

THE ROLE OF CRITICISM IN THE DYNAMICS OF PERFORMANCE EVALUATION SYSTEMS

Annick Bourguignon
Associate Professor

Eve Chiapello
Associate Professor

ESSEC Business School
BP. 105
95021 CERGY-PONTOISE Cedex
France

HEC School of Management
78350 JOUY-EN-JOSAS
France

bourguignon@essec.fr
tel. 33 1 34 43 30 12
fax. 33 1 34 43 28 11

chiapello@hec.fr
tel. 33 1 39 67 94 41
fax. 33 1 39 67 70 86

Working paper
Please do not quote without the permission of the authors

Acknowledgements

We gratefully acknowledge helpful comments of David Dugdale, Michel Gordin, Alan Jenkins, Carlos Ramirez, Arline Savage and participants of the 2002 Critical Perspectives on Accounting Conference, New York, where a first draft of this text was presented.

THE ROLE OF CRITICISM IN THE DYNAMICS OF PERFORMANCE EVALUATION SYSTEMS

Abstract

Drawing on the concept of “trial”, developed by French sociologists, this article analyzes the dynamics of employees’ performance evaluation systems, particularly those involving accounting performance measures. A case study is presented as an illustration of our proposal to consider these systems as one of the major trials in the business world, that is, social arrangements organizing the testing of people and resulting in ordering them, and further in consistent social goods allocation. This analysis emphasizes the role of criticism in the dynamics and evolution of performance evaluation systems and enables us to revisit concepts like controllability or objectivity which have been presented for decades as cornerstones of performance evaluation systems either in management control or in human resource management fields.

Key words

Criticism, Performance Evaluation Systems, Fairness, Objectivity, Controllability, Legitimacy, Bonus



She says: "Hey, can you tell me why the new boy you've just hired earns four times my salary?"

He replies: "Gosh, I'm surprised you earn that much!"¹

The female employee shown in this scene displays a strong feeling of unfairness, which leads her to protest and demand an explanation. She is calling into question the underlying criteria of justice that determine her remuneration and the remuneration of the young man. To what does her criticism relate? Does she consider herself to be a victim of gender discrimination? Is she deploring the fact that young new entrants earn more than people already working for the company because they benefit from a favorable labor market? Is she suggesting that the young man's salary is not justified by his skills? We can see that an argument about justice is about to break out. If this argument is continued, the complainant will doubtless try to show that the criteria of justice that must be used to determine remuneration (competence, for example, or seniority, depending on the type of organization) are not being observed, since some people earn more than others who are just as (or more) competent (or senior). She may

point out that the winners benefit from what we are going to call “illegitimate strengths” (their gender, their age, the state of the labor market at the time when they are hired, etc.). “Remuneration is determined on the basis of strength relationships that are profoundly unfair”, the complainant will say, and she will go on to call for the reestablishment of fair remuneration that excludes all “illegitimate strengths”, leaving only “legitimate strengths” (competence, seniority, etc.) as bases of remuneration.

But during the initial skirmish that we can see here, the criticism of our complainant is quite simply ignored. Her boss even replies in a provocative way by actually pointing out that he is not just aware of the salary differential giving rise to the complaint, but that furthermore he thought that the differential was even greater than it actually is. In other words, he does not see why a significant salary differential should give rise to a complaint. It is possible to imagine without difficulty that this reply will significantly increase the discontent of the complainant (and to wonder whether this is what the boss wishes to achieve...).

However, the criticism expressed may have risks for the company. The discontented employee may disengage herself from her work and create around her a climate not conducive to the interests of the company. It is also possible to imagine that the company trade unions would seize on this “flagrant injustice” and turn it into a matter for industrial dispute. The employee might also resign. Furthermore, if this feeling of injustice is not an isolated case, the criticism may be magnified, increasing its negative consequences (demotivation, industrial conflict, exodus of skilled staff, etc.). It is also possible to imagine that the boss will change his attitude and will cease to be “deaf” and undertake a revision of the criteria for evaluation and remuneration currently in use.

The feeling of injustice is well known in management research. Since the founding studies of Adams (1963), a prolific research trend has studied many aspects of the question of “organizational justice” (see Greenberg [1987, 1990] for a review). These studies highlighted the following points: factors leading to the perception of justice (or injustice) (Landy *et al.*, 1978, Dulebohn, Ferris, 1999), the consequences of this perception on various attitudes (for example, different levels of satisfaction) (Tremblay *et al.*, 2000) and the individual or organizational consequences of the feeling or attitude induced, such as employee’s morale and performance (Adams, Jacobsen, 1964), intended or effective absenteeism, turnover, and various aspects of behavior (Harder, 1992; Carraher *et al.*; 1992, Covin *et al.*, 1993). All these consequences influence more or less directly companies’ overall performance. In other words, research into organizational justice generally falls into an instrumental perspective: the feeling of (in)justice is beneficial (detrimental) to the company’s efficiency, i.e. the realization of its objectives.

This instrumental perspective perhaps explains why there is an aspect of organizational justice that has been little studied, namely the expression of protest and critics and its consequences (which are partly exemplified in the scene above). However, this aspect is central if a sociological, rather than managerialist, perspective is considered. A feeling of fairness is essential for companies in that it legitimizes and therefore ensures the continuation of the social system embodied. Max Weber (1922) emphasized that no system of domination, even of the most brutal kind, governs solely by force and that our consent and cooperation are always needed. However consent and cooperation are not achievable unless domination is legitimate. As the Weberian typology of forms of domination shows, an order is only legitimate if it is included in the system of constraints that is inherent in each form of legitimacy. For example, in a rational-legal system of domination, an order is legitimate if it

complies with the regulations and the law, and if it is issued by a carefully selected competent authority. In some way this conception of legitimacy assumes that the dominated evaluate the order given and declare it to be in keeping with its claims to legitimacy. Legitimate domination operates normally within a general context of consent, without conflicts and/or refusal to obey. However, a feeling of injustice gives rise to criticism, which calls into question the social system and places at risk the legitimacy of the domination. The latter will only survive if it is able to restore its legitimacy.

In this article, we focus on criticism as a major aspect of organizational justice questions. This approach is interesting for four reasons. In the first place, it explicitly recognizes that the expression of a protest is a consequence of a feeling of injustice, as were the different levels of satisfaction previously researched. This recognition then enables diverse forms of wage-earners' critical activity to be recorded, an activity which tends to be disregarded in the majority of management approaches. Indeed, still within the classic instrumental approach, it is commonly admitted that a dissatisfied worker is a worker with low motivation (see for instance Hertzberg, 1968), leaving it to managers to regulate the organizational parameters so that motivation remains at a sufficiently high level. This quasi-organic representation of workers attitudes is evidently reassuring, and the unappreciated (because unrecognized) character of their critical activity erases any conflictuality. The avoidance of conflict contributes in turn to the reduction of critical activity insofar as employees identify with the uncritical image of themselves reflected by the organization.

The third point is that a focus on criticism enables the study of the consequences of critical activity, in other words how those confronted react and the answers provided. Finally, the theoretical model we use to emphasize the role of criticism suggests that one of the answers to this criticism consists of modifying the "objects" criticized. The dialectics of protestation (workers) and of resistance (management) leads to compromises that result in change in management practices. Criticism thus appears to be an agent for change within companies and our analysis also proposes a grid analysing change in management techniques, which could usefully complete existing frameworks for organizational change.

In this article, we provide an illustration of the role and dynamics of criticism in organizations in relation to performance evaluation systems. These management devices are likely to arouse criticism, because in most companies (i) they are closely connected with rewards systems and people are highly sensitive to "distributive justice", that is the "fairness of the distribution of outcomes", (ii) they include various rules (ie. bonus and performance measures formulae) and people are equally attentive to "procedural justice", that is "how various outcomes (...) are determined" (Greenberg, 1987, p. 10).

This analysis thus contributes to research carried out in the accounting field which has already investigated questions related to organizational justice. Although fairness has long been claimed to be an important dimension for accounting research (Williams, 1987), until now relatively little research has focused on the question. The perception of fairness has been studied in relation to transfer pricing policies (Eccles, 1985; Ghosh, 2000), to budget practices (Cropanzano, Folger, 1991; Lindquist, 1995; Libby, 1999; Wentzel, 2002; Fisher *et al.*, 2002) and to evaluative style (Hopwood, 1972, Otley, 1978). However, as was the case for more general research on organizational justice, attention has been exclusively devoted to factors influencing perceived fairness (Ghosh, 2000; Libby, 1999; Hopwood, 1972, Otley, 1978; Fisher *et al.*, 2002) or its consequences (Cropanzano, Folger, 1991; Lindquist, 1995; Wentzel,

2002). We suggest that, as previously, this disregard of criticism can be interpreted as a consequence of the managerialist orientation of most accounting research.

In addition, our analysis points to “old” concepts, which have been presented for decades as cornerstones of performance evaluation either in management control (the controllability principle) or in human resource management fields (the objectivity requirement). Our analysis makes it possible to revisit these concepts and suggest considering some complementary aspects of these notions.

The analysis draws on the concept of “trial”, developed by French sociologists over the past twenty years. More specifically we use the conceptual synthesis developed by Boltanski & Chiapello (1999), which integrates elements borrowed from Latour (1984) (trials are strength relationships) and more irenic elements developed by Boltanski & Thévenot (1991). Boltanski & Chiapello define trials as social arrangements organizing the testing of people, and resulting in ordering and, further, allocating social goods to them. These authors also highlight the role of criticism in the dynamics of trials.

This article offers an application of this very general sociological framework to performance evaluation systems in organizations. The role of criticism in the dynamics of such systems is illustrated by the turbulent ten-year story of the bonus system of a French company, and we provide a theoretical discussion of the multiple fairness problems associated with various steps of the performance evaluation process.

The “story” has to be considered as an empirical illustration rather than a case study in the usual sense of the term. Indeed it suffers some methodological shortcomings which do not provide the usual scientific warranties of case studies. We did not directly observe the events traced in the story, which took place between 1988 and 1998. Furthermore the narration provided here relies on interviews carried out between November 2000 and March 2001. We are thus fully aware that the material available is an “*a posteriori*” reconstruction of events, which are subsequently likely to be “biased”, either by the selective memory of our informers or more simply by their lack of a comprehensive view of the question and problems involved. For instance, our story is incomplete on a number of points, which would be relevant to an extensive analysis of criticisms (for example, the role of trade unions, or turnover statistics). However, the available observations are highly illustrative of the role of criticism in the dynamics of performance evaluation systems, and we think that the case offers more on this point than a speculative general discussion on the dynamics of criticism and related answers. Additionally the situation described highlights the fairness-related questions that are discussed in this exploratory paper. These “benefits” explain why we have chosen to present these empirical elements in spite of their incompleteness and methodological shortcomings.

The remainder of the article is structured as follows. Part I presents the trial framework, from the French sociologists’ works, and it explains the role of criticism in the dynamics of institutionalized trials (sources, forms and consequences of criticism). Part II shows how this framework is relevant to the understanding of the changes that affected the bonus system of a French textile company over a ten-year period. Part III offers a discussion of the fairness-related questions associated with performance evaluation and of its overall dynamics. This discussion integrates elements from the theoretical trial framework presented in Part I and from the empirical situation exposed in Part II. Implications of the trial model are discussed with a view to more classical perspectives on performance evaluation systems. The fourth and

last part discusses the contribution of the model to a renewed understanding of performance evaluation systems and change in management instruments.

1. The trial framework

The concept of trial

The concept of trial (*épreuve* in French), refers to the social arrangements organizing any testing of people's abilities resulting in arranging tested people in order. This order makes it possible to allocate social goods (money, power, the authority to issue orders, etc.) (Boltanski, Chiapello, 1999).

Significant trials relating to distribution of social goods generally have claims to legitimacy because they organize and create the social system that, to be long-lasting, should not be called into question too easily (Boltanski, Chiapello, 1999). The term *institutionalized trial* refers to a trial that is considered to be important for the distribution of social goods within a society and that, for this very reason, has claims to legitimacy. If there is a certain amount of agreement about the way it is conducted, these claims result in a belief in its legitimacy. The institutionalized trial is therefore also a *legitimate trial*, i.e. it has claims to legitimacy and it is organized in such a way that there is an agreement on this organization (which gives it the claimed legitimacy). Finally, because it is legitimate, the legitimate trial tends to confer legitimacy on the social system instituted by it.

The first requirement for a trial to be considered legitimate is to present itself as a *test of something* that is precisely defined before the trial to occur. Examples are supplied by academic examinations and sporting contests. Persons entering them are measured in terms of their strength in French or mathematics, the javelin or the high jump. Imagine how little credibility an academic examination would have if nothing was known about it in advance: neither the subject nor the program, the evaluation criteria, qualifications of the examiners, the place or the duration. Is it conceivable that society would enable its future elite to be selected on this *ad hoc* basis and that the people under their control would agree to be governed by people chosen almost at random in this way? If a trial is to be legitimate, the strengths upon which it focuses must undergo a process of *qualification* and *categorization*.

The case of sporting trials provides a good illustration of this process. As Elias & Dunning noted, "The history of each sport is (...) fundamentally the history of the formation of sets of rules of an increasingly detailed and precise nature, which impose a unique code on forms of play and confrontation previously on a strictly local or regional basis" (1986, p. 16). In this case, rules are intended to specify the type of strength applied in the trial and the way the trial should be arranged in order to reveal this type of strength and not others. This aims at preventing competitors from applying other types of strength and therefore equalizing all contestants' chances, so that success or failure can be attributed to their merit alone.

But rule making does not cease as soon as the discipline is clearly established in its own right. The contestants' efforts to win result in the introduction of modifications to the techniques used, frequently of a subtle form, whether in using their bodies or in deploying the implements used (bars, bicycles, javelins, etc.). These changes may not be noticed for some time and be conducive to victory, but after a while, they become subject to rule. This mostly results from criticism of unsuccessful adversaries, who have not benefited from the changes, and who consequently consider that loss was unfair, since the conditions of the trial have been

unilaterally modified. For example, in 1956, a Spanish athlete, Érausquin, introduced a new way of throwing the javelin, described as “rotating”, inspired by a traditional Basque sport in which tree trunks were thrown. His success was striking. However this technique was prohibited fifteen days before the Melbourne Olympic Games on the grounds that (i) it was “dangerous” (it was easy for the javelin to leave the set course and to hurt spectators) and that (ii) it radically modified the “physical abilities previously expected” of a javelin thrower (Vigarello, 1988).

Therefore, changes in sporting rules are to a great extent due to the meritocratic requirement to equalize opportunities². The trial should be arranged in order to reveal the merits of competitors on as individual a basis as possible, while limiting inequalities resulting from chance or random factors as far as possible. For the same reasons, contestants should be of approximately equal strength before the event. In sports where success is determined by unchangeable physical characteristics, such as height or weight, this is brought about by introducing categories (as in the case of boxing) or more generally, by organizing a selection procedure including successive sets of trials; only those ranked best in the previous trials being admitted for the final contest.

Regarding the social selection process in general, the existence of an organized sequence of trials also has a practical aspect: it limits the number of contestants in each event. Thus the “judges” are not overwhelmed by a large number of contestants, between whom it would be physically impossible to judge, for reasons of mental load or simply of time.

The major problem associated with any institutionalized trial is the impossibility of designing a trial that is completely perfect and valid over time. This means that the organizers of the trial are required to modify it on a permanent basis, in order to take account of its criticized unfairness.

Academic examinations provide another example of this type of dynamics. Let us for instance consider France in the 1970s, bearing in mind that in this country, much more than in others (the United Kingdom or Germany, for example), the academic structure is responsible for the meritocratic selection of the elite, including the business sector (Bauer, Bertin-Mourot, 1996). The 1970s were years of intense criticism of the academic institution as shown by the effects of Bourdieu & Passeron’s (1970) and by Baudelot & Establet’s (1971) works. Schools were accused of failing in their task of meritocratic selection; of systematically favoring the success of dominant classes’ children, and therefore of organizing social reproduction; and of making the dominated classes’ children attribute their position to their poorer “intelligence”, as exhibited by academic examinations. Given that the French “laïque, free of charge and compulsory”³ schooling was conceived as an opportunity for working class children to escape their condition and socially elevate through a meritocratic structure, these assertions represented a very severe criticism of the educational world. In response to these criticisms, various measures were undertaken, so that children who had been *de facto* until then excluded from secondary and higher education levels could more easily gain access to them. The most salient measure was the Haby reform (11th July 1975 law), which merged the three existing types of secondary education institutions into a unique type of comprehensive school.

A trial can be regarded as legitimate, only if its arrangement specifies its purpose and if its implementation is controlled in order to prevent its illicit exploitation by unknown or at least unexpected forces. Criticism reveals what is unfair in the trials, namely the abilities mobilized by some of the examinees without the knowledge of others, giving them an unfair advantage.

In such cases, the most frequent claim and aim of criticism is to have the fairness of the trial increased (by what is referred to as a “tightening of the trial”), through an increased standardization and/or the extension of its regulatory or legal framework. Other less frequent strategies of criticism towards fairness will be examined below.

The diagram below summarizes the dynamics of the institutionalized trial.

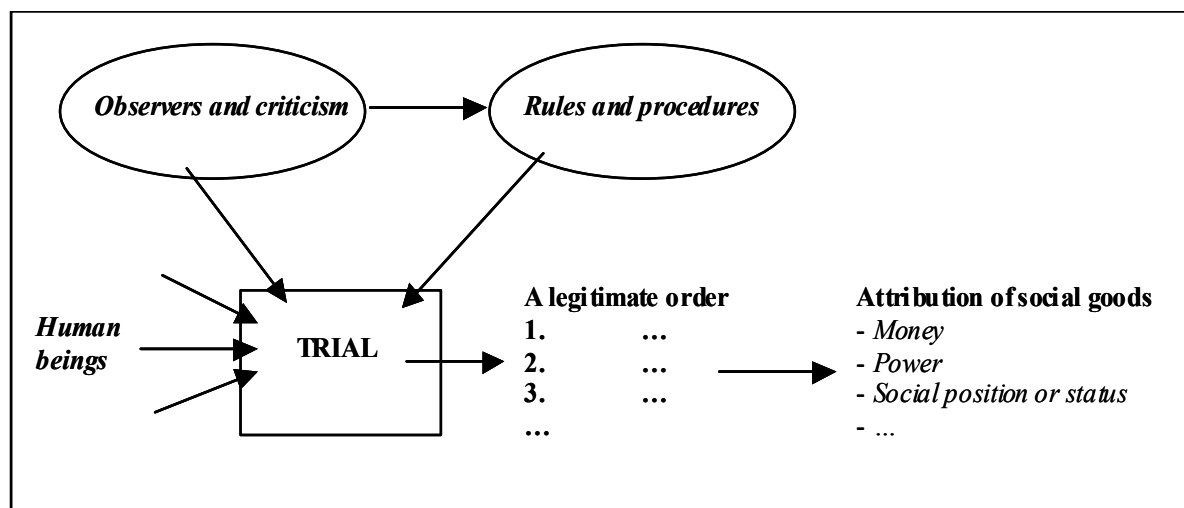


Figure 1. The dynamics of the institutionalized trial

Employee performance assessment can be considered as one of the most important trials of business life, at least for executive personnel. It is usually associated with long-term career decisions, with the granting of bonuses and pay increases, with social recognition, i.e. it directly governs the distribution of social goods. Because these goods are available in a limited quantity (especially money and positions), distribution is based on a more or less explicit ranking, which provides assistance for the decision of who should receive money or promotion. Performance evaluation systems are run according to rules and procedures (appreciation form and interviews, calculation rules regarding performance measures and bonuses...) which are supposed to grant fairness in the process and its outcome.

This fairness concern has already been emphasized by academics studying performance evaluation: for instance, performance assessment has been described as a “central practice in the development of the logic of fairness in human resource management” (Besseyre des Horts, 1991), or as “a symbolic management tool [which] operates as a symbol of equality in the treatment of individuals and of fairness (...)” (Bourguignon, 1993). The trial framework adds two elements to this common view. First, it emphasizes the legitimating function of the practice. Indeed agreement about the system and its rules provides legitimacy to the social order resulting from the fair distribution of social goods, and to their receivers: being repeatedly and positively assessed on the basis of fair rules provides a strong basis for a manager’s legitimacy. Second, it enlightens the role of criticism: disagreement, i.e. dispute about justice, hinders legitimacy and leads to amending the existing rules towards increased fairness.

Trial of merit and trial of strength

The concept of trial described above integrates two concepts of trial developed by French sociologists over the past twenty years. The first is the “trial of merit”, according to Boltanski & Thévenot⁴’s meaning (1991). In their work, a trial is what makes it possible to settle a dispute on the “merit” (value) of individuals in a given situation. The protagonists agree on an arrangement that will enable them to test this merit, the outcome of which will settle the dispute. A trial of merit makes it possible to compare particular individuals using *equivalence conventions* that only take into account a single aspect of their existence. A particular individual is declared to be “equivalent” to another individual in terms of his/her productivity at work, for example. The exercise of judgment assumes a two-way flow between a level occupied by specific individuals in all their variety and disparateness and an abstract and conventional level (the level of the values). This conventional level defines the criterion of judgment (in the above example, the productivity measure).

The second concept is the “trial of strength” as worked out by Latour⁵ (1984): the trial of strength is what happens when forces meet. Latour’s concept of trial is associated with a representation of the world as a network. One major characteristic of this concept is the desire to remain, as far as possible, at the level of singularity and therefore at the level of the multiplicity of individuals and relationships. In a world of this type, there is *a priori* no overhang position and every event or relationship is always situated on a more or less local level. We are in “the plane of immanence” (in the meaning of Deleuze), within a single-level space; what happens in the trial and the subsequent movements cannot be justified in general terms, as in Boltanski & Thévenot’s conception above. According to Latour’s approach, actors referring to values, laws, equivalence conventions, etc. are viewed as mobilizing resources to gain advantage⁶ in trials of strength in which they are involved. The world is only represented through strengths, with justice being just one of them. There are only relationships of strength and the very idea of the legitimacy of a social trial is nonsense.

The concept of trial as constructed by Boltanski & Chiapello (1999) therefore intends to combine the two concepts of “trial of merit” and “trial of strength”: it regards all trials as trials of strength within the meaning of Latour but it considers that some of these trials, which aim at legitimacy, tend to be clarified, organized and regulated so as to approximate to the ideal of the trial of merit. It is therefore possible to refer to *institutionalized* trials (as in the case, for example, of political elections, academic examinations, sporting events and negotiations between employers and trade unions), which are defined and recognized as such. Those involved in them cannot be unaware that their judgments and/or actions in such situations have lasting effects. However, as noted above, no institutionalized trial can claim to be so well regulated that no unknown force occurs in it. It is always possible to find elements which support a description of the trial, not in Boltanski & Thévenot’s terms, but in those of Latour. Criticism arises precisely along with the strengths which are not “filtered” by the existing rules of the trial, and therefore undermine fairness.

Latour’s concept of trial also makes it possible to define as trials moments of confrontation that are not institutionalized, controlled, codified or regulated but in which something nonetheless happens and results in the transformation of the confronted entities. These are not legitimate trials but moments of pure testing of strength. The everyday life of organizations is packed with this type of trial. A particular manager gives a talk on a particular subject on a particular day. It has not been established in advance that the day in question will be a key

date for the rest of his/her career. In his/her opinion, his/her career will rather depend on his/her ability to fulfill the targets imposed on him/her by his/her boss at the start of the period. Moreover, no one around the table realizes what is happening, not even those who will actually decide on the manager's next appointment. However, (s)he is so brilliant (or awful) that his/her future will be transformed as a result. There has been a confrontation; it is difficult to define what happened, but the everyone's position has been redefined.

Highly formalized trials have advantages but also drawbacks in relation to everyday life disputes. By limiting the number of participants involved and making the protagonists agree on the challenges and the aspects to be evaluated, they make it easier to avoid violence, exit from dispute and return to agreement. On the other hand, they place constraints on the individuals, who have to define and limit the grounds for their disputes and, therefore, have to sacrifice everything that is vague, ambivalent and uncertain. Categorization activities occur at all levels in the transformation of a trial into a trial of merit: these activities specify the classes of entities allowed to participate in the trial (cf. weight and age categories in sport) as well as the classes of abilities, the application of which is allowed, and those prohibited. Furthermore, measuring the outcome of the trial itself presupposes the application of categories of judgment.

Trials of this type are also permanently open to improvement and therefore to criticism. The work of refinement is actually without end, because the relationships in which individuals may be apprehended are ontologically unlimited⁷. Since it is still possible to unveil an element perceived as unfair in the running of a trial, in practice trials can be placed in a continuum between a "pure" trial of merit and a "pure" trial of strength.

Regarding performance assessment, it is also possible to identify a continuum between, at the trial-of-merit end, "institutionalized" assessments, based on relatively clear procedures (evaluation systems), and, at the trial-of-strength end, more spontaneous, everyday and informal assessments, like the one described above (cf. the brilliant [or awful] presentation).

Criticism of trials and its effects on governability and legitimacy

As mentioned above, criticism is responsible for continually keeping an eye on a society's main trials, ensuring that institutionalized trials are run according to the agreed format (and that deviance is denounced). In the latter case, criticism can ask for a "purification", a tightening of the trial regarding embodied fairness. Criticism ensures that strengths that should not feature in the trial are not in fact a part thereof – or else, promotes the implementation of procedures and regulations that make the trial fairer. Such a criticism can be termed "reformist".

Criticism can also try to get certain trials eliminated if there is a general desire to bring change to the current system of values. In this case, criticism works at delegitimizing the institutionalized trials. The underlying perception is that the principles currently used as bases for the distribution of social goods should be abolished. Such criticism can be termed "radical". Our study focuses on the role of reformist criticism, which aims at improving the trial, not at substituting it with another trial.

The level of criticism applied to a trial does not depend exclusively on its characteristics. Criticism is socially constructed and the outcome of a particular period of history. Some periods of history, like the 1970s, were dominated by an extremely high level of criticism,

while others, such as the 1990s were marked by a very low level of criticism (Boltanski, Chiapello, 1999). Some trials at the workplace that were criticized in the 1970s would perhaps no longer be criticized today. Furthermore the level of criticism varies between companies; it particularly depends on local trade union work. All companies do not experience the same level of conflict. The local presence of a structured and well established industrial movement is one condition for an increase in criticism, since such a presence “consolidates” reasons for discontent that would otherwise only result in a weaker form of criticism. Some individual factors also determine the intensity and type of criticism: for instance, the personal history of individuals, their place in networks (which favors social comparison, and then equity perceptions), their opportunities, etc.

Criticism may take the two main forms identified by Hirschman (1970): exit or voice. In the scene shown at the start of this article, the woman protests (**voice**) about her pay. The reaction of her boss makes it possible to imagine that she will maybe consider resigning (**exit**). The form of critical expression chosen by the people involved depends on many factors: opportunity for external mobility, belief that one will be listened to, relative position of strength, personality, loyalty to the company, etc. A third form of criticism may be considered: this is the “weak” criticism of people without the power either to leave or to protest without risk. They may take refuge in a form of **resistance**, a wait-and-see attitude, or cynicism, which pose real threats to the social system even though they are less visible than other types of criticism.

The consequences of criticism are multiple and in all cases result in difficulties in governing. Too many people resigning can disrupt work organization and leads to costly hiring, training and adapting of new personnel, not to mention the loss in expertise of those who have left. Too much resistance and bad spirit fundamentally undermine the company’s efficiency. And the expression of protestations frontally attacks the legitimacy of the order in place.

When a particular trial is strongly criticized, its legitimacy and, further, the legitimacy of the associated social system are undermined, causing problems for at least two types of protagonists: (i) the protagonists who have successfully completed the trial in question, (ii) the organizers of the trial. The legitimacy of the former depends mainly on the preservation of the legitimacy of the trial, while the latter feel responsible for the general acceptance of the trial and of the social order “produced”. Therefore, if a trial is strongly criticized, the risk in terms of loss of legitimacy and of authority (refusal to be dominated in Weberian terms) is likely to be perceived as high enough to suggest a reaction. Moreover, if the trial being criticized is considered to be legitimate (i.e. if its justification revolves around the same normative positions as those invoked by criticism), the administrators of the trial cannot be unreactive to the critical comments.

Reactions and responses to criticism and the dynamics of trials

The first reaction possible is to **ignore** the criticism, as does the man in the sketch that introduces this article. Obviously, if the criticism is recurrent and concerns lots of people, this reaction can be dangerous to the company. However, it is not rare, the protestor is perhaps not appreciated by his/her boss or (s)he is known to be dissatisfied or not much liked by the others. Ignorance can also characterize a certain style of management. Probably those exposed to criticism choose the remarks they prefer to ignore and those to which they will respond.

When criticism is recurrent and potentially damaging for the company, the trial's organizers must respond to it, for the legitimacy of the trial to be maintained. They can try and show how criticism is wrong, and provide convincing evidence. This operation aims at **denying** any basis to the criticism. For example, the protestors may be told they do not have a global vision of the situation and that with additional information, there are no grounds for an injustice argument.

The third type of reaction consists of "tightening up" the trial, cleaning it up in such a way as to restore its coherence with the underlying fairness model. This is what happens, for example, when following a series of criticisms, an examination originally revealing the names of the people being tested becomes an anonymous one.

Finally, there is a fourth possible reaction to criticism, in the form of trying to circumvent, instead of answering, it. It may be in some actors' interest for the trial to lose some of its importance, i.e. for it to become relatively marginalized. This is the case when it becomes too difficult to answer to constant and renewed criticism, which sustains permanent tightening up, and further, increased cost. This type of reaction involves a **displacement**, which leads to temporarily disarming criticism by presenting it with a world that it does not know how to interpret.

Displacement consists of establishing alongside the trial under criticism another set of tests not recognized as such but the importance of which gradually becomes central to the debate, without actually having been formally identified as such by the criticism. The latter concentrates on the former trial which falls into disuse without the fact that the rules of the game have changed being realized. The new set of trials is often not instituted, not visible and thus not very fair. An example of this type of displacement is the evolution seen in France over the past twenty years in the remuneration and selection methods of manual workers (Boltanski, Chiapello, 1999). Training seminars have progressively become new hunting grounds for recruiters, without becoming explicitly acknowledged as such. They have been used to reveal workers possessing good oral and written communication skills, i.e. skills hitherto not essential have suddenly acquired a new significance. Workers who did not come up to standard in these training sessions were rapidly put on the redundancy list. These new trials, centered around the identification and evaluation of new skills (and thus not featured in the old trials) lead to a change in the composition of the workforce over a period of time and to a social order based on a new type of legitimacy. Critics, and in particular unions, until then occupied in different struggles, were myopic on this point. This example clearly shows that one of the primary tasks of criticism should be to identify the main trials of a given society, and to clarify and/or incite protagonists to clarify the principles underlying these trials, in order to carry out effective criticism – depending on which options are available at the time, and on the strategies of the persons involved.

Naturally, nothing prevents the answers to those who criticize the trials from being multiple. Some criticisms are taken into consideration and lead to a strengthening of justice (tightening), others are ignored or denied, whilst at the same time new tests not well instituted or controlled (displacement) appear alongside.

The following diagram illustrates the dynamics of change in trials.

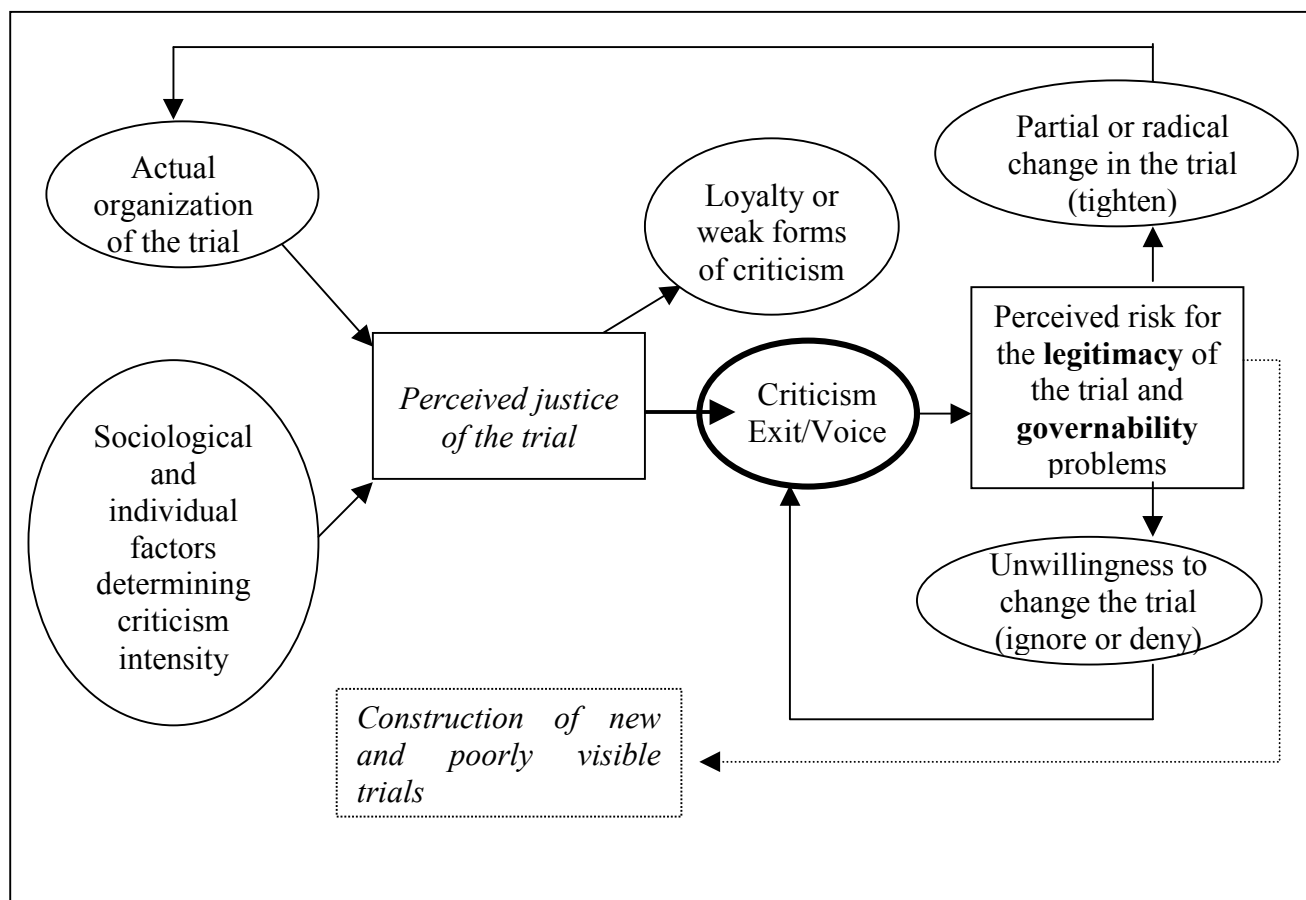


Figure 2. Criticism and the trial dynamics

Since performance assessment can be seen as one of the most important trials of business life, this model should be useful for revisiting the dynamics of associated measurement and management systems. However we suggest delaying such general considerations to focus now on the story of a French company which provides various illustrations of causes for and reactions to criticism, which are likely to enrich our understanding of the processes described in the theoretical trial framework.

2. An illustration

Textirem is the name given here to the Retail Division of a French textile group, within which the bonus system underwent various changes over a period of about ten years (1988-1998). The following sections describe the successive stages of the system, especially the reasons for and nature of change. As the case goes along, we reformulate what happened in terms of the trial model concepts presented above.

Textirem has a workforce of 1 200 and sales of EUR 150 million; it operates in France and other countries, where it controls 15 distribution subsidiaries. It sells supplies for sewing (thread, zips, scissors, needles, etc.) and embroidery (thread and materials for embroidery, embroidery patterns, etc.). It is organized along very traditional lines, with three functional departments (Marketing, Human Resource Management (HRM) and Management Control)

and four operating departments (Plant, Commercial Europe, Commercial United States, Commercial Rest of World) (see Figure 3.). Most of the products sold are manufactured by the (unique) plant, but some are purchased from other operating divisions of the Group and some are purchased externally. The commercial departments supervise the activities of the subsidiaries; commercial activities in France, which are treated as a distribution subsidiary (although they are not subsidiarized), are supervised by the Europe department.

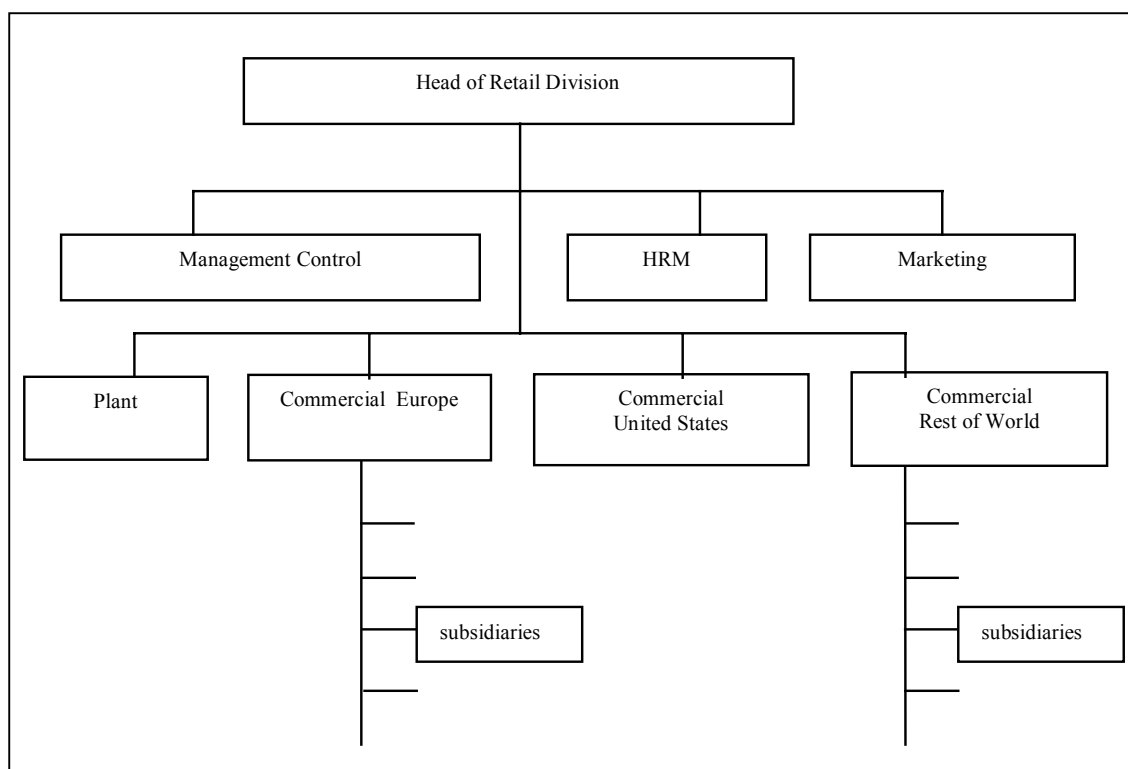


Figure 3. Organization of the Textirem Retail Division

At the very end of the 1980s, the Group experienced serious financial difficulties, which resulted in replacement of almost the entire management team. New management methods were introduced. Management control, in particular, was reinforced on the following bases:

1. All divisions were viewed as investment centers and the operating departments attached to them viewed as profit centers. In other words, within the Retail Division, the three commercial departments and the plant were considered to be profit centers.
2. The plant sold its production at standard cost, on the basis of what was described as normal activity. When the plant was under-utilized, the unabsorbed portion of structural costs was identified in a separate variance account, but nevertheless contributed to the bottom-line profit of the plant.
3. All executives of the Group received a variable remuneration (the “bonus”), which, to a large extent, was based on management control performance measures.

1988: the original bonus system

As regards HRM, the Group’s philosophy is termed “controlled decentralization”. The main features of HR policies are determined centrally, then applied within the divisions by the local HR manager, who is hierarchically subordinate to the central HR Manager and, for

operational purposes, reports to the Head of the Division. The bonuses are allocated on an annual basis by a “bonus committee” of four persons: the Managing Director of the Group, the HR Director of the Group, the Management Controller of the Group and, depending on the executives in question, the head of the division employing them. Decisions are made by the committee on the basis of local proposals. In addition, executives’ performance evaluation is articulated with MBO (negotiation of annual targets with the supervisor and *a posteriori* feedback on performance).

The term “controlled decentralization”, used to qualify HR policy within the Group, does not delude the executives of the Group. The Group exerts a very tight control on this policy, as shown by the composition of the bonus committee: the local representative is on his own before three headquarters’ members and the direct hierarchical superior of the executive in question is not a member of the committee. This central control of HR gives full meaning to the rite of passage constituted by the quarterly dinner, to which all recently hired executives are personally invited by the Managing Director of the Group.

In the Retail Division, 115 executives are concerned by the bonus. Depending on the executive’s hierarchical level (Director, his/her direct employees or others), the maximum bonus accounts for 20%, 15% or 10% of the fixed remuneration. At the beginning of the period being considered, i.e. in 1988, 60% of the bonus is based on a collective performance indicator (the outcome of local operations), the remaining 40% being based on two or three personal targets. This means that, within a distribution subsidiary, 60% of the bonuses of all the executives, including the head of the subsidiary, is based on the operating income of the subsidiary. The currency used for the bonus calculation is that of consolidated accounts (at that time, the French Franc). The number of personal targets (two or three) results from an interplay between three requirements: a) an economic requirement (increasing the number of targets makes calculation of the bonuses complex and expensive), b) a behavioral requirement (single targets always have the perverse effect of making opportunistic behavior seem very tempting), c) a requirement for fairness (“one single target would not give everyone a chance”).

Some of the personal targets are quantified (for example, absenteeism and waste rates, sales of new goods, etc.) and some are not. In the latter case, they are always derived from personal development targets worked out during individual appraisal and MBO interviews. Such objectives include for instance: improvement of the English language, provision of less lenient subordinates’ appraisals, improvement of the management of meetings, etc. No decision of giving to a development target the status of a bonus criterion is made unless there is a local agreement (superior-subordinate) about it.

The bonus is therefore made up of unequal parts (depending on the percentages assigned to each of its basic parts). For each of the parts based on a quantified target, the Group rule stipulates that 100% of the bonus is achieved if performance exceeds the budget by at least 5%. Below this threshold, no bonus is achieved. For parts of the bonus based on a qualitative target, the committee decides whether to grant some or all of the bonus on the basis of recommendations by the executive’s manager and of its own views.

Let us come back to our trial model. It is notable that the HR systems in place implement two performance assessments. The first one takes the form of the annual appraisal, coupled with MBO practices, and while there is no doubt that this appraisal impacts medium-term career decisions, and delivers social recognition, there are no explicit rules regarding the

consequences of this trial. Conversely, the consequences of the second assessment, embodied in the first stage of the bonus allocation decision (are bonus targets achieved?), are very explicitly regulated (composition of the bonus committee, calculation, allocation...). Both systems couple an evaluation of performance (on selected indicators or criteria) with the distribution of rewards. Both are institutionalized trials, since they test executives according to pre-determined dimensions and since the tests result in the distribution of social goods (money, promotion, recognition...). However, because the consequences of the first assessment are less clear and regulated than those of the second, the appraisal-trial appears less “institutionalized” than the bonus-trial. In other words, illegitimate strengths (such as education, reputation, social relationships...) are more likely to impinge on the fairness of the appraisal system, compared to the bonus system.

The fact that the annual appraisal is a less “tightened” trial than the bonus allocation could lead to the belief that there is more criticism towards the appraisal system than towards the bonus one. However, this is not the case at Textirem: executives’ criticism is mostly directed towards the bonus system. Indeed, the level of criticism may be low if the trial is not perceived as central to the distribution of social goods. This is the case of the annual appraisal insofar as executives’ careers in France are largely determined by their education (Bauer, Bertin-Mourot, 1996), and as there is an agreement between parties (companies and executives) about this point. Conversely, it is self evident that performance assessment is crucially important for bonus distribution. This point explains the strength of criticism towards the bonus system.

The high level of criticism towards the bonus system (which is reviewed in the next sections) can also be explained by the tightness of the link between the test and its outcome (the bonus): the relationship is direct, even sometimes mechanical, giving primary importance to the question of fairness.

1990: first criticisms and consequent reactions

Let us come back to Textirem. The bonus rules give rise to dissatisfaction for a number of reasons. First, in view of the centralization of bonus decisions, many managers claim to be “deresponsibilized”: they make recommendations regarding their subordinates, which are not always followed by the bonus committee. However, most criticism relates to the “unfairness” of the rules. The arguments are as follows:

1. Many executives working in foreign subsidiaries comment that “it isn’t fair” to make them “bear the cost of exchange rate differences”. However the management considers this argument as invalid on the grounds that “it’s the very job of a manager to accept risk”, including those of currency markets.
2. Other executives, working in the functional departments of the division, emphasize that the operating income of the division is “a fairly meaningless” measure of their performance.
3. Plant executives also claim that “the system isn’t fair”. When activity falls, the plant bottom line is a loss (which has a direct impact on the bonus), even though the drop in activity is attributable to the sales staff, not to production.
4. Many executives emphasize that the achievement of non quantified targets is evaluated in a highly subjective way, all the more as the bonus committee is very distant from local level (since three out of its four members are from the Group headquarters).
5. It is also observed that the 5% threshold results in a reward that is not in proportion with the contribution: a 100% bonus is received by those who outperform their budget by 5%,

but nothing at all is received by those who only outperform their budget by 4.5%! Moreover, the manager who outperforms his/her budget by 10% receives no more than the manager who outperforms it by 5%. Additionally this threshold effect encourages dysfunctional behavior. If the threshold proves too difficult to reach, it is better to wait for the following year to implement actions likely to improve the indicator. The same thinking applies as soon as the 5% threshold is reached.

These numerous critical comments on the fairness of the system exemplify the critical activity of actors which the trial model emphasizes. Four out of the five criticisms can be analyzed in terms of the ability of the trial to screen out illegitimate strengths properly, while the fifth one concerns the distribution of social goods. The denounced illegitimate strengths are however different in nature. In criticism One and Three (foreign subsidiaries, plant), the illegitimate strengths are factors that are beyond the manager's control (rate of exchange, drop in activity) which however taint the performance tested in the trial. In management control terms, this claim refers to a lack of controllability. Indeed according to this old principle, "a man should be held accountable for only that which he alone can control" (Dalton, 1971, p. 27). This convergence will be discussed in detail in Part III.

Criticism Two is more vague than criticisms One and Three. These latter criticisms point to specific illegitimate strengths, and it may then be assumed that there is no dispute about the other strengths involved in the trial. Conversely, criticism Three does not specify the illegitimate strengths: the whole force tested is under a suspicion of illegitimacy (the performance measure is "meaningless"). Again this can be reformulated in the more general terms of "construct validity" (that is the degree to which a measure faithfully represents the phenomenon) which we shall comment on further in Part III.

The illegitimate strengths conjured up by the fourth criticism are not specified either. Non quantified criteria are claimed not to "filter" enough strength, without specifying the nature of these strengths as in criticisms One and Three. This denunciation introduces another element of the trial: the evaluator. The margin of freedom left to evaluators puts fairness at risk, all the more so as most of them do not know the local situation well. Indeed criticism Four is also directed at the distance between evaluatees and evaluators.

The fifth criticism, which denounces a discontinuity between contribution and reward, relates to the distribution of social goods, not a weakness in the screening of strength, as the first four criticisms did. This type of criticism is enhanced by the presence of thresholds, which create strong discontinuities within a continuous phenomenon. This is detrimental to fairness, because what fundamental difference is there between the person who just reaches the threshold and the person who just misses it? Only the possibility of doing the trial several times at relatively close intervals would moderate the feeling of injustice. An annual evaluation like that of the Textirem bonus could be considered too infrequent to moderate the feeling of injustice, bearing in mind that a change in the economic situation, always a possibility, renders the next trial uncertain.

How does Textirem management react to these criticisms? Only the last criticism is acted upon. In 1990, the management decides to modify the bonus qualification rule as follows. Those outperforming their budget by 5% still receive 100% of the bonus; below this threshold, the percentage of bonus granted decreases. On average, one can hope to receive 50% of it if one has performed on-budget. No bonus is paid for performance equal to 90% of the budget. These new thresholds are on an indicative basis and can, if appropriate, be

modified in accordance with the retrospective views of the line management. For example, only 40% of the bonus (instead of 50%) will be awarded for a budget that has been met but *a posteriori* considered as very easy to meet.

With respect to the trial model, it is worth noting that the modification undertaken potentially opens a new path for dispute. The *a posteriori* appraisal of the level of difficulty of the target is highly subjective: on which basis does the bonus committee decide that the budget was difficult to meet, then that the percentage of the bonus granted should be reduced, and in what proportion? Thus, the change brought as a response to criticism simultaneously increases and decreases the fairness of the trial with regard to different aspects. The trial is tightened (by the change in thresholds) but at the same time relaxed (by reinterpretation of the target level).

It is also notable that four out of the five criticisms are ignored. Most probably the management perceives that the level of criticism is not high enough for the system, and subsequently the local social order, to lose their legitimacy. Maybe the number of protesters is not very high (what can be referred as to “the quantitative level of criticism”) and/or they use weak forms of criticism (i.e. “the qualitative level of criticism”). It is remarkable that the unique criticism considered for action does not only point to fairness, but also to strategic effectiveness. The threshold effect creates dysfunctional behavior, in having managers act to maximize their present or future bonus, and not the ongoing performance. The ignorance of the four other criticisms suggests that the fairness concern probably plays a limited part in the reaction to this fifth criticism, which might mainly be motivated by the gain associated with the elimination of dysfunctional behavior.

Another explanation of the weak level of change regarding the number of criticisms lies in the cost incurred by change. Indeed any trial is expensive for the body organizing it. Thus Textirem deliberately use a limited number of personal targets in order to contain administrative costs (which are apportioned to the number of targets). Trial modifications generate other kinds of costs, namely those of having the various parties reach a minimum agreement about the change to be made. This may contribute to ignoring criticisms, as in the case above.

1993: Further criticism and reaction

In 1993, an “incident” results in another revision of the system. Against the Head of the Division’s advice, the bonus committee decides not to award a bonus relating to the personal target of an executive. The target in question is a feasibility study regarding the setting-up of a finishing unit in another country. At the end of his study, the executive advises against establishment. The committee considers that he has failed to achieve the target; the Head of the Division takes the opposite view (whatever its consequences, the study has been carried out), but is unsuccessful in convincing the committee.

The Head of the Division considers the decision to be “unfair”. Additionally he keeps in mind the ignored previous criticisms, and namely the fourth one about the subjectivity of qualitative targets. Thus he decides to make use of his local margins of freedom, both to avoid repetition of such an “unfair” decision, and to contain the dissatisfaction which will doubtless arise from this incident confirming previous suspicion. He decides that, from that point on, his direct or indirect subordinates, i.e. all executives of the Retail Division will only be given quantifiable targets. Indeed such targets leave little room for subjectivity and therefore for that kind of unfairness.

The trial becomes tighter. This, however, poses a number of problems of application, in particular for executives working with medium-term timescales, such as development engineers and HR managers. Some performance indicators can be developed for these professions (for instance respectively, the number of new goods introduced and an absenteeism rate), but are they sufficiently valid representations of the contribution made by the individuals concerned? The executives involved doubt this: to what extent do the above measures reflect their effort and the action undertaken, most effects of which are only felt (and moreover, hardly identifiable) in the long term? Questions about the validity of the performance indicators used in the trial, already at stake in the second criticism, cast doubt on the fairness of the trial.

1996-1998: Further criticisms, consequences and reactions

Three years pass. Three changes occur over the following years:

First, the Head of the Group retires and is replaced by a man who decides to *decentralize* bonus decisions. From then on, bonuses are granted at the local level. The bonus committee, composed of the Head of the Division, the Management Controller and the HR Manager of the division, makes decisions on the basis of recommendations by local managers. Consistently, the decision of which criteria should be used as bases of the bonus is also decentralized at division level (bonus committee).

Second, the Head of the Division considers that the effectiveness of the division could be considerably enhanced if executives thought along more collective lines. According to him, everyone tends to think primarily “local”, although a number of management approaches can be shared across the division, at least across Europe as far as marketing is concerned. Therefore, in order to encourage his executives to think more collectively, the Head of the Division decides to introduce a slight change in the rule of “60% of the bonus based on operating income”. From now on, the operating income considered is not that of the Division, but that of the hierarchical level immediately above the level of the executive rewarded. For example, the European operating income is used for the heads of European subsidiaries (instead of, previously, the local subsidiary’s operating income).

Third, the Head of the Division is sensitive to the criticisms of executives who work to medium-term timescales, and who criticize the use of quantified targets as a basis for their bonuses. More generally, he is also convinced that “this type of [bonus] system must be changed every 3-4 years”, otherwise “people get used to it and the system loses its power to motivate”. Therefore, he decides to go back on the 1999 change and to re-introduce qualitative targets where relevant.

Insofar as they answer to criticism, the first and third changes have a positive impact on protest. Decentralization is positively perceived as a partial answer to the fourth criticism. Because it reduces the distance between the levels of operations and of bonus decision, evaluated executives anticipate more fairness in the appraisal of achievements, especially those expressed in qualitative terms. And the reintroduction of qualitative criteria reduces criticism by functional staff and all those dissatisfied with the previous system. However the changes generate new criticisms. While some executives are happy with the reintroduction of qualitative criteria, there is a new wave of criticism from all those who are afraid of subjectivity and therefore potential unfairness of assessments based on qualitative criteria

(cf. criticism Four). Finally, executives as local managers (who they are very often simultaneously) find that the decentralized system is much more “uncomfortable” to live with than the centralized one. Indeed, when there was a significant distance between the manager and the bonus decision, it could be invoked in order to justify a lack of or a disappointing level of bonus to a subordinate (and this was made even easier by the sometimes odd interpretations of the bonus committee, as in the example above of the feasibility study). Both the decentralization of bonus decision and the reintroduction of qualitative targets place all team managers in a very uncomfortable position. It is actually always difficult to give reasons for the assessment of qualitative criteria and/or to tell an employee that he does not deserve his/her bonus for a particular criterion.

As for the second change, which is motivated by collective effectiveness, and does not answer to any criticism, it leads to a criticism which, like criticisms One and Three, can be expressed in terms of non-controllability and, therefore, of unfairness perceived. The argument put forward is that it is very unfair to be rewarded on the basis of “others” performance. The denunciation of unfairness is stronger as executives perceive themselves as independent of others, as in the case of the head managers of European subsidiaries.

Summary

The story of the Textirem bonus system clearly shows the dynamic nature of criticism, and how management systems change in response to criticism. It is clear also that, from the company’s viewpoint, fairness is a constraint for strategic effectiveness, not an objective *per se* of the company. The main aim of a bonus system is to encourage/manipulate behavior, and this is only achievable if a minimum level of fairness is guaranteed. This means that, although fairness is, theoretically at least, at the core of the trial, it ranks second to strategic effectiveness when questions of design or change arise. Thus, as seen above, Textirem changes its system to enhance (supposedly) more efficient collective behavior, despite its being detrimental to fairness. The need for fairness is only included, by tightening the trial, when protests reach a level that seems likely to place effectiveness at risk. As shown above in the threshold change, the criticism is made all the more effective as it confronts the question of effectiveness. When it draws attention to rules that are not totally “effective”, criticism has all the more chance of being heard.

This illustration exemplifies the potential fairness shortcomings of assessment:

1. Lack of controllability of the measure: a performance indicator, the value of which results from my action, but also from external causes (colleagues’ actions, foreign exchange market, etc.) is unfair.
2. Qualitative measure: a performance indicator that permits “too much subjectivity” is unfair.
3. Lack of validity of the measure: a performance indicator that does not represent my contribution and action properly is unfair.
4. Assessor at an inappropriate distance: if the assessor is too far away from me, (s)he is not aware familiar with my situation and therefore is unable to interpret my performance in a relevant way, while if (s)he is too close, (s)he is tempted to be overindulgent to all his/her staff and therefore not to recognize my action specifically. Assessment is unfair in both cases.

In other words, from the person’s viewpoint, any differences between his/her action and the measurement or judgment of performance (i.e. strengths that are not “filtered” by the instrument) are sources of unfairness.

This story also shows that the various sources of shortcomings are not independent of one another and how tightening a trial in response to a criticism may at the same time result in its being loosened in another way. For example, improving the validity of performance measurement (i.e. complementing quantitative measures with qualitative ones) simultaneously tightens and relaxes the trial.

This story also shows the important and sometimes ambivalent role of agents at different points in the hierarchy. Perception of unfairness may differ depending on the hierarchical level (as the example of the feasibility study showed) and a manager at a particular hierarchical level may use his local margin of freedom and power to make the device fairer. Moreover, for all managers who are both assessors and assessed, the demand for fairness (which is first and foremost made by the assessed) sometimes comes into conflict with other objectives, such as the non-conflictuality of relationships with subordinates.

3. Performance assessment: a trial-inspired model

The case study provides various illustrations of processes, which the trial model describes in a very general way. In this third Part, we combine both empirical and theoretical contributions to elaborate an innovative framework for analyzing performance evaluation. We draw on a general conceptualization of performance assessment (Bourguignon, 1998a) in order to reorganize (i) our empirical observations, and (ii) the various fairness-related questions that can be associated with performance assessment as a trial. As we have seen above, fairness considerations sometimes “meet” existing concepts (i.e. controllability or validity); in such cases, we offer a brief review of the concept and explain the specific contribution of our framework to an extended understanding of it.

In organizations, performance assessment may be defined as “the process in which a duly appointed person makes, with the help of the appropriate instrumentation, a judgment followed by effects on the achievement of another person’s organizational objectives” (Bourguignon, 1998a, p. 96). Three steps may be identified:

1. *Instrumentation*. This first step of the trial includes the operations of qualification and categorization that are pre-conditions of assessment. As a starting point, this stage includes a definition of the performance to be assessed. Once performance is defined, categorization occurs and results in the design of instruments that will capture the information used in the next step (evaluation). Instrumentation includes the design of individual assessment grids, the choice of performance criteria and their definition, as well as the choice of the referents (past or budgeted data, general missions in the job...) that, later, will anchor the judgment on performance. This step may involve “technical” complexity, namely in accounting terms. This is especially the case when costs, the calculation of which is seldom a simple matter, are involved in performance assessment (Bourguignon, 1998a, p. 112 et s.).
2. *Evaluation*. The outcome of the second step is the production of a value judgment relating to the individual’s performance, through the comparison of his/her actual performance with that chosen as a reference. The value thus attributed to performance depends on three terms involved in the act of judging, i.e. (a) the information produced by instruments, (b) the reference, (c) the assessor. As seen in the Textirem case, the judgment can be automatic from figures produced or a certain discretionary margin can be granted to the

assessor.. In all cases, performance evaluation comprises a measure and a judgment on the value of the thing being judged (Bourguignon, 1998a, p. 135).

3. *Consequences*. Value judgments have numerous consequences, among them the distribution of social goods: income (variable remuneration of executive personnel, for example, as in the Textirem illustration), recognition and social positions (internal promotion) (Bourguignon, 1998a, p. 137). Indirectly the judgment on performance may also impact future financial resource allocation (a high performer's requests for funding are more likely to be listened to than those of a poor performer). This distribution phase is analytically different to the preceding phase of performance evaluation, even if the two are not always dissociated in time. In addition, in the attribution of a bonus as well as in performance evaluation, a more or less discretionary judgment may be made. Indeed, the bonus is not automatically awarded according to the produced evaluation, as shown by the first reformation of the Textirem system (1990), which allowed a discretionary reduction of the bonus. A contradictory example (i.e. a discretionary increase of the bonus) is given by a decision to encourage a new manager who is not (yet) obtaining good results by giving him a small bonus.

These various elements can be summed up as follows:

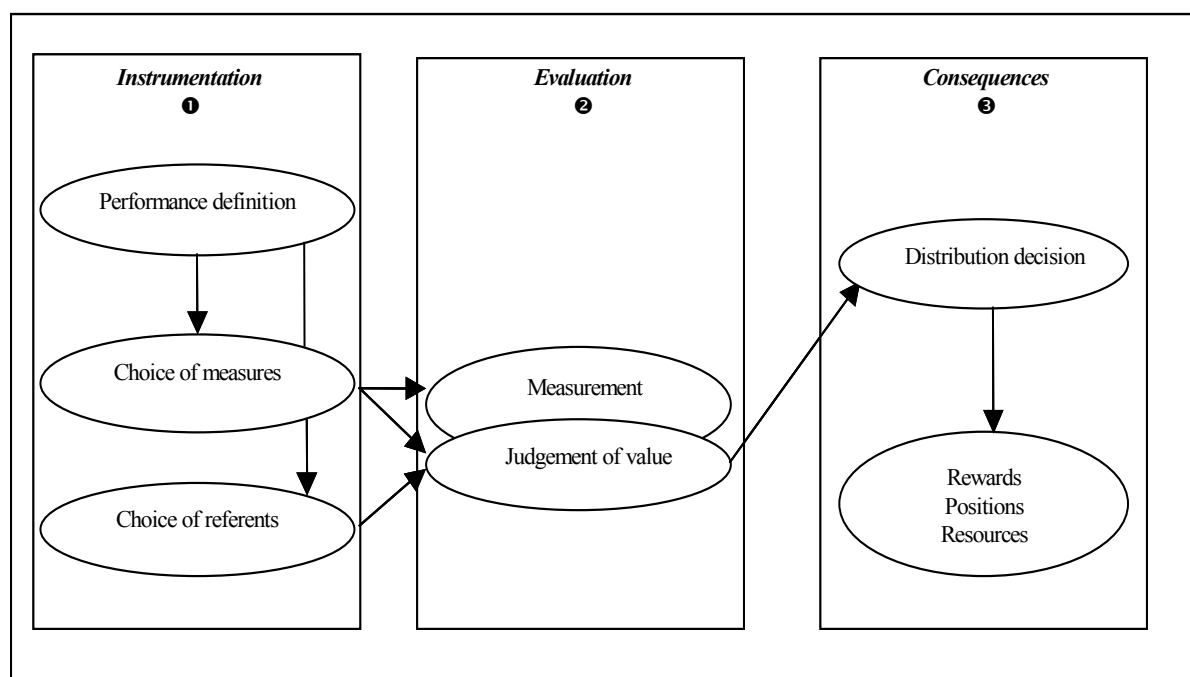


Figure 4. The performance evaluation process

We propose using this framework to review the questions of fairness raised at the three steps of assessment, namely those of performance definition and instrumentation (①), of assessment itself (②) and of distribution (③). However, obviously, some questions are cross-sectional, notably those associated with measurement (① and ②), which means that the structuration of the discussion below should not mask the interdependency of the questions.

Performance instrumentation and fairness: controllability and validity revisited

Performance instrumentation and justice meet around two questions, which have been recurrent themes either in management control or in measurement theory for a long time, namely (i) *controllability*, and (ii) *validity*.

Controllability. As Solomons claimed, “it is almost a self evident proposition that, in appraising the performance of divisional management, no account should be taken of matters outside the division’s control” (1965, p. 83). Controllability is mostly known as a normative “principle”, although there is no general consensus about it.

Indeed empirical studies show that this principle is seldom observed in the strict definition of the term (Merchant, 1987). Managers are generally assessed on the basis of indicators that include uncontrollable factors. The main reasons for not implementing the controllability principle are that it is a way to (a) have the managers share the shareholders’ risks⁸, and to (b) keep them aware that their decisions affect areas out of their control (*ibid.*). Companies differentiate along with different levels of risk which they ask their managers to bear (i.e. uncontrollable factors which however are involved in performance evaluation) (Hofstede, 1967; Vancil, 1979; Merchant, 1987). The Textirem case we studied provides several examples of these uncontrollable factors in the evaluation of performance (exchange rates, actions of other managers).

In textbooks, the controllability principle is generally introduced through its underlying rationale that “results measures are useful only to the extent they provide information about the desirability of the actions that were taken” (Merchant, 1998, p. 76). The emphasis is put on “noise” (*ibid.*, p. 77), not justice. Uncontrollables are undesirable because they “make it difficult to infer from the results measures whether or not good actions were taken” (Merchant, 1998, p. 76). In other words, the problem raised with the controllability principle is one of representation, not justice. However the problem of injustice is indirectly touched upon with the acknowledgement that holding individuals accountable for uncontrollable events can lead to dysfunctional behavior (game playing, loss of motivation and employee turnover) (Merchant, 1987). But again, it should be noted that justice is not considered for itself, but rather as a mediator and condition of effective behavior.

The situation is somewhat different in HRM. On the one hand, Murphy & Cleveland note that “performance appraisal has traditionally been viewed (...) as a measurement problem”, with a focus on scales and biases (1995, p. 1) – and there is no room for fairness in this perspective. On the other hand, fairness has been emphasized as a basic requirement for two reasons. First, if performance appraisal is perceived as unfair, the system is unlikely to be accepted and thus unlikely to meet its goals (*ibid.*, p. 310-311). Second, and this point mainly concerns the United States, a fair system prevents any legal contestation with respect to antidiscrimination laws (*ibid.*, p. 46). To sum up, although the fairness concern is certainly more common in HRM than in management control, the dominant conception of fairness is alike, that is instrumental and orientated towards the prevention of dysfunctional behaviors or attitudes.

The trial model increases our understanding of the type of “constraint” that fairness is. It is the very condition of the legitimacy of social order, which means that it should not be treated as a secondary and minor point. What is the use of refining instrumentation, from a representational viewpoint, if the whole system is delegitimized because it lacks fairness?

The trial model also enables us to reformulate one of Merchant's (1987) findings and further to suggest amending the controllability principle. Merchant observed that differences between companies in their application of the controllability principle did not lead to differences in perception of justice. In our words, the managers interviewed adhered to the trial in the form in which it was formatted in their company, i.e. they agreed with the local definition of the "good manager" as reflected in the trial. If everybody agrees about the thing to be tested in the trial (for instance shares the heroic vision of managers supposed to "get out of" any situation, even in the worst possible environment), there is no feeling of unfairness, and then no dispute. Thus in itself, a lack of controllability does not necessarily give rise to a feeling of unfairness, provided that everybody agrees with the idea that coping with uncontrollable events is one aspect of the manager's job.

This discussion suggests that controllability is not an absolute concept, but the always-contextual outcome of an agreement between parts of a given social order. Instead of looking for an evaluation aligned with the controllability principle, it would be more fruitful to come to an agreement on the definition of the manager's job.

Validity. Construct validity is the "quality of an instrument in which the constructed indicators are a good representation of the phenomenon studied" (Evrard *et al.*, 1993, p. 591). Although the term is not much used in management control⁹, it is commonly referred to, as shown in the following excerpt: "The results-control ideal would be to measure the value created by each employee. Because direct measurement of value creation is rarely possible, most firms use their management-level results controls on accounting measures of performance. (...) It must be recognized that even the best accounting measures are not perfect; they are only surrogate indicators of changes in shareholder values" (Merchant, 1998, p. 454). Validity raises the question of the correspondence between what is measured and what is meant to be measured. Thus it encompasses both questions of controllability (in the usual sense of the notion) and of completeness. Controllability refers to what is in the measure and should *not be*, and completeness, to what is *not* in the measure, but should be. Criticisms, like the second one at Textirem, which are expressed in terms of "meaninglessness" of performance measures, point to validity problems. They indicate that the performance measure tests "something other" than what should be tested, although it is not always clear whether that something refers to something missing (completeness) or something in excess (controllability).

The trial model makes it possible to reconsider validity, not only from a representational perspective, as is frequently the case in management control (*cf. supra*) and HRM, but from a social perspective. A lack of validity also causes unfairness, and, further, legitimacy problems.

Regarding validity, performance assessment in organizations presents a fundamental difference with some paradigmatic trials like academic examinations or sporting contests. Persons entering these paradigmatic trials can prepare for the event, which furthermore takes place in a space in which all possible disturbing factors (illegitimate strengths) are controlled. In the business world, the equivalent of academic or sporting trials are the assessment sessions designed for applicants, that is persons who are (still) outside the organizations. Conversely, the performance assessment of participants in organizations is not organized outside the everyday life of managers. And managers do not specially prepare for it in their everyday activities, unlike students or athletes. Instead, the trial takes the form of a review of past events, the purposes of which were not the manager's performance measurement, but the realization of organizational objectives. This type of trial is deeply embedded in economic life

and is therefore not “pure” at all and, by nature, only “purifiable” to a small extent. Embeddedness of this type automatically raises the thorny question of separating what is historically due to the particular merit of the manager (personal factors) and what is due to the context (environmental factors). It opens up potentially endless dispute about performance attribution and related bias, etc. This characteristic of performance assessment casts a new light on the controllability notion. Even if controllability can be locally defined by an agreement on what should be tested, controllability can never be achieved because the conditions of the test do not make it possible to disentangle the thing being tested from exogenous influences.

This specificity of the performance assessment trial also enables us to reconsider dysfunctional behavior, often referred to as the fact that people “manage” performance measures, instead of strategy (cf. the threshold effects at Textirem). This can be viewed as an unavoidable consequence of the embeddedness of the trial in “real” life. It should not be surprising that assessed persons anticipate the trial, and that, consistently, they behave according to this perspective, and not to the expected strategic one. Consequently, dysfunctional behavior should not be considered a side-effect of performance evaluation systems, to be corrected, but as a central and inherent aspect of such systems.

Finally one can wonder why all these embeddedness-related shortcomings have not resulted in a search for other types of trial that would be more satisfactory in terms of justice. Instead companies keep on organizing this “embedded trial” that constitutes a review of the past, while they attempt to tighten it up where it is possible (*ex ante* clarification of expectations, negotiation and agreement about targets, sometimes *ad hoc ex post* negotiation, with the manager, of the judgment about performance). A first explanation is that, as explained above, justice is not an objective *per se* in organizations, and that the shortcomings are not damaging enough in terms of effectiveness (the primary concern) for the organizers to undertake a radical (and costly) revision of the trial. A second reason is that often, the influence of the environment is not viewed as an illegitimate strength, insofar as it potentially impacts every manager’s performance in a similar way. A third reason is that environmental changes are viewed as relatively random, then difficult to anticipate, which means that there is no way of preventively eliminating these strengths, should they be considered illegitimate¹⁰.

A final question raised by the validity concern relates to the lapse of time between two assessments. Remember Textirem where the correspondence between the operational (long term) timescale and the (annual) rythm of evaluation proved difficult to match. In such cases, the definition of what should be tested, and beyond, the nature of the trial, are largely uncertain.

Performance evaluation and fairness: objectivity revisited

Performance evaluation (that is the second stage of our framework) and fairness meet on the question of objectivity, which is generally considered as a basic requirement of performance assessment, particularly by HR authors. However it is notable that *objectivity* is very often not defined, although it is systematically referred to, directly or indirectly, through its opposite *subjectivity*, as in the following excerpts:

“Does the instrument encourage objective assessment of performance?” (Gordon, 1986, p. 231).

“The problem which is likely to arise is that of the superior’s subjectivity” (Sekiou *et al.*, 1995, p. 275).

“Good performance criteria are observable and objective” (Milkovich & Boudreau, 1991, p. 170).

“Objective-type, non-rated and uncontaminated data should be used whenever it is available “ (Bernardin & Cascio, 1987).

“Problems of subjectivity are particularly evident when non quantifiable criteria are being used for assessment purposes” (Beardwell & Holden, 1994, p. 532).

Subjectivity (which is usually not defined either) is universally viewed both as inherent to assessment (Hall and Goodale, 1986) and detrimental. In North America, the detriment is generally associated with legal problems. Appraisal systems have to be “legally defensible” (Bernardin & Cascio, 1987) with respect to the anti-discrimination regulation, and subjectivity makes defense problematic. French authors sometimes mention other negative consequences, more concerned with social interactions: “Without constant vigilance, the thought producing judgments is constantly evolving; it is directed against people, but not openly, and thus influences our action” (Piveteau, 2000, p. 57).

Thus objectivity appears as the quality or state of being “just and unbiased” (Roget’s II, 1995) or “uninfluenced by emotions or personal prejudices” (The American Heritage[®] Dictionary of the English Language, 2000). While dictionaries refer to ‘fair’ for synonyms, it is worth noting that fairness questions do not appear to be an issue for HR authors, except when fairness is associated with legal requirements (Hall & Goodale, 1986). In this section we reexamine the classical considerations on objectivity with a view to our trial-inspired model, and we offer a synthesis which enables these classical perspectives to be defined and broadened.

As shown in the above excerpts, there are two main aspects to objectivity in its common acceptation, which can be associated with both components of evaluation, namely the (i) measurement and the (ii) judgment of value stages (*cf. supra*). These aspects respectively concern *the form of the measure used*, and the *biases affecting judgment*.

Form of measures. The measures used in performance assessment can be quantitative or qualitative. This distinction, which is borrowed from theoreticians of measurement (Carnap, 1966) is not the most common in management, where management control authors prefer to contrast financial measures with non-financial measures of performance (McNair *et al.*, 1990) or lag indicators with lean indicators (Kaplan & Norton, 1996, p. 32), while HRM appraisal specialists use typologies differentiating ranking, rating scales, critical incidents, behavioral scales, etc. (Milkovich & Boudreau, 1991, p. 173 *et s.*). The quantitative-qualitative opposition presents the advantage of being common to all assessment arrangements, whatever their functional source.

With a quantitative performance measure, all traditional mathematical operations can be performed (Carnap, 1966). This category includes all management control measures, whether they are of the financial type or not (i.e. costs, profits, breakdown-, waste-, or absenteeism rate, etc.), apart from satisfaction indicators. Comparatively, a qualitative measure only makes it possible to order observations (Carnap, 1966), as occurs in all assessment of skills, capacities or behaviors that are to be found in HRM arrangements. Satisfaction indicators are also included in this category. Please note that this typology is not in any way related to a very commonly mentioned contrast of “quantitative” and “qualitative”, which refers to

strategic aspects and therefore to the performance measures associated with them. In this acceptance, the adjective “qualitative” is generally synonymous with non-financial and an absenteeism rate is therefore considered to be a qualitative criterion. Although it is widespread, this usage seems to us to be inappropriate.

The form of the measure has important consequences for objectivity. When performance assessment uses a quantitative measure, the measurement actually precedes and objectifies the value judgment. On the other hand, if assessment uses a qualitative measure, the value judgment comes first, and the subsequent rating only formalizes the judgment¹¹. This difference explains why quantitative measures are commonly considered as more objective than qualitative ones. The Textirem illustration provides extensive examples of the recurring debate about the subjectivity involved by the use of qualitative measures. In terms of trial, subjectivity associated with qualitative measures means potential illegitimate strengths, about which all parties involved in evaluation feel nervous. For subordinates, illegitimate strengths threaten fairness, and for assessors, they constitute a threat for the quality of hierarchical relationship, inasmuch as it is always problematic to justify an assessment based on subjectivity-laden criteria.

Additionally it should be noted that the form of the measure used has implications on the possibility to separate both in time and analytically the two components of the assessment (measurement and value judgment). First, both steps occur at moments that are more or less distant in time. The measurement- and judgment-time are simultaneous when the measure is a qualitative one (the judgment is “inscribed” in the measure). Conversely, when quantitative measures are involved, generally, there is a first step of measurement (for instance, the computation of the costs incurred), and only later, the judgment about the value of the thing measured can take place, with reference to a norm (budget, past realizations, etc.). In such cases, the measurement step is clearly differentiated in time from the judgment step. Second, according to arrangements, the analytical differentiation of measurement and judgment (that is, the capacity of identifying them as autonomous operations) is more or less clear. In the case just mentioned, measurement and judgment are autonomous steps, and thus can be clearly differentiated in nature. Conversely, consider an indicator measuring the degree of realization of the budget. The reference is incorporated in the indicator, which means that the indicator does not only embody measurement, but also, in a kind of mechanical way, judgment. For instance, a performance equal to 70% of the budget is not a mere measure, but includes a judgment of value (performance is reasonably far from the target). This initial value judgment will serve as a basis for the assessor to make his/her final value judgment, with reference to another criterion (given the market conditions, is 70% an awful, a bad or an acceptable performance?).

This discussion shows that, although assessments based on quantitative measures appear more objective, thus more likely to be fair, than those using qualitative measures, they never totally eliminate subjectivity which characterizes the judgment of value stage of evaluation. Fairness is thus also at stake in the second aspect of objectivity, which is associated with the judgment stage, namely the question of biases affecting judgment.

Biases affecting judgment. Again this question is not new and a fruitful stream of research has highlighted a number of cognitive “biases” in evaluation. For example, it has been shown that evaluative judgments are subject to the influence of previous performance (Murphy *et al.*, 1985), the expectations of the supervisor (Hogan, 1987), the quality of the relationship between the supervisor and his/her subordinate (Duarte *et al.*, 1994), his/her previous

commitments to the ratee (Bazerman *et al.*, 1982), awareness by the supervisor of a more favorable self-assessment by his/her subordinate (Blakely, 1993), etc. Again it is notable that this type of research is not primarily sustained by a justice perspective, but by a metrological one – as in the above cases of controllability and validity, where the quality of representation was the core question. The very idea of “bias” only makes sense in a positivist-inspired perspective: it suggests that “reality” is different from its representation. Although cognitive research on appraisal also acknowledges that evaluation biases can be harmful to fairness perceptions (DeNisi, 1996, p. 128), it is not primarily focused on fairness. However these findings can very easily be reformulated in the terms of the trial model. What is called “biases” in cognitive research on performance assessment refers to our “illegitimate strengths”. They impair the trial and benefit (or harm) the ratee in an abusive way. Finally, although cognitive research on evaluation is not orientated towards an increase in justice, it is worth noting that its outcome is convergent with an improvement of fairness. Indeed a general recommendation of such cognitive studies is that evaluators be trained to develop their awareness of such biases, and further, to decrease the level of bias in evaluation. Reduced biases do not only mean an improved representation of performance, but also fewer illegitimate strengths.

Evaluative biases are also associated with *whom* is the assessor. According to Merchant, “(...) objectivity is low – meaning the possibility of biases is high – where (...) the actual measuring is done by the persons whose performances are being evaluated. Low objectivity is likely, for example, where performance is self-reported (...). Managers have two main alternatives they can use to increase measurement objectivity. They can have the actual measuring done by people who are independent of the process, such as on a controller’s staff, or they can have the measurement verified by independent persons, such as auditors” (1985, p. 26). This recommendation, that is having a third party, who is not involved in the supervisor/subordinate relationship, carry out the assessment, is implemented in the 360° appraisals, which are currently developing. Such appraisals combine appraisals from superiors, subordinates and peers. However they do not reduce subjectivity, they just offer a multiplication of subjective viewpoints and of illegitimate strengths involved.

A critical and integrative view. We would like to propose here a reunification of these different considerations of objectivity, based on the works by Megill (1994) and Porter (1995). This reunification places the assessor at the centre of the objectivity question, as it is his/her subjectivity that is implicitly at the heart of (i) the question of the form of the measure (qualitative measures providing a greater margin of freedom) and (ii) that of the different biases of the evaluation.

According to Megill (1994) and Porter (1995), objectivity has several meanings. One of these is “*mechanical objectivity*” (Porter, 1995) (or “procedural objectivity”, according to Megill [1994]). The previous debate about quantification refers to this meaning of objectivity. Mechanical objectivity focuses on impersonality and on standardization. Objectivity is guaranteed by the total reproducibility of the judgment procedure. Whoever applies it is supposed to arrive at the same result, making the judgment independent of the person making it. Quantification thus appears as a privileged means of mechanical objectivity. Judgments made in this way are objective in the sense that they are not dependent on the judge.

But for all this, are they fairer? Supposing the sacrifice of the singular and local to the advantage of a standardized procedure, these judgments do not take into consideration circumstances that only an expert’s judgment can encompass. We consider here a second

meaning of objectivity: “*disciplinary objectivity*”. This “takes consensus among the members of particular research communities as its standard of objectivity”(Megill, 1994, p. 1). It “refers to the claims by practitioners of a particular discipline to have authoritative jurisdiction over its area of competence”(ibid., p. 5). The judgments concerned are those of experts whose validity is guaranteed by their professional milieu, which provides training and the upkeep of members’ skills. In the Textirem case, when the bonus committee gives discretionary but consensual recommendations, it leans on a claim of disciplinary objectivity. If quantifications can be used here in support of the judgment, they do not determine it. The expert is the person capable of going beyond simple figures and considering in a global judgment the singular and non quantifiable aspects of the situation studied.

A final aspect of objectivity is *independence*. For example, when one expects a judge presiding a court case to be “objective”, one is in fact expecting him/her to be independent, i.e. “in the service of truth and justice” only and not under pressure from various groups (Porter, 1995). However, independence is rarely sufficient and it is often associated with disciplinary objectivity. A court judge is thus also supposed to be trained for and experienced in this work. His/her knowledge of similar cases, in other words his/her expertise, helps him/her form an impartial verdict. Audit literature is familiar with these two indispensable dimensions of judgment, competence and independence (Lee, Stone, 1995; Citron, Taffler, 1992; Richard, Reix, 2002).

When one is seeking a just judgment, one seeks objectivity in the three senses evoked above (independence, expertise, reproducibility) ... which is impossible as these three dimensions are partly contradictory. As seen with Textirem, there is no completely satisfactory situation. The mechanization of judgment (with quantification and automaticity) gives it a mechanical objectivity, but it loses some of its depth and disciplinary objectivity. In the same way, a judge who is too independent (far from the “reality” of the evaluated person’s work) lacks understanding of the situation and can be unfair, as a judge too close to this reality will lose in independence. The Textirem case is a perfect illustration of these contradictions.

One must thus give up the idea of finding an absolutely fair judgment to revert to a sociological perspective where the degree of justice attributed to a trial is the result of a compromise which temporarily stabilizes the arrangement of the trial. A fair trial is a trial that more or less satisfies the concerned parties, but it is always susceptible to being changed to introduce a little more independence, or a little more expertise or field knowledge, or an increase in reproducible procedures.

Distribution and fairness

The distribution phase is the last stage where the justice of the trial represented by performance evaluation can be altered. As seen with Textirem, bonus thresholds are arrangements that are likely to induce unfairness perceptions, inasmuch as distribution is discontinuously related to contribution. This can be associated with the distributive dimension of organizational justice, which has been extensively studied by the “content” stream of relative research (Greenberg, 1987, p. 10). The trial model bears little new light on this distributive dimension of justice.

Nevertheless, three points need to be underlined. First, where distribution supposes, except where it is mechanically induced by an arithmetical formula, the existence of another judgment than that of performance, everything we have previously said about the different

contradictory forms of objectivity applies equally to this distribution stage. Second, observations of the Textirem case suggest that in Greenberg's terms (1987), the stakes of distributive justice are linked to the stakes of procedural justice. When distributive justice is under threat (as in the case of threshold effects) the demand for tightening of the trial, in other words the elimination of illegitimate strengths before the distributive phase, increases. Third, the greater the value the actor attributes to what is distributed (bonus, promotion, etc), the more he is likely to be critical regarding the justice of the trial. One can thus suggest that the demand for tightening of the trial increases proportionally to the stake for the person being tested, and thus that this demand varies depending on the individual.

4. Discussion and conclusion

The contribution of this analysis can be summarized in the following propositions:

1. Performance assessment systems may be considered as one of the major trials in the business world since they provide justification for the distribution of social goods, on the basis of institutionalized rules.
2. Because the legitimacy of such systems is crucial to the legitimacy and continuation of social order, fairness is a major requirement of these systems.
3. Performance assessment comprises several different stages. All of them are potentially concerned with unfairness, which can be expressed in terms of illegitimate strengths. These strengths can be related to some concepts usually associated with measurement and/or assessment (controllability, validity, objectivity), but in a different way (see points 4. and 5. below).

These illegitimate strengths can also be associated with the various "transformations" inherent to the performance assessment process. Indeed, each step of the process operates a transmutation from one "level of reality" to another. Thus instrumentation turns action (performance) into its representation (performance measures); then the representation is turned into a judgment, and finally the judgment is "translated" into an outcome. Illegitimate strengths can potentially arise during each transformation. For instance, illegitimate strengths filter through when performance measures are not sufficiently controllable or valid with respect to the performance represented (instrumentation step). Other illegitimate strengths arise when the judgment about performance is said to be "biased", that is viewed as an inadequate assessment of the "real" performance (judgment step). Finally illegitimate strengths can arise during the distribution stage if rewards are not in line with the judgment about performance.

The correspondence between the various steps of performance assessment, the concepts associated with unfairness (or illegitimate strengths), and the "transformations" during which these illegitimate strengths can potentially arise are summarized in Table 1.

Stages in performance assessment	Related concepts	Illegitimate strengths arising during the change of...
Instrumentation (definition of performance, measures and referents)	Controllability and validity of measures	Performance (action) into its representation
Measurement and Judgment of value	Objectivity of measures and judgment	Representation into judgment, or Performance (action) into judgment
Distribution of social goods	Objectivity of the decision of distribution	Judgment about performance into distribution

Table 1. A typology of fairness-related questions in performance assessment

4. Controllability and validity, i.e. qualities of measures that are mostly considered in a metrological perspective (is representation faithful to “reality”?) are also prominent concepts in a justice perspective. Not only are they conditions for the strategic effectiveness of the organization, they are also conditions for the legitimacy of the performance assessment systems, and, further, of the legitimacy of the social order. Gaps in legitimacy are as detrimental to organizations as are gaps in performance.

5. Objectivity is also a major concept related to the justice of performance assessment systems. Objectivity means independence, and/or expertise of the assessor, and/or reproducibility of the judgment. This enlarged conception of objectivity goes far beyond its usual association with quantitativity.

6. Criticism is an agent of change of performance assessment arrangements. The latter are not only changed under strategic pressure, i.e. when strategic objectives change; they also develop as an answer to criticism regarding fairness, as long as the level of criticism is high enough to put legitimacy and governability at stake.

7. Performance assessment is embedded in everyday life. It is fundamentally a review of past events, not a special event as seen by most HR literature and practices. As a result, the search for controllability and reduction of dysfunctional behavior appears endless.

This model offers a significant contribution to the understanding of performance assessment inasmuch as (i) it provides an alternative perspective to the usual managerialist one, and as it is (ii) an integrative and (iii) a dynamic model.

First, the model places fairness at the core of the instrument, while traditional models primarily focus on questions of measurement, in a managerialist perspective. Most performance assessment literature is centered on the instrument and the assessor, the acceptance of the system by the evaluatees being at best treated as a constraint. This model reverses the perspective and places the evaluatee at the core of the model. This reversal is likely to improve understanding of the not unusual dissatisfaction about performance assessment systems. It offers a frame for criticism, likely to enhance its critical ability to discover new sources of unfairness, and further to unveil them.

Second, it offers an integrative view of systems that are generally described and designed by specialists from different functions (management control and HRM), which enhances our understanding of their levels of coherence (or divergence). This integrative view highlights the fact that controllability, validity and objectivity, which are generally considered independently (by, respectively, management control, theory of measure and HRM), are interrelated conditions of fairness, and that improving one of them is generally to the detriment of another of these aspects. Awareness of this interdependence is beneficial to criticism, and likely to inform the type of claims made for tightening the trial.

Third, it provides a dynamic model of performance assessment systems, which is not the case of existing models. For example, Bourguignon (1998a, p.145 *et s.*) has developed a view according to which performance evaluation implements the control of a basic rule of collective action in organizations (individual behavior should be orientated towards strategic organizational objectives). This view integrates management control and HRM perspectives, it explains the relationship between the different stages in evaluation (cf. Figure 4.); however it is not a dynamic model that enables one to understand why and how the system changes. At best it implicitly suggests that the system changes along with the change in strategic orientations – which corresponds to the classical managerialist perspective. The trial model develops a positive view of criticism: criticism does contribute to change in systems, as soon as it is strong enough and precisely targeted towards elements that potentially impinge the legitimacy of social order.

Additionally this model offers a contribution to more general questions such as organizational justice or the legitimating function of accounting. Our analysis contributes to enlighten various aspects of procedural justice, especially on questions which have not been studied much until now. For instance existing studies do not generally consider in detail performance measures and their dynamics, most probably because such a focus is perceived as a “technical” accounting matter, that is beyond the usual scope of organizational justice studies.

This analysis also complements the existing corpus of research investigating the legitimating function of accounting (Richardson, 1987). Up to now most legitimacy-related research has focused on the macro-level – see for examples, Tinker *et al.* (1982) or Neimark (1992) about how accounting theories or practices support social order. Regarding the micro-level (organizations), accounting measures have frequently been regarded as a source of legitimacy, which for instance increase confidence in the rationality of the decisions made (Boland, Pondy, 1983; Meyer, 1986). The legitimating role of accounting has also been studied within the neo-institutional framework (Ansari, Euske, 1987). Although our analysis might be connected with these perspectives, it also considers the relationship between accounting and legitimacy in a somewhat different way. It is because accounting measures are used in trial sequences aimed at organizing a legitimate social order by selecting deserving individuals and those who are less deserving that accounting produces legitimacy.

Finally, this study provides a contribution to the question of change in management arrangements. This question is mostly investigated using the contingency framework (organization changes under the pressure of various external or internal factors), or the neo-institutional framework (organization change under the pressure of external stakeholders). Our analysis shows that arrangements also change under the pressure of criticism.

As far as we know, this analysis is the first attempt to apply the trial framework, that is, a very general sociological model, in the management field. As noted above, our empirical

observations are incomplete on a number of points, which means that the induced propositions might also suffer lacks, myopia and other limitations. It should then be of interest to conduct longitudinal empirical case studies to complete our proposals.

It should also be of interest to conduct analyses to test the validity of our application regarding other performance assessment systems than the individual-orientated ones. Indeed arrangements for evaluating performance are everywhere in the economic sphere, whether they relate to the evaluation of companies' performance, their products, their organizational sub-units, their processes, activities or projects. For instance financial accounting could be considered as a trial for external funding (stock market and financial partners, including banks); and management accounting, as the trial basing internal financial resource allocation. Such hypotheses might be worth investigating to gain increased understanding of trials in business life.

Future research could also fruitfully investigate the cultural dimension and contingency of our proposals. Fairness perception is not associated with the kind of meritocratic distribution discussed here in all cultures. In most African countries for example, the recruitment or promotion of an incompetent family member is not perceived as unfair, even by competent and meritocratic participants who have not benefited from the opportunity. The trial of "merit" is likely to be very different according to places and the local sources of legitimacy. Moreover the widest possible definition of justice in the work place can evolve with time, as shown by the changing of legitimate representations in France since the Second World War (Boltanski, Chiapello, 1999).

In a world where companies are probably one of the most important areas of production of social hierarchies, and where globalization increasingly extends modes of management inspired by the Western *Weltanschauung*, such questions are of vital importance.

Notes

¹ The vignette above is reproduced from the French daily paper *Les Echos* (20th November, 2001).

² Other factors have contributed to the transformation of sporting rules. As a result of the new status of sports in our societies (Ehrenberg, 1991), rules are also changed in order to improve the attractivity of the "show" provided by the contest.

³ "Ecole laïque, gratuite et obligatoire" according to the ideals of Jules Ferry, who reformed and generalized schooling at the end of the 19th century. This motto is still in daily use in France.

⁴ The original concept of Boltanski & Thévenot (1991) is termed *épreuve de grandeur* in French, which literally means "trial of greatness". It is indeed possible to imagine trials which are not meritocratic in our present acceptance of the term, like for instance ordeals which were associated with God's judgement. However for the trials studied in this article, one can use indifferently "trial of greatness" or "trial of merit". See Boltanski & Thévenot (1999) for an introduction in English.

⁵ We have translated the French *épreuve* into the English 'trial', as did most commonly Latour. We nonetheless feel that 'trial' has more the connotations of a trial of strength (which is actually the type of trial considered by Latour) while 'test' which has been sometimes used (Chiapello, 2001; Chiapello, Fairclough, 2002) rather points towards a trial of merit.

⁶ In Latour's terms, the winner "reduces", "translates" the loser; (s)he can now "speak in his/her name" and "represent" him/her (Latour, 1984).

⁷ A completely perfect trial is a logical impossibility, since this would assume establishment of a specific procedure for each individual situation (and for each individual person), which would mean that it would no longer be possible to judge on equivalence and establish a justifiable order. A perfectly fair world would presuppose a type of prior coding of each situation and a negotiation procedure to enable the protagonists to tend towards an agreement on the definition of the situation. This is both physically impossible (the time required for the negotiation being greater than the time devoted to the action) and

logically impossible (since it would also be necessary to define negotiation situations by means of negotiations, leading to an infinite regression). In addition, nothing would guarantee that the *ad hoc* coding thereby obtained would really be adequate for the situation, because in the absence of precedents and trial-and-error learning, individuals would be unable to pinpoint parasitical forces and further to improve the trial.

- ⁸ This wish to keep managers' risks in line with those of shareholders' also tends to impact remuneration policies, in the form of an inclusion of a large variable component of the remuneration depending on the profit achieved. Conversely the company distinguishing between both kinds of risk will probably favor remuneration with a larger fixed component.
- ⁹ The notion of relevance, abundantly referred to in management control, is the closest one to validity. Thus performance measures are considered relevant inasmuch as they are aligned with strategy.
- ¹⁰ This view of environmental effects however is flawed. First, changes in markets or exchange rates are the same but only for managers working either in the same market or in the same currency area. Moreover, in some circumstances, organizational factors can introduce unfairness between managers. For example, a bank agency's manager who has "inherited" a particularly costly structure (as a result of the history of the agency, the seniority of his/her staff, etc.) perceives as "unfair" the comparison with other agencies, on the basis of the overhead ratio, involved by benchmarking practices (Bourguignon, 1998b).
- ¹¹ There is an exception to this general rule: when assessment is made at two different moments, as in the case of satisfaction indicators, the qualitative measure "behaves" like a quantitative measure. The measure used by employees or customers to evaluate the quality of the company's social climate or service is clearly a qualitative one (a scale). This measure is then used later by another assessor (a hierarchical superior, the management controller, etc.) who carries out a more general assessment of the employee's or unit's performance, whose activities are reflected by the measure (satisfaction index). In this case, because the value judgment is based on and objectified by the measure, the qualitative criterion "behaves" like a quantitative one.

References

- Adams, J. S., "Toward an understanding of inequity", *Journal of Abnormal and Social psychology*, vol. 67, 1963, pp. 422-436.
- Adams, J. S., Jacobsen, R. P., "Effects of wage inequities on work quality", *Journal of Abnormal and Social Psychology*, vol. 69, 1964, pp. 19-25.
- Ansari, S., Euske K.J., "Rational, rationalizing, and reifying uses of accounting data in organizations", *Accounting, Organizations and Society*, vol. 12, n° 6, 1987, pp. 549-570.
- Baudelot, C., Estabiet, R., *L'école capitaliste en France* (Paris: Maspero, 1971).
- Bauer, M., Bertin-Mouro, B., *Vers un modèle européen de dirigeants ? Ou trois modèles contrastés de production de l'autorité légitime au sommet des grandes entreprises ? Comparaison Allemagne/France/Grande-Bretagne* (1996, Rapport de recherche CNRS [Observatoire des Dirigeants]-Boyden [Research Center on Leadership]).
- Bazerman, M. H., Beekun, R. I. & Schoorman, F. D., "Performance evaluation in a dynamic context: a laboratory study of the impact of a prior commitment to the ratee", *Journal of Applied Psychology*, vol. 67, n° 6, 1982, pp. 873-876.
- Bernardin, J. & Cascio, W., "Performance appraisal and the law", In Shuler, R. S., Youngblood, S. A. & Huber, V. (ed), *Readings in Personnel and Human Resource Management*, 3rd ed, pp. (St Paul: West, 1987).
- Besseyre des Horts, C. H., *L'appréciation comme pratique fondamentale de développement de l'équité en GRH* (Jouy-en Josas [France], 1991, Cahiers de recherche du Groupe HEC, n° 404/1991).
- Blakely, G. L., "The effects of performance ratings discrepancies on supervisors and subordinates", *Organization Behavior and Human Performance*, vol. 54, 1993, pp. 57-80.
- Boland, R.J., Pondy, R.J., "Accounting in organizations : a union of natural and rational perspectives", *Accounting Organizations and Society*, 1983, pp. 223-234.
- Boltanski, L., Chiapello, E., *Le nouvel esprit du capitalisme* (Paris: Gallimard, 1999).
- Boltanski, L., Thévenot, L., *De la justification*, (Paris: Gallimard, 1991).
- Boltanski, Thévenot, "The sociology of critical capacity", *European Journal of Social Theory*, vol 2, n°3, August, 1999, pp. 359-377.
- Bourdieu, P., Passeron, J.C., *La reproduction. Eléments pour une théorie du système d'enseignement*, (Paris: Minuit, 1970).
- Bourguignon, A., "L'évaluation de la performance : du rationnel au symbolique". In Actes du 4ème congrès de l'AGRH, *Universalité et contingence de la gestion des ressources humaines* (Jouy-en-Josas (France), Groupe HEC, vol. 2, 1993, pp. 19-33).
- Bourguignon, A., *La perception des critères d'évaluation de la performance : l'importance de la relation au travail et des performances perçues* (Université de Paris I, 1998a, Thèse de doctorat en Sciences de gestion,).
- Bourguignon, A., "Benchmarking: from intentions to perception" (Brussels, Proceedings of the EIASM Workshop on New Directions in Management Accounting: Innovations in Practice and Research, vol. 1, 1998, pp. 179-192).

- Carnap, R. *An introduction to the philosophy of science*, (New York: Basic Books, 1966).
- Carraher, S. M., Hart, D. E., Buckley, R. M. & Sawyer, J. E., *Organizational implications of employees attitudes towards benefits* (Las Vegas, communication presented at the Academy of Management Meeting, 1992).
- Chiapello, E., *The role of criticism in changing organisational models. A theoretical framework based on the French experience between 1968 and 1995* (Paper presented at the research seminar of the London School of Economics' Department of Accounting and Finance, 18 feb. 2001).
- Chiapello, E., Fairclough, N., "Understanding the new management ideology. A transdisciplinary contribution from Critical Discourse Analysis and New Sociology of Capitalism", *Discourse and Society*, vol. 13, n°2, 2002, pp. 185-208
- Citron, D.B., Taffler, R.J., "The audit report under going concern uncertainties: an empirical analysis", *Accounting and Business Research*, vol. 22, n°88, 1992, pp. 337-345.
- Covin, T. J., Kolenko, T. A., Sighetler, K. V. & Tudor, K. R., "Pay and organizational effectiveness: empirical evidence and implications for management consultants", *Academy of Management Best Papers Proceedings*, 1993, pp. 185-189.
- Cropanzano, R., Folger, R., "Procedural justice and worker motivation. In Steers, R. M., Porter, L. W. (eds)", *Motivation and Work Behavior*, (McGraw-Hill: New York, 1991, pp. 131-143).
- Dalton, G. W., "Motivation and control in Organizations. In Dalton Gene W., Paul R. Lawrence (eds)", *Motivation and Control in Organizations*, (Homewood, Ill.: Richard D Irvin and the Dorsey Press, 1971).
- DeNisi, A. S., *Cognitive approach to performance appraisal* (London: Routledge, 1996).
- Duarte, N. T., Goodson, J. R. & Klich, N. R., "Effects of dyadic quality and duration on performance appraisal", *Academy of Management Journal*, vol. 37, n° 3, 1994, pp. 499-521.
- Dulebohn, J. H., Ferris, G. R., "The role of influence tactics in perceptions of performance evaluations' fairness", *Academy of Management Review*, vol. 42, n° 3, 1999, pp. 288-303.
- Eccles, R. G., *The transfer pricing problem: a theory for practice* (New York: Lexington, 1985).
- Elias, N., Dunning, E., *Sport et civilisation. La violence maîtrisée* (Paris: Fayard, 1986).
- Ehrenberg, A. *Le culte de la performance* (Paris: Calmann-Lévy, 1991).
- Evrard, Y., Pras, B., Roux, E., Choffray, J. M. & Dussaix, A. M., *Market. Etudes et recherches en marketing. Fondements, Méthodes* (Paris: Nathan, 1993).
- Fisher, J., Frederickson, J. R. & Pepper, S. A., "The effect of information asymmetry on negotiated budgets: an empirical investigation", *Accounting, Organizations and Society*, Vol. 27, No. 2002, pp. 27-43.
- Ghosh, D., "Complementary arrangements of organizational factors and outcomes of negotiated transfer price", *Accounting, Organizations and Society*, vol. 25, 2000, pp. 661-682.
- Gordon, J. R., *Human resource management* (Boston (Mass.): Allyn & Bacon, 1986).

- Greenberg, J., "A taxonomy of organizational justice theories", *Academy of Management Review*, vol. 12, n° 1, 1987, pp. 9-22.
- Greenberg, J., "Organizational justice: yesterday, today and tomorrow", *Journal of Management*, vol. 16, n° 2, 1990, pp. 399-432.
- Hall, D. T. & Goodale, J. G., *Human resource management*, (Glenview (Ill.): Scott, Foresman and Co., 1986).
- Harder, J. W., "Play for pay: effects of inequity in a pay-for-performance context", *Administrative Science Quarterly*, vol. 37, 1992, pp. 321-335.
- Herzberg, F., "One more time: How do you motivate employees?" *Harvard Business Review*, vol. 46, n°1, 1968, pp. 53-62.
- Hirschman, A. O., *Exit, Voice, and Loyalty* (Cambridge: Harvard University Press, 1970).
- Hofstede, G. (1967). *The Game of Budget Control*, Assen, the Netherlands: Van Gorcum, (Traduction française : *Le contrôle budgétaire. Les règles du jeu*, Boulogne-Billancourt: Hommes et Techniques, 1969).
- Hogan, E. A., "Effects of prior expectations on performance ratings: a longitudinal study", *Academy of Management Journal*, vol. 30, n° 2, 1987, pp. 354-368.
- Hopwood, A. G., "An empirical study of the role of accounting data in performance evaluation", *Empirical Research in Accounting, Supplement to Journal of Accounting Research*, 1972, pp. 156-193.
- Kaplan, R. S. & Norton, D. P., *The balanced scorecard*, (Boston (Ma.): Harvard Business School Press, 1996).
- Landy, F. J., Barnes, J. L. & Murphy, K. R., "Correlates of perceived fairness and accuracy of performance evaluation", *Journal of Applied psychology*, vol. 63, 1978, pp. 751-754.
- Latour, B., *Irréductions, following Microbes*, (Paris: Métailié, 1984). *The pasteurization of France* (trans) (Cambridge: Harvard University Press 1988).
- Lee, T., Stone, M., "Competence and independence : the congenital twins of auditing ?", *Journal of Business Finance & Accounting*, vol. 22, n°8, December, 1995, pp. 1169-1177.
- Libby, T., "The influence of voice and explanation on performance in a participative budgeting setting", *Accounting, Organizations and Society*, vol. 24, 1999, pp. 125-137.
- Lindquist, T. M., "Fairness as an antecedent to participative budgeting: Examining the effects of distributive justice, procedural justice and referent cognitions on satisfaction and performance", *Journal of Management Accounting Research*, Vol. 7, No. 1995, pp. 122-147.
- McNair, C. J., Lynch, R. L. & Cross, K. L., "Do financial and non-financial measures have to agree?", *Management Accounting (US)*, vol. 72, n° 5, 1990, pp. 28-39.
- Megill, A., "Four senses of objectivity, Introduction to Megill a. ed, *Rethinking Objectivity*", *Duke University Press*, 1994, pp. 1-20.
- Merchant, K. A., *Control in business organizations* (Marshfield, Ma: Pitman Publishing, 1985).
- Merchant, K. A., "How and why firms disregard the controlability principle? In Bruns W. J. Jr., Kaplan R. S. (eds)", *Accounting and management: field study perspectives* (Boston: Harvard Business School Press, 1987, pp. 316-337).

- Merchant, K. A., *Modern management control systems. Text and Cases* (Upper Saddle River (N.J.): Prentice Hall, 1998).
- Meyer, J.W., “Social environments and organizational accounting”, *Accounting, Organizations and Society*, 1986, pp. 345-356.
- Milkovich, G. T. & Boudreau, J. W., *Human resource management*, (Homewood (Ill.): Irwin, 6th ed., 1991).
- Murphy, K. R., Balzer W. K., Lockhart, M. C. & Eisenman, E. J., “Effets of previous performance on evaluations of present performance”, *Journal of Applied Psychology*, vol. 70, n°1, 1985, pp. 72-84.
- Murphy, K. R., Cleveland, J. N., *Understanding performance appraisal*, (Thousand Oaks (Ca.): Sage Publications, 1995).
- Neimark, M., *The hidden dimension of annual reports* (New York: Markus Wiener Publishing Inc., 1992).
- Otley, D. T., “Budget use and managerial performance”, *Journal of Accounting Research* vol. 16, n°1, 1978, pp. 122-149.
- Piveteau, J., *L'entretien d'appréciation du personnel* (Paris: INSEP, 4e éd., 2000).
- Porter, T.M., *Trust in numbers, The pursuit of objectivity in Science and Public life* (Princeton: Princeton University Press, 1995).
- Richard, C., Reix, R., “Contribution à l’analyse de la qualité du processus d’audit : le rôle de la relation entre le directeur financier et le commissaire aux comptes”, *Comptabilité-Contrôle-Audit*, Tome 8, vol 1, mais 2002, pp. 151-174.
- Richardson, Alan J., “Accounting as a legitimating institution”, *Accounting, Organizations and Society*, vol 12, n°4, 1987, pp. 341-355.
- Roberts, I., “Remuneration and reward”, In Beardwell, I. & Holden, L. (eds), *Human resource management: a contemporary perspective*, pp. 499-556 (London: Pitman, 1994).
- Roget’s II, *The New Thesaurus*,. (www.bartleby.com, 3rd ed 1995).
- Sekiou, L., Blondin, L. & Peretti, J. M., *Passion ressources humaines*, (Montréal: Éditions 4L, 1995).
- Solomons, D., *Divisional Performance: Measurement and Control* (Homewood, Ill.: Richard D. Irwin, 1965).
- The American Heritage®, “Dictionary of the English Language”, (www.bartleby.com, 4th ed. 2000).
- Tinker, T., Merino, B.D. & Neimark, M.D., “The normative origins of positive theories: ideology and accounting thought”, *Accounting, Organizations and Society*, 1982, pp. 167-200.
- Tremblay, M., Sire, B. & Balkin, D. B., “ The role of organizational justice in pay and employee benefit satisfaction, and its effects on work attitudes”, *Group & Organization Management*, vol. 25, n°3, 2000, pp. 269-290.
- Vancil, R.F., *Decentralization: managerial ambiguity by design* (Homewood, Ill.: Richard Irwin and Dow Jones, 1979).
- Vigarello, G., *Une histoire culturelle du sport. Techniques d’hier et d’aujourd’hui* (Paris: Revue EPS-Robert Laffont, 1988).

Weber, M., *Economie et Société* (Paris: Plon, 1922 -French translation, 1971-).

Wentzel, K., “The influence of fairness perceptions and goal commitment on managers' performance in a budget setting”, *Behavioral Research in Accounting*, Vol. 14, No. 2002, pp. 247-271.

Williams Paul F., “ The legitimate concern with fairness”, *Accounting, Organizations and Society*, vol. 12, n°2, 1987, pp. 169-189.

