behavior, three basic models can be reconstructed:

- models of deductive rationality based on "optimization" rules
- models of "heuristic" rationality based on acceptability rules
- models of "programmed" rationality based on automatic action.

They can be evaluated according to their capacity to link results to objectives in problems characterized by different levels of information complexity and by different degrees of conflict among interests.[1]

Deductive rationality and optimizing strategies

Consider the following problems:

- Find the sequence of operations a production department must perform to reach a desired level of output with the lowest costs and in the shortest time.
- Given a finite series of stocks and bonds with known interest rates, and an investment budget, define the best portfolio in terms of expected return on investment.
- Calculate the optimal production level of a standard good, given its market-determined price, knowing production costs.

Known objectives and alternatives

These problems are defined in a highly structured way. Interests and objectives are known, clear, and measurable. Alternatives are finite and defined; problem boundaries are defined by the specification of these relevant alternatives. Within these boundaries, an optimization strategy is not only feasible but superior to other strategies because it evaluates all relevant information for finding the best action in terms of the actor's utility.

An optimizing or value-maximizing

strategy can be defined as a combination of particular rules of search and choice. These rules prescribe the decision-maker to examine alternatives until the marginal returns from search are positive and, among the generated alternatives, to pick up those producing the maximum net utility (benefits minus costs). The rules do not require costs and benefits to be expressed in monetary terms, nor do they require that utility is measured "cardinally" (assigning numbers). Utility can be defined "ordinally," expressing a comparative judgment of superiority or inferiority in net benefits (Savage 1954). The decision-maker must at least be able to order alternatives according to preference.

Most times, the task is not easy. If one thinks even of a simple problem, such as buying a car, only where the quality of cars is perfectly standard and price covers all relevant information (say within a power category) can the buyer select the lowest priced car within the category. Almost no one buys a car in this manner, however. For cars today are now differentiated products, whose qualities and images are not perfectly comparable. In addition, people think that many evaluation parameters are significant. Many of them are subjective and not easily measurable. Space, speed, and security may matter, but the status the car confers and the association with the owner's personality may matter even more. How can we build a function which associates each alternative with an overall level of utility? In principle, each attribute can be ranked on a utility scale (as people are sometimes requested to do in questionnaires). Furthermore, it is necessary to specify what "weight" each attribute has in an actor's utility function. Suppose that the decision-maker is able to define these weighted utilities, and that the decision-maker can do so for "all relevant alternatives," assuming that the relevant alternatives are the cars of a certain class.

Further complications may arise. Suppose that used cars are also considered. In this case, the car's performance will depend not only on the decision-maker's choice, but also on external factors: how diligent the mechanic was in overhauling the car, what replacement parts were used, what the real state of those parts that cannot easily be checked, such as electrical components, will turn out to be. These two elements can be represented as in Figure 2.2. U (p_1 e_1) is the utility of an alternative expressed as a function of its sure attributes and of the possible states of uncertain attributes or of external factors that will influence actual performance (Feldman and Kanter 1965).

If this is the situation, applying a fully fledged, value-maximizing strategy would require one to specify with what probability the outcomes would assume different values – for example, what is the probability that the car will be a "lemon," a good bargain, or a gem in its class?

Multiplying utilities by the respective probability, a new single indicator to be maximized would be obtained: subjective expected utility (SEU). Many economic decision models currently in use apply this particularly sophisticated optimization rule, which allows

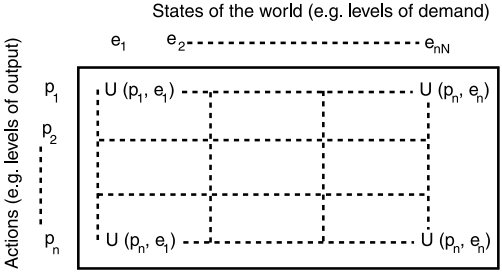


Figure 2.2 Expected utility matrix
Source: Feldman and Kanter (1965).

this strategy of decision also to be applied under conditions of *risk* – consequences are uncertain with a known probability – and even under a certain type of *limited uncertainty* – probabilities are not known but the relevant alternatives and the observable consequences are known. This may be totally appropriate in sufficiently structured problems, where decision parameters are known and quantifiable. Even though not all decision-makers actually decide on structured problems by value maximizing, those who do so can be expected to obtain superior results.

Limits of value maximization

It has been objected that a SEU model does not describe particularly well how people behave, or at least how they behave when facing unstructured problems. The core issue, however, is not that in practice people behave differently on average (perhaps mistakenly) but that it is doubtful whether the model prescribes well how they *should behave* in unstructured problems – or, in other words, whether it describes well the "best processes" under certain conditions. Why?

Optimization fails if uncertainty involves the definition of the relevant alternatives and consequences, rather than just the probabilities of outcomes, i.e. if the problem is unstructured, as in the case illustrated in Box 2.1 (see also the Palimpsest decider case).

Search costs and information losses

The usual defense of optimization rules for all seasons observes that, if information is difficult and costly to find and process, one should take those *search costs* into account in cost and benefit calculations: search should be stopped when its expected marginal costs are superior to its expected marginal benefits. This rule, however, can in turn be applied only when search is very structured and its costs and benefits can be assessed (such as in looking for a new spare part to repair a given machine). In more complex situations (such as in designing a new machine), those cost and benefit calculations would be even more difficult than the evaluation of alternatives. There is *a problem of knowledge (rather than just costs) which can bring an optimizing strategy to failure*. Compelling a quantification or semi-quantification of objectives and information on a utility scale or function can lead to *information losses*, at an increasing rate with the complexity of information to be processed. Utilities and expectations are not measured for choosing partners, mates, and not even for choosing cars, jobs, or new projects or products, because they are multidimensional entities, with aspects that are difficult to appraise and are highly incomparable, and because the relevant alternatives are potentially infinite. As a result, even though in principle it is possible to construct such measures and judgments, the *reliability and validity* of preference assessments and utility judgments such as those required by the SEU model would be extremely low in those situations.

Wealth effects

In addition to availability of information, a far-reaching condition for an SEU model to be applicable is availability of resources. Even if one knew how to assess the probability that a deal would be good or a "lemon," one might not have the resources to afford the failure. Resource scarcity easily leads to risk aversion. Whenever the level of wealth of a decision-maker influences the deals the decision-maker is willing to accept, it is said that there are

Box 2.1
Failure of a hyper-rational system

Let's take into consideration a very specific city service: fire fighting. Its goal is to reduce losses caused by fires and its results will be calculated based on these losses.

The number of losses caused by fires is calculated on the basis of several indicators. Among these there are natural ones (wind, heavy snowfall, hard winters, hot and dry summers, thunderstorms, hurricanes, earthquakes, and floods), structural and environmental factors (buildings, population density, and types of construction and rooftops), factors related to moral character (negligence and arson), and lastly either the efficiency or inefficiency of the fire department. Losses therefore depend on all of these variables including the work standard of the fire department. The fire chief needs to be aware of how the activities of the fire department can affect the losses whenever the fire chief is in a position to make a decision.

How does the fire department carry out its task? It performs inspections of various buildings to reduce the risk of fires, it promotes education campaigns against negligence, it fights fires, it trains firemen, and it will carry out research whenever there are arson charges.

But we could analyze this further. How is the battle against fire carried out? The necessary equipment has to be driven to the site, the hoses have to be ready, the water needs to be pumped and targeted toward the flames, the ladders need to be set in such a way so as not to be damaged by water. Again, each of these activities could be further analyzed. What does it take to unroll a hose? The rubber hose needs to be purchased and maintained, the equipment to carry it also needs to be purchased and maintained. Firemen need to be hired and trained. Firemen also have to spend time and energy in unrolling the hose.

One could analyze each of these factors involved in carrying out one of the tasks described above.

Efficiency can be measured by calculating the cost of each single factor involved in the task as well as the contribution made by that element to meeting the goal of the department. Whenever these costs and contributions are clearly known, the elements of the process can be combined so as to reach the maximum possible reduction of losses caused by fires.

The complexity of this decisional process is clearly such that even if it were possible to standardize it, its full capacity would be subject to unforeseeable events. And, in the meantime, most probably the city would burn down.

Source: Simon (1947).

wealth effects. An "expected value maximizer" is supposed to be risk-neutral and not subject to wealth effects (Raiffa 1968; Milgrom and Roberts 1992).

Conflict resolution limitations

And what if there were multiple conflicting actors and objectives? If these objectives were

measurable and comparable, it would be possible to make trade-offs, to build "indifference curves" between objectives (combinations of valued resources bringing the same level of utility to the actor), and to find solutions that maximize utility.

If objectives are not directly comparable – for example, because they belong to different actors – the application of value-maximizing rules becomes more difficult, but not impossible. Consider, for example, some alternative ways of restructuring a plant, evaluated in terms of implications for employment and for structure costs: the lower the generated unemployment and the greater the reduction in costs, the better the solutions, as represented in Figure 2.3.

Figure 2.3 Comparing alternatives without comparing objectives

If comparison of the two parameters is believed to be difficult, it becomes problematic to establish a preference order among alternatives. It would be possible to say that D is preferable to A and B, and that E is preferable to B, but it would be necessary to establish more complex rules in order to compare D, E, and C with each other. They are all, in fact, *Pareto-optimal solutions*, i.e. they are such that no improvement with respect to them is possible for both players simultaneously (or for one with equal benefits for the other). To compare multiple Pareto-optimal solutions among themselves, some additional criterion for maximizing joint utility is needed, such as an equity criterion.

Nevertheless, there are particularly conflicting configurations of interests that cannot be resolved within a value-maximizing scheme.

For example, preferences may be patterned in a cyclical way so that no one of them is preferred by a majority of players, as in the following case (a case in the "voting paradox"). Suppose that there are three alternatives, A, B, C, and three players, 1, 2, 3 and that the preference order for actor 1 is A over B over C, for 2 is B over C over A, and for 3 is C over A over B. Then a majority prefers A to B, a majority prefers B to C, and a majority prefers C to A.

There are also other types of "encounters" between "optimizing players" that are not solvable, owing to the presence of particularly strong conflict among interests. For example, in "Prisoner's Dilemma" games (see Figure 2.4), two optimizing players who have no

		B	
		Cooperate	Compete
A	Cooperate	Good, Good	Awful, Optimal
	Compete	Opitmal, Awful	Bad, Bad

Results for A before the commas, for B after the commas

Figure 2.4 Ordinal structure of the generic pay-off matrix of a Prisoner's Dilemma

chance to play repeatedly and to acquire specific information about a partner's behavior, typically end up in the dominated (non-Pareto-optimal) outcome brought about by the combination of the two competitive strategies. In fact, for both players, the choice to "compete" – which can result in an "optimal" or "bad" result – is superior to the choice of "cooperate" – which can result in a "good" or "awful"result.

It can therefore be remarked that the use of optimizing strategies by multiple actors may contribute to creating situations of unresolvable conflict. It could then be argued that in those situations a shift to non-optimizing strategies may be advisable and valuable, even though objectives and alternatives are clear.

Heuristic rationality and "satisficing" strategies

Consider the following problems:

- Which new product is it better to develop and launch?
- Which person is it better to hire as a marketing manager?
- Which public education program is it better to finance?

Objectives and alternatives as "hypotheses"

These are unstructured problems. Relevant objectives are numerous and not easily measurable, and it is wise to test them during the process. Alternatives must be sought. Problem boundaries are potentially infinite, in terms of series of sets of potentially relevant alternatives. For example, if a car acquisition problem is redefined as a broader time-saving and life-management problem, "the sets of relevant alternatives" – hence problem boundaries – become very numerous, potentially infinite, and cannot "all" be considered simultaneously. For example, instead of choosing the vehicle with maximum utility, one could "evoke a different set of alternatives" (March and Simon 1958) and discover that it is better to buy a house close to work and no car at all, or to change one's job.

In these circumstances, the problem cannot be effectively and efficiently addressed by using an optimizing strategy. The best alternative decision strategy is a heuristic one, based on research. It is the best among other possible strategies (described below) in terms of its ability to link results to objectives (hence to fulfill actors' preferences).

A heuristic strategy formulates objectives as hypotheses to be empirically checked with regard to the existence of acceptable alternatives. Decision-makers can and should be willing to modify all decision inputs in the course of the process, thereby improving their hypotheses. Faced with evidence that falsifies or contradicts their expectations, decision-makers can modify cause-and-effect theories, objectives, the sets of considered alternatives, or the tools for information gathering. For example, not finding a candidate possessing the hypothesized profile for a managerial position, the decision-maker can: change the evaluation parameters (add or eliminate some, lower their level); look elsewhere for alternatives (for example, abroad or in other industry sectors if this has not already been done); change the test being used to measure candidates' characteristics (the possibility of measurement errors is often neglected, but always present).

This decision strategy can be defined as "heuristic" (Grandori 1984) because its basic rules consist of research methods and procedures (Lakatos 1970; Newell and Simon 1972; Kahneman *et al.* 1982).[2] It is able to treat and solve problems of very high information complexity, while maintaining a

rather strong tie between actions and objectives. In fact, each potential action is evaluated in terms of the acceptability of expected results with respect to those aspired to by the decision-maker.

Research activities may concentrate on different phases of a decision process and be more or less intensive, as illustrated below.

Research strategies

An important distinction among research strategies regards the extent to which search is innovative and problem boundaries are allowed to change, or not. Research and learning have often been distingushed as "normal" or "revolutionary"; "incremental" or "radical"; "local" or "not local"; "exploitative" or "exploratory" (Kuhn 1962; Cyert and March, 1963; Argyris and Schön 1978; Tuschman and Romanelli 1985; March 1992). These dichotomies underline the difference between a limited research that focuses only on alternatives of the same kind, which look "close" to known solutions; as against research which calls into question cause-and-effect theories, the content of utility, and "paradigmatic knowledge."

Box 2.2
Incremental and radical learning

An example of alternating forms of incremental and radical learning is given by the evolution of car models produced by Ford, and specifically the improvement between the model T and the model A.

The model T represented the first example of a mass-produced car at a reasonable cost and contributed to Ford's great success. The innovation linked to the success of the model T represented a successful radical learning experiment: the goal of the car producer was to move from traditional scale production to mass production. This intuition was the basis of the thorough study of the industrial transformation process that later created an incremental learning track based on the steps highlighted by the scientific management school. Learning stimulated by this innovation gave birth to several other process innovations including the almost forced choice to standardize single parts of the car (for example, nuts, bolts). Ford's original intuition and the learning process that followed rapidly allowed the producer to save on costs and develop mass car production.

The introduction of the later model A followed the thinking that enabled the development of the model T, i.e. standardization and cost reduction without taking into consideration, however, the constantly changing needs and tastes of the consumers who had begun to appreciate the car not only for its usefulness and relative cost, but also for its immaterial characteristics such as its color or its optionals. Therefore, the model A, although based on the model T's success, was a failure and was overtaken by the competition from General Motors' cars that, while mass-produced like Ford's, were nevertheless supplied in different colors. General Motors' success marked the beginning of a new radical learning process.

Source: By Luca Solari.

The case described in Box 2.2 shows the importance of this difference in an economic decision-making process. If the problem is stable and important consequences can be predicted in advance, analysis of the problem as far as possible before acting is likely to be an effective heuristics. Examples are strategic problems in which there is time for analysis, but once action is taken long-standing and poorly reversible effects are generated.

In other problems this research strategy is not effective, because much relevant knowledge can be generated only by initiating the process of analyzing and comparing alternatives; even by experimenting with them, action can be separated into "small" steps, and consequences are not irreversible. Human resource management decisions, for example, are often of this kind. Personnel search and hiring are especially illustrative of decision processes that should be started if relevant information is to be generated on what is "easy" or "difficult" to find at a certain time. These experiences reorient the search for alternatives and allow objectives to be readjusted: if it is difficult to find the hypothesized alternatives, the objectives will be reduced or qualitatively adapted to what can be found; if search is easy, they can be raised and become more ambitious (Simon 1955a).

Ex post learning

Ex post learning may be the only resort if ex ante hypotheses-making is very hazardous. In decisions regarding people, this is often the case with promotion to new and complex jobs, because the reliability of ex ante information on the relevant characteristics of people is fairly low and the hypotheses on the relationships between these characteristics and performances are even more tentative. More generally, it is typically the case of undertakings of a new kind, from which it is important to draw significant positive results, but it is not very clear which classes of result will materialize. For example, a frequently encountered problem in the evaluation of public administration social programs is that they often prove to be ineffective in attaining previously stated objectives. Only if new objective dimensions are created ex post, on the basis of observed consequences, can many programs receive the deserved evaluation (Chen and Rossi 1981).

Research is not only a very costly activity but is also intrinsically unbounded. Research should be "stopped," but when to stop is not a trivial question: in particular, it is not only a matter of costs but also a matter of validity of knowledge. In a structured, narrowly defined problem (such as the search for a needle in a haystack), the search may be truncated according to the optimizing principle of marginal expected return from the effort. But in unstructured problems (such as the search for ways of improving product quality or of increasing product innovation capacity), the main issue is whether a reliable and valid model of the problem has been produced. Acceptability rules play a fundamental role in this judgment.

Acceptability rules

In heuristic decisions, any hypothesis is judged in terms of its acceptability. As applied to alternatives, the processes of search are "truncated" by comparing the attributes of found alternatives with the actor's hypotheses on acceptable attributes (or "*aspiration levels*").

As a rule of choice, acceptability poses lower information requirements than an optimizing rule. A precise evaluation of the expected benefits of an alternative is not

required; a comparison between the alternatives is not required; and, above all, a trade-off between objectives is not required. Only ordinal comparisons between the levels of aspiration and the levels of expected pay-off from each single alternative are necessary. For example, if two objectives are considered, the logical structure of the acceptability judgment can be represented as in Figure 2. 5.[3]

For example, if one were evaluating candidates for a managerial position and the considered objectives were the expected sales performance and the potential for professional development, the necessary judgment would only be an evaluation of whether the performances expected from the candidate profile meet or surpass the levels deemed acceptable.

Quasi-resolution of conflict

Acceptability rules are also able to resolve difficult conflicts and incomparabilities between objectives, even when they refer to different actors, in a more efficient and often more effective way than optimizing rules can. Indeed, decision-makers should evaluate whether there are alternatives that satisfy a series of aspiration levels or constraints, rather than identify the alternatives that maximize a joint utility function reflecting all objectives. Naturally, conflicts between objectives are solved in a weaker manner; and this has in fact been called a "quasi-resolution of conflict" (Cyert and March 1963).

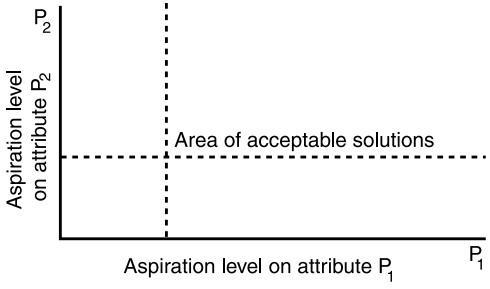


Figure 2.5 Structure of acceptability judgments
Source: Simon (1955a).

In concluding this assessment of the properties of heuristic and satisficing strategies it should be mentioned that economists have observed that "satisficing" is a form of optimizing in which search costs have been factored in, and that a repeated application of acceptability rules and aspiration level adjustments will slowly discover "all relevant information" and "converge" to optimizing, i.e. it will produce similar decisions (actions) (Baumol and Quandt 1964). This can be demonstrated for structured problems. But in complex problems, experience and learning may lead to the definition of a more comprehensive set of objectives, to improved research heuristics, to a greater ability to predict effects, and to solutions that are superior to those previously generated – but not "optimal" (and even by chance if they were so, nobody could know it).

Non-calculative rationality and the logic of appropriateness

A large area of economic behavior is guided by non-calculative decision logic, which does not imply a forecast of costs and benefits and not even the definition of articulated preferences. Actions are taken by matching the observed situation to the "appropriate" action for the acting system (Nelson and Winter 1982; March 1994). This type of "programmed rationality" involves at least three judgments, which may be more or less complicated and subjective:

- A pattern recognition judgment – What kind of situation is this? What is the state of the relevant "world?"

- A self-recognition judgment – What kind of actor am I? What is the state of the deciding system?
- A rule of correspondence between (1) the match of the two former elements and (2) a set of "appropriate" actions: what a certain type of actor, in given circumstances, "has to do."

The above judgments can range from the very simple and mechanistic application of a repertoire – such as the "right way" to fill out an administrative form – to the very complex and ambiguous assessment of "identity" and of "correct conduct" – such as the "right" type of intervention of a personnel director in a work dispute. As in the case of heuristic and optimizing decisions, various sub-models of programmed rationality have been identified. Two particular configurations that have been shown to be diffused in important decision fields and have been well specified in their cognitive components are the "incremental" and "cybernetic" models.

Incremental models

Consider the following descriptions of decision processes mainly found in economic activity in particular settings, such as public administration, universities, and research:

- The best predictor of a year's budget for a program or activity in local government is the budget of the preceeding year (Wildavsky 1964).
- How are sales budgets formulated? See if demand is expanding or contracting and modify the current budget by a small percentage in that direction.
- When a new company is acquired, especially in a foreign country and in a different industry sector, organizational changes drawn up by the parent company run the risk of "breaking the toy." If it is known that the current arrangement works, but it is not known exactly why, it is a good idea to change little and slowly.

The above examples illustrate decision processes that have been defined as "incremental" (Lindblom 1959; Davis *et al.* 1974; Quinn 1980). The most interesting property of an incremental strategy is that it generates sensible action "without" clear objectives. The decision-maker is assumed to have only rather rudimentary experience and preferences concerning the problem: either because the problem itself is too vast and complicated (as in collective political action) or because the decision-maker does not have sufficient experience.

The decision system is able to discern where it wants to go: increase rather than reduce production capacity, improve levels of education, increase well-being. However, no reliable theories are available on the cause–effect relationships that regulate how the action system works and changes. A reasonable decision strategy in similar circumstances is to try "incremental" solutions that differ marginally from those in use (Lindblom 1959).

A rational justification of this rule could be that, since they entail small variations, incremental alternatives also create small consequences and, in particular, "small risks" defined, as decision-makers often do, as small possible negative consequences. In this interpretation, the incremental decision process is a particular variation on a "satisficing" strategy – for example, accept only alternatives with small consequences or "bland alternatives" (March and Simon 1958) – or even a particular case of an optimizing strategy – for example, choose alternatives that minimize risk. Cyert *et al.* (1978) have suggested that in a new area of investment, where it is difficult

to assign probabilities to different levels of return on investment, an incremental approach is advisable, when investments are divisible, so that at each step it is possible to observe consequences and to update the probability judgments to be used in the next step. But, as in the case of interpreting satisficing as a particular case of optimizing, it is a reductive interpretation in that it only grasps a problem of information costs and not the problem of the limits of knowledge. An incremental rule can be applied *without* any calculation or prediction of results, not to mention calculations of risk and process costs; hence it is applicable where these calculations are unfeasible altogether, and both optimizing and heuristic strategies are too demanding.

Only if the accumulation of knowledge is based on addressing repeatedly "the same" problem can the problem itself become more structured and the incremental decision process become a more aware, satisficing (or optimizing) kind of process (Padgett 1980).

Linear rules of choice

In other areas of action, it is the type of conflict and the incompatibility between diverse objectives that is difficult to resolve with more comprehensive decision strategies. Is it better to allocate resources to research on cancer or on cardiovascular disease? Is it better to expand the faculty of engineering or that of economics? In these problems, experience probably does not lead to a greater clarification of objectives, which remain essentially elusive, conflictual, and incompatible. As a result, it is not objectives and their consequences which are learned, but decision rules for dividing and allocating resources to different uses. For example, what establishes the share of resource allocated to different university departments is often a simple linear rule such as that of maintaining the relative shares constant as the budget grows. The result is that current actions depend simply and linearly on past actions (Davis *et al.* 1974).

Tacit objectives

This property of incremental decision rules makes them particularly efficient in resolving conflicts between many, incompatible, or unclear objectives. As Lindblom (1959) noted, an incremental logic does not require that alternatives be evaluated with respect to multiple conflicting objectives. In reality, in an incremental process, objectives come into the game relatively little, and consequently there is little risk of the process being blocked by their incompatibility or vagueness.

The limits of incrementalism

The weakest point of incremental rationality is that *incremental actions do not always produce incremental results* (Padgett 1980). It is quite possible that small variations cause a reaction chain that leads to "major variations" (Weick 1979a). To take a simple example, think of the stability of a boat whose exact carrying capacity is not known when another person gets on: the effect will be incremental if the limit is far off, but could be disastrous if the boat is close to the limit. Similarly, in economic action, a manufacturer's decision to join other producers in a new industrial sector (an incremental variation) will have an incremental effect if saturation is far off, but the closer the sector is to its maximum "carrying capacity," the larger the effect on firms' death rates (Hannan and Freeman 1989).

Cybernetic models

An incremental strategy does not require knowledge and agreement on objectives but does require the capacity to formulate judgments of "similarity" and "marginal difference" between new possibilities for action and current actions, as well as the ability to generate (find, define) such alternatives or increments.

A "cybernetic" strategy (Steinbruner 1974) envisages a fully automatic model of decision. In fact, it can be used not only by human beings, but could be followed by simple machines such as thermostats or by animals (Weiner 1948; Ashby 1952).

A cybernetic decision strategy implies only the following kinds of judgment: the capacity to recognize situations (a certain temperature; a configuration of costs); the capacity to recognize performance gaps with respect to a standard (works/ does not work; positive/ negative); possessing a repertory of possible actions that are applicable to eliminate the gap or respond to the situation as in the example in Box 2.3.

As Steinbruner (1974) notes "in the cybernetic paradigm, values are articulated at a minimal level . . . Therefore, the cybernetic decision-making criterion is not the highest value nor an approximation of it. Rather, the essential criterion is simply survival." Soldiers knew how to identify a "non-working" state from a "working" state of a jeep. This is what defines the problem. Alternatives are not evaluated ex ante, they are simply applied or tested sequentially, in as far as they belong to a repertory of potential solutions. As a result, cybernetic processes can be applied in extremely uncertain circumstances; but it is necessary to be aware that they handle uncertainty by avoiding it, to a much greater degree than do other decision models.

> Cybernetic mechanisms that are able to handle uncertainty do so by focusing on a few entry variables and completely eliminating all serious calculations of probable results. It is assumed that the decision-maker has a smaller set of "answers" and decision rules that predetermine the course of action . . . that are of the nature of "recipes" established by preceding experience.
>
> (Steinbruner 1974: 66)

Box 2.3
A cybernetic decision process

A simple illustration can be taken from informal observations of an army maintenance unit staffed with men who knew next to nothing about the vehicles they were charged with repairing, and hence were decision-makers under uncertainty. They responded with a cybernetic decision process. Faced with a broken-down jeep, they replaced the battery and tested the jeep to see if it then ran. If that did not work, they would change the spark plugs and test again, then the distributor, then the carburettor. If all these actions failed, they declared the jeep inoperable and junked it. They proceeded thus for a substantial period of time, with the order of the sequence of actions reflecting roughly (by the principle of reinforcement) the frequency with which each action proved successful.

The men never did develop more elaborate causal understanding of the operations of jeeps or internal combustion engines.

Source: Steinbruner (1974).

The principle of reinforcement and the rule of imitation

Cybernetic learning is driven by the "principle of reinforcement." Actions that produce positive effects (a car starts up again; in behavioral experiments on animals, food arrives; a tentative and exploratory proposal or attitude elicits praise or rewards) are held in the memory as correct and are repeated (imitated) on subsequent occasions – they become recipes and routines. Those actions that produce negative effects are discarded for solving the problem at hand and lose importance in the repertory to the point where they are no longer repeated even on subsequent occasions. This principle allows a certain degree of dynamics and learning, fundamentally based on the *imitation of successful solutions*, without ever understanding why success is achieved. An imitative choice rule only requires that situations be recognized as belonging to certain classes (pattern recognition) and that a particular repertory of actions corresponds to that class of situations (matching rule). This principle can be applied to one's own direct experience, or, vicariously, to the experience of other actors believed to be similar (Cyert and March 1963; Bandura 1986). Rules of choice such as: "do as the most successful firm in the industry does," "do that which was done successfully in the past," "do as the average actor similar to us does under similar conditions," are imitative choice rules. Surprisingly or not, important processes in economic innovation, especially the diffusion of observable and codified technological as well as organizational innovations, can be predicted fairly well as simple imitative processes (Rogers 1962; Hannan and Freeman 1977; Aldrich 1979; Teece 1980b). Imitation is all the more relevant for explaining many facets of individual economic behavior as work choices (Lomi 1997).

Homeostatic processes

A *stationary version* of cybernetic behavior is also possible and not so rare in economic action.

If a deciding system obeys the principle of maintaining a stable state, the result is a stationary version of a cybernetic model (Ashby 1952; Beer 1972). Its logical structure is represented in Figure 2.6. Not only repeated structured decisions such as inventory management are taken on the basis of these types of programs. As will be seen in the next chapter, even elusive phenomena such as motivational processes can be modeled, in some respects, as homeostatic cybernetic processes.

Assets and liabilities of cybernetic models

The great advantage of cybernetic decision-making lies in the economies of cognitive energy and in the applicability of cybernetic

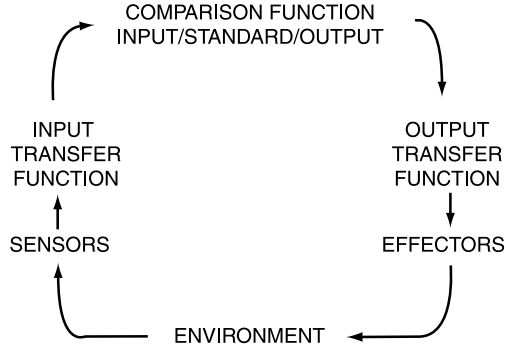


Figure 2.6 A homeostatic cybernetic model
Source: Adapted from Ashby (1952), Beer (1972), and Klein (1989).

rules even when the decision-maker possesses very little relevant information with respect to the complexity of the problem. Causal explanations of why tried and imitated actions should lead or have led to positive results could sometimes be reconstructed, if looked for. Nonetheless, a cybernetic logic is interestingly different and sometime advantageous with respect to other decision strategies, precisely by not looking for these explanations in problems where it would be too costly or too difficult to do so. On the other hand, the effectiveness of cybernetic strategies is limited to situations that repeat themselves over time and across subjects with similar characteristics and in which actions are easily reversible – so that learning by doing is not destructive.

MOTIVATION PROCESSES

The various models of motivation processes that have been proposed can be presented and understood as different models of decision processes. Although they have been elaborated largely independently of decision-process research, and have been applied mainly to work activities, their underlying logic repropounds the basic types of cognitive processes that seem to characterize rationality in general. Hence, the comparative assessment framework developed for evaluating decision strategies can be applied to motivation models as well.

Maximizing expected valence

This perspective is commonly defined as "expectancy theory," and the best known of these models is that developed by Vroom (1964). "Expectancy" is an actor's subjective probability assessment concerning the chances of obtaining certain results, through the allocation of the actor's efforts and competences to certain tasks or activities. "Valence" is the subjective utility, the value assigned by the actor to results. Vroom suggested measuring valence on a utility scale ranging from +1 (maximum utility, value of the best outcome, at a given time in a given place – for example, for a white-collar worker becoming a manager) to –1 (the consequence of maximum disutility, such as the loss of one's job). An alternative approach to measuring valence could be to quantify the monetary value of results (costs/benefits), providing measurement is not too difficult. More generally, subsequent work in the area has indicated that, if the model is to be reasonably predictive, it should include separate evaluations not only of the extrinsic valence linked to results (the value of monetary and non-monetary rewards), but also of the intrinsic valence that can be generated by performance (sense of competence, interest in the activity), as well as of the negative valence that can be a consequence of effort (House and Wohba 1972).

As regards the operationalization and measurement of expectancy, Vroom distinguished judgments on the probability of producing a given performance by applying certain efforts and competences (effort–performance expectancy) from judgments on the probability that certain results will be produced by that performance (performance–outcome expectancy). The first type of evaluation includes a self-efficacy judgment and a weighting of the incidence of external factors on performance, while the second evaluation concerns the probability of obtaining rewards.

Consider, for example, the situation of a sales executive who formulates all the above judgments as follows. The sales executive can decide whether to go to visit some clients to

promote a given product line or remain in the office. Each alternative (not to make a sales call, make an average, run-of-the-mill sales call, or make a carefully planned and well-prepared sales call) entails a different expenditure of effort (quantified as 0 ML, 3 ML, and 5.5 ML respectively) and a different probability of increasing sales by a targeted amount (0.1, 0.4, and 0.7 respectively). If sales do increase by the specified amount, the sales executive will earn a 10 ML bonus (extrinsic valence), with an estimated probability of 0.8 (this expectancy may be less than one, for example, because there might be competition with other sales executives for the bonus). In addition, if the sales executive makes a carefully planned, well-prepared visit, the executive could gain intrinsic benefits in terms of social relationships with clients and of personal image that could be capitalized (assessed current value: 1 ML).

To solve a problem perceived in these terms, the sales executive can apply an expected value-maximizing logic. As a result, the sales executive will not be motivated to make mediocre sales calls (whose expected value is low), but neither will the executive be particularly motivated to make well-prepared ones, because their expected benefit is equal to the expected value of making no sales call at all, as shown in Figure 2.7. A prescriptive conclusion that can also be drawn from the analysis is that, given these valence and expectancy judgments, only sales bonuses over 10ML will motivate high levels of performance.

The *situational conditions* where one can expect motivation processes to be configured according to expectancy theory can be derived from the general applicative conditions of value-maximizing decision processes. Consequences should be evaluated on a utility scale. Alternatives must be known and comparable. There must be enough information to be able to estimate a probability for each single outcome value. In addition, the analytic process itself may entail appreciable additional costs that may discourage its application if consequences are not important enough. Therefore, processes of this kind can be expected to be applicable and effective in structured activities, and to be enacted by actors sensitive to economic benefits and extrinsic rewards (which are more measurable) and endowed with relatively high competences and responsibilities (owing to process costs).

Action alternative	Cost of effort	Effort → performance expectancy	Intrinsic valence of performance	Performance → outcome expectancy	Valence of outcome	Net expected valence		
1 No visit	0	0.1	0	0.8	10	10 (.8) (.1)	=	0.8
2 Average visit	3	0.4	0	0.8	10	10 (.8) (.4) –3	=	0.2
3 Well-prepared visit	5.5	0.7	1	0.8	10	10 (.8) (.7)		
						–5.5 + 1 (.7)	=	0.8
						1 (.7) – 5.5	=	0.8

Figure 2.7 An application of the valence/expectancy model (costs and benefits in thousand dollars)

Goal-setting

A second type of motivation process can be retraced to the general characteristics of decision-making based on aspiration levels and acceptability judgments (March and Simon 1958; Chapter 1). Instead of taking into account utility functions to be maximized, actors can allocate effort and competence according to targets and goals to be reached. Hence, the informational requirements of this strategy of effort allocation are less ambitious than those of an expectancy-based strategy. The core question about motivation then becomes: *Are performance levels related to the type of goals actors formulate*? Originally, March and Simon formulated this problem as one of "optimal tension": low aspiration levels reduce search and lead to accepting low results; very high aspiration levels lead to lower success probability estimates, so that above certain levels action is inhibited. Therefore, the relationship between the goal difficulty and performance can be represented by a curve of the type represented in Figure 2.8, where the highest performances are stimulated by objectives of "reasonable" difficulty. Subsequent studies in the field of "goal-setting" (Locke and Latham 1990) have repeatedly confirmed this hypothesis.[4]

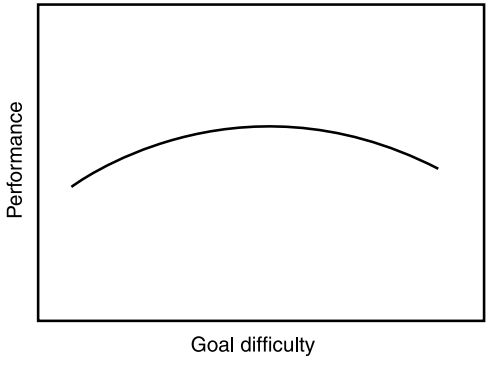


Figure 2.8 The relationship between goal difficulty and performance level

Goal difficulty and precision

An objective or goal is defined as a "prespecified performance level" that guides the choice of actions for reaching it (Sims and Lorenzi 1992: 117). Numerous laboratory studies have shown that performance is systematically influenced not only by goal difficulty, but also by goal specificity: i.e. the extent to which a measurement scale for results is specified (quantitative or nominal, i.e. an enumeration of things to be done) and a specific level to be reached on that scale is set (for example, a firm's division results are measured by rate of increase and a 5 percent increase should be attained next year). In other words, it has been shown that generic purposes such as "do your best" (or let us do our best) often pave the way for actually doing worse than one could.

The relationship between setting high-level objectives and getting high-level performance has been shown to be particularly robust. Nonetheless, it is mediated by competence and by self-efficacy judgments; and its form depends on the nature of activities. In a task where the actor is clearly perceived to have the required competence and uncertainty is perceived to be low – for example, log-loading on a truck by trained porters – the relationship between goal difficulty – reaching at least 60 percent, or at least 80 percent, or at least 95 percent of a truck capacity utilization – and performance is positive and linear (Tosi *et al.* 1986; Locke 1996). On the contrary, raising a sales budgetary target will improve performance

levels only up to a point beyond which they are offset by the downward adjustment of success probability judgments (the relationship has the inverted U shape represented in Figure 2.7).

Task complexity has a significant impact also on the optimal degree of goal precision or specificity. The more complex an activity (i.e. the less the path-to-goal is known), the less a priori information on attainable outcomes will be available, and the more arbitrary and risky the setting of specific goals will be, because it will engender a "tunnel vision which inhibits effective search procedures" (Locke 1996). For example, if the task is to produce innovations in applied marketing techniques, it is not a good idea to set precise and specific goals (how much to improve market share, contribution margins, client portfolio, or the size of discounts). Or, if the task is uncertain because it requires getting other actors' consent, as in sales negotiations, behavior will become rigid if precise and specific targets are defined. Effective goal-setting in complex tasks should not, however, imply reducing difficulties or ambitions, because this would simply lead to low performance. Rather, it should imply a more general, less detailed, and operational definition of goals, a shift from setting goals as specific-result parameters, positions, and targets to setting goals as interests that lie behind those parameters, going a step higher in the logical hierarchy of preferences. Examples and cases abound – especially in fields like budgeting, sales force motivation, and product division performance evaluation – of how narrowly defined performance and rigidly pre-defined targets can become a restricting and distorting factor in the learning of attainable objectives and an appreciable consideration of the multiplicity of consequences of complex economic action.[5]

Participation in goal-setting

Lastly, the process through which goals are set is obviously important and has been much studied. The results, however, are less obvious and clear than one might expect. The relationship between actors' performance and their participation in goal-setting processes is very complicated. The ties connecting the specificity and the difficulty of the objectives with the performance described above are valid both when the actor autonomously sets personal objectives, and when the actor accepts objectives set by others. Participation generates two contrasting effects: on the one hand, self-set objectives may not be as high as those which, in equal circumstances, would be set by others (a self-serving bias); on the other hand, the self-determination may solicit stronger conviction and dedication to goals (commitment) (Locke 1996). But it has also been demonstrated that high levels of commitment are also obtainable when the objectives assigned by others are convincingly explained and understood and are connected to interesting rewards, and regular feedback is provided on the progress of the performance toward the objective. Participation is fundamental, however, when the performers themselves possess the relevant knowledge for formulating valid and accurate hypotheses on attainable objectives – i.e. "participating in goal setting is necessary for cognitive reasons and not motivational ones" (Locke 1996).

The reinforcement of behaviors

A third class of models hypothesizes that motivation processes are regulated by automatic, ex post adjustments upon action feedback, rather than by expectations and predictions of outcomes. These models –

deriving from the behaviorist tradition or even from cybernetic control theories – have been harshly criticized by cognitivist scholars. Nonetheless, accepting the idea that a decision process proceeds automatically does not imply that there is necessarily little or no cognition, nor that the subject is not aware of the process. Automatic motivation is compatible with cognition, albeit with a particular form of cognition (see above, on programmed rationality).

Types of reinforcement

Reinforcement theory, as applied to motivation, maintains that when consequences are attributed to one's own actions and are perceived as positive, the probability that those actions will be repeated increases; whereas the perception of negative consequences of one's own actions diminishes this probability. Reinforcements can be direct (rewards or punishments connected to actions) or indirect (abstention or absence of rewards and punishments). Hence, the model considers four typical situations, called positive reinforcement, negative reinforcement, punishment, and extinction, as represented in Table 2.1, which shows the model applied to an example of secretarial work and feedback supplied by the boss.

Applying reinforcement theory to motivation has contributed notably to explaining apparently irrational behavior and developing "positive-reinforcement programs" oriented toward correcting such behaviors and improving the relational climate (Komaki *et al.* 1996). In fact, one of the characteristics of reinforcement processes is that of regulating (often inadvertently) behaviors that are given little explicit attention and analysis, such as those requiring quick interactions: interpersonal relationships, aspects of work that have not (yet) been analyzed because they have never constituted a problem, and habitual actions. The fact that these processes take place automatically does not, however, mean that they have consequences of little significance, whether in terms of the quality of inter-

Table 2.1 Motivation through reinforcement

	Action	*Stimulus*	*Response*
Positive reinforcement	Letter in English taken down in good shorthand.	Boss: "Great work. You ought to think about taking a course."	Increases the probability that the behavior will be repeated.
Negative reinforcement	Letter in English taken down in good shorthand.	Boss, who normally criticizes anything that is not perfect, is silent.	Increases the probability that the behavior will be repeated.
Punishment	Letter in English with many errors.	Criticism.	Reduces the probability that the behavior will be repeated.
Extinction	Letter in English with many errors.	Boss, who normally praises anything well done, is silent.	Reduces the probability that the behavior will be repeated.

Source: Adapted from Tosi *et al.* (1986).

Box 2.4
Reinforcing A hoping for B

The deadline for presenting the project to the client nears. The project group procrastinates and underestimates the timing. When the deadline is already too close and it is clear that they cannot make it, the group is given extra staff. (*Positive reinforcement of undesirable behavior*)

The rule was clear and written on a poster: wear protective gloves when working in the vicinity of the ovens. Despite this warning, no-one did so. The workers remembered when many wore gloves back in the plant's early days, then some began to get fed up, and nothing special seemed to happen: the bosses did not object and which of the workers were aware of the damage to their hands that only now begins to be manifest? (*Negative reinforcement of an undesirable behavior*)

Young surgeons in a famous clinic are constantly exhorted to study, attend conferences, and develop new surgical techniques. Nonetheless, when they try to experiment with some of their new ideas, they face tiresome discussions explaining or trying to convince their equals and their superiors. And heaven forbid saying no to an extra shift on duty or to an operation commitment, because they want extra time for studying. (*Punishment of desired behavior*)

Everything went along smoothly in the maintenance office. As a result, no one had anything to say. The other departments made no comments. The production department did not protest. The head of the plant rarely dropped by the office because he had other "irons in the fire." It was as if maintenance did not exist. (*Extinction of desirable behavior*)

personal relationships, or in terms of direct impact on economic results.

Pathologies and unintended consequences of reinforcement

The examples in Box 2.4 illustrate some of the typical distortions generated by uncontrolled reinforcement mechanisms, in particular the inadvertent reinforcement of undesirable behavior or the unintended punishment or extinction of desired behavior (Kerr 1975). The reinforcement model helps to make visible other important pathologies, such as the trap of governing mainly through punishment and negative reinforcement, thereby giving indications only of what should not be done, creating negative "frames," and giving no clear indication of what to do.

Control models of motivation

A second type of automatic model of motivation applies cybernetic theory to motivation (Klein 1989). This kind of model assumes automatic adjustment processes oriented toward maintaining a stable state, the positive and negative deviations from which are perceived as performance gaps to be corrected

(as in the case of correct temperature). This type of motivational process can effectively explain some types of work actions. For example, if a person has accepted the standard time of 1 hour for completing a particular task, and if the worker observes that after half an hour, less than half of the work has been completed, then the worker will put additional effort into bringing the performance back on track. Homeostatic motivational models, then, can be effective and efficient only in stable activities where maintaining a normal or standard result is more important than striving for better results.

A condition for the effectiveness of reinforcement models in general (homeostatic or not) in a cognitive interpretation is that actors see their own action as a deterministic and systematic cause of the observed consequences. Whenever this hypothesis does not hold, as easily occurs in uncertain and unstable activities, reinforcement models no longer predict behavioral responses well (Kelley 1971; Weiner, 1974).

EFFICIENCY, EFFECTIVENESS, AND EQUITY

The approaches to decision and motivation analyzed so far focus on the relationship between preferences and outcomes. Effectiveness and efficiency, therefore, have been implied as regulatory principles capable of orienting action: the capacity of actions to produce desired consequences (effectiveness) and their ability to do so while saving cognitive and material resources (efficiency). However, these two criteria are insufficient not only for explaining but also for designing organization structures and behaviors well. For example, suppose that a plant manager has the opportunity to participate in a project to open a new production facility, and the manager is evaluating it by looking at the expected economic benefits, the amount of travel requested, the career potential, and the additional work commitment. What level of overall benefit is sufficient to motivate the manager to take part in the project? Weighing up incentives and contributions (efficiency) alone is not enough; in general, this will not produce enough information. Even assuming that contributions and benefits can be quantified and compared, and that the positive values balance out the negative ones, how can the manager evaluate what level of net benefit is sufficient or adequate to induce the manager to act? Different forms of rationality can be used in making this judgment, but it generally implies at least the following two elements:

- a process of social comparison or interpersonal comparison: actors not only compare their benefits with costs, but will compare their expected net benefits both with those of comparable others and with the net benefits of their exchange counterpart
- an equity judgment on how "just" or "fair" is the share of resources they are getting with respect to other actors' shares.

There are also other reasons why equity criteria should be systematically used in the analysis of economic organization and behavior, even without considering the empirical fact that people often employ them in practice, and even without considering an ethical dimension. The point is that *efficiency and effectiveness criteria are not sufficient, logically speaking, to determine one superior solution (act, contract, structure) in most action problems involving more than one decision-maker.* Most often there are many "superior" (Pareto-efficient) solutions, with

respect to which no further improvement of benefits for one or more actors without losses for anyone else can be found (Luce and Raiffa 1958). Therefore, an additional choice criterion which says something on the distributional problem is needed for selecting an action.

It helps to consider all the various branches of research that have analyzed the nature of equity and fairness judgments together in order to reconstruct a variety of equity criteria and their reasons and application conditions.[6]

A first result of general interest has to do with the very existence and importance of fairness concerns in economic action. Consider the problem described in Box 2.5. A game-theory, profit-maximizing solution to this experiment would be for the allocator to propose a minimal "token" payment to the recipient, who should accept any positive offer. However, empirical evidence shows that many allocators offer much more than a token (minimal positive) payment, and that recipients sometimes reject positive but small offers. Cognitivist researchers conclude that fairness criteria are applied by decision-makers even in structured problems where market exchange and unilateral profit maximization criteria were applicable in principle, and even at the price of renouncing part of their pay-off when they have no obligation or compensation for doing so (Kahneman *et al.* 1986a, b).

An interesting question at this point would be when, and under what conditions, should we expect this behavior to be likely. Clearly it is not to be expected that fairness will always obtain, just as we cannot expect behavior always to be efficient or effective. The discussion conducted in this paragraph on the different possible fairness criteria and of their properties will help in answering this question. An introduction to the issue can be provided by an analysis of a work situation, among the many with which the reader may be familiar, that is considered unmotivating because it is unfair, and a discussion or elaboration of which fairness criteria could be applied in order to improve it (see, for example, the exerciser at the end of the chapter).

Outcome-based equity

Various criteria for fair resource allocation have been elaborated both in economic and sociological perspectives. They all try to translate into applicable criteria a philosophical concept of justice according to

Box 2.5
The ultimatum game

The critical laboratory experiment used for discovering to what extent fairness considerations come into play in economic decisions is an "ultimatum game": one player (the allocator) is asked to propose that a sum of money be divided between him/herself and another player (the recipient), who in turn can either accept the offer or reject it, in which case both players receive nothing. The experiment can be run on various dyads with varying amounts of money.

Source: Experiment described in Kahneman *et al*. (1986a).

which a just solution is not necessarily a "Salomonic" one – in which resources are divided equally – but is one that all parties could accept if they did not know what their own share will be, if they agree on the criterion for division behind a "veil of ignorance" about their specific position (Rawls 1971; Brennan and Buchanan 1985).

In an economic perspective, all benefits and costs should be reflected in the utility evaluations that actors make about the outcomes of available alternatives (bundles of resources or combinations of actions). If those different evaluations could be weighted and summed, a possible criterion would be to select as a fair option that for which the sum of utility is higher. This would be a way of taking into account the utility of all the parties involved, and in that sense it involves a notion of fairness. However, what weight is it fair for each party to have? Is one "point" of utility on one party's value scale worth the same as another actor's point? The problem is only shifted and it is usually accepted that there is no legitimate and rational way to perform those "interpersonal comparisons of utility".

Fairness criteria not involving interpersonal comparisons of utility have been proposed by game theorists and economic analysts.

The most widely used fairness rule in economic analyses is probably the Nash (1950) criterion, or the *maximum product of the parties' utilities*. According to this criterion, actors take into account the preferences of their counterparts in exchanges and, among all available Pareto-superior solutions, will give priority to those solutions where benefits are allocated in balanced ways – for which the product of the utilities is generally larger – over those in which some parties turn out close to the maximum they could achieve and others turn out close to their minimum (see also Chapter 6).

Among the limits of these fairness rules based on the pay-off for the different parties, there is the possibility that they will increase the initial inequality among partners' resources (which does not adequately achieve a philosophical principle of justice) (Sen 1992). Allocating resources according to the subjective utility function of players, in fact, may end up by giving more resources to the "rich" party (which assigns less utility to a marginal increase in resources) and less to the "poor" one (which assigns high value to even a small improvement in their position), thereby reinforcing the initial differences.

In addition, these criteria need a lot of justification, discussion, calculation, and bargaining. Therefore, more egalitarian and less calculative solutions can reasonably be preferred by all parties because they save information-processing costs and losses of "atmosphere," at least up to a certain level of importance of the pay-offs and consequences of the decision.

Lastly, outcome-based criteria require a clear assessment of preferences and interests. As extensively discussed in these first two chapters, this condition is by no means ubiquitous. A master of game and utility theory himself wrote

> To some extent, the complexity of the real situation softens the intensity of the bargaining dynamics. The parties are not clear about what is in their own interest, and their knowledge about the interests of others is likewise vague. Compromise is often easier to arrange in a situation of ambiguity . . . many real world negotiations are happily not as divisive as starkly simple laboratory games, because in the real world it is difficult to see clearly what is in one's own best interest.
>
> (Raiffa 1982: 274)

In sociological studies an outcome-based "theory of equity" has also been developed, asserting that people consider a resource assignment to be fair if the pay-off received by each party is proportional to the contribution given (Adams 1965). This principle was made operational in the following fair division rule (Walster and Walster 1975):

(Output – Input)/Input (for the actor) = (Output – Input)/Input (for comparable other actors).

In an early interpretation, this principle of equal returns on investments was thought to be so strong as to be applied by people even against their own interests: i.e. actors try to modify the elements of the equation in order to restore balance both when they believe themselves to be "underpaid" and when they believe they are "overpaid." These "altruistic" behaviors have been observed and are more likely if the actor and his/her "comparable others" perform similar and structured activities (so that social comparison is easy) and the actor who is overcompensated perceives that his/her extra benefits are taken out of the fair share because of other members of the group (Campbell and Pritchard 1976) – for example, in a group of blue-collar coworkers. In fact, in such cases, the losses in status, social acceptance, or even just self-acceptance may well compensate for the extra gain.

In other situations, it is probable that the actors will modify different elements according to whether they perceive their own position as a deficit rather than as a surplus. It is easy for "overpaid" subjects to react by raising their opinion of their own competence in order to justify what they are paid, and for those who believe themselves to be underpaid to reduce their contribution. Studies conducted by Staw (1982) have illustrated other ways of reacting to the perception of being insufficiently or redundantly compensated, guided by the cognitive tendency toward self-confirmation and self-justification. For example, in order to "justify" staying in a low-paying position at work too long, one tends to re-evaluate other elements and advantages of that job (for example, social relations, intrinsic interest). Symmetrically, if an actor is overpaid for a job that still offers sufficiently high intrinsic compensation to repay efforts and contributions, the actor could become convinced that pay (rather than intrinsic rewards) is necessary to motivate him to perform.

Fairness rules based on proportionality to contributions are all the more demanding in terms of information required. To be applicable, both contributions (inputs) and results (outputs) must be distinctly measurable and comparable. If work contributions and rewards are considered as examples of "inputs" and "outputs," one can easily see that this is often not the case. Only where resource and competences are relatively standardized and measurable throughout a group of actors and where actors can assess each other's contributions, can those perceived equity judgments on contributions and outcomes be reliably formulated. For example, in interfirm – rather than interpersonal – allocations of resources, whereas the value of the contribution of each firm can be measured (for example, the value of technical and financial assets contributed in a joint venture), the rights to rewards can be made proportional to investments (Grandori and Neri 1999).

Need-based equity and mutual acceptability

A different way to assess interests and preferences, rather than ordering all possible

outcomes according to preference, is to consider one's own needs and aspirations. As seen above, need is considered to arise for matters the lack of which would threaten balanced existence. The assessment of needs may then be less difficult and ambiguous than the assessment of the value of contributions and outcomes (Albin 1993). Indeed, if not for their clarity, needs may be easier to assess because they are rooted in the actor itself rather than in what the environment can offer as reward, or with what other subjects receive. In addition, needs and aspiration levels call for "being satisfied" rather than for being maximized. Therefore, fairness judgments based on need satisfaction (rather than on the maximization or equalization of parties' returns) will require less information processing.

Given that needs and aspiration levels are quasi-independent of contributions, need-based equity should be legitimized by the players' mutual acknowledgement of what is essential to the existence of the relationship or to the satisfaction of others (Pruitt 1972).

In economic thought, Sen's definition of needs and distributive justice comes quite close to that used in organization studies: needs are vectors of "acquisitions" that "can vary from elementary things such as being adequately nurtured, being in good health, escaping avoidable morbidity and premature death, to more complex acquisitions such as being happy, having self-esteem and participating in the life of the community" and *justice resides primarily in actors' freedom of access to these acquisitions* (Sen 1992).

Among the advantages and application conditions of needs-based criteria, therefore, there are:

- a redistributive effect to the advantage of the parties who are more in need of resources, which seems to fit with a general concept of justice, especially in conditions of relative scarcity and deprivation
- a reduction of calculation and bargaining costs with respect to joint utility maximization criteria.

These advantages, however, are likely to be realized only under circumstances that facilitate the mutual recognition of needs. They include personal acquaintance and the longevity of relationships (Albin 1993) and the extent to which parties are poorly substitutable or contribute resources that are highly specific to their relationship and critical for its existence (Grandori and Neri 1999).

Non-calculative fairness

In repeated and/or ambiguous situations, learned behaviors and automatic decision rules may substitute for too costly or cognitively complex calculations, of either the optimizing or the satisficing kind. Decisions about fair division of resources make no exception. Two kinds of fairness rules that have the nature of "blind" heuristics are especially important in practice, and they also exemplify two different major sources of legitimacy of automatic rationality: the reduction of information-processing costs and the availability of applicable past experience.

An *equal share rule* expresses an absolute egalitarian principle: different actors receive the same share of resources, regardless of their utilities, contributions, and needs. Among the advantages of this apparently insensitive rule are: a capacity to improve a climate of harmony and trust between actors, a drastic reduction of information-processing costs, especially when these are

not justified by the amount of differential expected benefits, and the capacity to produce solutions where it is difficult to measure inputs and preferences over outcomes are not so clear. For example, what are the equitable rewards for each member cooperating in a group where there is no division of labor, and the task is complex, such as developing a new product or a new transformation process? In cases like individual actors cooperating in a team or firms cooperating in a joint venture, we can observe that rewards are often distributed evenly.

An alternative form of automatic and non-calculative fairness is the use of custom and "good practice" as a source of "good right." In this case, the decision is made using historical precedent (Pruitt 1981) as a reference point and making adjustments from status quo situations. On the basis of our previous discussion and assessment of non-calculative rationality, we would expect these "*fairness heuristics*" to be frequently and effectively practiced on many recurrent matters in a stable, long-lasting relationship.

"Substantive" and "procedural" justice

All the equity criteria discussed thus far are criteria that serve to find a point of agreement on how to allocate or apportion resources. They are therefore called *distributive justice* criteria (Greenberg 1987). Each criterion has its own limits and, when there is a decision that is uncertain and important, on the one hand actors could reasonably wish to avoid relying on blind rules, but on the other hand they might not have enough information to divide resources according to inputs or outcomes. In conditions of high information complexity, the fairness of the procedures and of the process followed to determine what each actor will get, can become especially important, even more important than the amount of resources received. This type of fairness has been called *procedural justice* (Greenberg 1987). Its effectiveness under uncertainty can be supported with the general arguments that make "procedural rationality" (how to arrive at a decision) more interesting than "substantive rationality" (methods prescribing what solution to choose) in solving uncertain problems (Simon 1976). The notion of "fair procedure," inspired by that of procedural justice in law, includes the following traits or properties: fair procedures are consistent across parties and time (the rules of the game remain the same during the process), represent concerns of all parties, provide opportunity to exert influence and revise decisions, and are based on accurate information (Leventhal 1980).

In addition to answering the question as to when these various criteria of fairness should be applied, one may now also legitimately inquire when fairness criteria can be expected to be actually applied to regulating economic activity. Based on all the studies considered, the following list of conditions is the minimal answer to the question. It is more likely that fairness rules will be used, independently of individual propensities toward being fair, when:

- *the set of possible effective and efficient solutions to a problem is wide*, so that there is still a choice to be made among them (as in complex transactions or joint actions involving several matters)
- there is *uncertainty regarding the counterparts' best alternatives* to an agreement, and therefore it would be cognitively difficult to push a counterpart "close" to its minimum acceptable pay-off

- the *exchange relationship is multilateral* rather than bilateral, and is repeated many times with the same features, so that the "comparable others," who invite and sustain equity judgments, can actually be compared and even call for comparison (as in processes involving equal opportunities or equal treatment issues)
- the *relationship is repeated* and expected to be long-lasting, so that each party has incentives to adopt rules of resource allocation which can be accepted independently of its own particular position in an individual exchange
- actors are acquainted with and have a *personal knowledge* of each other
- the *need level of actors' preference structures is at stake* (i.e. the problem concerns resources that nurture actors' continued balanced existence).

For example, the field of labor relations fits all of the above conditions and is, in fact, the economic relationship in which equity concerns are probably densest (see Chapters 9 and 10).

SUMMARY

The first section presented a process model of economic actor rationality. On the basis of available interdisciplinary research on decision processes, the model of rationality on which economic behavior seems to be founded is a multiple or pluralistic model in which different forms of rationality can be expected to be applicable under specifiable circumstances. Therefore, none of these forms is treated in this book as an "assumption" about human rationality (as in mainstream economics and utility theory) or as a universalistic theory of how people behave in practice (as in classic behavioral decision theory).

It was shown how different combinations of search, choice, and learning rules configure different "strategies" of decision. Three main types of strategies – or models of rationality – are defined (articulated in various submodels):

- a *deductive model*, that can be articulated in various "*optimizing*" *strategies*, applicable and superior in the solution of *structured problems*
- a *heuristic model*, that can be articulated in various "satisficing" and *acceptability-based strategies*, applicable and superior in the solution of *unstructured problems*, where objectives, cause–effect relations, relevant alternatives, and possible consequences are treated as hypotheses to be tested
- a *non-calculative model*, in which behavior is determined by rules indicating "*appropriate*" *actions* contingent to the situation, to the identity, and to the state of the system. Non-calculative learning processes are sustained by the principle of reinforcement (in direct learning by doing) and the rules of imitation (in vicarious learning). The distinctive advantage of this decision mode is that it is *applicable in the absence of defined objectives and of consequences prediction.*

The second section presented the available process models of motivation, showing that they are no different from decision processes applied to the allocation of one actor's own efforts and actions. Motivation process models considered include:

- "*expectancy theory*" as a model of acting according to the maximization of subjective expected value

- "*goal-setting theory*" as a model based on setting aspiration levels to be satisfied
- "*reinforcement theory*" and "*control models*" of motivation as variants of automatic and non-calculative models of decision-making.

This connection between motivation and decision models has enabled us to specify the *conditions* under which we can expect the different types of motivation processes to occur, as a function of the clarity of preferences and the state of knowledge on action alternatives – whereas they are usually presented just as "different" (or "rival") models of how motivation works.

Lastly, equity theory and the concept of fairness were introduced. They had a special development in the analysis of motivation to work, but it has been argued that *economic behavior in general cannot be completely assessed or designed employing effectiveness and efficiency criteria only*. Using notions of fairness elaborated in different disciplinary fields, it was shown how different conditions of uncertainty can be dealt with by these different equity criteria, which incorporate different forms of rationality: *criteria incorporating some principle of maximization of joint utility; criteria based on the satisfaction of needs; and "blind" or automatic rules of fairness based on symmetries, repertories, or precedents*. The importance of "*procedural justice*," especially in uncertain and important decisions, in which the fairnes of any substantive distribution of benefits is difficult to judge, was highlighted.

Exercise: BluCer

BluCer is located in the ceramic district of Sassuolo, and produces large ceramic items, primarily tiles and vases.

At the beginning of 1998, the chief executive officer (CEO), Franca Gabrielli, decided to create an autonomous Information Technology (IT) department. "We want to computerize all the ceramics," she said at the time of the decision. The objective of the new department was (in the words of the CEO): "injecting information technology throughout the firm, designing systems to manage the warehouse, the orders, the billing, etc., based on requests by every single department." This needed to be done without forgetting the management of the electronically controlled machines that were already used in production activities (machines regulating movement, the kilns, etc.). This last activity of management/technical assistance (up to then outsourced) had turned out to be particularly costly and inefficient, and this had slowed the process of improving the quality of the product/service to the final client.

Carlo Rossi (an ambitious computer engineer) had been named the director of the new IT department. In total, the new unit included twenty-two people, of whom seven came from different parts of the firm where they had performed IT support activities, even if they had not been effectively coordinated, and instead worked as trouble shooters on problems and/or projects of immediate concern, in collaboration with external technicians and consultants. New staff chosen by BluCer were technicians, mostly with engineering backgrounds, who had been hired primarily to perform research activities. The initial compensation was not high, but they were told that there would be an incentive plan based on the development of new informational projects that would take them to the pay levels of specialized technicians.

At the end of the third month, the results seemed to confirm the appropriateness of the plan. The firm's management took comfort in several indicators, including the more than twenty

weekly interventions in the production department that had been performed quickly, and the launch of *Project Blue* (aimed at creating an integrated information system).

After little more than 10 months, the situation seemed to have definitely deteriorated: although the interventions had remained at more or less the same level, *Project Blue* had not taken off and there were conflicts inside the department, based on the distribution of the workload. In addition, in the production department (as in others), there was some grumbling about the quality of the services offered.

For their part, the technicians seemed to demonstrate intolerance and a lack of motivation related to the way in which things were going. First, there was the problem of the "economic return" or lack thereof, from the activities performed. With respect to pay, this had remained unchanged, i.e. it was stuck at a level below the average for information technologists, and the prospects did not seem encouraging.

The technicians complained: "We have not heard any more talk of incentive plans. We earn very little, both with respect to the workload and to our colleagues in other firms. There are no career prospects and we are certainly not treated like researchers," and said confidentially, "we now understand that the evaluation of our work depends on how quickly we intervene, and does not depend so much on whether the work is done or explained well . . . It does not seem to matter if they need to call us the day after to resolve problems caused by yesterday's intervention."

In contrast, regarding the need of various departments for new programming, the slowness of the response was interpreted to be a sign of work that was well done: "On delivery of a program finished 2 months after being requested, I heard that because the preceding one had taken a year, the new one could not possibly be better. It is useless to even talk with people like that!"

The technicians in the IT department were therefore encouraged to overlook the requests of the different departments, keeping busy with simple work instead of doing the complex work which meant delaying delivery of the complex work and giving priority to the frequent demands of the production department.

There was also great frustration about the recognition of the needs – material and otherwise – of the IT department. The budget to buy information technology seemed insufficient but, more importantly, ineffective in the management of technicians' time: "From the day I was hired, I have not had two days in a row to develop new projects," and "One has to understand that this is not a production activity like the others. We need more flexibility in our working hours and opportunities to participate in training courses and to work with other firms (clearly not our competitors) to learn new competences."

Finally, there was a strong sense that the relations among the members of the department were very fragile, in contrast to what had always characterized BluCer, which had always been perceived by its employees as a big family despite its large size. This dissatisfaction was fed by the fact that in other departments the firm had traditionally considered and resolved problems of individual workers, as long as they were compatible with the firm's workload (from shift work to vacations, advance pay to buy a house, "sponsored" hiring). In contrast, the attempt to consolidate the IT department and to make it an autonomous laboratory had led to an "aseptic" management that seemed distant from the problems of individuals.

In brief, the situation seemed critical after only a few months from the start of the project.

The CEO, looking at the data on the activity of the IT department, was asking herself whether to close the new department or to reconvert it for the exclusive management of the productive process and then to evaluate it only in terms of productivity and responsiveness. At the same time, Carlo Rossi had heard that top management was aware of the problems, and had begun to act in defense of his position: he was ready to "save his head" by attributing responsibility for the failure to the lack of information technology culture diffused throughout the firm, and the lack of motivation of the employees.

Questions

- How was the problem of setting up the new information system defined? How could it have been improved? How can the definition of the current failure problem be improved?
- What type of decision process could the CEO effectively follow in dealing with the current problem?
- Why are people "dissatisfied?" How could the equity of the process and of the solution be enhanced?

By Massimo Neri

Conclusion to Part I

An Actor with Multiple Rationalities

The model of actor and actor's behavior that can be assumed as a basis for the development of more complex models of organized economic action is a meta-model composed of many. These submodels are conceivable as salient and "discrete" on a continuum of feasible behaviors. They have been shown to be comparable, provided that they are considered as feasible alternative models, whose informative requirements and application domains can be specified. An advantage of this approach is that it captures an important but neglected capacity of decision-makers: to shift from one decision strategy to another according to the nature of problems. A theoretical implication of the developed framework is that it goes beyond the contrast between "global" and "bound" rationality, and between prescriptive and descriptive theories of choice, which has become an obstacle to dialogue and cross fertilization between studies of organization with economic and psychological underpinnings (and to a more general and empirically based explanation of economic behavior).

The plurality of rationalities

Three basic forms of rationality have been reconstructed based on the existing vast theoretical and empirical literature on decision processes in general, and on decision processes relative to work in particular. It has been shown that the general cognitive model of actor knowledge and behavior developed here is powerful enough to encompass as particular cases the main models of decision and motivation that have been developed in a partially independent way.

The fundamental traits of this model are summarized in Figure I.3 – as far as process models are concerned – and in Figure I.4 – as far as knowledge and preference structures are concerned. The two elements are related, in that decision and motivation processes can be activated only starting from a structure of knowledge, competencies, and preferences that are in turn fed and modified by each process.

Three basic forms of rationality can be distinguished and assessed:

- an expected value, valence/expectancy-maximizing form, which is "deductive" in the sense that solutions are derived logically from sufficient a priori knowledge of relevant objectives, alternatives and consequences
- an acceptability-based, goal-setting form, which is "heuristic" in the sense that the relevant alternatives, objectives, and consequences should be searched and tested
- an appropriateness-based, reinforcement-driven form, which is "non-calculative" in

<table>
<tr><td></td><td colspan="3">Decision strategies</td></tr>
<tr><td>State of knowledge and preference</td><td>Deductive value maximizing</td><td>Heuristic problem solving</td><td>Non-calculative appropriateness</td></tr>
<tr><td>• Known objectives alternatives and consequences
• Complete preference orderings</td><td colspan="3">All applicable
• Strategy selection, a function of:
— effort/accuracy trade-offs
— existence of Pareto-optimal solution</td></tr>
<tr><td>• Conjunctural objectives alternatives and consequences
• Incomplete pref. orderings</td><td>Not applicable</td><td colspan="2">• Both applicable
• Strategy selection as a function of:
— repeatedness and importance of decisions
— existence of intersections among acceptability judgments</td></tr>
<tr><td>• Unknown cause–effect relations and/or unclear preferences</td><td colspan="2">Not applicable</td><td>• Applicable, in repeated situations
• Unrepeatedness leaves with random trial and error</td></tr>
</table>

Figure I.3 A decision failure framework

the sense that given a recognition of the type of situation and the identity of the actor in it, actions to be taken follow from a rule of correspondence, which is adapted ex post upon observation of results.

The comparative framework developed in Chapter 2 and summarized in Figure I.3 specifies the maximum level of uncertainty that each strategy can deal with. As such, it is a "*decision strategy failure framework*": it asserts that if it is cognitively unfeasible or too costly to acquire the requisite knowledge about objectives, alternatives, and cause–effect relations ex ante to a decision process for a given strategy to be applied properly, alternative strategies become superior. This does not mean that, conversely, simpler strategies cannot be applied when more information than required by them is available. However, it does imply that less analytic strategies will usually be inferior in those circumstances because they do not make use of the available information (taking into account the costs of information). More generally, in structured problems all decision strategies are applicable, and their selection can be based on trade-offs between effort and accuracy; or tested against the existence of Pareto-optimal solutions. In unstructured problems, value-maximizing approaches fail; heuristic strategies will be superior on important or new issues, provided that an intersection among the sets of acceptable alternatives for different actors exists. Non-calculative appropriateness is applicable even if the actor is not farsighted and has unarticulated

preferences, but the formation and effectiveness of "programs of action" are conditioned to repeated action and consistency among relevant rules: uniqueness of problems or conflict among rules are expected to call for analysis and ad hoc problem-solving.

If those are important configurations of decision processes, they are not the only ones. *Other effective combinations of search, choice, and learning rules can be defined*. For example, in many complex and important problems, a heuristic approach is necessary to generate relevant alternatives and to envisage possible consequences. Once these elements are defined, a value-maximizing choice rule can be adopted for selecting the superior solution (among those available) rather than just the first acceptable solution. The design of organization itself, for those aspects in which it can be designed, can typically be addressed following this mixed, two-stage, decision strategy (Part III).

The fallibility of judgments and their improvement

The admission that human judgment is fallible, and that this is relevant in economic action as elsewhere, has led to the problem of improving judgment being taken seriously, irrespective of which decision strategy is used. A wide inventory of systematic cognitive biases has been presented, including framing effects, local knowledge traps, overconfidence and self-confirmation distortions. Examining the possible "antidotes" and remedies, some aspects of *organization structure and systems*, in particular the use of teams, structured checklists and decision support systems, control systems which do not conceive performance as lack of errors, formal evaluation systems turn out to be *important leverages for sustaining the generation and use of more valid and reliable knowledge.*

The hierarchical structure of knowledge and preference

Streams of study as different as the theory of scientific and technical knowledge, the economic and organizational analysis of competences and the theory of needs converge in the identification of a logical hierarchy in the structure of actors' competences and preferences, as suggested in Figure I.4, deriving from the need for "operationalizing" them in order to act, and from the different processes that lead to the formation of their different layers.

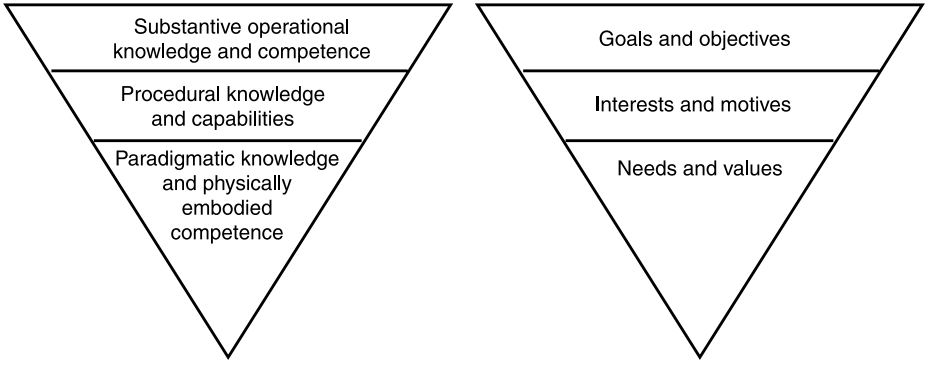


Figure I.4 Logical structure of actor knowledge and preferences

Both the inter-related and interpreted sets of information that make up an actor's knowledge and the wider set of skills that concur to make up an actor's competence have profound, highly embodied layers, in part inherited (socially and genetically) – in which case they are rather difficult to change – and in part accepted by convention. These elements are quite removed from action. There are many ways of applying them. At least two other components of knowledge and competence link paradigmatic competence to action: a substantive, domain-specific "database" of know-that propositions or repertoires (for example, content notions in the field of chemistry, repertories of action in the work process of steel transformation) and a set of "programs," of know-how notions, of procedures that govern the application of substantive competence (for example, experience in chemical research, learned sensibility about when to apply which productive correction as a function of a diagnosed state of melted steel). The relative incidence of the different layers of knowledge and competence have important consequences for the degree of change and discretion that is present in a system of action (Chapter 7).

A preference is a value-laden judgment, expressed (explicitly or implicitly) over alternatives. As a form of knowledge (about reality and about the self), a set of preferences exhibits a hierarchical structure, in which fundamental values and needs can be distinguished from assessments of what an actor's interests in a situation are; and interests can in turn be distinguished from the setting of objectives and goals to be reached. These distinctions are important because the less abstract and more operational the definition of preferences becomes, the more action is constrained. The *degree of precision and operationalization* of goals is therefore an important determinant of the degree of discretion and of the possibility to be creative. Another relevant dimension of the preference structure of an actor is the *ambitiousness or difficulty of goals*, the setting of which is fundamental in sustaining motivation.

Effectiveness, efficiency, and the need for equity judgments

The content of preferences is difficult to predict, and "content models" of motivation can be easily criticized. It can be more reliably supposed that actors are interested in linking actions' consequences to their ends – whatever they might be. This is a criterion of *effectiveness.* Most often, in economic action, actors are also interested in pursuing their interests while saving resources for other possible interesting uses – a criterion of *efficiency*. In many situations these two criteria are not sufficient for evaluating possible actions – the core reasons being that the resources employed and the benefits gained may not be comparable, and that the utilities of different actors should not be assumed to be comparable in a straightforward way (for example, through sums and differences) in most conditions. This creates an indeterminacy as to what the best action is. The notion of "best action," if one wishes to retain it, should include some "joint best" principle, i.e. some criterion of *equity*.

Part II

Coordination Mechanisms

Chapter 9

The Organization of Work and Human Resources: Systems and Contracts

Actors, both individuals and groups, are considered here as possessors of "bundles" of human resources (competences); and as holders of preferences, and of rights and obligations deriving from the more or less complex contracts that regulate their work relations.

- A secretary is hired by a firm for an indefinite period of time.
- A migrant farm worker sells his labor for a season.
- A professional sells consulting services.
- A lawyer specializing in employment law is hired by a firm's legal office.
- A ceramics craftsperson works for several firms in an industrial district.
- A jazz guitarist plays in a series of nightclubs and restaurants on an evening contract basis.
- An independent weaver works exclusively on materials and designs furnished by a single fashion house.
- A firm rents contingent work for discharging activities from a work agency.
- A group of masons form a work cooperative.
- A group of professionals form a partnership.

The contracts that regulate these work relations are very diverse. Each of these different contracts can be effective, efficient, and fair in specific circumstances. This chapter aims to explore these various circumstances and the contracts appropriate for them. It explores the different mechanisms governing work contributions, grouped into three main classes: *evaluation, reward, and mobility/development*. The fourth and concluding section examines and evaluates the combinations or configurations of these mechanisms that are embodied in some salient forms of *work contract,* and explores the conditions under which these can be considered effective, efficient, and fair. Enlarging the usual focus of human resource management, and integrating organizational economics contributions, the treatment encompasses both the "internal" and "external" organization of human resources and labor services with respect to the firm. The general scheme of the chapter, applying the general framework outlined in Chapter 8 to the problem at hand, is summarized in Figure 9.1.

To introduce the theme, consider the exercise in Box 9.1. What criteria can be used in deciding on the wage increases for the employees in the exercise in Box 9.1? There are many answers that are legitimate, in principle. Here are some typical criteria that can be identified and used in resolving the exercise.

- A performance evaluation, by superiors or peers, based on results or behaviors.

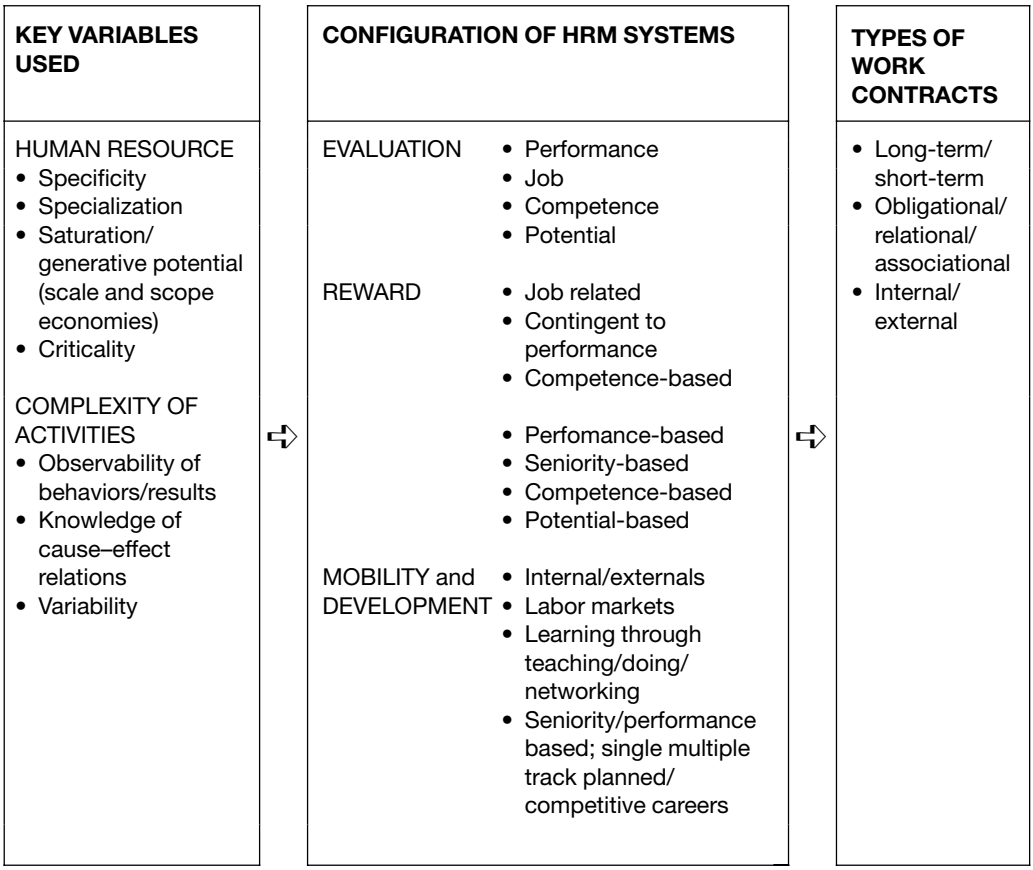


Figure 9.1 Configurating human resources organization

- A competence-based evaluation of the level and substitutability of employee's skills and professionalism.
- A job evaluation, in terms of the cost and value for the person performing it, and for the system of action in which it is embedded (How difficult is the work? Is it dangerous? Does it involve responsibility and discretion?).
- An evaluation of the employees' preferences and needs relative to the amount and form of compensation (in this case monetary).

Even the simple information given in the exercise opens a fairly wide range of possible evaluation methods. The above questions also show the possibility of conflict between different criteria (e.g. results, behavior, type of work performed, personal situation, starting pay). The relative weight of each criterion can of course vary in different compensation systems. We start with the analysis of the evaluation system because an evaluation of the contribution of the human resources constitutes an anchor and an input for effective and fair compensation and career and development systems. Furthermore, the responses to the question of "what" and "how" to evaluate – performance or results,

Box 9.1
Motivation through compensation

You have to make salary increase recommendations for eight managers that you supervise. They have just completed their first year with the company and are now to be considered for their first annual raise. Keep in mind that you may be setting precedents and that you need to keep salary costs down. However, there are no formal company restrictions on the kind of raises you can give. Indicate the size of the raise that you would like to give each manager by writing a dollar amount next to their names. You have a total of $17,000 available in your salary budget to use for pay raises. Current net salaries are indicated next to each profile.

$——*A.J. Adams*. Adams is not, as far as you can tell, a good performer. You have checked your view with others, and they do not feel that Adams is effective either. However, you happen to know Adams has one of the toughest work groups to manage. Adams's subordinates have low skill levels, and the work is dirty and hard. If you lose Adams, you are not sure whom you could find as a replacement. Salary: $20,000.

$——*B.K. Berger*. Berger is single and seems to live the life of a carefree swinger. In general, you feel that Berger's job performance is not up to par, and some of Berger's "goofs" are well known to the other employees. Salary: $22,500.

$——*C.C. Carter*. You consider Carter to be one of your best subordinates. However, it is quite apparent that other people don't agree. Carter has married into wealth, and, as far as you know, doesn't need additional money. Salary: $24,600.

$——*D. Davis*. You happen to know from your personal relationship that Davis badly needs more money because of certain personal problems. As far as you are concerned, Davis also happens to be one of the best of your subordinates. For some reason, your enthusiasm is not shared by your other subordinates, and you have heard them make joking remarks about Davis's performance. Salary: $22,700.

$——*E.J. Ellis*. Ellis has been very successful so far. You are particularly impressed by this, since it is a hard job. Ellis needs money more than many of the other people and is respected for good performance. Salary: $23,500.

$——*F.M. Foster*. Foster has turned out to be a very pleasant surprise to you, has done an excellent job, and is seen by peers as one of the best people in your group. This surprises you because Foster is generally frivolous and doesn't seem to care very much about money and promotion. Salary: $21,800.

$——*G.K. Gomez*. Your opinion is that Gomez just isn't cutting the mustard. Surprisingly enough, however, when you check with others to see how they feel about Gomez, you discover that Gomez is very highly regarded. You also know that Gomez badly needs a raise. Gomez was just recently divorced and is finding it extremely difficult to support a house and a young family of four as a single parent. Salary: $20,500.

$——*H.A. Hunt*. You know Hunt personally. This employee seems to squander money continually. Hunt has a fairly easy job assignment, and your own view is that

Hunt doesn't do it particularly well. You are, therefore, quite surprised to find that several of the other new managers think that Hunt is the best of the new group. Salary: $21,000.

Source: Lawler (1975).

what one does or what one knows, past performance or expectations about the future – allow one to answer questions about how to compensate whatever has been evaluated – what mix of monetary and non-monetary benefits are to be provided? should benefits be related to jobs or performance? should they be provided immediately or deferred? should they be provided according to the absolute contribution of the individual or to his or her contribution relative to others? These questions are formulated in a prescriptive way (i.e. how human resource systems should be configured) but they could equally be formulated in a descriptive way by asking why certain solutions are adopted and are effective in certain circumstances.

EVALUATION

Types of evaluation

What forms of evaluation are theoretically possible and applied in practice? Figure 9.2 illustrates the difference between evaluation systems, based on the distinction between human resources, actual activities performed, and the expected activities relative to a work position or job. The main forms of evaluation considered in the Human Resource Management (hereafter abbreviated to HRM) literature are evaluations based on *positions*, *skills* (or *competences*), *performance*, *and potential* (Costa 1997). These elements can be ordered in a causal attribution chain that runs from resources and competences through behavior to present and future results.

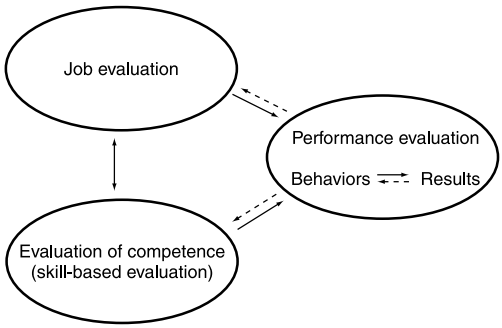


Figure 9.2 The causal attribution chain in evaluation systems

The evaluation of results and behaviors is defined globally as *performance evaluation*. One of the core problems in designing a mechanism of performance evaluation is the choice between the evaluation of behaviors, that constitute the input in productive processes, and the evaluation of results, that constitute the output. The scheme proposed here highlights *the knowledge problem and the causal attribution judgments underlying the choice between evaluating results or behaviors, or competences and professional qualifications.* The problem consists in the fact that the further one moves away from the characteristics and attributes of actors toward results, the less the performance is clearly attributable to the actor being evaluated. This is especially true in activities subject to high levels of uncertainty. This recurrent problem explains many of the actual choices made in the design of effective evaluation systems.

The basic function of performance evaluation is that of revealing and measuring the

value of the contribution of identified actors (single or groups) to the system of action. This data base is then, if necessary, used to make decisions about reward and mobility based on the contribution of each specified actor. Implying a problem of empirical measurement of a complex concept ("performance"), for which valid and reliable indicators are not always easily available, performance evaluation can be very difficult and often biased, as will be illustrated.

An alternative type of evaluation, that can complement or substitute for performance evaluation is the *evaluation of work positions or job evaluation.* For, even from a historic point of view, job evaluation has become diffused, often with the involvement of trade unions, as a mechanism to make the evaluation system less arbitrary and more transparent, through a standardization of the judgments and rules linking work and reward. More specifically, the contribution that is made by a position, without regard to the specific person who occupies it, has been considered a fair and efficient basis to design the structure of wages across jobs in an employment system. A possible justification for this belief is that the evaluation of a job, in terms of the competences and responsibilities it implies, can be seen as a measure of the average contribution expected from a given collection of activities (i.e. "the job") to a system. Thus, this can be seen as a substitute (or proxy) variable for a more specific and precise but more difficult and unreliable evaluation of the actual performances of each single actor.

Another important function of job evaluation is its capacity to consider important aspects of work that do not depend on the worker's actions, but on the context or the nature of the job. If evaluation were not done (also) on the basis of jobs, these contextual factors would not receive sufficient attention and care, and the lack of specific incentives and compensation for certain types of work would make the work unattractive for job-seekers. In fact, the responsibilities and risk that some activities can require of a person to the advantage of the system are job attributes rather than performance or people attributes. For example, the driver of a school bus may be perceived as being in a position of high social responsibility, involving substantial risk, even though it involves competences that are commonplace. If school bus drivers were paid only for their driving skill or performance, this would be neither fair nor efficient for the system (in the long run, it would be hard to find and motivate good school bus drivers willing to assume the necessary responsibility and take the necessary care) (Lazear 1995).

On the other hand, the competences and the responsibilities of an actor cannot be adequately evaluated only from his position. Even in role systems that are highly prescriptive and formalized, competences may be underutilized or hidden. The more jobs involve discretion about how to interpret one's own work, the more difficult an appraisal of the value of a work role independently of how it is played becomes. A response to these difficulties can be provided by *competence-based evaluation*, to the extent that it is intended as an appraisal of *actor-specific competences* and not of the standard competences required by a job description.

Finally, if evaluation processes were oriented only to existing positions, existing resources, and observed performances, they would shape a very static and past-oriented action system. Such a system would not generate very relevant information for the governance of mobility even in static structures (hiring and promotion decisions), much less

for the development of new competences or new activities. In response to this problem, a series of instruments has been developed for the analysis and appraisal of human resources' *potential*. This is a future-oriented evaluation, involving conjectures and expectations about the possible performances of a person in new roles, rather than something observable such as performance or competence. Measuring a person's potential involves estimating the possible variations and future development of this actor's competences and performance in different tasks and positions, even if these positions cannot be (or have not yet been) completely defined. Therefore the evaluation of potential is generally more subjective and less reliable, based on hypotheses and judgments under strong uncertainty about possible future combinations of resources and activities. Because of this, when evaluating potential, it is particularly important to rely on validated techniques and methods to minimize the likely errors. Upon this premise on the nature of evaluation and appraisal processes, we can enter into the merits of each system.

Performance appraisal

Some form of performance appraisal is implied in all types of work relations. Evaluating past and expected performance is necessary for a variety of reasons: to establish the price for a given package of work services so that those services can be purchased on a labor market; to allocate rewards across employees inside a firm, to hire personnel, or to staff a work group.

Observability

To the extent that evaluation is primarily a process of acquiring knowledge and information, the efficient and effective use of the evaluation of behaviors or of results is conditional to the *observability and measurability of the inputs and/or the outputs* (Ouchi 1979). As widely pointed out in Part II, many economic behaviors are not directly observable and measurable either for logistical reasons or owing to asymmetry of information and competence (e.g. the behavior of a salesperson or a highly specialized lawyer). Many types of results may also be difficult to judge or evaluate, because they can only be seen over a long period, or because they are multidimensional or because the information is very difficult to obtain (e.g. the results of a trainer of executives).

The mere observability of behavior or results, however, should not lead the design of evaluation systems. For results may be observable but not causally attributable to certain actions, while actions may be observable but the causal relation with results may not be well known. In general, the advantage of using behaviors as performance indicators is their greater correlation with the resources that generate them; while the advantage of evaluating by results is their greater correlation with the final objectives of the action system.

Knowledge of input–output relations

If behaviors are observable and the link between actions and results is known, then performance can be evaluated on behaviors. Evaluations can be based on the deviations from standards and expectations, because it is known what the best model of action is. For example, one can evaluate how "good" the activities involved in producing beer cans are, because it is known what operations will produce good results.

Imputability of results

On the other hand, even if results are observable and measurable, a further problem is represented by the extent to which they can be *causally attributed* or "imputed" to the evaluated actors. For example, in sales, the results are, typically, easily observable and measurable via the generated income, market shares, client portfolios etc., while the behaviors are hard to observe and often not well understood and not standardizable. This contingency should not lead to exclusive evaluation on results, unless the *level of external uncertainty*, and of *dependence on other actors' behaviors* allow these results to be attributed to the salesmen's actions and efforts. The effects of these variables are explored below.

If difficulties in causal attribution originate from the fact that there is significant variance in measured results, potentially due to exogenous factors, or from the fact that results are observable only after considerable time, the appraisal of result performance can be operated over an extended time series of observations – averaging out random variations (Campbell 1969; Milgrom and Roberts 1992) – or over extended time periods (Lawrence and Lorsch 1967).

If difficulties in assessing performances stem from dependence on other actors and actions, they can be responded to in yet other ways.

Suppose that actors can behave independently, but that their results are influenced by what other actors have done. An example is the sale of new and used cars described in Box 9.2. This condition makes an evaluation of the results of each actor difficult. The problem can be responded to in various ways:

- by integrating the evaluation of results with an *evaluation of input behaviors*, if they are observable (for example, in sales activities, the respect of territorial zones, the use of specific sale techniques, the delivery of specified post-sale assistance services).
- by *widening the considered result parameters* so as to make the composite result indicator more clearly attributable and more specific to each actor considered (for example, by considering qualitative parameters as served client satisfaction).
- by *evaluating the collective results* of the interdependent actors (through indicators at the group or even the firm level).

Interdependence among inputs creates further problems. Suppose that activities are linked in a series with others upstream and downstream (e.g. different operations involved in a continuous-process technology such as rolling steel) or are linked in parallel with other activities performed on the same object being transformed (e.g. operations performed by different maintenance specialists on a large production machine). In these linked activities, the main objective is often not to maximize productive results, but instead to optimize a composite process, which will involve quantitative, qualitative, and time elements. In these cases, evaluation of behaviors is also called for.

In general, therefore, effective evaluation systems will often be mixed and composite in situations of interdependence rather than one-sidedly based either on inputs or on outputs.

There are also activities that involve very high information complexity in which neither inputs nor outputs are observable and causally attributable at reasonable cost. In these circumstances a resource-based evaluation (see below) can substitute for a performance-based evaluation.

The combined application of the above design criteria can be experimented upon by

discussing the case illustrated in Box 9.2, describing an automobile dealership. The example also enables one to see how developing evaluation parameters is an exercise in the "operationalization" of the concepts of "results" and relevant "behavior" into observable indicators.

Evaluation processes

A good architecture of an evaluation system according to the above criteria does not guarantee that the system functions well. The process and technical instruments are also important. Especially in cases where the evaluation parameters are not "objective" (i.e. they are completely dependent on the estimates and judgments of an evaluator), the process may be subject to strong biases. In the most frequent cases, the evaluation is done through questionnaires directed to an evaluator containing items about the performance of the employee being evaluated. The evaluator is usually the employee's superior, but could be, and increasingly is, a user of the employee's services, or a colleague who has worked with the employee on a project or activity.

As in all research based on questionnaires, evaluators should be concerned with the *validity* of the performance measures (do they actually measure what they are intended to measure?) as well as with the *reliability* of the scales (do they yield stable results over repeated administrations in the same context?).

In addition, it should be remembered that the responses to the evaluation questions may be subject to specific distortions due to underlying conflicts of interests, which are particularly strong if appraisal is linked to reward. In the specialized research on evaluation processes, the inventory of biases includes: "*representativeness biases*" (allowing stereotypes and prejudices to influence the evaluations of the employee's performance in a role or position); "*availability biases*" (giving greater weight to emotionally charged events and interpersonal familiarity); *anchoring errors* (evaluating based on small changes to previous evaluations, or giving similar judgments on all the evaluation parameters); and *moral hazard errors* (deliberately distorting judgments either to damage or to avoid damaging the employee being evaluated).[1]

With reference to the management of the intrinsic conflicts of interest underlying the evaluation process, procedural justice has naturally been stressed, given that an evaluation system is, in an aspect, a judicial system. The essential characteristics of a procedurally just evaluation system have been specified as shown in Table 9.1.

Job evaluation

Jobs are partitions of activities of a larger system that can be "held" by people or groups as collections of rights and obligations to perform certain actions. Job evaluation is a comparative judgment of the relative contribution of these collections of activities, independent of the performance of the specific job incumbents.

The first problem to resolve in a job evaluation is to obtain accurate and comparable job descriptions of all the jobs in a given system. This might seem simple, but the activity of obtaining job descriptions may require extended and systematic organizational research, based on a specialized repertoire of investigative techniques such as interviews, questionnaires, and field observation. These job descriptions generate a certain level of formalization and are considered particularly important in large systems (or in interfirm

Box 9.2
The Matt dealership

Matt has been for twenty-five years the dealership of Beta – one of the largest automobile manufacturers in Europe.

Mattioli is a typical auto dealership operating on four "lines": the "new" line, the "used" line (used vehicles purchased or traded in for new ones), the "parts" line (parts, accessories, and oils used by the internal repair shop or sold to clients), and the "repair shop" line made up of two sections: the "mechanical" one and the "body shop" one (prepares new vehicles for sales, performs repairs on clients' cars – both under guarantee and not – and reconditions used cars before selling them).

Because of the numerous conflicts among the company's employees, the top priority seems to concern clarification of responsibilities and thus the criteria for performance evaluations of the various divisions. Examples of conflicts among employees include: the used car salesmen, who often sell at a loss, claim that the repair shop raises the price of the car too high by performing too many reconditioning touch-ups, while at the same time the repair shop manager blames the parts division for the high prices it imposes. Below are the characteristics of the divisions of Matt.

The "New" Line

The strategy of this line is relatively rigid given that products, markets, and prices are set by Beta. On new vehicles there is a fixed gross margin of 14 percent so the division's profitability needs to be evaluated mostly in terms of efficiency. In order to stay competitive it is very important that the dealership maintains a good image, based on continuous and intense work by the salespeople, who must grant small discounts and above-market prices on vehicles traded in (which makes it harder for the "used" line to make a profit), as well as additional services such as financing options and free maintenance (which jeopardize the repair shop's profitability). The management of these tools is assigned to the salespeople; it is therefore necessary to set specific rules and check that they are implemented properly.

The "Used" Line

The management of the "used" line allows for more reasonable actions: purchase and sales prices are not set by Beta, and product quality depends on the reconditioning and guarantee policy followed. Because there is not the minimum demand "guaranteed" by the trade mark on new vehicles, the marketing strategies are of crucial importance. However, a large percentage of purchases of used vehicles comes from "trade-ins" for new vehicles; this means that the estimated value of the used vehicle made in such instances influences the profitability of the line very heavily. Another important cost element – which determines the quality of service – is the reconditioning phase (besides the possible guarantee). The cost of repairs determines the financial result of both the "used" line and the repair shop (which fixes the amount). Finally, in order to better understand this activity, it would be correct also to consider the cost/opportunity ratio of the inventory of cars from the date of their purchase to the date of their sale.

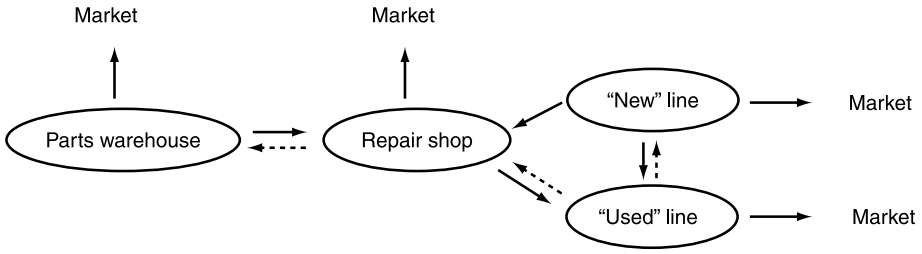
The Repair Shop

In the Matt dealership, the repair shop works on new cars (preparing them for delivery, inspections, and guarantees), on old cars (reconditioning and guarantees), and on vehicles not necessarily bought there (repairs on external customers' cars). The management of the repair shop directly influences the profitability of the other lines, and is influenced by them, so there are strong interdependences across lines. Income earned by working on new cars should therefore represent a revenue for the repair shop, and a cost for the new line. The same is true for reconditioning work: in these cases, though, the determining factor is the transfer price that the supply room applies to parts. While internal transactions represent a "captive market," the repair shop can be more reasonable with external customers and the profitability of this division depends on the marketing strategies that are adopted.

The Parts Department

The management of the parts department may be quite different depending on whether its main objective is providing service to the internal repair shop or to the "external" market made up of auto parts dealers and mechanics. When dealing with the outside market, the parts department can operate as a "separate enterprise" (in this case the key to success consists in always stocking a very large selection of parts), while when dealing with the internal repair shop the goal is always to stock the most frequently used parts (ordering stock from the distributor based on single orders). As far as the relationship between the supply room and the repair shop is concerned, it is important to keep in mind that profits generated for the parts department are costs for the repair shop. In addition, in evaluating the results of the auto parts division, one must keep in mind the cost/opportunity connected to the rate of turnover on parts, which is currently too high compared to the average of the Beta dealerships. One must also keep in mind the quality of the service provided, which does not always correspond to the expectations of the internal and external customers.

Given the interdependences among the four lines (summarized in the Figure below) and the different parameters of reference of the activities, putting some order into the management of the dealership appears to be no easy task. The existing relationships among the lines of the Mattioli dealership are as follows:



Note: Arrows indicate transfers of goods, services, or information.

Source: Adapted by Massimo Neri from Airoldi (1979).

Table 9.1 Due process and its relationship to the practice of performance appraisal

Element of due process	*Descriptions relevant to performance appraisal*
Adequate notice	Objectives and standards are established in advance, published, widely distributed, and explained.
	Employees have input into formulation of objectives and standards, or at least opportunities to question the content of the standards and objectives, the process whereby they were established, and the manner in which they will be implemented. Feedback is given on a regularly recurring and timely basis.
Fair hearing	Standards for the admissibility of PA "evidence" include the appraiser's familiarity with the appraiser's performance based on sufficiently frequent observation of behavior or work products.
	Employees have means to indicate their own viewpoint concerning their performance.
	Employees have opportunities to explain their own interpretation of PA "evidence" and present arguments supporting that interpretation.
Judgment based on evidence	Steps are taken to have the appraiser apply standards consistently, without external pressure, corruption, or personal prejudice.
	Evaluations show efforts to use principles of honesty and fairness (employees have opportunities to question evaluations, and the explanations provided reflect such principles).
	Evaluations withstand scrutiny, including that which might be engendered by an appeal or other type of opportunity for recourse provided to employees.

Source: Folger *et al.* (1992).

systems, such as a franchising structure)., both for reasons of transparency and fairness and for reasons of informational efficiency in making comparisons among several jobs.

The second problem to solve is the identification of the evaluation parameters. The relevant question should be: what aspects of the job are more correlated to the creation of value by that activity, independently of the particular job incumbent? As already observed, there are at least two components to be distinguished in this respect: *knowledge and competences* on the one hand, and *responsibilities and risks* on the other.

The most common method of identifying relevant parameters and for attributing them a weight or value is based on empirical correlations among the presence of certain dimensions or factors and the level of compensation generally observed. On this basis the evaluator can construct a scale on which a position scores more points (and is thus eligible for higher compensation) to the extent that the particular dimension or factor is

present. The most widespread and applied method has been developed by the consulting company Hay: it has identified in the *competences required to cover a job*, understood as knowledge to be mastered, *the intensity of problem-solving capabilities required to perform the job, and the level of responsibilities* – the three main parameters on which to assess the value of a position. These are in turn operationalized in very detailed measurement scales that are supposed to be sensitive to variations over hundreds of points.

After a period of notable diffusion, this type of approach came in for criticism. Beyond the criticisms addressed to the content of parameters – which can be remedied through the addition of other dimensions – there are more methodological caveats. One criticism is that the use of points to express standard values to assign to the contents of a job type, conceals the differences in value that similar or identical activities may have in different systems – that is, standard scales tend to confound job specificities. A second criticism is that whenever the construction of the points in the job evaluation scale is based on statistical correlation analysis among various types of positions and the average compensation observed for those positions, a circularity problem arises if job evaluations are used to determine pay levels. One way around this second criticism is to use the average market compensation for the job to establish a *lower bound* on the fixed compensation for a generically defined job, since it is a good indicator of what a job incumbent could get in a similar alternative job, without consideration of any possible quasi-rents or specific value of that job in the focal system. In practice, in fact, wages are partly determined on the basis of job evaluations and the average salaries for similar jobs, and partly negotiated based on the particular surplus created by the particular contribution of jobs and people in the specific context.

Competence-based evaluation

The *evaluation of human resources as collections of competences* can constitute an alternative to job evaluation, especially in dynamic or complex activity systems in which the contents of jobs depend largely on actors' competences rather than vice-versa. Techniques of skill-based evaluation are in fact being more and more widely used. An important distinction in these techniques is the extent to which they are geared to capture the advantages of the specificity and specialization of competences in given tasks; or the potential of competences in the generation and definition of tasks (Chapters 1 and 8).

"Task-driven" approaches

One approach – developed in the field of human resource management (Spencer and Spencer 1993) – has aimed at discovering which mix of competences can be linked to superior performance in a *given* activity. The method of competence analysis generally involves the use of observations and structured interviews[2] that reveal which knowledge, behaviors, and action procedures the employee has used to get the best results. These elements are then used to construct a *model of competence* that can be copied and diffused, and with respect to which the level of competence of job incumbents can be measured. It is supposed that these evaluations can be used for staffing, hiring, and promotion decisions, as well to match reward with competence levels. The effectiveness of this procedure, however, involves certain conditions. It implies an imitative learning process, that is appropriate for fairly large

systems of stable and similar activities.[3] It is "*task-driven*" – that is, it requires that tasks are given in order to define relevant competences. It offers a "standardized" approach to the problem, that may be successfully applied if competences are codifiable and it is known that they are common to many jobs.

"Resource-driven" approaches

Knowing that imitation and diffusion is not the only model of learning possible (Chapters 2 and 3), one can envisage modes of learning of superior practices that are better suited to complex and differentiated activities. Instead of simply "cloning" models of competence, one can construct causal models of the results obtained, specifying the conditions under which certain competences and actions produce certain effects, and taking into account the effect of exogenous and specific factors. In this way, these models may help to generate new profiles of competence instead of transferring models that are already observed. Hence they may be defined "*generative*" and "*resource-driven*." For example, if a consultancy firm has to organize a project group for a specific client/issue, the process, explicitly or implicitly, involves these evaluations:

- assessments of who possess relevant competences.
- estimates of which of these competences can develop more effectively and efficiently in providing relevant services (generative potential).
- judgments on which combinations of competences can manage the problem (combinative capability).
- calls for participation to relevant competence owners, who may evaluate the interest of the project (or participate according to previously set agreements and plans).

The case of competence-based evaluation described in Box 9.3 gives some material for identifying elements of both approaches, and for discussing some of the prospects and problems of approaching and using competence assessments in a non-standardized and generative way. In particular, it shows that – if evaluations are used for compensation purposes and not only for resource development – this may require a move toward a non-standardized, negotiated, and ad hoc approach in reward systems too.

Assessment of resource potential

The *evaluation of human resources potential* is aimed at estimating the capacity of the development and future performance of people in activities that they have not previously performed. It is an exercise in estimating the generative potential of resources (Chapters 1 and 8). As applied to human resources, this constitutes an important input for the design of mobility and development systems.

The difficulty of precisely and reliably evaluating potentials increases with the complexity and uniqueness of the past and future activities. In fact, when the contents and the characteristics of the job tasks are known in advance, so that the causal relationship between competences and results is clear, then it is possible to predict future performance with valid and reliable indicators. For instance, university admission tests can be based on correlations between the presence of certain capabilities and obtained grades, for very large numbers of subjects. Large hiring processes in banks and insurance companies or promotion plans within the same functional area are also examples. Objective indicators such as the possession of titles or degrees or professional qualifications can be combined with measures of past

Box 9.3
Competence-based evaluation at Andersen Consulting

Human resource management at Andersen Consulting is characterized by the company's focus on the concept of competence, which is defined by the firm as *the collection of know-how and skills necessary to successfully carry out activities related to business objectives*. Each individual's personal growth is focused on the acquisition of competences that are in line with Andersen Consulting's marketing strategy, and is personalized in a way that takes into account various interests, orientations, and individual needs.

Once the five core competences had been defined (Architecting Business Processes, Architecting Technology, Setting Business Direction, Changing People and Organizations, and Managing Complexity) the first step was to evaluate the portfolio of competences existing at the time.

In the Andersen case, unlike the classic method by Hay-McBer of analyzing the "best performer" in order to identify winning competences for the organization, the competences that were taken into consideration were "drawn-up" by the partners. The reason for choosing a "deductive" method is explained by the need to change the direction that the company was going to take; the best performers of yesterday would not necessarily be the best performers of tomorrow.

Once the collection of best competences had been defined by the Competency Head, the next step was to survey the existing portfolio of competences. The method used starts with each professional filling out a self-assessment questionnaire about his or her competences. These data are then double checked and standardized by the Competency Head, who aggregated and tabulated them.

The analysis of the gap between the actual competences currently possessed by the organization and the future needs (then defined as strengths and weaknesses) has made it possible to develop personalized training programs and competence schedules.

At the individual level, there is a "performance appraisal." At the beginning of each project, roles and objectives are defined, and every four months there is an evaluation of the skill domains that have been developed on the job. This performance appraisal thus leads to the determination of the proficiency level reached, along with additional comments in case the evaluation differs from the standard level of proficiency expected (i.e. what the company expects a person "to know and to be able to perform" in order to qualify for a specific position).

The method of using an evaluation scale consisting of seven levels (where 0 = no competence and 6 = acknowledged leader in the area in question, even considering experts outside the company) seems to be an attempt to introduce a quantitative element in a company where everything is based on qualitative evaluations by one's boss – evaluations that are influenced by his/her expectations.

In addition, one could wonder what actually counts more in an evaluation – having accomplished a task or having demonstrated specific competences?

As to compensation implications, at Andersen Consulting the "pay for competence" concept is not strictly applied even if the actual competences are used when

deciding on bonuses. The fixed component of pay is strongly anchored to one's hierarchical position, while the variable component depends on the level of proficiency that the individual has demonstrated in the projects in which he or she has taken part. They refer to *proficiency* rather than results because the profile of the person changes in line with his/her *competence* (both content skill domain and professional qualities) independently of the results of the project.

Just before promotion times (September and March), area manager meetings are called and the final personnel evaluations take place. Based on such evaluations, people are categorized into four bands (top performer, full performer, performer, and major problem) to which correspond specific compensations and career plans.

Careers are closely dependent on the performance bands; if someone is a junior top performer it will take him/her two years to become senior, while if he/she is in a lower band it may take him/her longer. Although compensation is also influenced by the bands, the linkage is not deterministic; not necessarily do all top performers receive the highest compensation. In this case, in fact, variables such as age, seniority, and the importance of the work performed play an important role in distinguishing one person from another within the same performance band.

In summary, at Andersen Consulting performance evaluation is done by the line managers, and the salary raises by the practice managers even though they need to be approved by the partners. However, the partners decide the compensation for each position while the positioning of the single individual within a band is some kind of a negotiation between partner and human resource manager.

As to the compensation, given that the "market" does not pay for competences, Andersen Consulting relies on benchmarking, looking at a set pool of companies (not necessarily all consulting companies) and within them at specifics such as the earnings of a college graduate of the same age in a certain position. From data revealed in a small survey, it does not look as if the staff are unhappy about their compensation.

However, once it is recognized that additional competences need to be rewarded financially, the problem remains to decide whether to pay for the competences actually used in job performance (as happens in, Deloitte & Touche, for example) or to pay for the additional competences developed, even if not used on the job (as for Andersen Consulting USA for very specific and specialized professional figures). The choice of Andersen Consulting Italy seems to lie in between the two positions.

Source: By Christian Montermini.

performance, and tests designed to measure one's expected performance in a job or a progression of jobs.

However, things are not so straightforward when the job is difficult, the competences are complex, and the activities are very different from anything performed by a particular subject in the past. If the evaluation of potential regards newly created jobs or involves a different context such as a new firm, a new work group, or new markets, difficulties proliferate. In these cases the evaluation of resources' potential becomes more subjective and is based on complex certifications and professional judgments. Informed judgments about the competences and the general characteristics of the subject will carry more weight in the evaluation process. Such informed

judgments will often be based on direct observation and field experiments, or on in-depth interviews, assisted by "clinical" and "expert" judgment as well as by referees and presentations (as in fact is the custom in admission to post-graduate studies) and on the self-assessment of the persons themselves.

REWARD

The *evaluation* of competence and performance is an input that – combined with an appraisal of *preferences* – enables us to define rewards. Keeping their competences and performances constant, different actors may attribute different values to different reward packages. In particular, people can attribute a value to both monetary rewards (or rewards that are easily convertible to money) and non-monetary rewards such as the possibility of professional and career development or the intrinsic content of the work (Herzberg 1953; Chapter 2). The rewards derived from career and development will be treated in the next section, while those deriving from the contents of jobs will have to wait until the next chapter. Here we concentrate on the monetary or quasi-monetary elements of the reward system.

Job-related compensation

Consistently with the major distinctions among types of evaluation systems, monetary compensation can be linked to jobs and time (also called "fixed" pay) and/or contingent to performance (also often called "variable pay" or "incentives").

The efficiency and the effectiveness of a pay scheme linked to a job (rather than contingent on performance) rest on several assumptions. The most important of these are (Chapter 4): that the actors contributing work are risk averse; that they sell to other actors (work employer) their work services and some rights to decide on their use over time; that the employers can either observe the workers' behavior or confidently infer it from the results. In other words, *job-related salaries* can be justified as *part of an exchange-based authority relation or "dependent employment relation."*

The problem of designing or explaining the appropriate *levels* of these kinds of wage can be solved in steps, considering the different *variables that may affect pay levels* and the *different decision processes* (Chapter 2) that may lead to their determination.

Optimal fixed pay levels

A minimal requirement that an efficient wage level should satisfy is that it be sufficient to motivate a worker to *enter the transaction* and to *produce at the desired level*.

This problem can be solved as an optimization problem if certain information requirements are satisfied. If the worker assigns a positive utility to monetary compensation and some negative (and/or positive) utility to the effort expended, and he can compare them; and if the employer can foresee the expected utility that can be derived from different actions and levels of efforts of the employee, then an efficient level of pay should satisfy the following conditions.

- The net utility for the worker – the positive utility of the salary minus work-related costs (opportunity costs, effort, transportation, material consumed) – must be a positive value.
- The worker's net utility should be greater than or equal to the net utility offered by alternative jobs available to him or her.
- For the employer, the optimal set of actions

to demand from the worker, is that for which the difference between expected earnings for the firm and the cost of labor required to obtain it is maximized (Levinthal 1988; Douma and Schreuder 1992).

Uncertainty, quasi-rents, and negotiated salaries

The above conditions are far from providing a complete framework for predicting or calculating wage levels.

In the first place, it is probable that in most situations this type of calculation will be done more heuristically, owing to the difficulties involved with the required evaluations and predictions. For example, it is not likely that the worker will know how to explicitly compare the costs of production of work services with the utility of rewards. In general, it is difficult to compare work contributions and rewards, because while benefits come partly in money, it is hard to think of effort in monetary terms. Empirical research has shown that workers evaluate the acceptability of their compensation more on the basis of comparisons with the compensation received by similar workers in similar jobs (Lawler 1971, 1973) than of comparisons with their own costs – that is, based on equity judgments more than on efficiency judgments.

In addition, the above criteria would set pay close to the best alternative salary that a worker can get, or at least do not provide guidelines on how much higher a fixed pay can be over that minimum level. To that extent, they neglect the likely presence of firm-specificities and quasi-rents in the long-lasting work relations in which fixed pay is typically efficient (Marshall 1890; Aoki 1984). In Marshall's words,

> the point of view of the employer however does not include the whole gains of the business: for there is another part which attaches to his employees. Indeed, in some cases and for some purposes, nearly the whole income of a business may be regarded as a quasi-rent, that is an income determined for the time by the state of the market for its wares, with but little reference to the cost of preparing for their work the various things and persons engaged in it. In other words it is a *composite quasi-rent* divisible among the different persons in the business by bargaining, supplemented by custom and by notions of fairness.

In the labor relation context, a *rent* is *the proportion of compensation that a worker receives in excess of the minimum necessary to be induced to accept a particular job. Quasi-rent* is defined, instead, as *the portion of compensation that the worker receives in excess of the minimum necessary not to leave a particular job* (Milgrom and Roberts 1992).

Rents and quasi-rents derive from the presence of some monopolistic elements (Chapter 8), which, in the case of work, can be the possession of very scarce and valuable competences and talents, or to competences that have a very high value in a particular relation (specificity). For example, the compensation of opera singers involves rents derived from the possession of scarce and useful resources (i.e. talents that can create value in the marketplace). A different source of surplus compensation is specificity. A carpenter with firm-specific competences can enjoy a quasi-rent that includes the greater value he generates in that firm with respect to alternative candidates, minus the costs of search for and transfer to a new job.

Wages should include and often do include part of the quasi-rents that work contributes to creating (Masini 1978; Aoki 1984). It would not only be unfair, but is unlikely in practice, that these quasi-rents will be appropriated entirely by employers. The more uncertain reservation prices are, the less

controllable is workers' behavior; and the more long-lasting the relation is, the less likely it is. If the parties lack clear information on counterparts' reservation prices, they are unlikely to get very close to them (Chapters 6 and 8). In addition, workers can win back what they lost in accepting a deemed unfair contract, by raising the level of free-riding and non-compliance at work. Therefore *a fair division of work-related rents has self-enforcing properties that other allocations would not have.*

Transactional advantages of unions

These conditions seem to apply particularly well to firm-specific human resources, that in fact usually enjoy highly protected and negotiated work conditions. One can ask, then, why wage negotiation, assisted by unions, within long-term employment contracts, is such a widespread system of industrial relations governance and is not confined to firm-specific and poorly monitorable work.

Among the reasons are the following.

- If there are significant information asymmetries (e.g. the value of alternative jobs for the worker is known to the employer but the value of alternative workers to the employer is not clear to the worker), or resource asymmetries (e.g. the relative amount and specificity of resources invested in the relationship are much lower for the employer than for the worker) then there is a substantial risk of unilateral appropriation of value by the employer. Hence, the formation of a "*countervailing power*" coalition on the side with less bargaining power is rational (Galbraith 1952; Goldberg 1980).
- The negotiation of labor contracts has high transaction costs; therefore the unification of job requests and the interaction with one unionized counterpart rather than with many individuals, *especially if jobs are similar and standardized*, makes the process much more efficient (Freeman and Meadow 1984; Costa, 1990).
- In negotiations likely to involve a lot of conflict such as those between workers and employers, the *longevity of the relationship* with the same counterpart facilitates an agreement (Chapter 6). So even if many individual workers do not have a long history with the firm, the union can represent a stable negotiating partner for the firm.
- The value assigned to a contract, especially in the case of labor contracts, and its acceptance, depend not only on the amount of rewards obtained, but also on the *justice of the procedure* through which the agreement is reached. Trade-union bargaining over pay can be justified as a fair voice-giving procedure on the matter.

Reward contingent on performance

"Incentives" and "variable pay" elements refer to those components of compensation that are contingently determined by rules of correspondence between pay and performance. Its distinctive contribution to the regulation of work transactions is that through adequate incentives it enables motivation of work activities that (1) cannot be, at least in part, governed under an authority and job-related wage regime; and (2) cannot be totally regulated by instantaneous exchange contracts. Typically, therefore, variable or contingent pay is useful in compensating work services *that are to be delivered over time, in which workers have discretion about what action to take, when these actions cannot easily be observed and evaluated, provided that performance can be reliably measured and can be confidently attributed to the worker's behavior.*

In fact, pay based on results is especially relevant in compensating those work services that are provided under an agency relationship both within and across firm boundaries; and the determination of the level of incentives has been one of the areas of application of agency theory. From this perspective, the problem is solved with a value maximizing strategy.

Optimal incentives

The core problem is that compensation contingent on results *transfers risk* from a less risk-averse actor (the employer) to a more risk-averse actor (the worker). The employer can generally be supposed to be better able to bear the risk owing to more diversified investments and the lower impact of any single action and contract on total wealth; while the agent, for the opposite reasons, is supposed to be typically risk-averse. There is thus a *trade-off* between greater incentives to produce, generated by the risk transfer, and greater labor costs, due to the greater sums required to compensate for risk acceptance. The basic formulation of the problem supposes, in addition, that the agent evaluates possible actions according to the costs and benefits they yield, where costs are operationalized as "efforts" and benefits as compensation. The optimal intensity of performance-related pay will then be a function of the following variables (Milgrom and Roberts 1992):

- the value created by the extra effort;
- the incidence of exogenous variables (not under the control of the agent) on results;
- the degree of risk-aversion of the agent;
- the rate at which the marginal cost of effort increases for the agent.

The basic structure of the argument can be summarized in the trade-off between the two functions shown in Figure 9.3. The marginal benefits from the greater effort, net of the costs of these efforts, is a function that decreases as the incentives increase (because the marginal costs of the intensity of work increase while the marginal effects on the results decrease). The marginal cost of the incentive is instead a function that rises with greater risk transfer to the agent, and rises more steeply the more the worker is risk-averse and the greater the variance of results due to exogenous factors.

This formulation and solution to the problem can be extended and qualified in various ways. In the first place, the assumption of a strictly quantitative and continuous shape and content of utility functions can be modified. Agents, and managers, in particular, are likely to order business actions according to utility scales that are different from those of their principals, for example shareholders (Marris 1964, 1997). However, managerial preferences are usually linked not so much to the level of effort required, as to qualitatively different views of business policy or to social and power returns rather than monetary ones.

In addition, the feasible decision process may be less close to an optimizing strategy

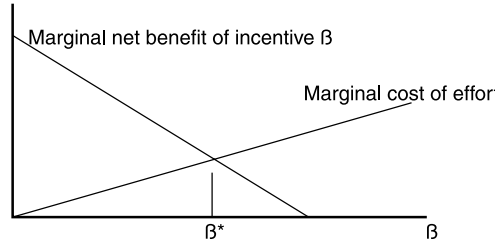


Figure 9.3 Optimal intensity of result-based incentives

Source: Adapted from Milgrom and Roberts (1992).

than the agency model expects and prescribes. Firms and managers do take into account the types of variables highlighted by agency theory, but they do so in a more heuristic way, based on levels of aspiration and performance targets to be reached, especially for complex and poorly structured activities involving a high level of discretion. A brief illustration of some of the most widespread practices and forms of result-based compensation, and of the conditions under which they have been found to be effective, supports this argument.

Management by objectives

One of the most common systems applied at the managerial work level is *management by objectives (MBO)*. In MBO specific levels of outputs to be reached are forecast and then used as criteria to evaluate the levels of performance reached by the responsible actors. MBO is based on a tight linkage between three subsystems: the planning and control system (that formalizes the quantitative economic objectives to be reached); the performance evaluation system (based on a comparison between planned and achieved results); and the compensation system (a contingent incentive scheme). From a cognitive point of view, MBO can be seen as a systematic and formalized version of *goal-setting* (Locke 1996; Chapter 2), that is particularly suited to activities with a lot of discretion about the type of objectives to reach and the best ways to reach them.

An important feature in MBO is the requirement that the process through which objectives are set be participative. A specific and economic justification for this requirement, if activities are uncertain, is that participation can help in overcoming the information asymmetries and the opportunism potential that are typically present in MBO. For, if both the amount of variable incentives and the formula to calculate them are established in a participative way, the incentives to set easily reachable targets, on the part of managers who enjoy exclusive information about the difficulty of reaching various possible objectives, could be diminished. The process of joint definition of objectives and of the rules for contingent pay can indeed be seen as the result of a proposal made by the employer of *a "menu of contracts"* and of the choice from that menu by the manager of the contract that s/he expects will lead to the best result (Milgrom and Roberts 1992). If the manager is confident of reaching a difficult objective, then s/he will choose a risky contract (i.e. with a higher variable compensation component); otherwise, s/he will choose a more conservative scheme (i.e. with less incidence of incentives).

Although compensation on results is particularly important for employees in positions with wide responsibilities and decisional autonomy, various forms of incentive can also be used effectively for operative positions, especially those involving discretion about work sequences and procedures and the use of time.

Piece-rate pay systems

One form of contingent pay traditionally used for blue-collar workers is the piece-rate system. Piece-rate systems aim to create an incentive for individual workers to an efficient use of time and resources so as to maximize the number of units produced and compensated for. The rise of this form of incentive has been connected with the diffusion of Taylorism, as a complement to jobs that were parceled out and studied to find the most efficient method of obtaining the greatest

productivity in the shortest time. Its decline can be attributed partly to the diffusion of automation in industrial transformation processes, partly to the vicious circles inherent in the system, and partly to the shrinking of the set of industrial activities that can be "Taylorized." For while, on the one hand, automation has often reduced the control on work paces and systems on the part of workers, on the other hand, vicious circles have stemmed from the incentive for employers to raise normal production standards in response to increases in productivity, thus leading to an escalation of the efforts required to obtain premiums contingent on the high productivity generated. Finally, less specialized and individual models of labor organization, such as job enrichment and teamwork, have come into more widespread use. With less work being divided among individuals and more work done by teams, the efficient worker must not only be efficient in the use of time, but must also learn to manage relations with coworkers, solve problems, and work effectively in a team.

Gainsharing

A form of contingent reward that is more appropriate to work taking place in teams and requiring problem solving is *gainsharing*, which involves paying members of a group a share of the productivity gains earned from the increased performance of that group. Gainsharing can be claimed to have positive effects if group members have concrete chances to influence results through the exchange of relevant information and participation in decision-making – for example, through "quality circles" and other applications of group problem solving to production quality improvements (Mitchell *et al.* 1990).

Profit sharing

The two forms of incentive just analyzed (piece rates and gainsharing) are based respectively on individual and group results. It is also possible (and becoming more common) to offer compensation contingent on the residual economic results realized after sale by an economic unit. Mostly known as *profit sharing*, this compensation method is often applied at the level of entire firms or of sub-units for which profits are measurable (e.g. divisions or profit centers). Many of the debated conditions for an effective application of profit sharing have to do with the cost of transferring risk to employees. Also on the basis of the previous discussion of this issue, it can be hypothesized that profit sharing is efficient (beyond a symbolic function) when:

- the group is not too large, in order to reduce free-riding problems and to distribute notable benefits;
- workers have enough responsibilities and ways to affect the firm or unit results;
- profit sharing does not also involve loss-sharing, so that the problems of inefficient allocation of risk may be attenuated;
- it is possible to agree on fair and transparent procedures on what profit indicators to use and how to measure them.

Reward policies

The combination of job-related and performance-based pay elements as distributed across work positions can be represented synthetically as a function of the type shown in Figure 9.4. The curves plot the distribution of compensation levels as a function of the evaluation of the positions held.

The different shapes that these compensation curves can assume, as well as their comparisons across units or firms, or with

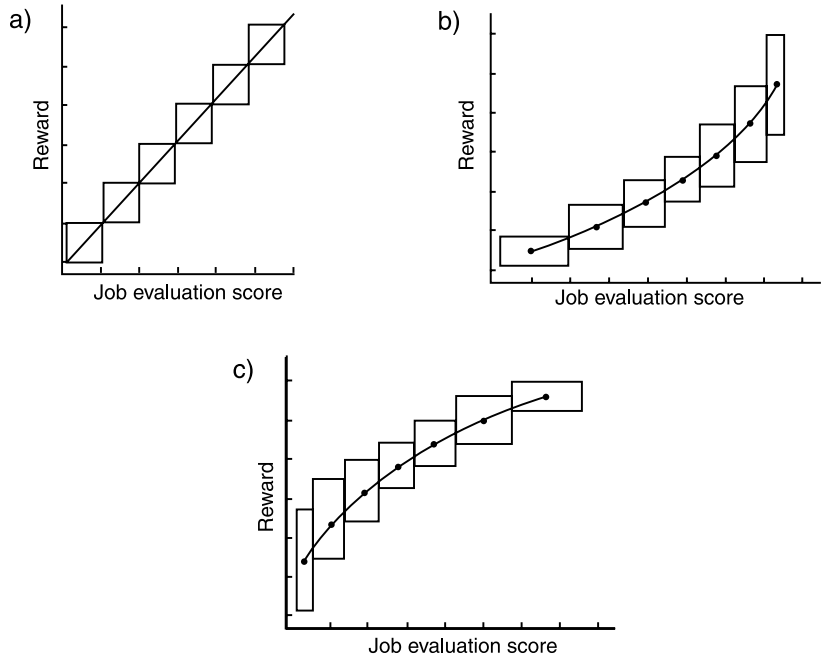


Figure 9.4 Compensation curves

learning and productivity curves, yield useful information about what compensation policy is followed and indications on what should be followed. For example, a slowly rising curve describes a policy of pay increases that responds slowly to increases in the competence and responsibility represented by the job evaluation ratings on the x axis. This type of policy creates a "hostage effect" because employees must make investments in their competences and produce value today in order to be compensated in the future (Itoh 1994). This system is therefore sustainable if the firm is prepared to invest in the formation of those competences and if both parties are adequately protected from sudden termination of the work contract. By contrast, compensation curves that respond more elastically to the growth of experience and responsibility are compatible with systems involving greater job mobility.

Lastly, the compensation curves of single firms or units can be usefully compared with the average curves for their sector, or with comparable firms and units, especially in order to diagnose whether their compensation schedules over jobs are sufficient to hold and attract the desired types of human resources.

Procedural justice of reward systems

The use of explicit compensation criteria that are transparently connected to jobs, competences, and/or performances should increase both the procedural and the substantive fairness of a compensation system (Folger and Greenberg 1985). Procedural justice research (Greenberg 1987; Folger and Konovsky 1989) found that providing opportunity for employee participation in pay determination can be expected to lead to a

perception of fairness and satisfaction with the pay system. Miceli and Lane (1991) applied the general procedural justice framework to the reward system such as:

- performance appraisals used to determine merit pay should be uncontaminated by unfair discrimination or personal bias
- when pay surveys are used to set rates, data should be complete, representative, and up to date in relecting other organizations' pay rates
- employees should have the opportunity to provide information that is used by the supervisor to adjust pay rates
- members of the compensation committee should be selected from a variety of locations, units, and stakeholders.

The case described in Box 9.4 enables one to apply the conceptual tools on evaluation systems and reward systems discussed so far, to assess the shortcomings of the proposed scheme of performance appraisal and contingent pay as applied uniformly to all the employees of a bank, and to devise a differentiated system that takes into account the different values of the relevant variable for reward system design in the various units.

MOBILITY AND DEVELOPMENT

Processes of mobility and development can be usefully viewed as those of match and co-evolution among individual actors, the resources they command, and the activities they generate. The effective, efficient, and fair configurations of these matching processes and the systems that govern them, depend heavily on the characteristics of the human resources and of their possible employments

Box 9.4
The CRC incentive system

The CRC is a bank which is being forced to drag itself out of a difficult situation brought about by a policy which has placed little emphasis on competitiveness and proved short-sighted with regard to internal organization. The management's strategy for recovery therefore involves a large commitment to customer relations (contacts, visits, and advertising initiatives to back up the relaunch of the bank's image) and with staff (efforts to improve competences and increase involvement in the bank's new approach). As part of this strategy, the bank has created a system of incentives with varying remuneration levels, in line with the strategy of customer-orientation, to be applied to the whole of the bank's staff of just under one thousand, who are subdivided between central functions and branches in a proportion of 30 to 70.

The objective of the system of annual incentives proposed, in the words of the General Director, is "*to create a strong team spirit amongst staff and to provide executives with a means of direct dialogue with their subordinates, without creating orders of merit since each branch only has to measure its results against its own previous performance; it must reward those who deserve it, those who have attained special heights during the year, stimulating dedication without guaranteeing anything else; it must allow clear communication of the bank's annual targets, facilitating their achievement.*"

The incentive applies to each organizational unit (branches and central departments) individually and is awarded (in differing measures, see Table 2) to the executives and clerks working there; however, clerks are not awarded the bonus unless their individual performance (evaluated annually by their respective managers) is rated at least as *good*.

For the branches, the incentive system is based on the comparison between budget targets and results with regard to the following parameters

- direct collection of funds (current accounts)
- indirect collection (securities and investment funds)
- lending
- overall profit margin, including a share of overheads.

The system is therefore closely linked to the planning and control system. The levels of the target parameters, ambitious but realistic, must be set so that the incentives system is self-financing and a proportion of profit is redistributed to employees. On average, the bonus envisaged is between 10 percent and 40 percent of gross salary.

The variations over each target parameter are translated into scores (see Table 1).

Table 1 Ratio between parameters and scores

Target parameters	*0/+4%*	*+5–15%*	*+16–25%*	*>+26%*
Direct collection	5	10	15	20
Indirect collection	3	6	9	18
Lending	2	4	6	8
Branch margin	2	4	6	8
Total score	12	24	36	54

As can be seen, in order to receive the bonus, the branch must have achieved a score of over 12. The score is then converted into the bonus sum on the basis of the following table.

Table 2 Calculation and distribution of the incentive

Scores / *Ranks*	*12/23*	*24–36*	*36/50*	*50/54*
Branch manager	×	2 ×	3 ×	4 ×
Deputy manager	×	2 ×	3 ×	4 ×
Clerk with evaluation *"good"*	1/6 ×	2 1/6 ×	3 1/6 ×	4 1/6 ×
Clerk with evaluation *"very good"*	1/3 ×	2 1/3 ×	3 1/3 ×	4 1/3 ×

Note: Where x is the basic level of the incentive bonus.

So for clerks, a positive evaluation by their superior is the essential precondition for "getting a piece of the action," the size of which is established in close connection with the planning and control system.

For Central Department staff, participation in distribution of the bonus depends on the results achieved by the branches. So, for example, if not all the branches have achieved their minimum increases, this affects the internal departments in terms of a smaller bonus. On the other hand, if the results are positive, the distribution of the bonus to the internal departments reflects the split in staff numbers between the center (30 percent) and the branches (70 percent), while the allocation in relation to rank follows the procedures already described above. This solution means it is possible that not all staff with the same evaluation (better than "good") will receive the bonus, and so the HR management will have to make a choice on the basis of the comments on the evaluation forms.

In conclusion, although it has been introduced with a fair level of success the incentives system leaves two important questions open. The first concerns the attitude of the trades union; opposed to the differences which the system institutionalizes, the union suggests that it should be modified through the creation of a productivity bonus identical for all, to be calculated on the basis of a single performance parameter.

The second relates to the feeling of frustration expressed by some central department clerks, whose statements can be taken as representative:

"In the same city, some of our branches may tend to compete with each other."
"In the internal departments there is more uncertainty about the bonus; for example, here in the HR department we may do a good job, but if the branches don't sell successfully we don't get a penny, is that fair?"
"And what about the managers' final word over who receives the bonus in the internal departments?"

Source: By Luigi Golzio, adapted by Massimo Neri.

(as well as on people preferences). Two characteristics are particularly important: *the specificity of resources to uses and users, and the complexity of activities.* All systems that are relevant to mobility and development (*search, selection, training, career, exit*) are considered here together, in their interrelationships and in their aspects internal and external to firms.

Systems of search and selection govern the access of resources to jobs. The problem is often rather unstructured, involving questions as: How can relevant alternative candidates be defined? Where and how to look for them? Should the allocation and use of resources be thought of as a single job that currently exists, or a series of jobs (defined or to be defined), or a set of connected jobs such as a firm in its entirety?

Search, selection, and labor markets

Personnel searches involve, in principle, a huge number of alternatives. To make it even more difficult, the desired results and the criteria by which the alternatives should be evaluated regard future performances in as yet inexperienced combinations of people and activities. To that extent, it is reasonable to expect that personnel search and selection

are heuristic processes, where learning by trial and error is not pathological but rather a way to discover effective combinations. On the other hand, exactly because of their heuristic character, research and selection processes can be highly improved by debiasing techniques and decision support systems. In addition, an appraisal of the different types of difficulties and uncertainties faced in searches for different types of resources helps in accounting for the differences between the systems that govern search and selection processes.

Internal labor markets

A basic distinction among systems of search, selection, and allocation of human resources runs between "internal labor markets" and "external labor markets" (Doeringer and Piore 1971). An internal labor market is constituted by a finite and definite system of jobs the application for which is reserved to people that belong to the system.[4]

The use of the term *internal labor market*, intentionally or not, effectively conveys the idea that the processes of reallocation of resources that occur internally to a firm or other contractually closed system, can still use "market-like" mechanisms: an internal labor market is a system of competition between internal candidates for possible jobs, regulated by a system of negotiated prices that are only weakly connected to the external system. Among the reasons why these protected labor markets arise, it is supposed that internal candidates are more likely to possess or be able to develop competences that are *specific* to the job system considered. Internal candidates also have the advantage of being easier to evaluate, because of the availability of direct information about past performance, especially if it is not easily quantifiable and if it can be observed only over the long term. Therefore, the *complexity of work*, especially in terms of low measurability of performance, is also expected to favor the formation of internal labor markets; and even the formation of several internal labor markets, with limited mobility between one and another, in the same system, to which only groups of candidates with comparable competences have access (Milgrom and Roberts 1992).

Information costs

Even if they are generally costlier than the external market in the case of specific and complex labor transactions, internal labor markets are also not without costs, which in some circumstances can surpass those of external markets. The usually underlined costs of internal search and selection include the possible loss of control and loss of access to performance information as the number of system members increases, the cost of internal personnel management services, the costs of renegotiation of contracts especially if they are "typical" employment contracts incorporating extensive formalization of procedures and job descriptions; the escalation of pay and staff size.

Learning costs

A probably more important but less considered cost of internal labor markets, however, derives from the likely reduction in the sources of innovation, the higher organizational inertia, and the risks of obsolescence of the system's competences, no matter how much linked by specificities, due to the long-term stability of people, the homogenization and routinization of organizational culture and know-how, and the lack of input variety.

As a result of all the costs of internal labor markets, it happens that one turns to the

external labor markets not only to acquire simple or standard resources, but also for complex and/or co-specialized resources. For example, innovative firms, or firms under pressure for change, extensively use the external labor market for executive and professional positions.

The tools and actors of effective and efficient search and selection processes, even if external, will however be different from those of an external market of standard and simple labor. The configuration of the system in the latter case is nearer that of a classic market: the generation and comparison among alternatives is feasible owing to the availability of good objective indicators of the competences required (e.g. a diploma or professional training) or by reliable tests; the encounter between demand and supply is facilitated by intermediary institutions that operate on large numbers of candidates; and finally the expected consequences of performance differences are often relatively low in standard and simple jobs, thereby providing an economic criterion for limiting the investment in the search and evaluation of candidates.

In the case of reliance on external markets for complex or specific jobs, instead, the effective instruments usually constitute a hybrid between an external and internal market, as can be expected given that the process should exploit some advantages of external acquisition of resources while respecting the internal need to evaluate candidates on the complex competences and their capability of performing synergistically within the context of the existing system.

Selection

While search represents a phase of variety generation, *selection* is a process that reduces variety by the elimination of candidates until an acceptable set is chosen. This process of evaluation is characterized by substantial uncertainty on both sides. In most cases the employer, at the moment of selection, lacks certain information about the productive and professional capacity of the candidate (Spence 1973), even if that employer is using the internal labor market. On the other hand, the candidate does not possess all the useful information about the real nature of the job. Both face the risk of adverse selection biases. Both are faced with the trade-off between investing ex ante in analysis or relying on ex post learning that experience will provide.

The structured instruments to evaluate human resource potential and competences partially reduce the uncertainty owing to the lack of *ex ante* information. As already observed by examining evaluation systems, these instruments will be more effective to the extent that the job descriptions are clear and specialized and to the extent that the evaluation concerns a specific job rather than a career or development path. For specific and complex jobs, owing to the lack of reliable predictors about the job performance in single activities, it will be more effective and efficient to focus the ex ante evaluation on competences and professional capabilities and to emphasize an ex post evaluation in the internal labor market, about performances and the degree of fit between specific people and specific jobs and activities.

Independently of the nature of the job, there are various reasons why efficient and fair search and selection systems should be quite structured and formalized.

- Selection interviews conducted in an intuitive and narrative mode are likely to be subject to strong and systematic cognitive biases. To that extent, the depersonalization

and codification of the process should improve the quality of decisions.
- The application of transparent criteria to the evaluation and selection processes is a principle of procedural fairness that is highly appreciated by the candidates, and is in some aspects required by laws designed to ensure equal opportunities (Singer 1993).
- A decision process based on valid and reliable research and data gathering methods (such as questionnaires and simulations) yields better and more comparable knowledge, and permits economies of scale if selection activities occur on a regular basis.

Furthermore, a clear and precise declaration of the rules and administrative procedures that regulate an internal labor market (including the systems of evaluation, compensation, careers, training, etc.) activates mechanisms of *self-selection* that help in finding good job–people matches (Salop and Salop 1976; Milgrom and Roberts 1992). For example, if it is made clear that the spirit of a labor contract with a firm includes willingness to change the place of residence, to accept some compensation contingent on performance, to study to acquire new competences, this "offer" is more likely to be accepted by people whose resources and preferences are consistent rather than diverging. This mechanism permits the system to economize on the cost of search and to better use the information that both parties possess.

Nevertheless, this effect is by no means automatic. Its functioning depends especially upon demand/offer ratios. Box 9.5 illustrates the failure of self-selection mechanisms in both the opposite conditions of almost full employment and high unemployment.

Training

Training has an important role in the generation, development, and maintenance of individual and collective competences. For the firm, or other system of action, it is a possible source of distinctive competences and competitive advantage (Itami 1987), and for the people it is a form of reward, a capital increase that will remain forever in their possession (Pfeffer 1994).

The nature of the training system is tightly linked to the specificity and complexity of the competences that it is intended to develop. One of the reasons why competences are specific to a use or user is because they are formed through direct experience or learning by doing that can only be accumulated in specific contexts (Doeringer and Piore 1971). As to complex competences, they are by definition characterized by long learning cycles and multi-dimensional development paths (e.g. multi-functional, multi-firm, multi-sectoral, and international).

These propositions can be used as guidelines for assessing or designing an effective and efficient training system, in terms of the allocation of training costs, the effectiveness of in-house training versus external training, and about the level of training (e.g. individual or group, intra-firm or interfirm).

Locus of training

Using the distinction between specificity with respect to the uses (activities) or with respect to the user (e.g. a particular firm), one can predict that firms have an interest in investing in training that is to some extent specific to the user firm, and not in undifferentiated abilities that can benefit all firms (or all firms in their sector) and can be furnished by external training institutes. However, firms are often

Box 9.5
Selection systems and levels of employment

The Agensud case. A group of teachers at a successful management school in northern Italy is engaged in founding and launching a management training institute that can be developed and will in time operate independently in the south of the country. Among the various activities, the most critical is the constitution of a group of local teachers. The first problem is the selection of people of sufficient potential who will then be trained and assisted in the first years of activity. The qualifications required, the type of motivations and abilities felt to be suitable to the kind of activity, the characteristics of the work, and the prospects offered for professional advancement – all these are carefully predefined. Considerations of transparency and fairness require a public competition for access to the admission concours. Despite the purposely restricted criteria used in defining the figures sought, the result is disconcerting: 2,000 applications for 5 places.

The Agensoft case. Specialized engineers, with some years' experience and qualification, are rarely encountered "at large." They are very hard to find and very easy to lose. They possess high, very specialized skills, which can be applied in several contexts and several firms. How to attract them? Various consulting firms have begun to offer support in "candidates' marketing," where the personnel is treated as a consumer to be attracted in highly competitive conditions. Market studies and analyses of the motivational profiles of engineering graduates are performed, they are segmented by branch of specialization and nationality, "employment packages" are offered, designed ad hoc on the basis of expectations as well as employment prospects. For example, the French firm Altex has perfected a solution that offers the engineers a permanent contract, high investment in training and development and variegated international experience through detached work on projects with firms requiring the specialized services of the engineers. The search for candidates is done through channels differentiated by country and by field of specialization, so as to maximize the probability that the message will reach its audience (for example, the periodicals most read). But even in this way – as Altex's research and selection staff emphasize – in this "hunting" the "quarry" will scarcely select itself; and one can never really be certain of having "caught" it.

Thanks to University of Modena trainee Federico Mazzoli for information on the Altex experience; and to Massimo Pilati, Bocconi University, for information on the Agensud experience. The names of both companies have been changed.

(and increasingly so) interested in investing in the formation of flexible knowledge that is not very specific with respect to particular uses within them, but in core competences that can generate economies of scope and can support learning and diversification processes.

Further, analyzing the level of the action system to which competences are specific one can assess what is the best level at which

training can occur. For example, if certain competences and professional skills are specific to clusters of firms, rather than single firms, for example regional clusters of firms specialized in particular transformation processes, they can be acquired efficiently through processes of knowledge sharing and exchange at the interfirm level, for example through associational structures (Provan and Heimer 1999).

Training, learning by doing, and learning by networking

The complexity and innovativeness of the know-how to be transfered or generated in the training process, is the single most important variable for designing an architecture of knowledge transfers. Traditional school-like training is a mechanism of one-way knowledge transfer by means of communication. As such, its effectiveness is bound to codify well established knowledge. The more the relevant expertise is tacit, the more effective transfers call for *learning by doing*, or by observing other people doing. The more the relevant knowledge is innovative and should be generated within the formation context itself, the more a critical and selective sharing of experiences and knowledge becomes a superior mode of learning. For example, the recent diffusion of *learning by networking* in multinationals or professional firms – realized by connecting people with differentiated experiences – can be explained by a need not only for vicarious learning and imitation processes, not only for the transfer of sophisticated techniques, but also by processes of confrontation and exchange of knowledge, problem-solving, and reciprocal learning. Box 9.6 describes only some of the varied mechanisms for knowledge sharing and knowledge generation that are present in one of the most notoriously innovation-oriented companies in the world. As can be seen, "training" systems in the traditional sense become a part of a wider set of "knowledge management" mechanisms for transferring and developing knowledge and competences.

A useful way to discuss the case is to locate the *different mechanisms governing knowledge transfer and sharing as a function of the characteristics of actors' knowledge* (as noted, sophistication and computational complexity, tacitness and discovery intensity – Chapter 1) *as well as of the configuration of interests*. In fact, barriers to knowledge exchange and sharing do not come only from knowledge complexity and lack of proper channels, but also from the risks of expropriation and competitive use and lack of proper incentives. Electronic networks are set up but sometimes poorly nurtured. Inter-divisional groups or supplier–customer mixed groups are constituted, but members retain jealously their most relevant information. The higher the conflict potential and the risk of hostile expropriation, the more explicit incentives to exchange and protect property rights on information and knowledge are required (see, in the case, the internal patenting system, the knowledge sharing awards, the "pathfinder program").

As to the *processes* of design of training and learning systems, some form of "*training needs*" analysis entered early in the repertoire of techniques is widely used in this area of management. On one hand, training and development needs analysis can be anchored on the expected and desired evolution of activities and competences. For this (just as for the selection process), it is important to identify the method of action that the system is aimed at sustaining (e.g. a specific job, a career, a firm, or a profession). On the other

Box 9.6
Knowledge management systems at 3M

3M has over 80,000 employees, including almost 9,000 researchers in more than 60 countries. How do they keep such a machine together? How do they transfer innovation, ideas, technologies and competences? How can distant and often anonymous contacts among experts (with problems, unrealized opportunities, and solutions). become transformed into close relationships? There is the need for pollinators – small bee "integrators" able to distribute pollen.

Space

All the functions within the New Life Sciences (mainly health products). have been gathered under one single roof for the first time. More generally, in the Austin, Texas, facility, all the units of a single business are located on the same floor. 3M considers space very important; and provides plenty of well-equipped meeting places in order to encourage people to take breaks from their routine for conversation.

Mobility

The most important way that knowledge is spread is definitely through the mobility of human resources. In R&D, mobility is considered a company tradition, with exchange programs on various levels. Some job assignments involve formal education or training but others are aimed at establishing comparisons and collaboration among diverse research units. These assignments can vary from a quick visit to a two-year commitment; but it is the people who assess the quality of the relationship and value of prolonging it.

However, 3M has recently taken a relatively unusual step further, by giving new meaning to the word "self-development" as a process to be valued and sought within every individual. There is a program called EJIS that displays on every European employee's computer the current vacancies in continental Europe. This program exists in the US as well and forces each boss who needs to fill a vacancy to advertise the position through EJIS. The announcement contains a description of the position, the job responsibilities, goals, and objectives, position site, etc., and remains on the system for 15 days. Any person interested in applying for a position can simply press a key on the keyboard, without having to inform his/her boss, and is then automatically included in the selection process. The ultimate goal of EJIS is to fully exploit a system that is marked by broad-ranging and trans-national competences.

Forums

Space and mobility are not enough. For this reason, 3M is set up as a spider web of functional forums, including some Technical Forums (TF) and Lab Forums (LF). The Technical Forums include the entire technical "macro-family" – researchers, product developers, and technicians. The purpose of the Technical Forums is to provide informal and formal communication tools to encourage the cross-fertilization of ideas. Activities developed by the Technical Forums are all carried out on a

voluntary basis and include (for example) organizing specialist seminars with Nobel Prize Winners to bring all technicians up-to-date with the latest scientific knowledge.

Technical Coordination

Corporate-level technical coordinators exist to support the sectors and divisions, but remain under the direction of the senior vice president for R&D. Technical coordinators provide internal consulting. They manage a huge database, providing all 3M employees with access to the various experts in the scientific disciplines and functional fields. The database is fed by all the work carried out by 3M researchers around the world (e.g. Internal Technical Reports). The feedback time of the Technical Coordinators is estimated to be an average of two working days.

The Pathfinders

Within 3M there is a program called Pathfinder whose goal is to encourage the development of ideas into industrial and commercial products.

- The original Pathfinder is aimed at all business units that develop new activities based on the creation of new market opportunities based on innovations. To be considered for the prize, products must generate profitable gross sales of over $400,000 or 25 percent of the total local annual gross sales. In 1995 there were 654 nominees for the Pathfinder program, 82 of which won a prize.
- The Pathfinder Merchant is reserved for companies that have successfully adopted Pathfinder programs that were originally developed in other countries. Although not explicitly stated, the objective of this program is to improve "listening skills" and thereby improve the potential to assimilate existing ideas and avoid the *not invented here* syndrome. The goal is to improve the capacity to combine markets and products in new ways by stimulating the diffusion of winning solutions.

Source: C. Turati, *Minnesota Mining & Manufacturing*. Case study SDA Bocconi 1997.

hand, a survey of preferences and expectations of potential participants has been revealed as crucial. Successful training requires a receptivity to communication and a willingness to change the structure of one's knowledge, skills, or even values and motivation. Hence, expecially in adult training, people should perceive that it is in their interest to accept training or to invest in learning or to exchange knowledge.

Analyze, for example, the intersection between the competence needs of the firm and people needs in the "guest courtesy program" at Walt Disney, described in Box 9.7. What incentive structure is at work? What motivational structure can be hypothesized? What difficulties could be anticipated if the same program were offered, say, to aged life-long employed traveling personnel in a state-owned railway or post office?

Careers

Career systems regulate horizontal and vertical development of the people/jobs matches toward superior responsibilities and competence, and usually also toward superior rewards.

Box 9.7
How to make dreams come true

Walt Disney said that it takes people to make a dream come true – dreamers as well as doers – and he founded Disney University as a place for developing Disney people. When Disneyland was launched in 1955, its creator wanted customers to feel like invited guests. Disney believed in training staff people to have good guest relations practices – and Disney pioneered a highly successful concept.

Today, guest relations programs in service companies are common –and Disney is working to build on its experience to maintain its service edge. Disney University continues to play a key role in this process, with its focus on "cast members," Disney's term for employees who are expected to play their part, and play it well.

"Disney University serves multiple functions," Bill Ross, manager, Disney University, said. "HR planning and development, cast communications (internal publications, formal communication programs), cast activities (social, recreational, and interpersonal communication programs), and audio visual programs are all handled out of the University.

"Each Disneyland unit has a dotted line relationship to Disney U, so that we can provide centralized HR services, yet have a strong link to individual entities. At Disneyland, we have about 6,000 employees."

International Focus

"Putting together our PHRM planning group is one of our newest HR efforts. Within the last two years, we've computerized information on our salaried employee population so that we can submit the criteria for a particular position, and generate the names of candidates through the computer. This system has also allowed us to pinpoint some of our development needs.

"We've found this capability particularly helpful since we've launched Tokyo Disneyland and our Euro-Disney project.

"We relocated about 200 executives to Japan – some for short stays, some for long. Given the expanded scope of our operations, we need to know all we can about the talent we have to draw on.

"We've been focusing on our need to become more internationalized. We are undertaking extensive training of Japanese managers, and our managers in the US, so that we can better understand cultural issues. We know that we can cross cultural lines – but we want to understand the issues to enhance our chances for success. We have established an International Fellowship with that goal in mind.

"We've also worked on providing a support system for those who return to the US after a tour abroad. We recognize them for their efforts, we listen to their descriptions of their experiences, and we review what took place here while they were gone. After all, they are not returning to the same organization that they left – things change. And the manager who has served abroad has changed, too.

Guest Courtesy – inside Disney

"We believe that, for our cast members to treat our guests in a friendly and helpful way, they themselves must be treated that way. We look at guest courtesy as something that must extend to those within the organization, too.

"When cast members join Disneyland, they are treated as VIPs – they are personally greeted, and everyone is on a first-name basis. We reinforce good guest relations through our orientation process, training, performance appraisal system, and we circulate guest compliments and complaints. Our biggest challenge is to stay in touch with the changing values of both our guest and our cast members.

"We look at how our guests define service – both first-time visitors and repeat visitors. We look at the environment itself, since elements such as temperature have a definite effect on people's perceptions of their experience. We encourage those behind the scenes to be conscious of courtesy, too – for example, we have a campaign called 'Put a Smile in your Voice' that emphasizes telephone courtesy. And for our Christmas party, cast members and their families come to the park – and management mans the park for that day. Cast members experience the park as a guest – and management experiences the cast members' jobs. There is a management program that focuses on guest courtesy, as well.

"We have a two-person team whose sole job is to evaluate the level of courtesy that exists in the organization. We poll guests daily, and circulate results.

"We show examples of good and bad guest relations in our training program, based on the information we gather. We teach cast members to understand outcomes – their goal is to focus on what they want the guests to experience. There are many ways to get to that goal – but the end result is what counts.

"Our training teaches them to enhance skills to initiate a relationship, to take the first step in approaching guests who might look puzzled or in need of help. In the service business, we are fortunate to get a second chance when something goes wrong: a guest may have an unfortunate experience in the ticketline, but a helpful interchange in a restaurant that helps compensate. We want to avoid the first mistake – but make sure we take advantage of all opportunities for a second chance."

Source: "Making dreams come true," *HR Reporter* (Jan. 1987).

Internal careers

If a career system is governed as an internal labor market, it contributes to increasing the longevity of the work relationship, through requests for contributions that will be rewarded – in terms of position and/or connected superior compensation – at a later time. The prospect of a career thereby allows contributions and incentives to be balanced over the long term instead of in every single transaction.

However, considered as an internal labor market, careers are only partially isolated from external competition for jobs. As Milgrom and Roberts (1992) have noted, workers that perform extremely well in one activity can make themselves visible and attractive for alternative employers, who might offer salary increases in order to convince them to

change jobs. External offers therefore create incentives for the current employer either to use variable pay as a reward for high performance, or else to promote the worker. Promotion is often an attractive alternative compared to increases in incentives, for several reasons. First, as discussed above, there are limits to the efficient use of incentives (e.g. the difficulty of accurately measuring performance, variable risk aversion, etc.). Second, particularly high performance constitutes a signal that the competences and capacities of the person might be greater than those required by the current role, and therefore the full use of human resources would require activities of greater complexity and responsibility. To that extent, using promotions as a form of reward contains an element of organizational learning that monetary compensation does not include. Further, it is likely that employees with high goals and good results might assign a higher value to the intrinsic rewards represented by the increased opportunities to assume responsibility and apply competence that a promotion offers, than to pay increases.

External careers

The external labor market may however be better able to reallocate human resources to their best uses, especially in the case of jobs involving general rather than specific skills, or competences that can no longer be combined with other skills to create value in a specific system. For example, if a firm concludes that this is the situation for a given employee, it would be in that firm's interest to invest some resources to help that employee to look for better suited alternative jobs in other firms. To give this support may be convenient for several reasons: first, it may cost less than side compensating exits in other ways (for example, with lump sum money); second, it can make the firm more attractive to people who are either employed elsewhere or looking for work, and to those who remain within the firm (Ulrich and Lake 1991; Pfeffer 1994). The recent diffusion of individual and collective *outplacement* processes can be explained by these advantages. Outplacement normally involves the firm in providing consulting services, logistical support, and useful contacts to its employees so that the latter can more accurately evaluate their own competences and expectations, and more easily find a job that is congruent with these.

Performance is not the only criterion on which mobility is based. The practice of adding other criteria, most commonly job tenure and the evaluation of potential, can be explained on the basis of the complementary properties of efficiency and fairness that such other criteria possess.

The seniority system

Internal promotions can be based on criteria of seniority. These tend to be related to other variables such as experience and competence, although not necessarily in a deterministic way. As a form of deferred time-based compensation, the seniority system has the property of creating disincentives for turnover by rewarding those employees who stay longer. In addition, the criteria underlying a seniority system are objective and non-discretionary, which makes such a system both efficient and fair. Because it is applied more or less automatically, the so-called "*influence costs*" are reduced. In fact, influence costs are incurred when people have the incentive to dedicate time and effort in presenting and misrepresenting information and in negotiating so as to influence the decisions of the evaluators; this is likely to occur when the

promotion system is subjective and not transparent (Milgrom 1988). In contrast, a seniority system implies a standard rule for promotion that lowers the managerial discretion in career choices, thus reducing the efforts directed to influencing promotion decisions.

Performance-based promotions, following repeated positive performance evaluations, can generate also other types of distortions, that can be reduced with the use of other mechanisms.

The Peter principle

It can happen that the person with the best performance at his or her current level might not be the best candidate for the position at a higher level. In other words, it is not always the best blue-collar worker who could be the best production manager, and the best product manager is not always the best marketing director. Promotion on the basis of performance in the preceding role can lead to the so-called "Peter principle" where everyone is promoted to their level of incompetence, and for which many firms under- or mis-utilize talents and competences.

The evaluation of potential and the design of multiple career paths can contribute to the correction of these biases. In the first place, merit-based careers can and should be fed by the evaluation of resource *potential*, where people are promoted not only according to the level of performance at their current level, but based on their possessing the best potential for achieving at the next level.

Multiple careers

In the second place, it is possible to create more than one career path, even within one firm or unit. For example, in a practice originating in service firms, it is possible to follow a *professional* career path (e.g. researchers, teachers, doctors) or a *managerial* or *administrative* career path (e.g. project manager, sales manager, controller). Among the advantages of these *dual ladder careers* (Heskett 1986), are the capacity to facilitate synergies among employees and tasks. For example, dual career paths allow people with strong technical and professional competences such as researchers and scientists to continue to develop these competences, instead of reaching a "plateau" and being required to learn management competences in order to win further promotions. Dual career paths also can reduce the level of competition in promotion games multiplying and differentiating responsibility positions, at the same time making good career opportunities compatible with less hierarchical and vertical organizational structures.

Random careers

Furthermore, when the performance information that would be relevant for promotion is multidimensional and estimates are uncertain, complex performance and resource potential appraisal may give only an illusion of control. The evaluation data may turn out to be weakly correlated with actual performances and competences, making performance-based promotions subject to a significant random factor (March and March 1977).

Tournaments

A partial remedy to this problem may be to institute "tournaments" among people under reasonably controlled conditions, so that it is not necessary to measure absolute performance, but only the relative performance of the candidates (Milgrom and Roberts 1992). These races have disadvantages, however,

such as prospecting zero-sum games that may damage the organizational climate and demotivate the losers, who may have had good performances anyway, and end up being underutilized.

Career and risk

Seeing a career as the assumption of a series of positions with greater responsibility and requiring more competences implies not only higher compensation due to the increased contribution, but also a change in the form of compensation. In fact, an increase in responsibility implies more decision rights and obligations, greater discretion, and greater consequences of choices. Therefore, career development may imply, beyond certain levels, an increasing *assumption of risks*, a change from the status of employee under authority relations to a status of an agent, or to a status of a partner or of an internal (or independent) entrepreneur (Pilati 1995). It is also likely to imply a change in the form of compensation, toward a greater incidence of rewards that are contingent on results, due to higher discretion. However, it is likely that positions of greater responsibility will also involve greater uncertainty about results and greater dependence of results on exogenous factors. Furthermore, the results are likely to become more complex and multidimensional and observable only over long time periods. Therefore, the relation between the level of a position and the form of reward that can effectively be attached to it is not likely to be linear and straightforward. A heavy reliance on variable compensation can either paralyze managers with excessive risk transfers, or at least distort managerial actions and investments in favor of short term and conservative policies; or, alternatively, cause executive pay to skyrocket.

Among the possible remedies to this problem are the following. First, it is possible to propose variable compensation that is only contingent on particularly positive results and not also on negative results. As Milgrom and Roberts (1992) have observed in their analysis of managerial compensation, stock options offered to managers can be interpreted as such – i.e. as opportunities to participate only in increases of the value of the firm.[5] If managers are given an option to buy their firm shares at a pre-defined price, they are encouraged to work in ways that can increase the shares' value, and buy them when this is conveniently better than the old fixed price they have to pay. If such a win-win system were thought to be too indulgent, it should be considered that there are other systems beyond compensation that can discourage errors and incorrect investments in high-level positions. One of the most important resources that is usually at stake for a manager is, for example, his or her *human and relational capital*, that can exhibit very sensitive variations as a function of observed performance. Prestige is very sensitive to failure, and it can seldom be salvaged by parachutes based on subtle caveats about where the "real" causes of success or failure were lying. Box 9.8 shows the notable and tangible importance of this factor, in a case in which a manager explicitly put his own human capital at stake, offering in advance to conduct action in a very public way and to accept the losses of prestige and reputation (and even of position) in case of failure of an important industrial restructuring project, as a pledge and guarantee offered to the shareholders that the investment that he was proposing was in effect a good one.

A more radical solution, used in cases where it is particularly hard to measure performance, is that of reducing the distinction

Box 9.8
Human capital pledges in industrial investments

The interview in this case focuses on the restructuring of the Fabbrica Pisana, an Italian firm owned by a French MNC. Mr Scaroni was CEO and General Manager of the company at the time.

Q: What was the company's economic situation when the restructuring took place?
A: Fabbrica Pisana produces flat glass for uses in floats, cars, and the construction industry. It is fully owned by the St Gobain Group, which has been operating in Italy since 1881. Fabbrica Pisana faced the 1980s after living through a decade of labor union battles (all lost) that, in addition to forcing them to go on a hiring spree, had created delays and obstacles in launching the process of restructuring the firm.

To make things worse, foreign investors, still feeling insecure and uncertain after the unusual economic climate of the 1970s, did not feel comfortable about making further investments. As of 1981, the productive assets of the firm had been seriously compromised.

The French shareholders were seriously considering the possibility of leaving Italy and selling the company to the state (at the time this kind of transaction was still possible). Over the summer I prepared a restructuring plan that was then presented to the shareholders in September.

Q: How were the relationships with shareholders and headquarters handled?
A: Both shareholders and headquarters are located far away in Paris and did not have specific details on the situation. As far as they were concerned, the investment in Italy represented only one alternative among many different kinds of investments in different countries.
Q: In practical terms, how was the situation handled with headquarters?
A: First of all, by carefully working out all the technical details of the restructuring plan, and by explaining all possible investments alternatives, and the reasons why some were eliminated. Second, by carefully analyzing the evaluation criteria of headquarters and demonstrating that the proposed project met those criteria. In addition, we involved all offices (sectoral divisions, finance departments, and human relations) and we committed ourselves to "maintaining" the agreement by fostering continuous relationships, involving all parties in further developments of the plan, etc.
Q: Were there different positions with regard to the necessary investments to be made?
A: Conflicting interests existed on at least two dimensions:

- Headquarters was interested in choosing the investment that was the most profitable while to the Italian CEO the critical issue was that his project be accepted at any rate;
- Once the investment was approved, the Italian managers' main goal was to expand the plant and make use of the unexploited resources, while headquarters' aim was to reduce the gap between resources and results.

Q: How did you persuade headquarters that your project was the best, or at least better than the others?

A: Headquarters in Paris was initially particularly sceptical due to the labor union climate in Italy. My first task was to explain that Italy was changing fast especially as far as unions were concerned and because of my Italian background I could understand those changes better than they did. How did I persuade them that I was right? Well, I was betting on the project more than I was asking them to do; my name and my head were at stake with it.

At the end of the restructuring process, which lasted four years and called for over 150 billion lire worth of investments, the cost of labor to the company – which in 1985 had gross sales for about 220 billion lire – had decreased to about 65 billion. Such reduction contributed to shift the negative margin of 25 billion per year in 1981 to a positive margin of 15 billion per year in 1985, even after taking into account the high financial burden of paying for the new investments. The gross margin of the company went from −8 percent in 1981 to +22 percent in 1985 thanks also to a substantial reduction in transportation costs.

Source: Scaroni and Grandori (1989).

and the possible diversity of interests between principal and agent actors through a more intense sharing of property rights. Particularly common in professional firms, such as consulting companies, advertising agencies, and accounting firms, and generally typical of activities conducted through long term projects, this solution often implies that to make a career means to become a partner in the firm.

Procedural fairness in career systems

In practice, the career system is one of the organizational processes that is most influenced by unwritten norms and habits, and by implicit contracts (Pfeffer and Salancik 1978). Moreover, it has been shown that these are greatly influenced by the capacity of managers to strategically construct and manage their own "social capital" of relationships and contacts (Burt 1990). One could ask the extent to which these characteristics of career systems are justifiable.

From an efficiency standpoint, a promotion process based on criteria that are not very clear and explicit is extremely costly in terms of influence and negotiation costs. The quality of promotion decisions made without structured supports is surely threatened by strong availability, representativeness, and anchoring biases, even if these are only involuntary. The occurrence of stereotypes and discriminations (whether based on race, gender, or other prejudices) threatens not only the fairness of the process, but also an efficient and full use of human resources. All these considerations support the position taken by scholars of organizational justice, who have argued the superiority of personnel systems that are as explicit and transparent as possible.

On the other hand, it is fair to note that, both in promotion and compensation decisions, games are often very competitive, both because they involve dividing scarce resources and because people have a tendency to over-estimate their own contributions with

respect to others and to believe that they receive an unfair share of those resources even if the system is well designed (Lawler 1966). To that extent, it is likely that rules and procedures alone, even if clear and fair, are not sufficient to govern these processes, and that it is inevitable that there be a component of discretional judgment, especially on the part of arbitration authorities. Another possibility, more applicable in compensation than promotion decisions, is to communicate the procedures clearly, but to maintain a certain confidentiality on single decisions (for example, the job responsibilities and compensation of single employees).

Manpower planning

If this is the nature of the main processes whereby job mobility and career development are managed, it follows that even in cases in which these are regulated inside the firm, they can be regulated by "planning" only to a very limited extent. *Even if internal, labor markets behave to a considerable extent like markets* – i.e. a system of decentralized, interactive, and multilateral decisions. The mechanism of negotiation, whether it is collective or individual, institutionalized or not, is required to define both the procedures and the individual decisions involved. This assessment of the effective mechanisms that can coordinate human resource mobility and development would suggest de-emphasizing the importance and applicability of "personnel planning" techniques, which received considerable attention in the 1970s. Managerial recipes during that period were in fact celebrating integrated, systemic, and strategic approaches to management in all areas. In personnel management, these approaches sought to foresee the evolution of the overall human resources of firms, through accurate estimates of turnover (both departing and entering personnel) among various positions and professional families, across organizational units and the boundaries of the firm. However, the idea of being able to plan the evolution of the firm – from strategy to personnel management to information and control system – over long time periods has been revealed as being at best only possible in particular situations where development is linear or predictable, typically in large firms with protected markets, and where the firm is governed primarily by programs and plans and by hierarchy. This vision has not stood up particularly well, even in large firms, to the challenges posed by competition through innovation and through people and competence.

However, it is possible to identify some types of useful planning instruments, limited in scope, valid for the short term, and applied to systems of moderate dimensions and complexity, for which it is possible to estimate reliably the information inputs required. Those inputs are fairly numerous and include:

- a selection of activities to be performed and of positions entitled to conduct them;
- the type of competences required for sustaining those activities;
- estimates of the average man-time required for perfoming the various activities;
- the modes of acquisition of the human resources (internal or external labor market; times and costs; channels);
- the training interventions required;
- the normal exit rates from positions (turnover);
- a planned promotion or tranfer rates across positions;
- the availability of candidates.

On these bases it is possible to calculate or simulate, eventually for various possible

business scenarios, the evolution of the population in the relevant organizational layers or families of positions and to assess the needs for hiring or staff reductions. Box 9.9 provides an example.

FORMS OF EMPLOYMENT CONTRACTS

The systems governing work relations shown above can be combined in various modes, thereby configuring different types or forms of work contracts. Some characteristics of human resources, and work activities and relations have been shown to be particularly useful to explain and predict which configurations of human resources organization are effective, efficient, and fair: the *specificity of human resources* with respect to other technical and human resources with which they are combined in transformation processes; and the *information complexity* of the activities and relationships, in particular the *observability of performance* and the *knowledge of its causes and consequences*. In addition, the variables more directly affecting the structural costs of organizing labor internally or externally (the economies of scale in the utilization of resources, the economies of specialization that can be achieved by the focal firm or other firms, the frequency of transactions) should be taken into account. Therefore, it can be expected that different effective forms of work contracts are correlated with those dimensions. Figure 9.5, illustrated below, proposes a typology of contracts based on these dimensions.

Contracts can be *complete or incomplete*, primarily because of the variable amount of *uncertainty* (unforeseeable circumstances) that they are designed to regulate (Chapter 7). These can incorporate complex clauses, procedures, rules, and systems to manage interdependence and resolve conflicts, including authority (*obligational contracts*) without necessarily taking the form of a permanent internal dependent employment contract. In the presence of uncertainty about changing circumstances and of difficulties in the evaluation of performances, contracts can become very incomplete in their written and formalized part, and are integrated with agreements and expectations of "good faith" and "good conduct" according to the norms that are prevalent among those who practice certain activities. This informal regulation of behavior is also assured by mechanisms such as reputation, social control, and mutual adjustment (*relational contracts*). In situations of maximum complexity of activities, characterized by difficulties in evaluating both input and output performances as well as the value of positions, contracts arranged around a logic of exchange – based on the measurement of some of those values – cannot be effective. In these cases, contracts implying the association of resources (*associational contracts*) can be effective because they create direct incentives for the actors to behave effectively in their own interest, such as partnership and gain sharing contracts. Objective alignment and intrinsic motivation can be sustained by the sharing of property rights in their various forms (the rights to decision, control, and residual compensation), and by a strengthening of the coordination mechanisms that rely on shared values and fundamental objectives. These conditions tend to occur frequently in brain-intensive and personality-intensive systems.

In a second dimension, contracts can be ordered according to their *time frame* and to their *degree of* "*internality*" or "*externality*" (Chapter 7). These features of contracts are particularly related to the comparative *administrative and productive economies of*

Box 9.9
Why does everybody run away from InfinityNet?

Mr Morrison, Sales Director of InfinityNet, the third leading cellular phone operator in the country, could not understand the reasons for the increasing turnover in the call center, which was worrying him. The call center was the main organizational operating unit within the company.

Once signed-up, the customer was managed through a call center, a complex center for receiving customer toll-free calls that is equipped with the latest technology and is open 24 hours per day all year round. It should be able to solve all customers' problems, and at InfinityNet employees were proud to have created procedures that would guarantee the so-called "one call solution," where the customer's problem was solved in the first phone call.

The call center was also at the heart of the marketing and sales strategies of InfinityNet: with a toll-free call, anyone could receive a complete and polite answer about any of the now 20 charge plans or the 50 additional services (some included and others not) in the contract. The call centre was therefore overwhelmed by customers' or potential customers' calls any time the company launched a new promotional campaign or a new calling plan. This strategic choice was what differentiated InfinityNet from the other two cellular phone network providers.

The third task performed by the call center staff was one of support to the sales activities. Specifically, this involved telemarketing (for promotional activities toward customers or potential customers), sales management (recording of new contracts and set-up of new phone lines), payment solicitations, and the management of notices.

The call center employed 1,200 people who were organized on various shifts that would guarantee staff coverage 24 hours per day, 365 days per year. The typical staff person was a young college graduate or a college student in his or her last years of study. The motivation for hiring staff with these profiles was that the company wanted to convey a sense of safety and competence, as well as an image of dynamism and freshness. Because of the complexity of the procedures, training costs for new staff were on the increase: the minimum length of training was now 6 weeks. Shifts were scheduled with a maximum of 7.5 daily working hours 4 of which (all continuous) were spent providing phone assistance. Task rotation was ensured by the fact that, when traffic permitted, operators could be re-directed by their supervisors to different kinds of tasks: the recording of contracts, telemarketing, payment solicitation, assistance to business customers, sorting of recorded calls.

It is because of this variety of tasks that Mr Morrison could not understand why the annual turnover had gone from an acceptable 35 percent to a worrisome 70 percent. This meant that, lately, staff at the InfinityNet call centre had actually worked there for barely over a year. Even though a percentage of the turnover was explained by the fact that many students would finish their degrees and then find less stressful and better paid jobs in other companies, Mr Morrison thought that the rate of turnover was still high. In order to study the problem, he created a task force.

The task force concluded that it was necessary to improve the selection process and the overall management of the call center staff which accounted for half of the total number of company employees. The task force made the following suggestions:

- after a two year period at the call center (or earlier for those students who finished their degree) all staff would receive a performance evaluation. If the evaluation was positive, there would be an offer to work in a different division within the company. The human resources manager had stressed that the professional profiles of the call center staff were the same as the people sought by the other divisions of the company (network engineers, business managers, corporate lawyers, etc). Being able to count on the call center as the human resources candidate pool for the whole company would certainly reduce overall recruitment and selection costs.
- all call center staff would receive more in-depth training on all other company divisions with the goal of creating more competent and prepared staff when dealing with customers. The additional training would be of an informational nature in order to encourage the staff to consider other jobs in other divisions of the company in the future.

Mr Morrison accepted these recommendations, but stated that they would complicate the management of the call center because the scheduling of shifts (modified now to allow for attendance at training courses) had become extremely complicated. At the same time, Danny Scott, VP for Finance, was worried about the increasing personnel costs and in particular those of the call center, as it employed a large number of company staff. On the other hand, decisions about call center staff directly impacted those about customer service strategy: fewer staff means longer waiting time for the customers in receiving services. Hence, it was absolutely essential to find a tool that would allow for an annual human resource planning among the numerous tasks in order to be able to forecast the human resource budget. This is the reason why a human resources planning model was created (and nobody would have accepted it at the beginning) that, based on certain technical parameters necessary to provide the service required, would allow them to determine the staff needed, excluding those staff who were attending training sessions. Examples of parameters include: a standard number of minutes spent speaking with a client on the phone, the time necessary to activate a new contract, and average number of incoming monthly calls.

Source: By Ferdinando Pennarola.

scale and specialization that can be realized through a continuity of association, and to human resources' *specificity* to the system.

Even considering, simplifying, only some combinations of the two dimensions (and of their sub-dimensions) one can explain some relevant forms of efficient contract, such as those indicated in Figure 9.5.

If *competences are not specific or co-specialized, or the scale at which activities are conducted do not saturate the resources generating them*, then contracts are likely to be *short-term*:

- almost *complete* and *obligational*, if information complexity is also low, and activities can be clearly defined (as in the case of "temporary work");

LABOR CRITICALITY AND COMPLEXITY	ECONOMIES OF SPECIALIZATION SCALE AND SCOPE IN HR CONTINUOUS ASSOCIATION AND HR SPECIFICITY		
	LOW	HIGH	
SIMPLE COMPETENCES • Observable outputs • Observable inputs	• *External contingent contracts* (seasonal work) • *Internal short-term contracts* (temporary workers)	• *External long-term* (insurance or car agents) • *Internal long-term* (clerical work)	*OBLIGATIONAL*
COMPLEX COMPETENCES • Observable outputs • Observable inputs	• *External short-term* (purchase of advertising services) • *Internal short-term* (project-based engineering work)	• *External long-term* (insurance co-makership outsourcing of components) • *Internal long-term* (complex production work, managerial work)	*RELATIONAL*
HUMAN RESOURCES CRITICAL • Unobservable inputs and outputs • Human capital principal asset at risk	• *External associational contracts* (profit sharing association of work-service providers, e.g. "contract managers")	• *Internal associational contracts* (property sharing association of work providers, e.g. professional partnership)	*ASSOCIATIONAL*
	SHORT-TERM OR EXTERNAL	*LONG-TERM OR INTERNAL*	← *TYPES OF CONTRACT* ↑

Figure 9.5 A typology of employment contracts

- *incomplete and relational*, but *external* or *short-term* if competences are complex but standardizable, and can be more efficiently employed in many firms rather than one (as in the case of a contract for advertising services or in the temporary employment of engineers in complex projects);
- *associational, but external or short-term*, if the work services are highly discretional and poorly observable (as in "contracted out" managerial work).

If *human resources are firm-specific or co-specialized, or there are economies of scale and*

scope in employing them in multiple activities, efficient employment contracts tend to be long-term:

- if *uncertainty is low* (work conditions are foreseeable), then contracts can be *obligational*; internal and linked to behavior and time if those are observable (as for "typical" "dependent" blue- and white-collar work relations) or external and linked to results if those can be evaluated (as in the case of the outsourcing of material handling services, or of agency contracts in insurance or car sale);
- if *uncertainty is high*, contracts will be *relational* rather than obligational; internal if processes and inputs are observable and critical for obtaining good results (as in research, complex production, staff activities), external if good results are clear and could be achieved without tight process integration (as in the outsourcing of industrial good maintenance or sale activities);
- if *information complexity is high in all its components*, *associational contracts* entailing property rights sharing would be appropriate (as in professional partnerships).

Notwithstanding the variety of employment contracts considered here, these do not exhaust the variety of possible effective modes of regulation of work relationships. Other combinations are also possible. For example, can temporary work be combined with property right sharing? This is what happened to govern highly professional but highly standardized work in the case described in Box 9.10.

SUMMARY

This chapter focused on the "nexuses of contracts" linking actors contributing their work to the economic action systems in which they participate.

The final section of the chapter explored the variety of contractual structures that can provide the link. The typology of work contracts proposed there highlights the peculiar features of work contracts, and their varying degree of "completeness" in formal terms, as well as the different configurations of the basic human resources management systems that they incorporate – evaluation, reward, and development systems. The proposed framework enables one to assess the conditions under which different types of contract can be expected to be effective (in particular, the conditions of information complexity, and the conditions of specificity, specialization and scale economies).

The first sections of the chapter were devoted to analyzing the various systems governing human resource contributions. In addition to exposing some core and corroborated tenets from human resource management research, the conceptual framework for describing and designing these systems has been developed also with the help of other perspectives and contributions. In particular, it has been shown how evaluation systems provide the cognitive foundation for human resource organization, through the observation and appraisal of resources, performances, and activities (jobs). The relative effective use of the different methods of evaluation has been shown to be dependent on the clarity of available information and possible causal attributions relative to resources, behaviors, or results. Compensation methods and forms related to jobs, to performance, or to attributes of human resources themselves (competences) have been linked to evaluation methods and assessed as a function of the levels of uncertainty and competence complexity, of discretionality, of added value and

Box 9.10
Microsoft, freelances, and stock options

In one of Microsoft's latest appearances in a court of law the debate centered on *freelances*, who work in the computer sector as independent entrepreneurs. The reason for the law suit was stock options, which play an important part in the remuneration for employees in both hardware and software sectors. If a product is especially successful, the value of the shares will rise, as, at the same rate will the remuneration of the engineers and programmers, who have elected to be paid partly in shares. Some of them have even become millionaires overnight.

Microsoft USA introduced its program in 1986. Every six months, all the employees can exercise the option to be paid partly in shares at 85 percent of their current value. This program led eight foreign collaborators, who had been "hired" as freelances between 1987 and 1988, to file a suit. At the time, they had been told that owing to the limited duration of their residence permit they had no right to this option. The court ruled in their favor, to the effect that freelances with a fixed term residence permit must also be considered as full employees. Microsoft and the entrepreneurs' association tried to take a tough line, but in vain.

In actual fact, anyone in the sector requiring specialized workers tends to find them through firms specializing in information technology and electronics. These firms offer very flexible contracts, by which the programmers can be "rented like cars: they don't need to be filled up with petrol on return" (says the advertisement of a firm in Seattle). It may be understandable that Microsoft didn't care to use these firms, but the court did not forgive it for obtaining the same result by creating two sets of rights, for first-class and second-class employees.

Thanks to the development of the Internet, these problems will probably sort themselves out. For some time now, many software houses have been letting programs be written in other countries. In numerical control programming, the offshore programming share may have already reached 85 percent. But the use of such telecollaborators is not without its disadvantages: they are not available for the marathons that go on for several days and in which a team of programmers creates software under terrible time pressure in order to deliver to the client on the stipulated date. Projects of this kind require group work and the carrot-cum-stick of shares.

Source: By Giuseppe Delmestri, adapted from Detlef Borchers, *Die Zeit: Bulkware*, February 1998.

of the costs of efforts. Development and mobility systems have been examined in their dimensions of internal and external labor markets; of learning by formal training, by doing and by networking; of performance-based, potential-based and seniority-based promotion; linked to knowledge complexity and innovativeness, human resources specificity, and specialization, scale and scope advantages.

For each system, we have examined *both effectiveness and justice issues*, and applied them both to the system *architecture* (for example, the optimal incidence of contingent reward) as well as to the *process* through which evaluation judgments and compensation and mobility decisions are made.

Exercise: The Chimney Sweeps of Finance

Pursuit Credit Sp.A., based in Rome is a debt recovery agency. Through the Istituto della Cessione del Credito it purchases high insolvency loans from the banks and starts the procedures allowed by law for at least partial recovery of the sums owed.

The managing director is worried about the activities of the section responsible for preparing the client's file for debt recovery, a crucial function. This activity precedes the judicial one (that is, the representation in court by lawyers or prosecutors) and consists in preparing the briefs (analysis of the debtor's situation, evaluation of the mortgaged property, verification of conditions required for legal action) and in contacting the debtors (verification of debtor's willingness to repay part of the debt). Profiting from the firm's growth over the last few years, the idea had been to upgrade the professionalism of this activity by hiring young graduates in jurisprudence, on training contracts. But the atmosphere in the briefing department seemed very fraught.

What follows is the text of some interviews performed with operators in the aforesaid unit:

"Well, I studied hard for four years. There's a crisis in the market, so I took the first thing I was offered. Of course, I didn't want to be a clerk. Now that thirst for cash I felt immediately after university has faded. I've saved up a bit of money, and I'm going to wait for better times and then leave. Soon, I hope. But I certainly shall leave." "Work is a problem, and this job isn't at all interesting. Of course, I never expected to become a manager on taking my degree. You can accept an apprenticeship, as you can make do with 1,570,000 lire a month – if you have prospects, a future, the chance to grow. But that's the salary, and the prospects of getting anywhere are non-existent. Will anyone tell me why I should come happily to work, and sometimes stay on till eight in the evening, with no overtime pay?"

"But there's no career here. In order to become a prosecutor you must first pass the exam, and to have access to the exam you need two years' practice in an office and presence at a certain number (prescribed by law) of court cases. Here, none of us has ever set eyes on a court, a judge, a conciliator. Letters, yes, we write tons of letters."

"It's plain," says one of the firm's lawyers, "that if we had more motivated staff, not only would we be able to recover more debts, but we could also reduce our costs. Very often I find myself before the conciliating judge who wonders why there is all this bother with legal recovery when heavier pressure on the debtors would have done the trick. Moreover, the briefs that are sent to me by the unit are often inaccurate or quite wrong, and I sometimes have the feeling that they conceal something like a taste for making life difficult."

By Giuseppe Soda

Questions

- What is the present configuration of the systems of selection, training, career, assessment, and remuneration in the situation described?
- Under the assumption of maintaining and enriching the work content in the department, and to continue staffing with recent graduates, what would be a better configuration of the system?

Chapter 10

The Organization of Work: Structures

The problem examined in this chapter is the "division of labor" and the coordination between divided activities, which a person or group can be responsible for ("*job*"), on the basis of a set of rights and obligations. This level of analysis is often defined as the *microstructure*, as opposed to the *macrostructure* – the higher order aggregation of jobs in organizational units. The microstructural problem has sometimes been defined as the problem of the *organization of work*: forms of organization analyzed at the level of what people do and have the right to do (Friedmann 1963; Butera 1979; de Terssac 1992; Williamson 1980). *A form of organization of work will be defined here as a particular configuration or distribution of rights over the domains of action, control, decision, and ownership.* In this way, the discussion can integrate the classic contributions from the sociology of work and the socio-technical systems approach – focused mainly on the allocation of tasks and the rights of decision and control – and the tenets of organizational economics – looking principally at the allocation of property rights and the incentive mechanisms associated with a given distribution of tasks.

There are three parts to the chapter. In the first section, the principal problems that have drawn attention to the theme are recalled in a historical perspective. The second section offers a procedure for designing job boundaries and work coordination mechanisms, based on all the key variables that have been suggested by relevant research, as indicated in Figure 10.1. As also evidenced in the figure, in the third section, an overview of some particularly important alternative forms of organization of work, and a framework for their comparative assessment are given.

A LONG HISTORY OF ACTION RESEARCH PROJECTS

Job design is probably the oldest theme in organization studies. Perhaps also because of this, history is particularly important in this field. Early studies in the organization of work emerged from the problems posed by the diffusion of "Taylorism." Notwithstanding the productivity gains permitted by the specialization of jobs and in many industrial sectors during the first half of the twentieth century, many exceptions to the rule, important unintended consequences, and some changes in conditions, have stimulated work redesign studies and interventions. Among the problems dealt with in early studies there were: the loss of attention and productivity linked to fatigue and boredom; the disaffection and alienation toward work linked to

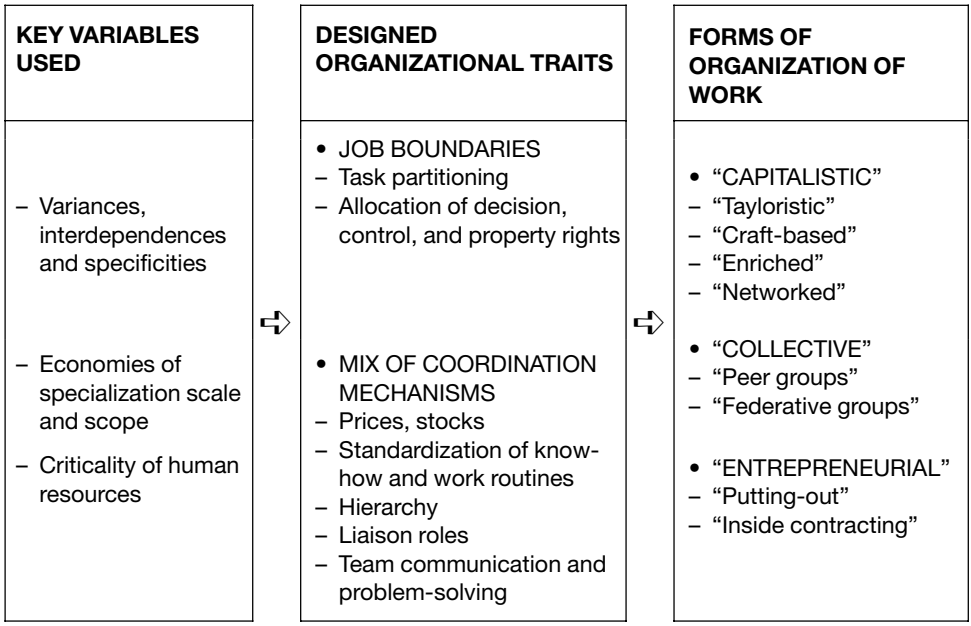


Figure 10.1 Job analysis and design

the lack of a sense of contribution; the low "quality of working life"; and the union opposition that Taylorism sparked from the very beginning.

Despite these numerous problems, the first wave of studies attempting to redesign the organization of work was characterized by many difficulties and failures (Trist 1981). It may be that these early studies, in the 1950s and early 1960s, were characterized by an approach that some later claimed was too "universalistic," attempting to diffuse new "philosophies of management" valid for all seasons and giving scant consideration to the different productive needs of various sectors. In addition, the economic conditions may not yet have been mature, while, later, the conditions that had favored the success of the Tayloristic model changed in many if not all sectors. Furthermore, research and intervention were initially focused precisely on those mass industrial production sectors where tasks could be divided and programmed profitably.

Later, some instructive studies started focusing on technologies and sectors that were either impermeable or unsuitable for the implementation of the principles of Taylorism, with its massive division of labor (both vertically and horizontally), and extreme formalization and programming. Examples include steel transformation, the chemical industry, and continuous processing technologies (Miller and Rice 1967; Murray 1960; Trist and Bamforth 1951).

Furthermore, even in sectors with divisible technology and standardizable processes, the 1960s saw the emergence of conditions of uncertainty caused by changes in the characteristics of demand, competition, and competitive strategies. This cluster of factors led to conditions – to use the label of the time – of *environmental turbulence* (Emery and Trist 1963). In sectors such as automobiles and

household appliances, the increase in the number of competitors, the increasing buying power and needs of consumers and the acceleration of the innovation of processes and products all implied productive systems very different from Taylorism, with its standardization and specialization of technical and human resources.

In addition, some of these studies and others highlighted how the development of automation enlarged the spaces of discretionality and *"organizational choice"* on how to organize work even under the same technological and market conditions, hence the likely "*equifinality"* of different models of work organization (Touraine 1955; Rice 1958; Trist *et al.* 1963).

Finally, in practice and history, the second postwar period was marked by a generalized development and legitimization of unionization, or otherwise in the weight of workers' interests in firm decisions in work-related matters, and the extension of the labor relation issues regulated by negotiation with respect to those submitted to a regime of authority. In organization theory and method this led to the recognition that *organizational arrangements at the microstructural level have a particularly strong and systematic impact on primary individual interests,* and to the inclusion of explicit surveys of the preferences of job incumbents in job analysis and design models (Hackman *et al.* 1975). More recent models have also typically included an analysis of the impact of the organization of work on the *quality of working life* (Davis and Cherns 1975), as inclusive of the broader consequences on health and family life.

Together, these studies have generated a fairly coherent body of knowledge and methodologies for work analysis and design. These will be discussed in the next sections, together with other more economic design models of work arrangements.

KEY VARIABLES AND ANALYTICAL STEPS IN JOB DESIGN

"Primary work system" and "unit operations"

A basic error that one could make in a job analysis would be to start from job descriptions as they are currently constituted, and/or to consider each of them as isolated from neighboring ones (both vertically and horizontally). For example, if we are interested in understanding how the job of a secretary could be efficiently and effectively designed, we should go beyond the activities that are currently assigned to the position, identifying which other activities these are interdependent with (even if they are currently assigned to colleagues or superiors), and understanding what criteria for aggregating these activities are currently applied and what alternative criteria could be used. To that extent, the two initial levels of analysis are the super-ordinate level of the "*primary work system*" that comprises the job being redesigned, and the subordinate level of the *elementary unit operations* that can be comprised in the job.

Primary work systems

A *primary work system* is a collection of interdependent activities that leads to an identifiable result – typically a unit of product or a service provided (Susman 1976). The reason for using this unit of analysis derives from the following hypothesis: if a system of activities is potentially able to regulate itself, an effective and efficient organization of this system should use this self-regulatory capacity. By doing so, the costs of coordination and control can be reduced and at the same time the level of satisfaction of the workers' identity need can be increased. A system that is

capable of self-regulation should "contain" within its boundaries the main sources of interdependence and uncertainty that can significantly influence its output. Therefore, a primary work system would typically have to include at least the following two types of related activities:

- The interdependent transformation activities that are involved in the realization of an identifiable and valuable output (for example, all the assembly activities that lead from the components to a finished automobile; or all the administrative activities that are needed to process a request for a bank service on the part of a client).
- The activities of support, maintenance, control, and regulation of the transformation process (for example, maintenance activities if we are analyzing a production process; human control by exception of numerical control machines in an automated process; decision and planning activities regarding the type and sequence of operative activities).

Unitary operations

If the primary work system bounds job design externally, *unit operations* set a lower bound beyond which tasks are no longer technically separable. The question is to what extent and according to what criteria these elementary operations should be aggregated; i.e. what the *"efficient boundaries" of jobs* are. Figure 10.2 shows a simplified matrix of the unit operations in a work system, identified by researchers in one of the most well-known large-scale socio-technical job design projects, the Shell project (Hill 1971; Trist 1981).

Once a work system and its unit operations have been identified, some key attributes of activities and of their relations can be analyzed. Among the variables that generally

Variances		1	2	3	4	5	6	7	8	9	10	Unitary operations
Temperature	1	K										Bitumen storing
Gradation	2											
Level	3	*		K								Mold filling
Mold position	4				K							
Speed of flow	5	*										
Extraction of separators	6	*		*		*	K					Extraction of separators
State of separators	7					*	*					
State of blocks	8				*		*					
State of molds	9						*			K		Cleaning
Action of cleaning agent	10	*					*		*	*		

K = key variance;
* = interdependence

Figure 10.2 Example of variance matrix

influence effective economic unit boundaries and inter-unit coordination mechanisms, the following have shown themselves particularly relevant at the micro-structural level.

Variances, interdependences, and specificities

Variances

In job analyses, the exceptions, the unforeseen events, and the uncertainties that can affect a transformation process are usually called "*variances*." Enriching the statistical meaning of the term, in job analyses *"a variance" is defined as a deviation with respect to a "normal" transformation process, that has a non-negligible effect on the output, and that creates a need for human intervention in the form of corrective action.*

For example, in the Shell case, for each elementary operation, some characteristics of the material or the processes can involve unforeseen variations in the conduct of the activity, affecting outputs and through them, eventually, other activities. The temperature and the other characteristics of bitumen such as the level of fill in the molds and the positions of separator grids could become abnormal and turn out to be key variances in the process. In fact, if that happens and someone does not intervene, the effect of the variance is felt on output through a waste of materials and wrongly shaped blocks.

Interdependences

The unit operations in the example of the formation of bitumen block molding may seem at first sight linked by simple sequential interdependence. However, the presence of a high level of variance in the process complicates and intensifies interdependences with successive operations, conditioning aspects such as the speed of flow, the withdrawal of separators, and the action of cleaning chemical agents at the end of the process. The analysis of downstream transmission of variances highlights how the operators upstream need a considerable amount of information about downstream activities in order better to regulate variances. The type of interdependence between activities can therefore be defined as reciprocal rather than sequential. In turn, this implies that effective coordination mechanisms should be based more on mutual adjustment between parties, than just on programming.

In general, *holding other conditions constant*, *the greater the variance and the more intense the interdependence between unit operations, the less efficient and effective it is to divide work into specialized jobs assigned to different workers.*

Initially, in the Shell molding unit, the work was divided in various different jobs, assigned to different people with different training and professional qualifications, including: a "filler" position (specialized in the mold filling operation), a chargehand role (responsible for storage tanks), an electro-magnets operator (removing the separator grids), and cleaning personnel. There were no mechanisms built into the process for the upstream workers to regulate the sources of variance ex ante; instead, the downstream workers would try to fix the problems how and when they could ex post, and many of the problems that they could not fix were transformed into defects in the final output.

Specificities

An additional factor that raises the level of interdependence and therefore the need for integration between technically separable activities is the reciprocal specificity of the

resources generating those activities. *In general, holding other conditions constant, the more activities are linked by specific transactions, the less efficient it is to divide these activities into different jobs.* For example, if the activities needed to operate a particular set of machines, to maintain and repair them, and to decide how to allocate them to different productions all require reciprocally specific competences, then the activities of production, maintenance, and production planning can be effectively and efficiently conducted by the same worker or by the same group of workers. In certain studies some evidence can be found to support the hypothesis that where there are job-specific competences, jobs tend to be richer and wider than otherwise (e.g. O'Reilly 1993).

Variance, specificity, and interdependence therefore favor relatively more "integrated" jobs.

Specialization, scale, and scope

Horizontal specialization

If interdependences between elementary activities typically work in favor of the aggregation of tasks in wide jobs, then often economies of specialization, and sometime also economies of scale, work in the opposite direction. Herbst (1976), in particular, has specified this important step in job analyses. He proposed analyzing various matrices of people and tasks such as those illustrated in Figure 10.3 (they could be also constructed as competences/activities matrixes) (Chapter 8).

The situation defined as "maximum specialization" indicates that significant economies of learning and specialization in single tasks are present and that the "bundle of resources" of a single individual can be fully used in a single task. For example, in a surgical operation, it is efficient and effective that the tasks of cutting, sewing, and anaesthetizing be done by different people with specialized jobs, both because they require very different knowledge and learning paths and because the operations must be done in parallel.

A situation of maximum polyvalence can be derived either from the facility and from low "entry barriers" to different tasks, or from the presence of economies of scope among different activities. For example, one who has acquired competence in a task (for example, writing a computer program to automate a procedure) may be the best candidate to use this acquired know-how in other work (e.g., programming other procedures).

Vertical specialization

The same type of analysis is also relevant for the vertical dimension of the micro-structure – that is, for the choices of vertical aggregation

	α	β	γ
A	1		
B		1	
C			1

Complete specialization

	α	β	γ
A	1	1	1
B	1	1	1
C	1	1	1

Complete polyvalence

Figure 10.3 Specialization matrices

Source: Adapted from Herbst (1976).

between operative activities and related activities of decision, control, and regulation. For example, the operative activity of traditional manual work at the lathe did not create barriers to the aggregation of decisional activities concerning the sequence of work and the re-equipment and maintenance of the machine. Rather, the knowledge and skills regarding the state of use and the performance characteristics of the machine and the materials generated many of the competences relevant to the programming of production and the regulation of variances. This is different for numerically controlled machines (for example, the automatic lathe), and even more for complex automated equipment, where the activities of surveillance and feeding of the process require low level and fairly generic competences, while the activities of planning and regulation of the variances (possible problems and production anomalies) require high-level and very specialized (electronic and electro-technical) competences. These competence and specialization problems can explain some of the observed consequences of automation in terms of "polarization" of occupational and job structures between a band of execution and machine surveillance jobs and a band of planning and designing jobs (Pollock 1956; de Terssac 1992).

For other types of activity, we find, on the contrary, that it would be absolutely ineffective and inefficient to separate the operative activities from those of decision and control: because operation and control activities are based on the same competences; because operative activities involve very high variance and cannot be planned; or because activities are hard to inspect and guide without actually doing the work (Alchian and Demsetz 1972; Susman 1976; Butera, 1979).

Observability and criticality of work inputs

It may be noted that the vertical aggregation of action, decision, and control rights could create problems of conflict of interest that might make it inadvisable even where it responds to the need to control uncertainty and the existence of common competences. Leaving the agent to decide what actions to undertake can lead to his or her reducing effort and following particularistic objectives, whereas allowing the agent to control his or her own actions could lead to self-serving biases and moral hazard problems (Milgrom and Roberts 1992). As distinct from the horizontal aggregation of tasks with different specialization, therefore, the vertical aggregation of the activities has implications for the allocation of property rights.

Agency theory studies (Levinthal 1988) have been paying particular attention to jobs in which significant decision and regulation rights are associated with the right and obligation to take action. In fact, by definition, an agent is not only entitled to act but also to decide which action to take.

In the case of observable activities, a possible solution is a division of rights: an allocation to the agent of the rights of action and decisions over activities, but an allocation to the principal of the rights of control (for example, the activities of a skilled worker with high qualifications and discretionality, such as a chief laminator in a non-automated process, or the activities of a nurse in a hospital). If activities are not observable, then incentive- rather than control-based solutions to the problem are in order (as it is often the case in sale activities). This result can be obtained through a reunification of the rights of action, decision, and control with a more or less ample quota of residual

reward rights and other property rights (Chapters 4 and 9).

A corollary of the previous proposition on the diversity of efficient and effective systems of work organization with activities with high or low observability is that *activities with different levels of observability should not be united in the same job* (Milgrom and Roberts 1992). The integration of tasks differentiated in that respect is likely to bias behavior toward those aspects of performance that are measured and rewarded at the expense of other tasks. Corrective measures should be taken if the aggregation is advisable for other reasons.

With regard to the conditions of efficient allocation of property rights to the workers, other studies in organizational economics have underlined the importance of other variables, beyond the degree of observability of work. Further elements are:

- the degree of substitutability of human resources
- the value added by work contributions with respect to other inputs
- the risk to which human capital is put with respect to other resources
- the extent to which the knowledge and competences on which activities are based have been accumulated by the workers and are tacit and poorly transferable.

If this set of variables is synthesized in an indicator of *human resources criticality*, keeping other conditions constant, it holds that if human resources are critical, a fair and efficient form of work organization should involve the allocation of property rights to the actors contributing those resources. In fact, arrangements under which "*work hires capital*," or associational contracts among the providers of different types of human, technical, and financial capital (Williamson 1980; Grossman and Hart 1986; Hart and Moore 1990; Milgrom and Roberts 1992) are often observed in human-capital-intensive firms (law, consultancy, education, health care, etc.).

Analysis of preferences

A complete procedure for job analysis and design should either make assumptions on worker preferences or empirically survey them. Organizational economists usually opt for the first methodology – making assumptions. For example, some important economic treatments of job design and the organization of work assume that workers' preferences are distributed randomly and therefore do not affect efficient arrangements in a systematic way (Williamson 1980). Others, like agency theorists, assume that workers preference orders over actions and work arrangements conform to a principle of maximum economic benefit with minimum effort.

However, this approach has important methodological drawbacks: first, it binds solutions to a particular configuration of preferences, and second, it does not provide evidence that interested people would actually "vote" as assumed if requested to express preferences about work organization arrangements.

In this respect, the socio-technical tradition, in tackling the practical problems of implementing redesigned jobs, ended by developing an empirical approach to the analysis of preferences that can be claimed to be more rigorous not only in the terms of scientific research standards (it is more falsifiable) but also in terms of its accuracy in the representation of interests (a utility theory criterion).

Socio-technical research has shown, first, that the preferences that workers define regarding their own work activity *consider not*

only the actions to undertake, but also the modality by which such activities are organized, and second, that *these preferences are not distributed randomly* but that their configurations are to a great extent predictable. Below, the main job attributes that are systematically used by job-holders to evaluate their jobs are reviewed, together with some substantive results on how they might be rated under what circumstances.

Variety

Individuals usually have definite preferences about the variety of activities that their jobs allow. That does not imply that in all cases more variety is preferred to less variety. For example, it has been found that in cases of uninteresting work with little autonomy, the horizontal aggregation of more tasks to augment the variety and reduce the monotony may not be appreciated by the workers. In fact, the increased variety in these situations augments the levels of attention required and fatigue for work that remains, which is in any case, uninteresting, and reduces the possibility of dedicating one's mind to other activities while one works in an automatic mode (for example, social relations, planning one's family life). On the other hand, more variety is often appreciated in the activities that have a high discretional content – for example, by line managers (Mintzberg 1973).

However, the horizontal aggregation of different activities assigned to the same person can be judged positively, even in situations of low discretion, where monotony reaches health-damaging levels. Conversely, variety can be judged negatively in high discretion and rich activities where high concentration and specialization are needed for problem-solving (such as for planners or researchers instead of line managers).

Autonomy

The degree of autonomy characterizes the vertical dimension of a job. Autonomy is the extent to which decision, planning, and control activities are attributed to the job. The degree of autonomy therefore expresses the element of discretion, self-control, self-determination, and the degree of freedom possessed by an actor in a job. For example, the job of a salesperson is usually characterized by a greater level of autonomy as against that of a production worker.

It has already been emphasized that the degree of information complexity and of variance in activities is positively related to the effective amount of autonomy in the job. Workers' preferences can be expected to act in the direction of extending the level of job autonomy required by task characteristics, because people fairly systematically prefer more autonomy to less autonomy, especially if they are currently in positions that are highly constrained and formalized (Crozier 1964; Crozier and Friedberg 1977; Salvemini 1977).

A dynamic factor and the starting level of autonomy explain the exceptions to this rule. If a very large increase of responsibility and autonomy is planned with respect to a preceding situation, and if the initial job is already challenging, then workers may perceive excessive risk and stress – especially if self-confidence and self-efficacy in the task are not yet high. For example, a young person in a staff position in a personnel department might not appreciate the premature assumption of personnel management responsibility. Or again, there may be a perception of too much autonomy and overstress in very flexible organizational systems, where jobs are not well defined (Lawrence and Dyer 1983).

Identity and identification

The need for identity, for a positive image of oneself and the meaningfulness of one's contribution, is considered one of the basic needs of human beings. In work contexts its realization is influenced by the possibility of identification: with the output of work or with reference groups. Research shows that the lack of significance perceived in what a person does negatively affects job satisfaction (Hackman and Lawler 1971; Grunenberg 1976) and that a perception of significance is sustained by the clarity of connection between workers' contributions and identifiable and valuable outputs. For example, a job that is limited to monitoring an electric circuit in an assembly line for an office machine would be classified as a low meaningfulness job; while a polyvalent job inside an assembly group that has an identifiable output, such as a typewriter, would give the worker a greater sense of contributing something meaningful.

Studies on the meaningfulness of contributions have highlighted how this can be enhanced by either increasing the interest, image, and identifiability of the output (Bergami 1996) or by increasing the clarity of the relationship between the worker's individual and partial input and the final result (Salvemini 1977).

In a wider perspective, identification is however possible even where outputs are ill-defined. Identification with a social group – a craft, a firm, a profession – can perform this function (Gouldner 1957/8; March and Simon 1958). A work position belonging to a system with high identification potential (high social status, strong cultural identity, "a name") may be preferred to positions similar in tasks and better rewarded in low identification systems (Albert and Whetten 1989).

Social interactions

The possibility of engaging in social relationships and satisfying needs for closeness, sociability, power, and emotional or affective exchange at work is not uniformly appreciated by people. Factors that can influence the preferences of a worker for relationships with colleagues are diverse, and include age, cognitive style, how interesting and technically difficult the job is, and the interpersonal "chemistry" – how well people get along who are not free to select each other merely on the basis of personality fit (Tosi 1992).

However, even the early studies on job satisfaction revealed that extreme conditions such as isolation (the technical impossibility of speaking and interacting with others) can be systematic causes of dissatisfaction and stress (Walker and Guest 1952).

Development

A job attribute that is often under-defined is that of the dynamic prospects for professional development, which prepare the jobholder to assume other more qualified or attractive jobs in the future. Most workers assign a positive preference to good prospects for career development (meant in the sense of progressive development of the individual either within or outside the current firm), but the type of development sought is not uniform. For example a classic empirical distinction between employees with regard to the preferences for different career development paths is that between "locals" and "professionals": the former are people identified with a particular organizational system and oriented to the career within that system, while the latter are people identified with a profession, in search of career development within that profession – such people tend to have a much stronger propensity to

geographic and organizational mobility (Gouldner 1957/8).

In addition "development" may refer to the content of jobs or also to the type of contract and the type of rights attached to it. For example, in traditional or handicraft industries in areas with strong local identity (e.g. in some industrial districts), people attribute a positive value to tradition, and to the continuation of social relationships and of craftsmanship. To this extent, the evolution in the content of work is not of primary interest. Instead, people often look for development from the point of view of the rights of control and ownership over their activity, with the objective of doing the same activity "on one's own behalf" instead of "working for others' (Inzerilli 1991).

Self-actualization

Both theory and empirical research have demonstrated that work can be a source of intrinsic reward for those who do it, rather than merely a means, and an expenditure, for receiving extrinsic rewards. Such intrinsic rewards, and the sense of self-fulfillment that these lead to, often consist of the psychological benefits from doing one's job, which can be manifested as interest, amusement, a sense of competency, or merely using all of one's capacities for a useful purpose. The primary characteristic of the job that can feed these intrinsic rewards from work is the correspondence of the job's activity with one's individual competences, capacities, tastes, and values.

It should be noted that, among all job characteristics, this one is the most volatile and subjective. Even if full recognition is given to the subjectivity of actors' preferences, one must note that it is very difficult for individuals to evaluate in advance how interesting an activity is likely to turn out. To that extent, for this characteristic of the job even more than for the others, not only should preferences be empirically elicited, but actors must have the time to acquire and learn about them in the course of action. It is not by chance that many projects of job analysis and design have taken an "action-research" approach, based on long processes of collective field learning about cause–effect relationships and the objectives and preferences of actors (Elden and Chisholm 1993).

Health, safety, and quality of work life

This collection of factors is more relevant to the external conditions and the context of the job than to job content. Emphasis has recently been placed on the opportunity for designing not only jobs, but more ample change packages oriented to a more comprehensive improvement of the quality of work life (Hackman 1977).

This need derives primarily from the impact that job design has on safety, health, and the work environment. For example, jobs that are extremely fragmented with little variation can have negative effects on one's psychological health; and jobs that are not sufficiently rich and coordinated can increase the risk of errors and accidents (Golzio 1985). In addition, the organization of work obviously affects private and family life, through the availability of free time and energy, the richness of stimuli, the influence on social identity and prestige, and on psychological equilibrium. The characteristics of jobs also have important consequences on the possibilities for two-career families and on equal opportunities for men and women. In the *Quality of Working Life (QWL)* programs (Davis and Cherns 1975), the "comprehensive well-being" of the person is treated as relevant.

Another enlargment of perspective

proposed in QWL studies is to ask explicitly not only workers but also employers about their preferences for different job design solutions (Tosi *et al.* 1986). In traditional socio-technical studies, in fact, the "preferences" of the entrepreneur or manager were assumed to coincide with the technical needs based on the analysis of elementary operations and their interdependence – which is not necessarily the case. Lastly, technology and technical needs themselves are seen as an element that can be re-designed and co-designed together with jobs, rather than as a given and unmodifiable independent variable (Davis and Taylor 1973; Edberg and Mumford 1978).

Approaches to "joint-optimization"

Designing or assessing a micro-structural arrangement is, as is evident from the above discussion, typically a multiple-actors multiple-objectives problem. In the socio-technical tradition, this was early acknowledged and some rules for a "joint optimization" of the frequently contrasting "technical and social requirements" were devised to solve job design problems.

Operatively, it was proposed to assign utility points to each alternative organizational solution identified, both in terms of "satisfaction" and in terms of "efficiency." Each utility evaluation should be the sum of grades obtained by a job design on each relevant job dimension, such as variety, autonomy, the meaningfulness of the contribution, etc. Then the solution with the highest score could be chosen (Lupton 1975). For example, suppose that a particular arrangement, say an autonomous working group solution, receives the following "votes" on a 5 point scale from workers and management respectively: 2 and 3 on variety, 2 and 4 on autonomy, 4 and 4 on meaningfulness, 2 and 5 on development and career implications. The work group solution would be evaluated with 26 total points $((5 \times 1) + (4 \times 3) + (3 \times 1) + (2 \times 3))$.

As an applied example, we can consider the implications that this would have in the Shell case, introduced earlier in this chapter. The analysis of variances and interdependences highlights how a recomposition of activities in larger and richer jobs would have many advantages: fewer errors, better final output quality, and a better utilization of personnel. On the issue of specialization, the filling phase and the electro-magnetic removing of grids required different training and timing, so that they could not be easily aggregated. To that extent, a technically efficient solution might have been to enlarge job boundaries to coincide with the two technically different phases of the process, to be assigned to two different small groups. Coordination tasks could be assigned to the most competent figure (the filler), whose status in the group was at stake.

The approach described employs a fairness criterion based on the sum of utilities. As such it could be criticized for implying interpersonal comparisons of utility – as well as for being an "armchair" integration of interests.

Indeed, both in theory and in practice these issues are conducive to solution by negotiation (or voting). If organizational solutions are negotiated, we could expect the parties to "weigh" their own utility for different jobs' attributes and find efficient exchanges between them (e.g. a greater flexibility in the allocation of people to tasks could be exchanged with greater task autonomy and meaningfulness); and that they would appraise their reciprocal "weight" or bargaining power through the negotiation process. A stylized reconstruction of a process of negotiated design of the organization of work in a harbor is presented in Box 10.1.

Box 10.1
Negotiated organization of work in a harbor

In 1987 a long series of strikes led to the closure of Genoa harbor. This action was taken by the Autonomous Group of Harbor Workers – a quite unusual organization which is a combination of a trade union and a cooperative which "sells" its work force time to the harbor authorities – a consortium of enterprises. The group of harbor workers was created from the older corporations typical of commercial harbors and with time it acquired a series of privileges, the most important of which was an exclusive right to perform work related to harbor activity. The group is organized as an association, its members share the profits and have a kind of job security in the form of a fund to ensure everyone a minimum salary. Among members there are clan-like relationships with strong norms of mutual social assistance, which turn out to be useful in the coordination of various stowage jobs that are sometimes considered dangerous.

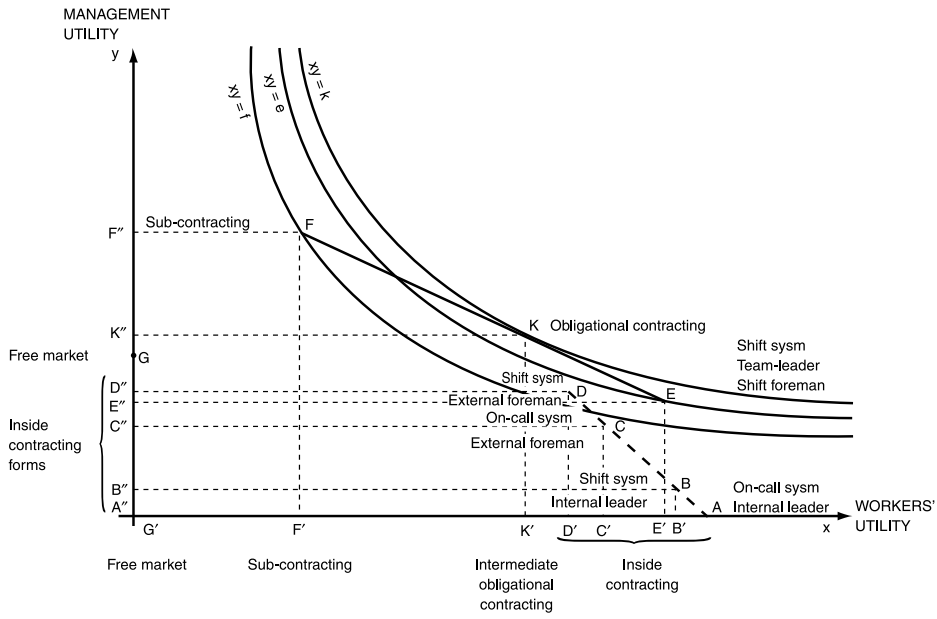
The Genoa harbor authority had been recently restructured, changing from a single functional hierarchy to a consortium of autonomous enterprises, with the goal of increasing awareness of local demands to utilize specific competences and reduce structural costs. The organizational change included the transformation of the role service system from an "on call" structure to one based on shifts. In the "on call" structure (which had been established when harbor traffic was much less intensive), workers were on duty and were called when a ship came in, based on the estimate of the number of people and skills needed to load or unload the cargo in question. Thus the group would work by the hour until the job was completed. The central figure of the group was the "caporal" who, usually, was the natural foreman; then there were the members of the group, who usually had worked together for some time. The foreman would assign jobs, supervise safety procedures and deal with the clients in order to ensure that a priority system was respected for loading and transportation. Both the "on call" structure and the role of the foreman were subjects of dispute by the recently organized harbor authorities. Management claimed full right to exercise control as far as relationships with clients were concerned, as well as the right to choose the foreman. In addition, management wanted to introduce a shift-based structure that would ensure the availability of resources present at the harbor, 24 hours a day.

Negotiations were difficult for both parties. The four different projects originally presented were nothing but the four combinations of "on call" and shift structures, and internal or external foreman (see line A–D in the figure). The negotiations finally failed and the harbor suffered a dramatic loss of traffic which was picked up by other ports.

It might have been possible to explore some Pareto-superior solutions (some of them were actually recently studied) to try to separate the issues in order to find efficient trade-offs, thus modifying negotiations from distributive to integrative ones. For instance, by separating the role of the foreman into its various components, it would have been possible to realize that the workers are actually interested mainly in an internal allotment of jobs and in the pursuit of safe working conditions; on the

other hand, harbor management gets the most benefits from controlling the external components of that role. A specific redesign of the role of the foreman, properly negotiated among the parties, could have led to splitting the role into two parts; the first part would be assigned to a foreman chosen by the workers' association, thus responsible for the internal organization of the group; the second part would be assigned to a shift representative appointed by the port authorities, who would be responsible for job planning and customer relations. This project could be associated with a shift-based structure in which the group members are reliable and selected in order to preserve an efficient social coherence. However, this organizational structure implies gains by both parties and is thus Pareto-superior compared to the majority of other solutions that are based on negotiating the distribution of a fixed sum. In addition, it involves minimal losses in terms of cost efficiency, given that 24-hour service, together with management control and the planning of the stowers' jobs based on direct contacts with clients (the lack of which in the past increased costs and caused an arbitrary and uneven service) were guaranteed.

As the figure shows, this is only one part of the story. If we do not take for granted that a quasi-internal system, currently protected by the law, lasts forever, there are conflicts of interest that are much more fundamental as well as cost efficiency problems. As a matter of fact, management could consider the idea of paying the price of a unilateral action aimed at repealing the statute that protects the exclusive right of the group of workers to the harbor work so as to reinstate a free work market. However, the market mode of management of the transactions would not represent a point of agreement given that the game would be a zero-sum one. If the choice were between a free market and any form of internal regulation of work relations, both parties could only act unilaterally to try to win.



Another solution is, however, applicable. The harbor workers group would have to assume that it is improbable that the workers could maintain such a highly protected organizational structure in an environment where there is strong competition among European ports. On the one hand, it is true that it could be difficult for the harbor workers to sell their labor to organizations other than the port authorities. On the other hand, owing to the high costs of replacing them, the former would have the right to fight for a highly integrated solution. The port authority (1) prefers a free market because the harbor workers' skills are not sufficiently complex or specific to distort or block competition in dealing with loading and unloading procedures; (2) can count on a stronger negotiation position, being less replaceable and being supported by the increasing pressure of competition. If the harbor workers considered the risk of the port's declining, they might decide an acceptable sub-contracting formula. In the case of such a solution there would not be any additional statutory rights of monopoly: rather, we would notice privileged long-term contracts and an enduring relationship as long as the needs of both parties were totally satisfied. The port authority, however, could prefer this alternative to that of the free market, given that it would have to take into consideration the long experience at the port and the social cohesion of the various groups of workers..

All this said, the structure of negotiations on organizational forms can be described as shown in the figure. Among Pareto-efficient forms, two are particularly interesting; solution E (the mixed form of internal contracts) and F, the subcontracting solution. According to the equity criterion (the product of utilities) they both rank high, as is shown by the position of the hyperbola touching those points ($xy = e$ and $xy = f$). The truth is that the inside contracting E form is "more equitable" (given that $e > f$) and reflects more closely a principle of balance among the parties; whereas the sub-contracting form is more cost-efficient and reflects more closely the relative structural power of the parties. Also, if we assume that there is evidence of a continuum of possible contacts, between E and F, each of them representing various degrees of intensity of commitment of the parties (based on merely informal agreements) we can interpret line E – F as a continuum of contract forms that are binding to various extents. Assuming hyperbole $xy = k$ tangent to this line, we can thus determine the degree to which the contract is binding as well as the specific internal organization form which tends to maximize the product of utilities (point K in the figure).

Source: Grandori (1991).

A critique that economists and games theoreticians would make about both the socio-technical approach and the negotiation approach to job joint optimization, is that many of the conflicts between preferences about the configuration of the jobs could be simply resolved through side payments and monetary indemnities paid to workers in exchange for their acceptance of non-preferred jobs and tasks. In effect, this happens in part. Monetary aspects are in fact usually negotiated and defined in close connection with job characteristics. It happens that jobs that are unpleasant, or dangerous, are compensated for with extra bonuses. However, the effective use of payments for this function has limits for at least three important reasons. First, the monetization of

the imbalance in the distribution of other desired resources would reduce the incentive to seek solutions that would be qualitatively superior. Second, not all the characteristics of the organization of work can be monetized without creating ethical problems and affecting the human rights of a worker (such as health). Third, compensation is typically a tool that is already overloaded with functions. Hence, compensation structure may send too many conflicting signals: payment should be correlated to the level of discretion and decisional responsibility; it should also (on the contrary) be an indemnity for jobs that are poor or hazardous; it should also guarantee a decent quality of life for everyone; it should also be correlated to results. The result of all this could be that of weakening the readability, the perceived fairness, and the transparency of any compensation system.

ALTERNATIVE FORMS OF WORK ORGANIZATION

In this section some salient discrete configurations of work organization, that have been shown to be effective under specified conditions, are reviewed. Such solutions involve different combinations and allocative configurations of the rights and obligations defining a job: the rights to action, decision, control, reward, and ownership; as well as different combinations of coordination mechanisms. *These are only a few combinations among the many that are theoretically possible*.

Williamson (1980) has proposed a distinction among forms of organization of work based on the allocation of property rights that can provide a useful starting point. The classification identifies a few broad types of proprietary structures: capitalistic, entrepreneurial, and collective. Within these general classes, various subtypes can be distinguished characterized by different allocations of other rights and obligations,[1] with the support of empirical organizational research.

Capitalistic forms

In the forms of work organization traditionally defined as "capitalistic," ownership over all of the technical resources – raw materials, intermediate products, finished goods, facilities, equipment, and know-how – is allocated to a single party. This actor or group of actors own the capital and "hire" the labor of others using various types of contracts. Primarily, then, within capitalistic forms, jobs are characterized by work assignments carried out wholly or in part "for" (in the interest of), and "according to the instructions of," actors who are not the agents themselves. In general, the effectiveness of this broad class of forms is linked to contingencies such as: the greater criticality of technical and financial capital with respect to human capital; large teams; and the measurability of agents' performances (Alchian and Demsetz 1972; Hart and Moore 1990; Chapters 4, 8, and 9).

However, many variants can be identified inside this general class, characterized by different degrees of division of labor, different allocations of decision and control rights, and different coordination mechanisms.

The "Taylorist/mechanistic" model

This model of organization of work is characterized by a maximum division of labor between different workers and by an allocation of the tasks of decision, coordination, and control to a super-ordinate authority. Conceptually linked to the thought of Taylor and Weber, it has been and still is widely

diffused in manufacturing sectors with stable technologies and competition based on costs, such as traditional metalworking or food processing.

It is often said, especially in organization and sociological studies, that the "Taylorist" model is out of date, and that by now we are living in a "post-Fordist" epoch. This is probably true as regards many cultural and ideological traits of Taylorism and Fordism. However, if we consider the more technically organizational attributes of a configuration "with maximum specialization and programming," we may doubt whether it has disappeared or whether it may not be efficient in some circumstances. Thus, just as "Taylorism did not stop at Prato" (and in many other technological "places") – i.e. it did not spread everywhere – so in others it has been preserved or has been rediscovered. Specialization may also be associated with the deepening and excellence of knowledge and skills in sophisticated activities (Chapter 9). The case in Box 10.2 illustrates the possibility of a professionalized, post-industrial Taylorism.

"Craftsmenship" models

Historically antecedent to Tayloristic models, "craftsmenship" forms still provide an interesting, more decentralized, alternative. Consider the type of organization described in Box 10.3. Specialization and qualification are both high. The material and know-how are not standardizable, knowledge is to a large extent tacit and embodied in specific people, and workers' discretion has a paramount impact on product quality. Knowledge is diffused and technical resources are not clearly owned by any "one side." Actually it would be rather irrelevant whether the craftsmen are "hired" by the entrepreneur who accumulated the commercial know-how or they were collaborating with him on the basis of long term contracts. In every case, property rights configuration would not change much and the key coordination mechanisms would remain the same: the standardization of know-how, work routines, communities of practice, norms, and codes of conduct.

Examples of this "craftsmenship" arrangement can be found wherever high-quality products must be made from non-standard raw materials, such as in the leather or woodworking industries and are particularly diffused and performing effectively in Italy and other European countries.

"Enriched" models

Many organizational change projects at the microstructural level have been aimed at modifying Tayloristic systems of work so as to make them more flexible and more motivating (Butera 1984). Three types of interventions about job descriptions have become widely diffused: *job rotation*, *job enlargement*, and *job enrichment*. Job rotation is aimed at increasing the knowledge of the worker about the entire cycle of work, increasing the worker's sense of offering a meaningful contribution, and the equivalence of resources through the periodic assumption of different jobs on the part of the same worker. Job enlargement is aimed at increasing variety through the aggregation of different tasks at the same decision level. Job enrichment is aimed at "vertically loading the job" with activities of decision, control, and planning, so as to guarantee a better regulation of variances, an improved capacity for local adaptation, and a better satisfaction of the workers' needs for autonomy and self-actualization.

Box 10.2
Professionalized Taylorism in eyeglasses

Optissimo is a young commercial distribution firm (not more than six years of age) selling eyeglasses (sight and sun, with more than three thousand frames per sales point) and contact lenses of the best makes, through a network of its own shops located throughout the country.

The Offer

The offer combines customization with rapid service. Optissimo holds rapidity to be an important competitive factor because it enables the customer to exploit the time of the visit to the shopping mall (average 2.5 to 3 hours) to best advantage, avoiding further visits that may be needed. The formula of Optissimo's offer is proposed to the market through the opening of the first four sales points in a like number of malls.

Subsequently, Optissimo proposes the same offer to the customer going into the town center, in the opinion that the value of time is a crucial element also for this second segment of its clientele.

The Shop: the Contact Area

At the Optissimo sales point, frames are displayed in transparent cabinets, with interposed mirrors, for ease of access and choice by the client. The cabinets are lit so as to enhance the frames and decorate the shop walls.

The customer is welcomed at a reception point at the shop entrance (provided with cash register, telephone, and counter sales items). In the center are little tables (the counters), each furnished with chairs and personal computer, where the client is assisted by the *sales optician*. This person (holding an optician's diploma and suitably trained by Optissimo in-house) has the job of choosing and advising suitable lenses and counselling/confirming the best frames, from the technical point of view (correction of visual defect). In choosing the frames, the client does not consider (not being competent to do so) the aesthetic and functional effects of the lens needed (more or less thick) on the spectacles when it is mounted.

If the client arrives without a prescription, the *optometrist* (holding a diploma) proceeds to test the client's sight in the refraction room, a specialized part of the shop.

Staff (men and women) performing the two above roles wear a dark-colored uniform and a tag showing their first name.

The Shop: the Back Rooms

Behind the counters are the laboratory for lens construction and fitting, and the refraction rooms and rooms for application of contact lenses. The laboratory uses advanced electronic equipment enabling many operations to be effected automatically and in quick time. The laboratory is *open to view*, so that the customer can personally follow all the stages of construction and fitting of lenses as they are performed by the *laboratory technicians* wearing white coats. In the shopping malls, the Optissimo sales point is located in such a way that passers-by can see inside the laboratory.

The Process of Purchasing and Production of Service

While conversing with the customer, the sales optician uses the computer to input the customer's data in the data bank of the sales point; to look for the correct lenses and items regarding the frames. The duration of the conversation depends on the customer's needs, perceived or implicit. The sales optician's skills in listening and asking the right questions are aimed at understanding the use value of the spectacles for the client and thus advising on the most suitable kind of lens; for example, a bifocal lens instead of two distinct lenses. The sales optician has to manage both the client at the counter and those waiting (through visual contact and welcoming or orienting speech), especially at peak moments.

If eye testing is needed, the sales optician accompanies the client to the optometrist, who does the test in the refraction room. Testing is done with a state-of-the-art machine (the electronic Phoropter), equipped with measuring lenses. The machine changes the lenses rapidly in front of the customer's eyes and guarantees very precise measurement. As against the manual equipment traditionally used in family-run opticians' shops, the electronic test halves the time needed (15–20 minutes compared with the usual 40 minutes), and has much greater precision. The results of the test are entered by the optometrist in the "technical" prescription that contains indications about the value and type of lens and is passed to the sales optician.

The latter adds to the prescription the type of frame and places the document (together with the frame selected by the customer) in a numbered box which is then handed to the lab technician for construction and fitting of the lenses in the frame to complete the spectacles. Standard working occurs in mounting the "finished" lenses, already having the correct power (for visual correction) in the "closed" frame (that entirely surrounds the lens). The lab technician is responsible for the following operations:

- Extracting the finished lens from the relevant drawer and finding the focal point of the lens manually instead of electronically – because the operation is then more precise. Time for both lenses: two minutes.
- Inserting the frame in the "electronic scanfor," which reads (by means of the palpator) electronically the shape of the frame and transfers it to a display. The technician introduces in the display the data needed to obtain correct centering and then superimposes the centered lens so that the points coincide. The technician then inserts the lens in the grinder. This shapes the lens and marks the border, making it round so that it can be fitted in the frame. Time: for each lens two minutes, for both lenses four minutes.
- Manual fitting of lenses in frames. Time for both lenses: two minutes.
- Final check of the spectacles, consisting in verifying the centering, and checking the setting (the earpieces must have correct inclination with respect to the front piece, must sit neatly against the temples, and the lenses must be flush with each other). This synthesizes all the work performed for the client's satisfaction. Time: four minutes.

Work on the closed spectacles therefore takes 12–15 minutes. Variations are due to some not strictly technical operations (e.g. fetching lenses from the stockroom, or calls from the counter).

Special working consists in constructing special lenses (diopter power, type of lens, bifocal, gradual, or type of material) in the laboratory at the sales point. Particularly as regards special working, Optissimo undertakes to deliver the spectacles within one hour after the frames have been chosen (and, if necessary, performance of a test). The special working represents the distinctive element in Optissimo's offer as compared with competitors. In constructing the lenses, the lab technician performs the following operations:

- Choice of the ideal *blank* (the unground lens) and defining the work plan. The technician inputs the technical prescription data into the computer, which performs the necessary calculation to define the curve of the lens and gives the work plan (border and center of lens), tools (grinders), and setting for the work. The data are transmitted automatically to the machine that carries out the work. Traditionally the work plan is made by the optician who does all the calculations manually. Use of the computer, however, requires that the plan be verified by the lab technician, who, on the basis of experience, may make necessary modifications to it. Time: two minutes.
- Working the *blank*. This involves operating only on the back of the blank, in order to obtain the necessary rear curvature. The working consists of the stages of finding the inner curve that will give the diopter power), lapping (eliminating the furrows left by smoothing the surface) and polishing the lens. Time: smoothing for both lenses, 15 minutes, lapping and polishing both lenses, 11 minutes.

The subsequent operations depend on the choice of frames and are those of standard working.

Source: By Luigi Golzio.

This model is widely and effectively used in sectors and activities with high variance and high uncertainty, high interdependence among elementary operations, and relatively low requirements for specialization in single activities. A combination of these conditions has also occurred in some sectors that have been traditionally organized according to the Tayloristic model, such as the automobile or office equipment sectors. Richer or enriched jobs are more widely diffused in sectors of activity strongly exposed to variable market demands and competition based on differentiation and innovation of products; in activities involving white-collar workers, in many service activities, in activities in which the work is qualified and firm-specific.

"Networked" models

A more radical solution aimed at a more "flexible" organization of labor is to allocate a job to a group rather than to an individual (Salvemini 1977). These solutions are effective in the simultaneous presence of specialization economies in single activities and high variance and interdependence. Positive preferences for social interaction and group work, and low individualism concur. The success of Japanese group-based organization of work in sectors such as electronic equipment can be taken as an emblematic example.

Various configurations may be conceived and are diffused in practice. Herbst (1976) for example, early identified some "alternatives

Box 10.3
Castello pipes

Pipa Castello was founded in 1932 by Carlo Scotti in Cantu' for the production of a top quality smoking pipe that could compete with the best producer of the time, the English *Dunhill*.

Over more than sixty years, *Pipa Castello* has become known for being the only producer that directly performs all the phases in the construction of the pipe, including the research, the work on the briar wood, and the construction of the mouthpiece. Every pipe is carefully hand-made from one single piece, owing to the characteristics of the raw material (the briar wood).

Vertical integration distinguishes the Castello Pipe from its competitors, who assemble and sell pipes worked by small artisanal subcontractors.

Among smokers, the Castello Pipe is a true cult object which is often purchased for the simple pleasure of owning one, or for a collection. The firm is managed by Mr Scotti's son-in-law, who also owns it.

The Product: the Smoking Pipe

At the beginning of the century, the pipe was used by lower income people to smoke tobacco. Apart from those made with briar wood, pipes were therefore manufactured with poor raw materials, e.g. cherry wood, terracotta, and meerschaum. Cigarettes, on the other hand, were in limited production and were reserved for a wealthy elite of smokers.

Italy soon became the largest producer of pipes in the world, with 35–40 factories located mainly in Lombardy, many of which worked as subcontractors for distributors, often British, who then would put their trade mark (the "punzone") on the finished product.

The importance that Italy gained in the pipe making industry is explained, among other things, by the fact that the raw material, briar wood (a form of tumor of the roots of heather) is a Mediterranean plant typical of the Italian and French coastal areas.

The Woodworking Process

A good quality pipe is obtained by processing the briar wood (the "stick") which ages in the stockroom for nine years from the date when it is picked. Briarwood is the most appropriate material because it conducts heat very well. The wood of the internal walls burns, while the external walls stay fresh. Briar wood is also a very variable material: only at the end of its processing can one tell whether the finished product is worthy or not, as the owner of Castello Pipe comments: "When I start producing fifty pieces, I know how much manual labor is necessary; however, I do not know what I will get out of it, whether they will be first or second quality pipes."

The pipe's mouthpiece is made of plexiglass, which is a material that is exclusively guaranteed and is a result of a secret formula created by an established supplier.

The pipe production process consists of several phases: sketching and shaping of the bowl, hollowing and creation of the smoke channel, working the mouthpiece, polishing it, etc.). These steps are easily definable and can be carried out by individual workers (each worker may build a whole pipe or all the workers may specialize in one step) or by teams (all workers jointly build the pipe and are accountable for the production of the pipe).

The Organization of Labor

In the owner's opinion: "Each artisan feels a little bit like an artist and wants to work on his own. It is very important to respect this individuality, hence each worker is responsible for one detail, not the entire piece, because the Castello pipe is purposely the result of team work. The seven artisans who work for me need to be able to carry out each step of the process: they are perfectly interchangeable and I rotate them often. This allows them to check on each other's work. The team ensures the best quality: fourteen eyes see better than two. If each worker works on one detail, the pipe belongs to Castello and not to Joe or Tom; this way no jealousies are created. When I sell an exceptional piece, I show it to all the workers for the last time. They all know which pieces are the best, so they remember those and keep them in their minds as examples, and so they work well together.

In truth, it is the group that controls the entire production process of the pipe. Each morning the group decides who is doing what. In addition, it is always the group who regulates the behavior of its members. In the mornings, there are workers who come in at 8.00 and others who come in at 8.15. At noon some go home for lunch and others stay in the factory. During the day, if someone is bored with his job, another agrees to do it. As the owner states: "Everybody knows each other for a lifetime (the youngest artisan has fifteen years seniority) and one look in the morning is enough to figure out who is in good shape and who is not. They all are asked to give their best every day and I think this is the best system to meet my needs." The group even decides on new employees to hire. The owner proposes the potential new members and the group, after an apprenticeship period, chooses and decides the new hires. Lastly, it is the group that controls the quantity and the quality of its own work, as the owner confirms. "I cannot interfere and control the work of my artisans. I can only check the final results, which is very subjective. The pipe must be weighed in the hand, touched, felt in the mouth, to see what emotion it sparks in you."

About five thousand pipes are produced annually, of which there are not more than thirty of the highest quality. The production of a single pipe is difficult to plan because it depends on the condition of the raw material – the stick.

The wage for each artisan is personal, and is paid in cash every fifteen days. The wage level is much higher (almost double) the average of the artisans in the area. The cost of labor counts for 85 percent of the sales price. There is no formal checking on the work hours; individual statements are enough, and those who want, work also on Saturday. The artisans are not unionized. As the owner summarises: "We are a family. The pay, which I decide, is varied, because there are real differences among my artisans and they know it. They are masters over whom one cannot lord it, but only reason. I have to know how to stimulate them, to work on their own love for their craft, because from them I expect new shapes and lines for the pipe."

The customer of Castello pipes, after the first purchase, becomes a pipe collector and is then certain to buy more pipes throughout his life.

Source: By Luigi Golzio.

to hierarchy" in the organization of work, including "*autonomous work groups*" and "*matrix groups*". An autonomous work group should be capable of regulating all relevant variances and interdependences internally. Therefore, it is effective if the external interdependences with other groups are not too high. Classical examples of this are "islands" of production or assembly (see Box 10.4a). The group is typically autonomous also in the flexible assignment of tasks to its members; ideally also in the selection of members. A more stable and specialized division of labor among the group members, by technical qualification and by type of output they contribute to, would configure a matrix structure at a micro-level, as is often used in the internal organization of the R&D function.

Finally, a fully networked form of organization of work can be even less pre-defined and more ad hoc. Work groups can be responsible both for assigning tasks internally *and* also endowed with the right to regulate external interdependences with ad hoc decisions (such as some experiences inspired by the *just-in-time* approach, as described in Box 10.4b).

Collective forms

The team-based forms of work organization considered hitherto are only so in a partial way. A more radical alternative is a group of associated workers entitled not only to rights of action and decision but also to rights of ownership of complementary resources. In collective forms of work organization, the workers share property rights over resources, including technical, commercial, and financial resources. Legal forms under which *peer groups* are actually constituted and play an important role in market economies include *cooperatives* and *partnerships*. They are diffused in sectors ranging from semi-artisan to professional work; and even in large and globalized firms – like professional partnerships with thousands of associates all over the world (Greenwood *et al.* 1990; Pfeffer 1994).

As with any other group of forms, a certain configuration of ownership rights is compatible with very diverse configurations of division of work and coordination. The case that follows can be used to reconstruct a discrete range of these, and also to introduce the discussion of certain *raisons d'être* of collective forms (see Box 10.5).

If the case of the institution and design of the cooperative structure is conducted like role-playing, various groups reach solutions that in different ways combine modalities of division of work, allocation of decision and control rights, and forms of reward. At least two salient alternative configurations can be compared.

Peer groups

In the extreme case, the peer group members are equally entitled to all the decision, control, and property rights and obligations associated with a certain action (Marglin 1974; Williamson 1980). Internal democracy can be representative rather than direct, in large size groups. Administrative responsibilities are

Box 10.4a
Semi-autonomous working groups: the Nobo Fabrikker AS case

In the 1960s and early 1970s, a team of researchers led by Prof. Einar Thorsrud of the Industrial Sociology Institute of Trondheim, Norway, completed a research project on "new organizational labor" experiments that many local enterprises were carrying out.

The goal of the research was to:

- get a picture of the top (in terms of the hierarchy) level workers and the various labor organizations they belonged to;
- carry out research on the possibility of the evolution of workers' participation in the job.

Among the various cases studied, some more successful than others, the best organization innovation experiment was that of Nobo Fabrikker AS. The main objective of this experiment was to increase opportunity for participation in the job among the workers and at the same time to empower those unutilized resources within the existing organizational structure. In doing this, there was an attempt to create a gradual transition from an organization based on principles of fragmentation of work to one based on autonomous groups.

The experiment took place in 1965 in the electric heater division of the headquarters of Nobo Fabrikker AS, located near Trondheim. Before the experiment, each worker was assigned a specific set of tasks with a high degree of redundancy.

In order to complete the experiment the workers involved had to be trained to be able to carry out at least three other tasks in addition to their own. All production was structured in five groups, each developing a significant portion of the task. Thanks to the close collaboration among groups (transfers of workers among groups was quite frequent and occurred based on the respective work loads), the workers managed to learn about the entire production chain.

Within the groups, each member was not assigned a fixed task, rather, all of them helped with the common task based on criteria of need and their own personal preferences for certain assignments. As a result, after about three years, half of the workers were able to carry out all tasks related to a certain division. Each year, each group elected the person who would coordinate relationships with the other groups and with the local management. The five contact people, who worked full-time just like the other workers, represented, together with local management, a committee in charge of overseeing a number of issues: production plans, budget, organizational and technical development, suggestions, etc. The main responsibility of the top management was to oversee boundaries between each group – that is, maintaining relationships with the top management itself as well as the other staff, and supplying materials as well as technical equipment and information.

The workers agreed that the new organizational model and the new technical and social working conditions were a definite improvement over the previous ones; they displayed a high degree of involvement in their job, as was shown also by the increased number of suggestions that were made by the groups.

The electric heater division had a minimum turnover thanks to the excellent relationships established between the workers and the local management. In terms of industrial democracy, the most interesting aspect of the experiment was that the workers were very eager to carry out the project.

The experiment has shown a dramatic change from individualistic behavior to a group oriented attitude; this change can be explained by the improved knowledge of the work and the working structure. In addition, while earlier workers were involved in the planning of labor from day to day, now they actually participated in a three-month planning horizon.

Thanks to all this the workers reached a level of competence that allowed them to be totally confident of handling any problem internal to their unit. On its side, the management stopped checking up on the individual workers and its relationships with the work groups started to be regulated by the payment system, by sales schedules, and by production techniques. Even quality control became a responsibility of the workers, that led to a reduction in customer complaints.

Source: Martino (1982).

Box 10.4b
Networked jobs: the Toyota kanban case

The kanban system represents an example of production organization in which the central office coordinating the transfers of materials among divisions is lacking; such an office is however a typical structure in western countries.

Kanban, which in Japanese means "wooden sign" (those typically put outside a store), identifies a form. At the end of the daily production, the divisions in the final assembly phase communicate their production orders – or kanban – for each kind of component or semi-finished product they need (engines, transmissions, headlights, etc.) to an office located within the warehouse. The kanbans report the kind and quantity of the goods picked up, as well as the schedule for future re-supply. The divisions that supply these components pick up the kanbans based on regular schedules at the warehouse office – a few times per day.Therefore, the kanbans work as order forms and are returned to the warehouse together with the actual delivery of the products ordered. In short, kanbans have two functions: order forms and delivery slips. The division that receives the kanban during the final phase of assembly communicates its orders – or kanbans – to the divisions directly above it; thanks to such circulation of kanbans, the relationship of orders to deliveries among divisions working closely in the production process is extended to external suppliers who have long-term contracts with the firm.

Such a system assumes that the divisions at the beginning of the cycle match their production based on the demand coming from the divisions below them which are written up in the kanbans. In such a circumstance, the single divisions do not have to strictly follow the orders given by the central offices which can easily be overridden

owing to issues arising between the planning and the revision process; rather, they can make necessary adjustments in a timely and coordinated fashion thanks to the use of kanbans.

A second and important prerequisite for meeting high levels of productivity (possibly linked to the use of the kanbans) is the development of flexible and multi-functional workers' skills; in fact, workers are assigned to various phases of the process based on work loads. The result is the reduction to a minimum of dead time for each worker.

A third important aspect is quality control: this occurs along the whole cycle and is not limited to the end of it. The divisions at the end of the cycle have the choice of refusing the semi-products that are produced by the divisions above them.

Finally, the direct link with the order issuing phase enables the company to substantially reduce investments in stock, justifying a "zero-stock" system. Such a structure resolves the hierarchical problem as well, given that production levels and product mix adjustments occur without any interference by the central offices within the factory.

In addition, thanks to team work, the interdependence of the phases and the typically Japanese culture, the kanban system enables potential opportunistic behavior typical of any non-hierarchical structure to be reduced.

Source: Aoki (1988).

allocated by election (and occasionally by rotation).

Specifying the conditions for the effectiveness and efficiency of these forms is not simple, to the extent that the debate on these has been strongly polarized by positions with an ideological or universalistic character in favor or against (especially between radical and neo-institutionalist economists). On the basis of the available empirical and theoretical research, it is, however, possible to outline the following variants of collective forms and the following evaluative considerations.

Employee-owned firms persist and are efficient in both productive and motivational respects in activities where human and social capital is the critical input, from the cooperatives of masons and construction workers to those of professionals and consultants (Zan 1992). In the case where the work is specific and not easily monitorable, there are even stronger reasons to align the interests of participants through the diffusion of property rights (Fama and Jensen 1983a, b; Chapter 9).

Federative groups

A milder version of collective organization is that of a group of co-workers who collectively possess some or all of the main means of production, but who hold separately the right to the rewards deriving from their own work (Demsetz 1967). This can be defined as a *federative group* to evoke the community of resources or property but the retention of the rights to residual rewards and self-regulation. In this case the activities must be separable, even if there is participation in the use of common resources (e.g. the commercial contacts, the plant, a brand name).

For instance, a training firm whose partners own the brand name and the offices, but where each selects his/her own collaborators, uses a partly personal network of contacts, draws on partly self-owned, not communal

Box 10.5
The St Francis Cooperative

Antefact

Twenty women living in a small community in southern Italy have decided to put an end to their permanently unemployed status and start a business. The motivations leading these women to take this initiative are as heterogeneous as their personal characteristics, educational backgrounds, and political ideologies. To some of them, work is a livelihood, due to the unemployed status of their husbands or fathers; to others work means dignity and autonomy as opposed to a situation of dependency on other relatives; to still others work enables them to build on their own emancipation in a social and economic context that is still very backward; for all of them, work is a dream come true. The business idea involves offering (at a very low price) to perform all the finishing work on knitwear produced by large firms in northern Italy.

Doing finishing work on knitwear is a very expensive process because it is considered high intensity work and a high quality clothing item is made of good quality collars, cuffs, etc. (the so called finishings). The worker-entrepreneurs know that the lower prices which northern Italian firms demand in order to offset the transportation costs will limit their personal earnings (to the point of being lower than the salary of a hired textile worker). However, as they like to say "it is better than nothing." There are two problems facing these women today: first, it is necessary to find the best arrangement for the management of the new company; second, it is necessary to define an organizational model and a compensation system. There are several questions that need answering. How do we allocate property rights? How do we distribute decisional rights? Who will decide whether or not to accept a contract? On what basis will employees be compensated?

The cost of the equipment, and the start-up costs of the firm have been divided equally among the twenty women, and there is an agreement that property rights should also be distributed equally. The new company is thus set up as a cooperative in which each "partner-worker" has the right to vote on company decisions. The first conflict among the twenty partner-workers arose when it was time to discuss for which decisions it would be necessary to opt for a democratic vote. Some partners proposed the creation of a few new positions (a director, a production manager and a customer service manager) who would be responsible for operative decisions and who would oversee contracts that required a working commitment of over 10 days. Others maintained that given the small size of the company, decisions needed to be made within some kind of a permanent "Agora." The second issue that led to a conflict is more serious and deals with the compensation policy. Again, there are two "groups" having conflicting opinions.

The first group proposes a form of compensation for individual labor based on the average number of knitwear items produced in a month. This hypothesis also implies that each worker would have to perform all necessary production steps in order to complete one clothing item. Unanimously, at the beginning of each week, the workers

would decide how to "rotate" themselves within the various tasks in order to ensure fairness. The second group instead believes that it would be better to specialize work among groups with homogeneous tasks, thus creating a sort of assembly line. The partners-workers would therefore be compensated on the basis of the work performed individually (here again by counting the items produced).

Epilogue

The St Francis Cooperative – the name was chosen for the "women's firm" to remember St Francis who would protect them along the road – became a shining example for many other initiatives in the area. In 10 years, twenty cooperatives have been created, supporting unemployed women. The St Francis Co-op decided to adopt a collective method for decision-making: all decisions took place through a voting process and were made based on a majority vote. No Director was nominated. The organizational model chosen was the assembly line one, characterized by high workers' specialization within the various tasks. In order to ensure continuous work, the co-op invested in keeping stock available at all times. However, this created some problems when the co-op accepted small contracts or when the customer requested a quick turnaround. Compensation was calculated based on the hours worked and not the number of items produced. Quality control was a group assignment given the close working relationships that the women had with each other. A sum of money necessary to pay ongoing expenses was deducted from each pay check.

This solution worked very well for the first few years. Quality control, the hourly based compensation, and the choice of accepting even barely profitable contracts permitted the co-op to work full-time for a few years. However, once the initial motivation died, conflicts arose concerning the company strategies. Some partners wanted to expand the range of activities which would, in addition to doing finishing work, also include higher value activities that would increase profitability. Another group wanted an increase in the hours of work. The system of compensation based on hours worked led to the co-op splitting into two factions: those who wanted and could afford to work long hours and those who by choice or need could not be away from home for more than 6–8 hours per day. This, together with the initial and irreversible choice of opting for specialization, led to conflicts among the partners-workers and to delays in the delivery of contracts.

The cooperative attitude that had made this dream come true disappeared. Besides more conflicts, there were boycotts, sabotage, and factions were created with little "bosses." The St Francis Co-op ceased to exist on December 31, 1997 without being able to celebrate its tenth birthday. The equipment, bought by an entrepreneur from the Marche region, is today being used in a small firm in Albania.

None of the partners became rich but all of them are convinced that it was worth the experience.

Source: By Giuseppe Soda.

knowledge, and has the right to the economic results of the projects performed, after allotting a percentage to the common structure, would be an example of federative group.

In the real case of the St Francis Cooperative, a solution was adopted having some features of the peer group, others of the federative group, and yet others of a classic Taylorist-bureaucratic model. The outcome of this was not positive. The solution of work organization adopted, beyond distribution of ownership rights, envisaged an impracticable combination of highly divided labor and the absence of coordination mechanisms, burdened by incentives to maximize the number of hours worked rather than the quantity and quality of the outputs.

Entrepreneurial forms

A possible alternative for organizing the work of finishing garments could be an entrepreneurial form instead of a collective one. For example, the firm producing the garments could hand over the finishing stages to independent outside workers. If there exists a sector where these forms have always been widespread, it is indeed the textile-clothing sector.

Entrepreneurial forms can be conceived as mixed arrangements, where actors owing and contributing technical and financial capital also contribute work, and actors who contribute work are also owners and contributors of human and technical capital (competence and task-specific equipments).

Among entrepreneurial arrangements, one more "external" and one more "internal" form can be identified as particularly relevant.

Putting out

This has been analyzed as an archaic and outdated form of organisation by some scholars (Williamson 1980). In their view, the advantages of flexibility are overridden by the costs of free-riding, negligence, waste, and delay. Ironically, the retention and revitalization of putting out in the textile sector has been a factor in the success story of Italian fashion worldwide (Box 10.6). The advan-

Box 10.6
Putting out in the fashion industry: the Benetton mills

The 80,000 clothing items that Benetton sells can be categorized in three groups: wool, cotton, and outfits and Capi Spalla. Their production occurs partially in Benetton's own factories, partially at subcontracting factories, but also at a number of independent factories: the mills. Over the years, the number of mills has increased based on the growth of Benetton itself. Today, the mills number about 500, if we exclude those directly managed by firms owned by Benetton. Their dimensions range from very small to medium, based on the production phases that they specialize in, and the technology they employ. Generally, each mill carries out only one phase of the production process, realizing economies of scale in that single phase. Within the entire transformation process (which for wool starts with the production of the fiber itself) the mills fit at the beginning, in the phase that starts from the thread or the material, transforming it into the final product. The mills concentrate their activity in tasks that demand high intensity labor such as cutting, outfit-making, ironing, and embroidery.

From a juridical-formal point of view, each mill is autonomous and independent. Today, as in the past, there are no written contracts between the mother factory and the mills. Relationships are managed through some tacit and informal rules that are the result of selection, inclusion, and dismissal procedures that the parties have developed over the years and that allow the mother factory to directly monitor the single mills without any binding contracts. The rules pertaining to the management of the relationships are very simple. Benetton acts as central agent and controls the entire production chain; it sets quality standards, the ratio between prices/costs, and the quantity to be produced. The mills are not asked for any planning commitment; the issue concerning production comes down to respect for the standards and the delivery schedules. Since production is fragmented among so many factories, it is critical that the mother factory be in charge of quality control, keeping statistics for each of the middle products; to this end, each lot is brought to Benetton before then being sent on to the mill that will perform the next working phase. Mills work exclusively for Benetton (except those specialized in embroidery and printing) who usually keeps them fully operational all year round. Together the mills can be considered as a relatively stable group of factories even though there is some rotation among them, which is encouraged by Benetton itself. Each mill is free to end the relationship at any time. The same can be said about the mother factory even though the stability of the system gives advantages to both of them..

The mother factory does not guarantee the mills a constant level of work; however the now commonly used practice of accepting more contracts then one can complete, has allowed the development of confidence about the reliability of the relation. In fact, what characterizes this group of factories is the strong and tight collaboration among them, around which also rotates the flexibility that the fashion industry demands nowadays. Collaboration however also brings competition among the mills to be "the best" – think for example at the "Benettoniadi" – true battles between different mill teams. The overall positive feeling that the success of each mill is based on the overall system facilitates collaboration and a sense of belonging, thus creating improvements to efficiency.

Source: S. Bagdadli, F. Buttignon, and D. Montemerlo, "The Benetton Case," in Airoldi and Ruffini (1993).

tages of specialization and flexibility in productive combinations as a function of demand, as against integrated capitalistic forms (Sabel and Piore 1984) have been concretized thanks mainly to coordination through very clear sectoral/local rules and routines, though not very formal ones (Brusco 1999) – which were probably absent from the putting-out systems of the nineteenth century (Kieser 1993).

Inside contracting

Instead of a commissioning firm entrusting production of proprietary materials (and, if necessary, with proprietary equipment) to outside entrepreneurs, one can envisage a situation where a firm lacking the technical equipment and knowledge specific for a given activity may make use of "inside entrepreneurs" able to perform the production –

by supplying complementary resources such as finance capital, buildings, or structures and competences for marketing and distribution. Thus, not only decision and control rights but also ownership is diffused and mixed, rather than entirely held by any single party. One party, the employer, owns some tools and resources such as the buildings, central facilities, sales outlets, and commercial brands. Despite this, the transformation and the generation of products or services are carried out by actors who own complementary assets, such as competence, technical instruments, software, and relational capital. In addition, they may hire and manage their own co-workers and are compensated on results according to an agreement negotiated in advance. Even though this system has been diffused in agriculture in an economically distant past – for example, in the farms of the beginning of the century, the milling of grain or milking of cows were entrusted to specialist entrepreneurs who went from one farm to another – it is not difficult to find modern and highly efficient examples, when economies of specialization and the differentiation of knowledge for conducting the different activities are coupled with site specificities and other sources of interdependence.

Box 10.7 describes the case of a professional service firm, but inside contracting has been resurrected also in industrial production, for regulating the outsourcing of material handling, maintenance, and other activities that are carried out "under the roof" of the externalizing production firm by independent entrepreneurs.

SUMMARY

The theme of the organization of work was introduced by recalling some elements of the history of the interventions and projects in this "contested terrain."

A basic model of analysis and design was displayed in the second section, including and combining both socio-technical and economic model variables. In particular, economic models have been useful for reintroducing concerns and concepts relevant for appraising the importance of *production costs and learning economies*, and of the *specificity and criticality of human resources*, among the antecedents of work organization forms – in addition to the much debated impact of *variance and interdependence* – and of *property rights* among their dimensions – in addition to *action, decision, and control rights*.

In the last section, a range of discrete forms of work organization – characterized by different ownership structures, different decision and control structures, and different mixes of coordination mechanisms – have been described and assessed. The relative superiority of "capitalist," "collective," and "entrepreneurial" proprietary structures is connected mainly to the relative criticality of technical and financial resources, or instead of human resources. Among capitalistic forms, the specialized and programed "Taylorist-like" organization – especially effective in divisible and sophisticated tasks – has been joined by "enriched" models (with more "generalist" and discretionary jobs, better suited to adapt to varying demands and to solve new problems), network models (providing flexible combinations of specialized competences and activities as conditions change). When human capital is critical (add more value, is more exposed to risk, is poorly transferable) "group-like" arrangements are effective, in the federative group form – if contributions are discernible – or in the full peer group form – in "team production" and

Box 10.7
Inside contracting in a business school: the Taj case

The Taj Business School was founded in the early 1980s by Peng Lee, a leading man in the electronics industry who believed that a top training school was vital for the economy of any country. The school owned a beautiful post-modern building facing a small artificial lake with ducks and lotus flowers and was equipped with the latest technology. Facilities included two auditoriums and 50 classrooms, each equipped with multiple sliding blackboards, video-beam, video camera, and personal computer, an excellent snack bar-restaurant and a gym with a heated swimming pool. All this infrastructure represented the "hardware" of the school. What the school lacked was faculty members and researchers. They were the actual "software" of the school, the people who held the scientific and technical knowledge to produce and deliver the service in question: training.

The division of labor reflected (or at least should have reflected) the specialized competences of individual faculty members. Each course was planned and coordinated by an internal faculty member (usually an expert on the course topic) who would choose to staff the course with assistants who were the most knowledgeable on the course material and were therefore a good match; for example, an executive training course required more experienced faculty whereas a basic course could be taught by junior members of the staff. When a faculty member was thought to be an expert on a topic, there was a true competition for his/her expertise, with the consequence that some faculty members were overloaded and others much less so. The products that Taj offered on the market were courses, research studies, and other projects that were planned with large (though not total) autonomy by the various faculty members who would "hire" the best resources for the job.

Shared norms and mechanisms facilitated the coordination of courses and faculty as well as standardization of the product. Faculty members were compensated based on specific categories. Teaching had a different rate of pay than other activities such as the preparation of course material. A fixed portion of the course was the responsibility of the coordinator, who was thus paid based on his/her planning and management activity. Compensation based on the success of a certain product could also be awarded. The remaining part of the profit was awarded to the "department" or "client group" of which the coordinator or the course were part. The school faculty was in fact grouped on the basis of professional families or sometimes client groups at whom the courses were aimed. Each similarly defined "area" consisted of a series of common structures that everybody had access to and its objective was to standardize knowledge though the development of material and specific know-how later made available to other faculty. The margins (profit) obtained from the different courses were used to pay for general and shared costs (advertisement of courses, research, faculty training) and anything left over was added to the central budget. In fact, in order to standardize ways of delivering services, as well as to safeguard the "Taj, training for the future" trademark, some central services in charge of marketing and communication were created. With the same objective, logistics, copying, and scientific research offices were established. These structures would guarantee, more or less

in general terms, equity among the clients, economy of scale within the services for the delivery of activities, and integration with the market.

The image of matrioska fits Taj well, even though the external shell started to be a little heavy and, with the increase of training, the client integration was not always total.

Source: By Silvia Bagdadli.

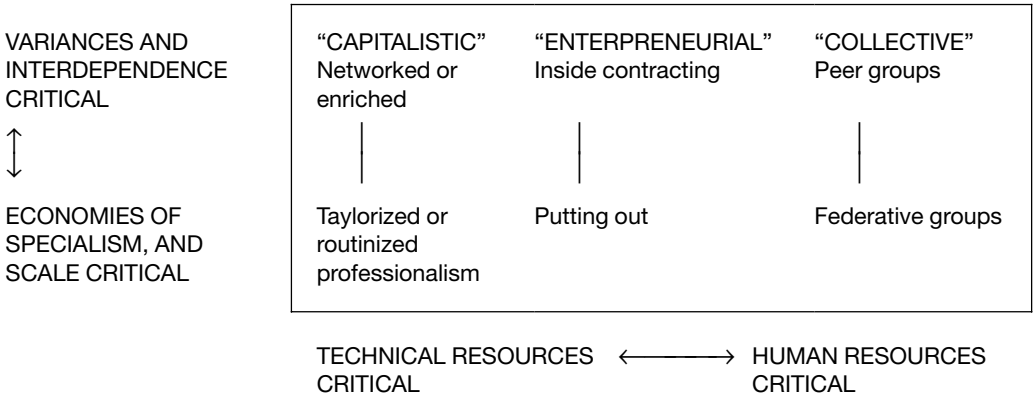


Figure 10.4 Comparative evaluation of work organization forms

"intensive interdependence" situations. In entrepreneurial arrangements, the same actor is the assets owner and provides critical work services. In the putting-out system, external entrepreneurs process materials which are owned and delivered ("put out") by another firm, thereby realizing phase-specific economies of specialization and scale. In the inside contracting system, entrepreneurs are "internal," in the sense they use common resources and structures owned by a firm, but they produce effective specialized services too thanks to their proprietary assets (knowledge, electronic equipment and programs, tools and instruments).

In very synthetic terms, some of the main variations between the considered forms can therefore be captured by a two-dimensional space of human resource criticality and task interdependence, as indicated in Figure 10.4.

Exercise: Westa I

The story

The Westa publishing company was set up in 1989 in St Petersburg. It consisted of three people: the director and two young women, Natasha and Marina. The only equipment was a computer and a laser printer. Westa's activity involved typing new books on file and printing one copy.

The division of labor was very simple. The director was responsible for seeking new clients; Marina typed the texts on computer and Natasha dealt with the editing and printing. They were paid for each printed page of the original copy. Payment was documented by part-time accountants.

As from 1980, the demand for this kind of service began to grow very quickly. Many daily newspapers and periodicals were springing up, and there was a large, constant request for thrillers, love stories, and Russian literature. New firms also wished to print advertising prospectuses and headed letter paper.

There were plenty of firms like Westa in St Petersburg, but there was work for all of them.

Early in 1990, it became plain that one computer and two operators were no longer sufficient. Marina hired three part-time typists; Natasha could no longer deal with all the work – editing the files, printing them, correcting the errors, and then printing them again; so she took on a part-time girl to correct the errors.

At the end of 1990, Westa rented three rooms, five computers and two laser printers. Marina headed the group of six typists; Natasha was responsible for editing and printing the copy; two girls worked with her; an accountant was hired. The director was responsible for the marketing and administration; and the group now included an engineer, who provided technical backup.

In 1993 there were two groups of operators headed by Marina and Natasha.

Marina's group dealt only with typing the texts on computer. There were three computers, with three girls working in the morning and three in the afternoon of each day. This group produced the files, indicating only the chapters, with no form of editing.

Natasha's group saw to editing the files, using appropriate software, printing them, correcting errors, printing them again and producing the text. They used a further three computers and two laser printers.

At a certain point, the group's productivity began to decline. Clients found many errors in the finished products and were dissatisfied. "We should like the files transcribed without errors, and we should like to have them by the deadlines fixed. If you cannot do a good job, we shall find another firm," said an angry client to the director on receiving his order two days over the deadline and with various errors on the first page.

The decision to reorganize the work

"What can we do? Who is responsible for all these errors? Why can't we ever deliver on time?" the director asked Marina and Natasha.

"The errors are typos, nothing to do with us," replied Natasha.

"Your job consists in finding and correcting them, and you're responsible for the time lost," retorted Marina.

"That will do, thanks. We used to be friends. Let's try and make some changes. We are losing customers and lots of money; if you don't want to return to poverty, put your heads together and think up something," said the director.

By Silvia Bagdadli

Question

- How would you reorganize the work in order to recover productivity and increase the workers' perceived benefits?