EconomEtica

Centro interuniversitario per l'etica economica
e la responsabilità sociale di impresa
promosso dalla Fondazione Italiana Accenture

N.22 August 2010

Lorenzo Sacconi

A Rawlsian View of CSR and the Game Theory of its Implementation (Part I): the Multistakeholder Model of Corporate Governance

Working papers

A Rawlsian View of CSR and the Game Theory of its Implementation (Part I): the Multistakeholder Model of Corporate Governance [1]

Lorenzo Sacconi

Department of Economics - University of Trento and EconomEtica, Inter university centre of research University Milano - Bicocca

1 Introduction

This is the first part of a comprehensive essay on the Rawlsian view of corporate social responsibility (in short CSR). CSR is defined as a multi-stakeholder model of corporate governance and objective function based on the extension of fiduciary duties toward all the firm's stakeholders (see sec. 2). A rationale for this idea is firstly given within the perspective of new-institutional economic theory in terms of transaction costs efficiency. From this perspective, abuse of authority in regard to the non-controlling stakeholders emerges as the main unsolved problem, and which makes it impossible to sever efficiency from equity within the domain of corporate governance (sec. 3). Intuitively, a Rawlsian principle of redress emerges as the natural answer to the legitimization problem of ownership and control rights allocations when, in order to provide incentive to one party (incentive to undertake important specific investments), they give it a disproportionate advantage over other non-controlling stakeholders.

Moreover, in accordance with the prevailing opinion about its voluntariness, CSR is viewed here as a model of corporate governance that companies may undertake by autonomous self-regulation in terms of the explicit adoption of expressed self-regulatory norms and standards. This is to be understood as an institution in Aoki's sense of the term: i.e. roughly put, as a rule in the behavior of a group of players which is maintained through the repeated plays of a given game, thanks to a system of mutually consistent beliefs by players predicting each

[1] to appear in *Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*, edited by Lorenzo Sacconi, Margaret Blair , Edward Freeman , Alessandro Vercelli, IEA series, Palgrave Macmillan, Basingstoke (in print)

other's behavior and that induces them to act again and again according to the same rule. Because such an institution is self-supporting, it does not need a statutory law to be enforced; but neither can it be seen as the gracious, arbitrary and occasional concession of management discretion. With respect to Aoki's definition of institution, however, proper understanding of CSR requires the addition of an explicitly expressed norm, including prescriptive principles and normative standards of behavior, which is to be accounted for in terms of the firm's stakeholders' *social contract* (see sec. 4).

The account of the social contract adopted here is Rawlsian. An impartial agreement is reached in an hypothetical original position by putting the parties 'under a veil of ignorance'. In our case, this is a matter of unanimous and impartial agreement among the corporate stakeholders that must be reached under a 'veil of ignorance' about the particular stakes that each of them holds (and with respect to any other personal traits). It takes place in the hypothetical bargaining that precedes the repeated non-cooperative game between the firm and each of its stakeholders. By this agreement, the principle of extended fiduciary duties and fair balance among different stakeholders is established as an explicit constraint on directors, managers, and in general on the party who controls discretionary decisions in the firm - a constraint that must prove to be effective throughout the repeated game between the firm and each of its stakeholders.

The bulk of this essay, in fact, is concerned with a game theoretical explanation of the roles played by explicitly expressed norms and standards in so far as they are based on the *stakeholders' impartial agreement* (the social contract). Put briefly, the social contract on an explicit CSR norm performs essential functions in solving the basic game theoretical problems faced in the implementation of the very broad idea of multi-stakeholder corporate governance (see sec. 5). These are:

- *construing* commitments to allow definition of a reputation game? such that reputation effects can be attached to compliance with the CSR normative model;
- *selecting* just one of the many equilibria possible in such a game as the unique equilibrium ex ante acceptable by all under the condition of impartial and impersonal agreement;
- *refining* the set of possible equilibria so that only those reflecting conformist motivations deriving from the ex ante social contract are retained as true candidates for the ex post emergence of the equilibrium to which actual individual actions will converge;
- and finally, to *predict* that the players' effective reasoning in the ex post implementation game will converge exactly to the equilibrium that would have been selected from the ex

ante perspective, so that the social contract proves to be essential also to the generation of a mutually consistent beliefs system supporting CSR as an equilibrium institution.

This part I of the essay focuses on the first role played by the social contract. First of all, the social contract works as a gap filling device with respect to the holes of incomplete contracts linking stakeholders (or the most essential of them) with the firm (sec.5). In a context of incompleteness of contracts and unforeseen contingencies, the repeated reputation game involving the firm (or those who control it) and each stakeholder would be badly specified because contingent strategies and commitment would be undefined with respect to unforeseen contingencies. Then the intention to accumulate reputation pursuant a strategy of stakeholders' fair treatment would be frustrated because there would be no standard of behavior whereby reputation could be assessed. Thus, at the outset of the stakeholders/firm interaction, a social contract must be established on a set of general and abstract principles of fair treatment, and precautionary (non contingent) standards of behavior, which can be adapted to unforeseen contingencies: that is to say capable of defining commitments neither meaningless nor void if unforeseen events occur. In the absence of such an explicit norm, no regularity of reputation-based behavior on the part of the firm could emerge through its interaction with stakeholders. In the presence of an unforeseen event, the only opportunity open to the party occupying the position of authority in the firm would be to take advantage of discretion. Abuse of authority would be the natural consequence. The ex ante social contract on a CSR norm is what enables completion of the *game form* of the reputation game involving the firm and its stakeholders through definition of the firm's types that carry out strategies with expected behavior in whatever state, even if unforeseen.

The further parts (part II, see Sacconi 2010b, *infra*, and part III, see Sacconi 2010c) of this essays illustrate other roles of a Rawlsian social contract over CSR norms. It may be useful to the reader to have here an overview on how the whole argument will be worked out. A Rawlsian social contract, as said, makes possible describing the game so that several types of reputations, based on the full or less than full respect of the CSR model may be developed even if unforeseen contingencies are involved (part I). But the Rawlsian social contract performs its main role in the second function discussed in part II of the essay: that is, the ex ante impartial selection of a unique equilibrium amongst the many possible in the repeated trust game involving the firms and its stakeholders. In this context it allows impartially selecting just *one* fair reputation equilibrium amongst the many possible. Elaborating on Binmore's *Natural Justice* (2005) (but see also Binmore, 1987, 1991, 1994 and 1998) and it

reevaluation of John Rawls egalitarian and maximin principle of justice within a game theoretical perspective, this task is accomplished again from the ex ante (under the 'veil of ignorance') point of view, but in a way that allows to find out a unique course of action that satisfies the requirement of incentive compatibility (i.e. a Nash equilibrium) (see part II *infra*). Further, an agreed CSR social norm aids reducing to *just two* the candidate reputation equilibria that ex post, in the real world interaction taking place beyond the "veil of ignorance", may be played after an agreement (maybe seen as cheap-talk and not-binding) over a general principle of fairness has been reached by the firm and its stakeholders (see part III Sacconi 2010 c, and Sacconi 2008). These equilibria are defined not as traditional Nash equilibria, but as psychological equilibria according to the theory of conformist preferences (Grimalda and Sacconi, 2005) developed along the lines of other behavioral game models (Geanakoplos, Pearce and Stacchetti, 1989; Rabin, 1993). It is argued that the behavioral model of conformist preference is nothing more than the development of Rawls' theory of the sense of justice, and hence is a constitutive part of a Rawlsian theory of CSR, able to include not just the theory of choice under veil of ignorance in the original position, but also the neglected theory of ex post social contract stability (Rawls, 1971; Sacconi and Faillo, 2010). Last, given the psychological equilibria that remain candidate as possible results of the game, the social contract admits to identify and to make credible the initial players' beliefs over the possible game solutions wherefrom an equilibrium selection dynamic (representing the revision process of mutual expectation) singles out the game solution effectively carried out (my favorite equilibrium selection dynamics is the Harsanyi's *tracing procedure* – see Harsanyi and Selten 1988). For a large array of situations, that are cognitively the most reliable in case the players have ex ante agreed on a social norm or standard (even if the agreement is not binding), the process selects an equilibrium corresponding to the normative model of multi-stakeholder fiduciary duties (see Sacconi 2008).

2 The definition of Corporate Social Responsibility (CSR) as an 'extended' corporate governance model

For many authors, corporate social responsibility is related to the stakeholder perspective in strategic management (Freeman 1984; Freeman and Evans, 1989). In light of a well-known classification by Donaldson and Preston (1995), it may be suggested that CSR is a concept that naturally fits the level of normative stakeholder theory (understood as a normative managerial theory). Taking the stakeholder theory seriously from a normative point of view,

i.e. from the point of view of the rights and legitimate claims of all company stakeholders, would imply that the company must be run in a 'socially responsible' manner. According to Freeman (Freeman, 1984; Freeman and Evans, 1989; Freeman and Ramakrishna Velamuri, 2006), however, 'social responsibility' is not the proper expression for normative strategic management within the stakeholder approach because it suggests a concern for 'society' which is collateral and not deeply integrated into the firm's proper economic nature and functioning. 'Stakeholder responsibility' would be the key concept, although many attempts to clarify what constitutes CSR could as well be considered ways to clarify the normative content of the stakeholder approach to strategic management of the modern corporation.

Nevertheless, even accepting that CSR essentially means corporate responsibility toward stakeholders, maintaining CSR only at the level of management (managerial values, methods, rules and practices) seems to be reductive (see also Trebilcock, 1993). Management works within the limits of some institutional corporate form, and under social norms concerning the firm's nature and the obligations. It is constrained, for example, by fiduciary duties and the institutional goals of the firm, and moreover by the exercise of residual control rights by owners (which may be more or less effective according to the company legal structure). I hence suggest moving up to the higher level of the firm's institutional form and its governance structure, which also involves the choice of the company's objective-function. Therefore, within the stakeholder approach, this essay will understand corporate social responsibility as the quality of an institutional form of the firm based on a norm (mainly an ethical norm, but which must nevertheless be complementary to the legal order) concerning its corporate governance and its objective function and - as a consequence - also its strategic management.

Let us therefore propose the following definition of CSR (see also Sacconi 2006a,b, 2009):

CSR is a model of extended corporate governance whereby those who run a firm (entrepreneurs, directors, managers) have responsibilities that range from fulfillment of fiduciary duties towards the owners to fulfillment of analogous – even if not identical - fiduciary duties towards all the firm's stakeholders.

Two terms must be defined in order for the foregoing proposition to be clearly understood.

a) Fiduciary duties. It is assumed that a subject has a legitimate interest but is unable to make the relevant decisions, in the sense that s/he does not know what goals to pursue, what alternative to choose, or how to deploy his/her resources in order to satisfy his/her interest.

S/he, the *trustor*, therefore delegates decisions to a *trustee* empowered to choose actions and goals. The trustee may then use the trustor's resources and select the appropriate course of action. For a fiduciary relationship – this being the basis of the trustee's authority *vis-à-vis* the trustor – to arise, the latter must possess a claim (right) towards the former. In other words, the trustee directs actions and uses the resources made over to him/her so that results are obtained that satisfy (to the best extent possible) the trustor's interests. These claims (that is, the trustor's *rights*) impose fiduciary duties on the agent who is entitled with authority (the trustee) which s/he is obliged to fulfill (Flannigan, 1989). The fiduciary relation applies in a wide variety of instances: tutor/minor and teacher/pupil relationships, and (in the corporate domain) the relationship between the board of a trust and its beneficiaries, or according to the predominant opinion, between the board of directors of a joint-stock company and its shareholders, and more generally between management and owners (if the latter do not run the enterprise themselves). The term 'fiduciary duty' therefore means the duty (or responsibility) of exercising authority for the good of those who have granted that authority and are therefore subject to it.

b) Stakeholders. This term denotes individuals or groups with a major stake in the running of the firm and that are able to influence it significantly (Freeman and McVea, 2002). A distinction should be drawn, however, between the following two categories.

b1) Stakeholders in the strict sense. Those who have an interest at stake because they have made specific investments in the firm (in the form of human capital, financial capital, social capital or trust, physical or environmental capital, or for the development of dedicated technologies, and so on). They are investments that may significantly increase the total value generated by the firm (net of the costs sustained for that purpose), and which are made specifically in relation to *that* firm (and not any other) so that their value is idiosyncratically related to the completion of the transactions carried out by or in relation to that firm. These stakeholders are reciprocally dependent on the firm because they influence its value but at the same time – given the specificity of their investment – largely depend on it to satisfy their own well-being (lock-in effect).

b2) Stakeholders in the broad sense. Those individuals or groups whose interest is involved because they *undergo* the 'external effects', positive or negative, of the firm's transactions, even if they do not directly participate in the transaction. Thus, they neither contribute to, nor directly receive value from, the firm.

It is now possible to appreciate the scope of CSR defined as an extended form of governance. It extends the concept of fiduciary duty from a mono-stakeholder setting (where the sole stakeholder with fiduciary duties is the owner of the firm) to a multi-stakeholder one in which the firm owes *all* its stakeholders fiduciary duties (the owners included). Classifying stakeholders on the basis of the nature of their relationship with the firm must thus be regarded as an important device with which to identify these further fiduciary duties.[1]

3 A 'transaction-costs-economics' rationale for extending fiduciary duties

This section argues that extending fiduciary duties follows naturally from a critical understanding of the new-institutional view of the firm (see also Sacconi 2000, 2006, 2007, 2009). The bulk of this theory is an answer to the question of 'why does the firm exist?'. It maintains that companies, and in general firms, are "unified governance structures" devoted to the reduction of transaction costs that would otherwise materialize due to the imperfection of contracts (Williamson, 1975, 1986; see also Hansmann, 1996). Specifically, three well-known sources of costs are specified:

(i) First of all, *contracts are incomplete in the sense that some relevant contingencies are unforeseen*, so that concrete and contingent provisos cannot be explicitly written or implicitly agreed with reference to such unforeseen events.

Contract incompleteness is sometimes tamed by a much less deep and troublesome understanding of the subject: for modelling convenience, non-verifiability by a third party (i.e. a form of information *asymmetry* to the disadvantage of the judge or the external arbiter) plus the parties' complete knowledge of what may unfold is substituted for unforeseen contingencies in the proper sense (see Hart, 1995; Grossman and Hart, 1986; Hart and Moore, 1990; Tirole, 1999; Maskin and Tirole, 1999). The result is that the cognitive and epistemological bases of contract incompleteness (*bounded rationality*) are swept under the carpet. On the contrary, it must be reasserted that the explanation rests on the empirically grounded assumption that the contracting parties are cognitively unable to represent, describe and forecast some possible states of the world, and that these states are *relevant* to their relationship, in the sense that the contract's outcomes and payoffs are not independent or separable in their definition from the states of affairs wherein they occur. At least sometimes,

unforeseen states shape the *meaning* of the outcomes that they obtain from the contract (for example, in terms of "good" or "bad" descriptions of such outcomes, and hence different preferences to the receiver).

(ii) After signature of a contract, *parties may carry out specific investments* which are also not contractible in any details: they may produce an unforeseen outcome, or their effects can materialize under unforeseen states of the world that cannot be ex ante described in such a concrete way that they are effectively includable in the contract through contingent provisos.

Specific investments change the contractual parties' relationship from one of indifference to one of strategic interdependence and bargaining over the surplus made possible by investments. In fact, what is typical of specific investments is that they increase (under some possible future state, not completely describable ex ante) the value of the transaction to the participant parties (to be precise, investments by a producer or a consumer, or both, may increase the value of the transacted item - a good, a service or whatsoever - to the consumer directly, and hence they increase also the potential value to the producer, in so far as he may claim a higher price or remuneration for contributing to provide it, and he is in fact needing, or preferring, higher remuneration if it is possible).

(iii) The parties' behavior under incomplete contract is to some relevant extent 'opportunistic': in a situation of contract incompleteness, they would try to renegotiate or change the terms of the contract or threaten - unless they are allotted a larger part of (or the entire) surplus - not to complete the transaction in the future if the profitable opportunity to do so appears.

Opportunism typically takes place when specific investments by some parties have already been carried out and an unforeseen state of the world materializes such that these investments have potentially important consequences on the transaction values, even though such values cannot be made available without some decision under the control of an agent (not necessarily the one who made the investment) whereby s/he may act opportunistically in order to extract as much rent as possible from control over this relevant decision variable.

To say that behaviors can be opportunistic is not to imply that people always behave opportunistically and that agents have no other motive to act in different situations. It is simply to say that, *ceteris paribus*, under incomplete contracts (and specifically in the absence of any other agreed *ethical norm* underlying the incomplete formal contract or any other *social convention* among participants (Lewis, 1969), with a surplus at stake as it is created by

specific investments, there is significantly positive probability of observing the onset of the typical selfish behavior called 'opportunistic renegotiation of an (incomplete) contract'. All together, these assumptions have important consequences as to the explanation of why the firm has emerged as an economic institution. Awareness of the possible renegotiation of incomplete contracts (which does not entail the prediction of concrete states of the world by the parties, but rather that they are *aware* of not being able to describe and foresee all possible future contingencies) induces the expectation that investments will be expropriated. This destroys incentives to make efficient investments, and hence a possible surplus value will not be created by intelligent prudent but cognitively limited agents (in the sense of their capacity to draw up complete contracts). Otherwise, if some party lacks even this basic degree of prudence, the instability of transactions generated by resentment at having been unfairly exploited will be observed. Note that the inefficiency effect of excepted opportunistic behaviors is closely bound up with the expectation by those making specific investments that they will be unfairly harmed. Harm is seen as deriving from *expropriation* of a fair share of the benefit to which they believe themselves entitled (whatever the holes in the contract) because of their contribution to the surplus's generation**.**

Against this background, the firm enters the scene as a unified *governance structure* able to alleviate the problem. Its institution, by giving ownership of physical assets to one party in the contract, also allocates to this party (and more in general to one stakeholder category among the many involved in a complex web of related transactions) the residual right of control, i.e. it gives that party the right to make discretionary choices on the *ex ante* non-contractible transaction variables. (For example, either the decision whether or not to carry out a specific investment or - once an investment has already been made - decisions essential for the investment to achieve its goal, which may affect the transaction value). Since these decisions may entail actions performed by individuals other than the right-holder, for a residual decision right to be effective it must entail *formal authority* over the firm, i.e. the owner's authority to see decision variables - residual with respect to those inserted in the written contract - carried out according to his/her will, independently of any specific agreement on the precise case in point and just because the right-holder 'says so'. Formal authority in fact provides those who undergo the authority relationship with preemptive reasons to act (Raz, 1999); reasons that (within the legitimate range of authority exercise) replace other reasons to act without any need to enter in balance with them. However (given that authority is not merely power exerted by means of a threat to use force and violence), it is not obvious how this could be so. The

explanation is that the preemptive nature of the authority's reasons to act results from some voluntary acceptance or *legitimization*. Thus, in order to enter into a formal authority relation, a party B must accept that another party A - who is in the authority position - makes decisions which are taken by B in general (within the range of legitimate A's authority) as the premise of B' deliberation process – i.e. neither executed for the convenience of the specific case in point, nor just because of the threat of punishment in case of non-compliance. This of course confronts the owner with the challenge of justifying (legitimating) the firm's authority structure, and explaining why a given residual right of control allocation should be accepted by those who will then be required to obey its exercise.

But before turning to this aspect, let us recall why the allocation of residual rights of control to a single party may be efficient. In essence, a party holding control over the non-contractible decision variables of the contract will be protected against the other parties' renegotiation threat, so that its investments are safeguarded against the other stakeholders' opportunism. This assurance of the party being able to benefit from its own investments is a sufficient reason to invest in some relevant aspect of transaction at an efficient level. Since the protection of specific investments enhances efficiency, this is the basis for a transaction costs efficiency explanation of the firm. If the specific investment of agent A is by far the most important in terms of specificity, A is the natural candidate for the allocation of ownership and control.

However, this is only a two-tier explanation of why the firm exists. In fact, even if this is an efficiency reason for the institution of a hierarchical relationship between the party making specific investments and any other party, it is not enough to cope with the fairness and distributive concerns that underlie the non-controlling stakeholders' decision to accept the authority of a party holding the right of control if also these stakeholders invest idiosyncratically.

Consider that only in very special cases can the firm be understood as a way to regulate transactions among stakeholders in a network wherein only one of them has an idiosyncratic relation with the transaction under consideration, whereas all others are indifferent about whatever transaction in which they may be involved. In general, the firm makes sense as 'team production', that is, as a team wherein many stakeholders cooperate by means of some joint and coordinated activity for the production of a joint surplus – which can be translated into the view of the firm as a productive coalition with a super-additive output function. Being part of the team or otherwise is not a matter of indifference to each potential team member.

An interesting result in the theory of the firm is the unification of team production with the new-institutional idea that specific investments are typical conditions for the emergence of the firm (see Blair and Stout, 1999 and 2006; Rajan and Zinagles, 1998 and 2000; but see also Aoky, 1984; Sacconi 1991, 1997 and 2000 for a previous formulation of a similar view). On this unified view, team production generates a surplus on each individual's production due to cooperation among the team members; but cooperation - and its joint output - arises from a joint activity made possible by their complementary specific investments, and especially by specific investments made at the moment of joining the team. Hence, the firm becomes a typical case of team production among many holders of specific investments (who are also stakeholders in the strict sense), with some other stakeholders potentially subject to the (negative or positive) externality deriving from it. Stakeholders in the strict sense are those who are materially in the position to make specific investments or, owing to their control over essential but non-contractible decisions, are themselves essential for the success of other stakeholders' investments. By way of example, consider employees, both highly qualified and otherwise, who develop and learn firm-specific skills, competencies and behavioral codes which make their productivity for a given firm higher than any others (and who may also be idiosyncratically related to a place where the team operated due to sunk costs already incurred to become productive in that location). Or stakeholders in the strict sense may be raw materials and instrumental goods providers or technology developers who sell materials, goods or equipment specifically devoted to a specific firm's production process (materials, goods or equipment that would not be provided by the general market). Or they may be capital goods investors who immobilize a large amount of money in the acquisition of complex equipment and technologies or employee training, all items with highly delayed returns on costs. Consider also consumers who invest time and effort in collecting information on goods and services that may be idiosyncratically tailored to their personal non-standardized preferences, and in developing trust relationships with sellers. They expect to profit in the future from this knowledge and social capital investment by being furnished with the idiosyncratic good or service on a trust basis, which prevents them from adding new information and search costs at any further purchase. All these investments attach surplus value to cooperation among stakeholders .

Note that team production is usually related to the idea of the firm as a *nexus of contracts* (Alchian and Demestz, 1972) with one actor (the owner) in the special position of a central contracting party with discretion over terminating any particular contract without terminating

the entire team's life. On the unified view, these contracts must be incomplete, so that the owner placed at the center of the nexus of contracts - *pace* Alchian and Demsetz - necessarily exercises authority over members of the team. In fact, s/he holds discretionary power over non-contractible decision variables essential for the possibility that each contracting party, after investing idiosyncratically in the team, may benefit from its participation.

But consider what is meant by having residual right of control and authority over decision variables that concern any stakeholder's relation with the team. According to the standard theory, the owner may terminate any stakeholder's relation with the team by excluding it from the physical assets if it does not perform the requisite actions and relinquishes any claim over the surplus. Actually, this may be an oversimplification of the reasons for a formal authority to be able to work. However, assume that formal authority annexed to ownership in one way or another entails that ex-ante non-contractible decisions are resolved in the owner's favour. These decision affect the surplus distribution generated by all specific investments. In brief, player A (the authority) will not allow player B (the non-controlling stakeholder) to benefit sufficiently from his/her investment to be able to repay its cost unless s/he accepts that A appropriates the surplus. Thus, the party holding residual control is in a position to claim the full surplus by expropriating other stakeholders' returns on investments.

Summing up, if fiduciary duties are only attached to ownership, while the non-controlling stakeholders are still left unprotected through incomplete contracts, then neither ownership nor contracts insure them against opportunism that will deprive them of any benefit deriving from their cooperation throughout the firm. Residual control, by affecting surplus appropriation, can then generate distribution schemes such that the surplus is entirely appropriated by the owner no matter what contribution other stakeholders have made to surplus generation – stakeholders which are left at the level where they barely cover investments costs. This is what I call 'abuse of authority'.

When stakeholders are sufficiently aware of such a prospect, they will prevent this risk by not entering the authority relation, so that the firm does not form even if 'team production' could be an efficient way to organize. Alternatively, once they have entered, stakeholders will under-invest in their specific contribution (note that standard theory assumes that residual control is relevant for decisions that affect the possibility for an investment to achieve its goal when the state of world is favorable, whereas the decision to invest as such remains up to any single stakeholder). This is why control structures are always second best: abuse of authority

induces some to over-invest, others to under-invest. Again a governance structure inefficiency is strictly connected with the expectation of unfair behaviour.

The threat of authority abuse does not forestall the need - just for incentive reasons - of giving residual control to the stakeholder responsible for the most important specific investment, granted that by assuming the governing role he does not incur governance costs so high as to dissipate the wealth created by efficient investing in the assets he holds. Nevertheless this should not prevent the non-controlling party from benefiting fairly form their specific investments and joint generation of surplus. Obvious here is a first reference to the Rawlsian maximin principle as the proper balancing criterion among different stakeholders claims. Owing to mere incentive reasons, those who are in the position to carry out the most important investment must be granted the opportunity to benefit form it by holding residual control, which in general will induce inequalities between them and other stakeholders to the advantage of the former. However, since the firm is a joint venture for mutual advantage, disadvantaged non-controlling stakeholders must also benefit from cooperation. This grants them the right to veto any control structure if it is not also the better one for the worst-off stakeholder with respect to all the available alternatives (including also the case that they take over control and the disadvantaged stakeholder position is taken by some other stakeholder). To legitimate a unilateral control structure, wherein ownership is held by the stakeholder undertaking the most important investment - which also gives him the opportunity to abuse non-controlling stakeholders - the implementation of a redress principle is necessarily required. This entails that also the non-controlling stakeholders can reach a position better than those possible under any other possible control structure arrangement. My suggestion is therefore to understand CSR as this Rawlsian governance structure.

When CSR is viewed as 'extended governance', it completes the firm as an institution for the governance of transactions (see Sacconi, 2000). The firm's legitimacy deficit (whatever category of stakeholders is placed in control of it) is remedied if the residual control right is accompanied by further fiduciary duties owed the subjects not controlling the firm and at risk of authority abuse. At the same time, this is a move towards greater social efficiency because it reduces the disincentives and social costs generated by the abuse of authority. From this perspective, 'extended governance' should comprise:

- the residual control right (ownership-based) allocated to the stakeholder with the largest investments at risk and with relatively low governance costs, as well as the right to delegate authority to professional directors and management;

- the fiduciary duties of those who effectively run the firm (directors and managers) towards the owners, given that these have delegated control to them;
- the fiduciary duties of those in a position of authority in the firm (the controlling owner and/or delegated directors and managers) towards the non-controlling stakeholders, that is
  - the obligation to run the firm in a manner such that these stakeholders are not deprived of their right to participate in the surplus distribution as it is cooperatively generated by their specific investments and their join actions – so that the company distributes to each *strict-sense-stakeholder* a 'fair share' of the surplus (acceptable by whatever stakeholder in an impartial agreement), while the broad-sense stakeholders are immunized against negative externalities;
  - the duty of effective accountability to the non-controlling stakeholders in terms of reporting relevant information in a veracious, transparent and understandable way about the accomplishing of tasks related to their legitimate interests and rights (as defined at the previous point),
  - and the right of these stakeholders to be represented in corporate bodies where they can exercise effective supervision over the owner's, directors' and managers' compliance with their fiduciary duties – as defined to the previous two points - owed to non-controlling stakeholders (for example representation through independent members of a supervisory body not appointed as representatives of shareholders but as advocates of the non-controlling shareholders' points of view).

According to this revision of the corporate governance structure, boards of directors or managers appointed by owners owe a *special* fiduciary duty to the 'residual claimants' who have directly delegated authority to them (*via* a narrow fiduciary proviso). This duty applies, however, only under the constraint that the more *general* fiduciary proviso relative to *all* the stakeholders is accomplished – which is specifically defined *via duties owed to non-controlling stakeholders*.

Moreover, the extended fiduciary duties model of corporate governance redefines the firm's objective-function (more about this in Sacconi 2006a,b,, 2009). This can be reconstructed by a three steps decision-rule which moves from the most general condition to the most specific one:

(i) Run any corporate activity in the way that minimizes negative externalities affecting stakeholders in the broad sense by preventing any corporate action from bringing about not repayable damages, such as those caused to the global environment, or compensating them in kind as they materialize, also before any legal suit for damages is started;

(ii) Identify the feasible set of agreements compatible with the maximization of the joint surplus and its simultaneous fair distribution, as established by the impartial cooperative agreement among the stakeholders in the strict sense (more on this in the Part II);

(iii) If more than one option is available in the above-defined feasible set, choose the one that maximizes the *residual* allocated to owners (for example, the shareholders).

The rest of this essay concentrates on an argument in favor of this extended governance structure and objective-function, taking seriously (at least from the abstract perspective of game theory) the challenge that any proposal for reform must prove to be implementable.

4 CSR as an 'equilibrium institution' based on the social contract of the firm.

A common tenet concerning CSR is that it should go beyond what can be required of companies by statutory laws and that it involves a certain degree of voluntarism and self-regulation. However, discretion is quite different from effective self-regulation, in that it does not entail any *rule* (either internal or external, enforced or self-enforced, legal or moral). Moreover, self-regulation may be understood in rather different ways: (i) as the case of an organism (the firm) endowed with its own 'natural' (so to speak 'unchosen') internal regularity of functioning, whereby its behavior is completely endogenously directed, without any need for interaction with other agents, either to agree on or at least to abide by any social norm at any time; or (ii) as the output of an agreement (explicit or implicit) among individual members of more or less extensive social groups - whereby they establish and adhere to an expressed (in language) set of principles or rules, with a normative content that they understand and which gives them guidance by vetoing some actions and recommending others, such a rule is *not enforced* by any external authority imposing sanctions because this is instead performed through the voluntary adherence of the individual members of the relevant

social group to the principles expressed (Posner, 2000). The self-regulatory nature of CSR is here understood in accordance with the second view. In particular, let us state the following definition of a CSR effective self-regulation (Clarkson, 1999; Sacconi, De Colle and Baldin, 2003; Wieland 2003):

a) CSR is established by social norms such as multi-stakeholder governance codes and management standards, not merely managerial discretionary decisions;

b) These include normative utterances: general abstract principles and preventive rules of behaviour concerning fiduciary duties, general statements of the fair treatment principle for each company stakeholder, principles of inter-stakeholder justice and fair balancing, precautionary rules of behaviour in any critical sphere of potentially opportunistic behaviour between the firm and some of its stakeholders - so that fiduciary duties and related rights are put in practice by standard precautionary rules of conduct that pre-empt opportunistic behaviour in typical critical situations;

c) Such norms are agreed upon by both firms and stakeholders through (voluntary) forms of multi-stakeholder social dialog (which simulates the idea of a 'small scale social contract' among them);

d) Nevertheless, these normative contents and standards of behaviour are self-imposed by firms on themselves without external legal enforcement, but instead by means of the internal adoption of statutes and codes of ethics reshaping the corporate governance and participatory structures, self-organization, training, auditing and control, which are compatible with voluntariness at the corporate level; and only on the basis of the consequences that non-conformity my induce for the stakeholders/firm interaction;

e) The previous self-enforcement approach does not prevent self-regulation from being monitored and verified by third-party independent civil society bodies (which do not have conflicts of interest with their mission of impartial overview over companies voluntarily subjected to self-regulation); this enhances the level of information and knowledge whereby stakeholders define their expectations about the firm's conduct. By contrast, this monitoring, verification and rating of conformity levels may be strictly necessary due to the typical information conditions wherein CSR social norms and standards are established.

Of course, effective CSR self-regulation is a viable option only within an institutional and legal environment that does not obstruct it. Such obstruction would occur in the case of too narrow definitions of the firm's objective-function such as that prescribing shareholder value maximization as the company's only goal – as today to be found in many company laws at international level.[2] If maximizing the joint stakeholder value conflicted even in the very short run with immediate shareholder value maximization, these laws would prevent the board from deciding to balance stakeholders' interests according to the social contract view, which implies a constrained maximization view (that is, constraining shareholder value maximization with the condition of the simultaneous maximization of other stakeholders' utility according to a bargaining solution) (for more on this, see Sacconi 2006a,b, 2009).

This is a good reason (in order properly to assess the implementation and stability of a CSR norm) to admit a sort of hypothetical 'state of nature' benchmarking into the assessment of institutions. It logically precedes historical legal constructs that without necessity may legally obstruct by design (or due to contingent historical equilibrium paths) the emergence of such a normative model. Thus, admitted that company laws do not obstruct proper self-regulation, the thrust of my argument is that the endogenous beliefs, motivations and preferences of economic agents (companies and stakeholders) are the essential forces driving the implementation of the CSR model of multi-stakeholder governance. If this is true, there will be a plenty of reasons - not only normative but also from the incentive compatibility and stability viewpoints - to promote reforms that enable companies to adopt governance structures, management systems and organization designs consistent with the CSR model.

Making sense of CSR as a self-regulatory explicit social norm requires a definition of *institution* different from simple consideration of existing formal-legal orderings. Here Aoki's *shared-beliefs cum equilibrium-summary-representation* view of institutions seems to furnish an essential part of the appropriate institution concept. According to this view, an institution is "a self-sustaining system of shared beliefs about a salient way in which the game is repeatedly played" which is a rule not in the sense of "rules exogenously given by the polity, culture or a meta-game", but in the alternative sense of "rules as being endogenously created through the strategic interaction of agents, held in the minds of agents and thus self-sustaining - as the equilibrium-of-the-game theorist do. In order for beliefs to be shared by agents in a self-sustaining manner (….) and regarded by them as relevant (…) the content of the shared beliefs" must be "a *summary representation (compressed information)* of an equilibrium of the game (out of the many that are theoretically possible). That is to say a salient feature of

an equilibrium may be tacitly recognized by agent or have corresponding symbolic representation inside the minds of agents and coordinate their beliefs" (Aoki, 2001, p.11)

The self-enforceability condition of Nash equilibria is implicit in the above definition. A *compressed summary representation* of information about the way a game has been repeatedly and regularly played is not a complete description of all the histories of the repeated game under any contingency. Nevertheless, it is a summarizing pattern (a model resident within the players' minds, i.e. a *mental model*) containing salient features of the players' equilibrium action profile that has been played in the game so far and which are sufficient to define reciprocal expectations and beliefs concerning each other's actions from now on. Given this mental compressed representation, boundedly rational players – without complete information - derive beliefs about how any other player currently plays the repeated game. And these beliefs are *shared* - in the sense that any two players make the same prediction about any other player involved - and consistent – in the sense that beliefs whereby any player derives his choice also cohere with his prediction of beliefs whereby other players derive their choices. These beliefs replicate the prediction that a particular equilibrium will be played among the many possible, and it is from such beliefs that all players derive their best actions. Because these actions are best against beliefs, and these beliefs correctly summarize current behaviors, these actions are also the best responses to the other players' actual actions as these are represented by beliefs. Then the derived action profile satisfies the typical Nash equilibrium condition.

This clarifies why the belief system is *self-sustaining*. The resulting equilibrium profile, as it is generated by best responses to beliefs, also replicates the same behavior that the compressed information summary in fact represents - i.e. it exhibits the same salient characteristics as summarized in that compressed information representation. Hence, it cannot but replicate the same summarized information on how the game is played, and hence support the same beliefs system.

The *beliefs /compressed information summary representation* pair is an institution *not* in the sense of a 'rule of the game' exogenously imposed on the players' choices by some physical or technological feature of the environment, or by any further external institution or authority. These rules are useful to define the *game form*, that is, the objective set of constraints and opportunities within which the game is played. But the *beliefs /compressed information summary representation* pair instead defines an institution as the endogenous rule of behavior emerging from how the game is played. In fact, *given* the game form, the beliefs system

describes a regularity of behavior resulting form the players' choices that they represent in their minds and replicate in response to that representation. Thus the *belief system* replicates itself *endogenously*.

An important consequence of Aoki's view is the following. A statutory law passed by a parliament or another legislative body, even though it may explicitly settle rights and duties, if there is *no* shared belief that it will be complied with by those who 'should', it is not to be considered an *institution*. Instead, the ongoing practice of violating the statutory law could be considered the 'true' institution of the relevant action domain (Aoki, 2001).

Nevertheless, at first glance, this definition has a major drawback. Institutions thus defined seem to be devoid of any significant normative meaning and force. On the contrary, institutions like constitutions or laws, ethical codes, shared social values, organizational codes of conduct and procedures have primarily a prescriptive meaning (in the case of ethics such meaning requires "universalizability" (Hare, 1981)) - i.e. they are action guides and not just description of state of affairs. They tell agents what must not be done or what is to be done in different circumstances. Institutions in the above game-theoretical definition may seem to give an indication about the best action of each player only *ex post* - that is, once the participants have chosen their actions and have shared knowledge that they have already reached an equilibrium state in their choices. The institution (beliefs system and the relative compressed information representation) tells players only to maintain the existing pattern of behavior because it is an equilibrium supporting the existing beliefs system. An institution such as this seems to have no normative content. It is based on a summary of how the game has been played in the past and consists of a set of mutually consistent predictions of how the game is currently being played and will be played in the future.

But why then would institutions be as they are? Why would they contain principles and norms (moral, legal, social or organizational) explicitly formulated in sentences through utterances whose meaning is not mainly a *description* of how people normally act (even though they can also contain descriptions) but a prescription of how they *must* or *must not* behave. There is no reason why what the addressee *must* do according to a norm corresponds to what - before the utterance of these prescriptive sentence - s/he *de facto* does. A norm (as a component of an institution) is not *falsified* by the observation that people do not conform to it, even though it can be thus recognized as ineffective (and discarded as an institution in the proper sense). The point is that a necessary component of the belief system defining an institutions must not merely replicate the description of behavior in a given action domain; it must instead

prescribe it independently of the description of the ongoing course of action. In other word, it rests on some *a priori* standpoint. Arguably, this is a *necessary* though not sufficient condition for an institution to exist (for sufficiency, the beliefs equilibrium definition must be met).

Moreover, a norm is sometimes explicitly introduced in order to change the received behavior and to set up an institutions that can regulate a given domain of actions. It thus provides guidance for action choices in the given domain when the players' summary compressed representation of information about how they have acted cannot replicate the required change. Because it is a theory of institutional change, Aoki's theory provides an answer to this question. The problem under consideration is twofold:

(i) the problem of equilibrium selection within a given game form, where an old equilibrium path (old institution) has been abandoned for whatever reason and a new equilibrium path (new institution) has to be reached by all the players even though it has not yet been stabilized among them; and secondly

(ii) the problem of achieving such a new equilibrium actions profile supported by a stable and shared beliefs system (a new institution), when the underlying action domain changes because environmental or technological changes have been introduced, or some further action opportunity is simply discovered by players and represented for the first time in their subjective mental model of the game.

To these distinct but interlocked questions Aoki gives an answer based on the idea of 'salience' of some game feature, which is not understood as mere description of a characteristic. That is to say, it is not confined to the condition that players' beliefs contain the description of a salient characteristic of how they have acted in the past and that they transfer into a prediction of how they will act in the future. Here, the genuine *guidance* function of a normative beliefs system emerges. And it is part of the explanation of why that beliefs system is widely accepted by every participant in the action domain, so that it is recognized as 'salient' or 'prominent' – i.e. so that everybody knows that also others accept it and use it to assess each other's behavior. It thus gives players reasons to coordinate (so to speak 'for the first time') on a specific equilibrium profile *inter alia*, given that many are possible, also in cases when the domain of action changes or is enriched by new opportunities.

"The point is that some symbolic system of predictive/*normative* beliefs [emphasis added] precedes the evolution of a new equilibrium and then becomes accepted by all the agents in

the relevant domain through their experiences. It could be 'unsettled culture or ideologies – explicit articulated highly organized meaning systems – that may establish new styles or strategies of actions (…), 'an entrepreneur's vision that may trigger certain action that eventually remove the limits of organizational capabilities and environmental constraints' (…) or even the political program of a subversive political party (…) bounded rational individual agents form their own subjective models of the game that they play" (…) so that the mechanism of institutional change is seen "a process of revision, refinement and inducement if mutual consistency of such model incorporating a (common) representation system" (ibid. p. 19)

These examples of symbolic systems of normative and predictive beliefs are introduced as possible empirical explanations of how an equilibrium may become focal before it is stabilized by customary behaviors and beliefs. Clearly, however, this view presumes that these beliefs exercise a *justificatory force* able to induce the *general acceptance* of a new equilibrium in a given domain, so that - but only later on - it becomes the 'salient' basis for reciprocal prediction of all of the participants' actions.

Thus, a second component of a proper definition of institution – integrating Aoki's definition - is the mental representation of a norm, necessarily expressed by utterances in the players' language (oral, written or simply mentally represented) concerning rights and duties, values and obligations, which needs to have a prescriptive and universalizable meaning able to *justify* its shared acceptance by all participants in a given interaction domain. Because it is *ex ante* accepted by all players, it enters their shared mental model (Dezau and North, 1994) of how the game should be played and hence becomes the basis for their coordination on a specific equilibrium under a given action domain. The key point is then explaining *how* a normative system of beliefs, preceding the evolution of the corresponding equilibrium, *becomes accepted by all agents* in the relevant domain. And to be useful for the purposes of this essay, this explanation should make sense of a CSR norm accepted by all the corporate stakeholders and those in the position of authority in the firm.

To my knowledge, the best justificatory account for norms on responsible exercise of authority, entailing ex ante shared acceptance, is the *social contract model*. Contractarian norms result from a voluntary agreement in an hypothetical original choice situation which logically comes before any exogenous institution is over-imposed on a given action domain, or before any institution (in the equilibrium sense) has yet emerged. Thus a norm (and the institution that may encapsulate it) arises and can be maintained only because of the voluntary

agreement and adhesion of agents. To define the agreement on a justifiable norm, any social contract model sets aside threats, fraud and manipulation resources that would render the parties substantially unequal in terms of bargaining power. Besides the normative reason for doing so, such initial conditions would need an explanation in terms of a previously reached equilibrium in a game of threats played in the relevant domain, or it would be seen as the effect of institutions already existing in some adjacent domain that give some players more strength than others. The hypothetical choice under the original position proceeds as if these contingencies were arbitrary and irrelevant to the proper calculation of the social contract.

The idea of a 'fair agreement' thus becomes intuitive: the agreement must reflect only each participant's rational autonomy, decision-making freedom and intentionality, which are assumed to be *equal* in weight among the participants in the contract. (This can be disputed on an empirical basis, but *in principle* the idea is to skip any morally irrelevant difference among participants). The agreement thus gives equal consideration and respect – i.e. equal treatment - to reasons, interests and decisions put forward by each participant in the contract, because a voluntary and unanimous agreement among autonomous choosers necessarily equally reflects the reasons to enter the agreement by each and all of them.

It is not only the initial creation of norms and institutions that is seen by the social contract model as a matter of unanimous agreement among autonomous agents. Also their implementation is understood as being a matter of voluntary adhesion. Thus the endogeneity of institutions with respect to the agents' strategic interaction is respected at both stages: an institution is endogenous to the *ex ante* players' strategic interaction understood as *rational bargaining* among equally situated rational agents, i.e. it can be started only by the unanimous individual players' decision to enter a voluntary agreement. Moreover, the *ex post* implementation of an institutional arrangement is also seen as the composition of the autonomous decisions that players make in their strategic interaction, whereby they chose whether or not to comply with the social contract by carrying out decisions that reflect the whole set of their reasons and motives to act.

In order to accomplish these tasks, the social contract model must operate in two different but necessarily related directions. Entering *ex ante* and adhering *ex post* to the agreement on principles and norms for institutions are distinct decision problems, with quite different logics of choice, but which nevertheless must be solved in a mutually consistent way and within a unified view. The choice of entering the contract must provide a justification for norms and institutions. The form of this justification is the impartial rational agreement of all the

concerned stakeholders. It is appropriate here to give weight only to considerations relevant to the rational decision to enter an impartial agreement, which is provisionally assumed to be possible since all the parties involved are hypothetically assumed to voluntarily participate in a *thought experiment*. Hence preventing cheating and defection is not the focus of the decision logic employed to calculate the agreement, even though these considerations may be essential in defining the feasible outcome set from which the agreement should be selected. What is relevant here is the opportunity offered by an unanimous agreement to improve to mutual advantage the state of affairs with respect to the "state of nature" that would result from cooperation failure. Moreover, such a mutual improvement and advantage must itself be recognized as acceptable by equally autonomous, free and rational participants in the bargain – so that it must not only be *mutual* in the sense that whatever improvement one party gains over the state of nature status quo necessarily corresponds to *some* improvement in another's. In addition, it must also treat participants symmetrically, so that they can accept such an agreement proposal of mutual advantage form an impartial standpoint.

Quite different is the decision logic of the *compliance problem*. When we move from the *ex ante* to the *ex post* perspective, we ask whether an agreement reached can also be complied with by the same players who agreed on it. This is a different problem because the game-logic of compliance differs from that of entering a cooperative agreement. It is instead the logic of an *ex post* non-cooperative game in which the players decide separately but interdependently whether or not to comply with the *ex ante* agreed contract. From this perspective, the question is not so much whether the contract provides reasonably high joint benefits and distributes them in an acceptably fair way; rather, the question is mainly whether there are incentives for cheating on the counterparty to the agreement, given the expectation that s/he will abide by the contract.

Social contract models convincingly answer the ex ante decision problem, but are typically at odds with the compliance problem. This difficulty also applies to the most elaborate social contract theories that have made significant steps toward a unified view of both aspects. (See Rawls (1971) and Gauthier (1986). Binmore also provides a unified view of the two problem according to the social contract model (see extensively Part II of this essay). On the other hand, Aoki's institution definition guarantees that, if the agreed norm is represented within the players' minds by summary information about a "salient" equilibrium profile and thus generates a system of predictive and normative beliefs, then also the compliance problem is amenable to solution, since it will satisfy the equilibrium condition. Thus, taking jointly the

two requirements - (i) acceptability of the normative content of an institution through a social contract, and (ii) a shared belief system based on the compressed representation summary of an equilibrium - seems to provide the comprehensive definition of institution needed here.

There are many different accounts of the social contract model. For example, both Rawls' and Gauthier' accounts are compatible with what has been said thus far. However, Rawls's idea of the original position is basic to the purpose of this essay. It is a choice condition requiring unanimous agreement under a 'veil of ignorance' concerning any detail of each participant's personal identity and social position. To be clear, I mean by a 'veil of ignorance' radical uncertainty about the mappings that would identify each participant in the original position with a particular set of personal attributes such as strategies and payoffs that would represent his personal characteristics and social position under different contingencies. The veil of ignorance creates an impersonal and impartial standpoint whereby an agreement is unanimously workable because each participant's separate standpoint becomes identical with that of all the others. In other words, behind the 'veil of ignorance' each individual is ready to take symmetrically the position of any other and to replace his/her initial personal standpoint with that of everybody else. Under these symmetrical exchanges of position, whereby everyone assesses acceptance of any given set of normative statements, they reach an agreement that reflects a reasonable impartial combination of all the reasons to act that they consider in turn. Importantly, the agreement accepted by each of them cannot but be unanimous, for the symmetrical replacement of personal positions is carried out in identical ways by all the involved parties, so that they are identically situated in their exercise of institutional assessment.

Thus, it is the agreement under the veil of ignorance among all the corporate stakeholders that should generate the shared acceptance of CSR as a social norm corresponding to a particular equilibrium among the many possible. Since it is a "thought experiment", it would impress the players' minds with a mental model of how the game should be played and generate an identical 'salient' aspect of their interaction that would favor effective coordination over a specific equilibrium point to be played by the choice of each actions. When the shared system of mutually consistent beliefs has been formed for the first time, it will allow for mutual predictions and the generation of an equilibrium that also confirms the same beliefs set. The summary information compressed into a mental representation of the regular players' behavior throughout the repetition of the game, generated by ex ante acceptance of the normative beliefs that a particular equilibrium is to be played, can then be understood as an

institution. Now argued is that CSR is the social norm in the corporate governance domain that satisfies this definition.

A social contract explanation is a *zero–level* explanation which in fact assumes as its starting point the "state of nature" hypothesis. It is more fundamental than, and prior to, any consideration of complementarities between a CSR model of corporate governance and institutions belonging to different domains. And it also logically precedes any assessment of how institutional changes in other domains – such as labor law, the industrial relation system, or in general the political system - may ease the introduction of CSR. In fact, assume that a social contract among all the company stakeholders induces them to build CSR as an institution which is not only impartially acceptable to stakeholders but also self-sustainable - admitted that it is neither obstructed by prohibitions in the legal system nor incentivized by other institutions or regulations. Such a normative model is the natural candidate for a legal reform of statutory company laws and corporate governance regulations because it has already proved to have endogenous forces of its own pushing toward its institution. [3]

5 The four roles of a social contract on CSR norms

To understand why the stakeholders' social contract on a CSR norm explicitly stated through utterances in normative language is so essential for the endogeneity and self-sustainability of the corresponding behavior and expectations (e.g. an institution in Aoki's sense), we must consider the *roles* performed by voluntarily agreed explicit norms. But let us first model the relationships between the firm and each of its stakeholders as a case of the well-known *trust game* (TG) – a formal context wherein these roles can be better situated (see fig. 5.1) (Fudenberg and Levine, 1989; Fudenberg and Tirole, 1991). A stakeholder A may or may not enter into a specific relationship with the firm. The firm is here identified with the particular stakeholder B who owns its physical assets and hence exercises control on some discretionary decision variables that affect the mutual opportunity to profit from the stakeholder's A (and maybe his/her own) specific investment and cooperative decision to enter the relationship. Hence, in the trust game, what stakeholder A may or may not enter is a fiduciary relation with those in a position of control (synthetically called 'the firm'). By entering, it is assumed that the stakeholder A makes a specific investment that renders his/her relationship with the firm idiosyncratic, but also makes possible a surplus deriving form this relationship. On the other hand, the position of the firm's owner in the game makes explicit the possibility that s/he may

abuse his/her authority toward the non-controlling stakeholder. The owner may or may not abuse the stakeholder's trust. In the case of abuse, the owner appropriates all the surplus generated by specific investments and gets 3, leaving the stakeholder with only the cost of its investment (-1). If the owner does not abuse, there is a mutually beneficial sharing of the surplus for both the players (2, 2) that reflects their joint contribution to 'team production'. As well known, this game has a single Nash equilibrium, the Pareto-inefficient outcome corresponding to the payoffs vector (0, 0). Since the firm B will necessarily abuse ('abuse' is its dominant strategy), the stakeholder A will not enter.

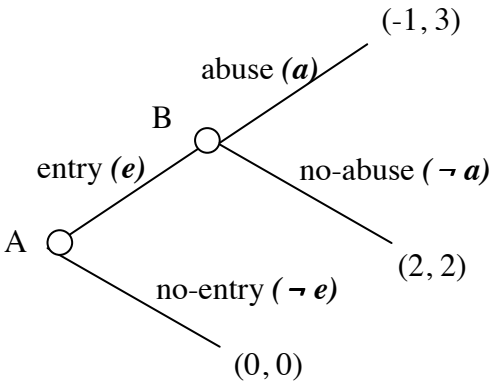

Fig. 1 One shot Trust Game in extensive form

But matters may substantially change if the TG is infinitely repeated between a single long-run player B, in the institutional role of the firm, and an infinite series of short-run stakeholders seen as players $A_1,\ldots,A_n$ (where n goes to infinity). At each stage game (repetition) a player in the role of A_i has a short-run strategy choice at hand: whether or not to enter, given the consideration of the previous story of how the game has been played until the stage where s/he is required to make his/her decision. On the other side, the long-run B player has to make a choice among long-run strategies which at each repetition select a concrete action (abuse, not abuse or a random mechanism to mix the two probabilistically) as a function of the story of the game until each possible stage. Note that because B chooses at each stage, a long-run player's strategy is a rule for making such selection at each stage given any story of the game at whatever stage. Thus, a long-run strategy considered as a whole accounts for every possible story of whatever length according to which the game might have been played at each stage. As a consequence, each mono-periodical short-run stakeholder A_i (for whatever value of i) has a payoff function defined on the outcome of the specific stage at which s/he participate in the game. Otherwise the long-run player B's payoff function is the

infinite summation of each payoff s/he gets at any stage multiplied by a discount factor δ ($0 \leq \delta \leq 1$) reflecting player B's impatience or short-sightedness. Under convenient conditions, such a payoff is the *limit of the mean* payoff associated with the loop of whatever length (going to infinity) into which player A's strategy enters again and again along its repetition, given the short-run players' strategy choices (i.e. loops generating identical series of stage game payoffs). Let us assume that the discount factor δ is not 'too small' with respect to the ratio between (i) how much player B in a single case forgoes by not abusing player A_i instead of taking the opportunity to exploit him/her, and (ii) how much s/he forgoes at each successive stage by receiving the payoff associated with non-entrance by player A_i instead of the payoff of mutual cooperation.

The game is qualified as 'incomplete information game' in a distinct sense. Short-run players A_i are *uncertain* about player B's rationality (i.e. criterion of choice) so that they take as possible different player B types, where types identify the long-run strategies played by B. This is to be understood in the sense that players $A_1,\ldots, A_n$ take it for granted that player B is irrevocably committed or disposed to play some specific behavior rule - which consists of a specific repeated strategy - but is also uncertain about what among the many possible such commitments is. Thus player B is deemed to be a not completely strategically rational agent because s/he would stick to a rule of behavior independently of player A's choice. This is only the way that players $A_1,\ldots A_n$ think about the game, however. Indeed, player B is nevertheless completely strategically rational and informed, so that s/he will decide his/her strategy without any sense of absolute commitment, and only on the basis of his/her best prediction of strategy choice by players $A_1,\ldots,A_n$. This in turn is based on his/her understanding of how the short-run players' beliefs change from one repetition of the game to the next.

Player B's reputations are the probabilities attached by players A_i at each stage to B's types, whereas types are stereotyped commitments on player B's rules of play (strategies). Changes in reputations are a function of the repeated observation of how stages games? have been played by B, and of the stage game outcomes and their comparison with what a given commitment would have entailed (contingently on also the behavior of players A_i). Each player A_i is assumed to update by means of the Bayes rule the initial probabilistic beliefs shared by all players A_i concerning player B's types. Repeated observations of 'not abuse' will augment the ex post probability of any B's strategy (pure or mixed) that does not abuse at all or abuses very slightly. Whereas such observations will falsify the hypothesis that player B

is the abusive type, or they will reduce the probability of any significantly abusive B's mixed type. Player B supports his/her reputation of being a given type by continuing to play stage game moves which are consistent with the type.

Under these not innocuous assumptions it is well known that a whole set of new equilibria becomes possible in the repeated trust game. In particular this set of equilibria (consisting of repeated short-run strategies chosen by players $A_1,\ldots,A_n$ paired with a long-run player B's strategy) is bounded from above by the equilibrium wherein player B plays his Stackelberg strategy, and from below by the equilibrium in which no player in the role of A_1 enters throughout the game repetition. (Fudenberg and Levine, 1986; see also Fudenberg and Tirole, 1991). It is important for understanding how spontaneous cooperation can arise between the firm and its stakeholder that if only pure strategies are considered, then a repeated B's decision not to abuse will eventually induce entrance by every short-run player A_i (after some periods spent on accumulating reputation). If the discount factor is not too low, continuing to play no abuse is also player B's best response, so that repeated non-abuse and substantial entrance by players A_i will be an equilibrium of the game. This is the typical 'good reputation' equilibrium which is typically advocated by those who are 'optimistic' about spontaneous cooperation between the firm and its stakeholder.

Against the background of this concise representation of the stakeholder/firm interaction, we may understand the *four roles* of a social contract on a CSR norm expressing player B's fiduciary obligation not to abuse player A's trust.

- The *cognitive-constructive role*, which answers the question about *how* the firm *works out* the *set* of commitments that it *can* undertake with respect to generic states of the world that it is aware of not being able to predict in any detail, and therefore *what* types of *possible* equilibrium behavior the firm can work out so that stakeholders may entertain expectations about them;

- The *normative role*, which answers the question about what (if any) pattern of interaction the firm and its stakeholders must a priori *select* from the set of possible equilibria to be carried out *ex post* (according to the answer given to question a), if they adopt an *ex ante* standpoint ('under the veil of ignorance') enabling an agreement to be reached from an impartial point of view;

- The *motivational role*, which answers the question about *what* and *how many* equilibrium patterns of behaviors, amongst those that may emerge ex post from the interaction

between firm and stakeholder, would retain *their motivational force* if firm and stakeholder were able to agree in an ex ante perspective on a CSR norm along the lines of question (b);

- The *cognitive-predictive role* concerning how the ex ante agreement on a CSR norm *affects* the beliefs formation process, whereby a firm and its stakeholders cognitively converge on a system of mutually consistent expectations such that they reciprocally predict from each other the execution of a given equilibrium in their *ex post* interaction (given that more than one equilibrium point still retains motivational force). The question to be answered by this function is 'does the norm shape the expectation formation process so that in the end it will coincide with what the *ex ante* agreed principle would require of firm and stakeholders?'

6 The cognitive /constructive role of the social contract

The second role is t*he focu*s of the part II of this essay, where the main contribution of the Rawlsian view is discussed (see Sacconi 2010, *infra*). I have discussed at length the first role elsewhere (Sacconi 2000, 2006a, 2007b, 2008), so here I may briefly summarize the main argument with reference to the repeated trust game.

To enable the reputation cumulative process, the firm should commit to a strategy carried out with specific unambiguous and verifiable actions at each stage game according to a conditional rule. *The* stage game choice induced by a strategy is specified with respect to every possible story of the game, that is with respect to all the possible state of the world wherein the game has been played till the current stage, for whatever stage. This means that, given a player B's strategy, every player A_i at any stage t is capable to predict how player B will play at any stage (given any previous possible story).

Consider, however, that modeling the firm like this entails assuming a context of incomplete contracts, which we interpret in its genuine nature as the existence of unforeseen and unforeseeable states of the world (Kreps, 1992). Complete contracts between two parties would be agreements on pairs of contingent strategies, one for each party. In our case these would at least make it possible to say how the firm will act in whatever state of the world that may unfold through all the game repetitions. With contract incompleteness, by contrast, some states of the world are unforeseen. Hence it is impossible ex ante to define how any *contingent* strategy will behave when an unforeseen state of the world arises at some repetition of the game. In fact, under incomplete knowledge, contingent contractual

commitments are mute, or not even specified, on the unforeseen states, and this implies that also commitments to specific contingent strategies that the firm B may undertake toward its stakeholders A_i will be unspecified.

But a type's reputation crucially depends on verification of the correspondence between the game outcome in a given state and the commitment to be fulfilled by the type in the same state, which entails an expected outcome for that state under the given type (also contingent on player A_i's choice). When a state of the world is unforeseen, a concrete contingent strategy cannot be ex ante specified as to its possible occurrence. Thus no contingent commitment can ex ante be undertaken with respect to unknown states of the world. From this it follows is that there is no basis for saying whether *"what had to be done has been done*" (Kreps, 1990). Commitments are emptied by cognitive gaps in relation to states that stakeholders and the firm cannot ex ante concretely describe. These cognitive gaps give *no* basis for reputation as modelled as the probabilistic updating of initial beliefs associated with commitments calculated in function of stage-by-stage observation of whether or not actions prescribed by commitments are performed at any stage of the game.

In more general terms, the problem is essentially one of incomplete specification of the *game form* and in particular of the strategy set (type set) and outcome functions (which map strategy combinations to payoffs for each state of the world at each stage). But without types uniquely related to commitments to strategies, no reputation effects are possible. Thus an "existence of the equilibrium" problem arises. Players cannot calculate the equilibrium strategies of the reputation game because their commitments are unspecified with respect to unforeseen states of the worlds. Put differently, they lapse into a state of cognitive unawareness of the equilibrium strategies that would support any level of mutual cooperation amongst the players.

The picture changes if the social contract has been introduced ex ante on a norm understood as the firm's constitution stating its fiduciary duties toward all the stakeholders in terms of general and abstract principles and precautionary rules of behavior. It predefines the standard conducts to be carried out if some principle is put at risk of violation by the occurrence of whatever (even if unforeseen) state of the world. What is crucial here is that the social contract introduces explicit norms (general and abstract principles and precautionary rules of behavior) that are established without ex ante complete knowledge of all future states of affairs. In general, this is the role of constitutional principles in legal orders, and specifically the role of universalizable principles in ethical codes.

Once a social contract has been introduced, there will be universalizable, general and abstract principles and precautionary rules of behavior to which stakeholders and the firm have agreed without being contingent on any concrete and complete ex ante description of future states of affairs; and these principles can be taken as benchmarks with which to assess the firm's behavior also when unforeseen states arise (as Kreps suggested concerning corporate culture principles but mistakenly restricted them to cultures rather than to ethics, see Kreps, 1992 and Sacconi, 2000). In so far as the agreement is worked out through counterfactual reasoning under a hypothetical original choice situation, and concerns general and abstract universalizable principles - by definition independent from any concrete description of details about the players' positions and any other concrete contingency – the principles agreed are adaptable to a wide array of situations. The social contract thus plays a cognitive role as a *gap filling device* (Coleman, 1992) which establishes the *types* of behaviors that stakeholders can expect from the firm in situations where contracts fail owing to the absence of conditional provisos constraining residual decisions.

This cognitive function is primarily *constructive*. The *game form* (Aoki, 2007) is badly specified under unforeseen situations, because contingent strategies for such states are unspecified. Norms nevertheless allow a default inference to be made on how the honest type of firm will behave under these circumstances. These 'strategies' are not defined contingently on states of the world that the parties are unable to write down in the contract or are even unable to foresee. These default rules are based on the satisfaction of a *fuzzy* membership condition of states with respect to the domain of abstract, general and universalizable ethical principles that are ex ante known (because they are agreed through the social contract) (Sacconi, 2000; Zimmerman, 1991; Sacconi 2007b). Membership is always ex post verifiable through a shared understanding of the inherent vagueness of unforeseen contingencies with respect to the principle. Once these norms have been stated ex ante in terms of precautionary standards of behavior, it is possible to say how the firm is expected to behave in whatever unforeseen state that may put a general principle at risk, until contrary proof is given that the principle does not apply to the new situation. In other words, the firm types implementing or otherwise strategies of conformity to norms are described. Explicit norms then complete the description of the game form by substituting default rules of behavior for conditional strategies. What is involved here is not inductive learning about the probability of an already given set of possible but uncertain set of types, but the conception of the type set itself that contributes to an (approximate) description of what may occur in the future. Accordingly, the

social contract role is *constructive*. Through the agreed statement of norms, firms and stakeholders *construct* an approximate model of the game that they will play in states of the world that they are ex ante unable to describe in every detail.

Nevertheless, the cognitive (and constructive) function of norms takes us only half-way into our argument. A well-conceived game form makes it possible to define the players' strategy combinations and equilibria wherein the firm may be described as acting in support of its reputation, so that after some time stakeholders will begin to trust it. Under the usual condition of the long-run player's non-myopia, these equilibrium combinations include the firm's continuing not to abuse and the stakeholders' continuing to enter the relation with the firm. Nevertheless, in general, this will be *just one* of the many possible reputation equilibria of the game. Other equilibria will entail strategies of random compliance with the norm by the firm (a mixed repeated strategy) such that the stakeholder's best response is to yield to the firm's strategy (entering throughout all the game repetitions and enduring consequences from the firm's partial abuse). Among these equilibria (see *Figure 2*, where the equilibrium set *X* of the repeated TG is depicted as the dashed area, and note in particular the equilibrium with average discounted payoffs (0, 2.66)), one is the *Stackelberg equilibrium*, this being the equilibrium that the firm would select if it committed unilaterally to its preferred *mixed type* and induced stakeholders to play their best responses to such an irremovable commitment. (Note that in a non-cooperative repeated game such an irremovable commitment can only be 'simulated' by the firm with the accumulation of a reputation of being such a type, so that stakeholders play their best responses whereby the firm must respond by fulfilling the commitment). Under such an equilibrium, the firm must have been able to accumulate a reputation for a mixed level of abuse which leaves stakeholders indifferent between entering or not entering – so that by entering a very large part of the potential surplus is appropriated by the mixed type firm.

There is no reason to assume that, because the Stackelberg equilibrium is one of the possible Nash equilibria, it must necessarily be the one selected. Yet there are also strong reasons to believe that in so far as no other element is introduced into the picture, player B will engage in maneuvers to develop a reputation that will allow him/her to select exactly this equilibrium, which gives him/her the highest payoff within the equilibrium set. To sum up, when a repeated reputation game is constructively defined in terms of strategies that abide or otherwise with the ex ante agreed CSR norm, the game will have too many equilibrium points, not just the 'socially preferable' equilibrium where the firm abstains from abusing

stakeholders and cooperates with them at any stage. Then the typical game theoretical problem of *multiple equilibria* arises.

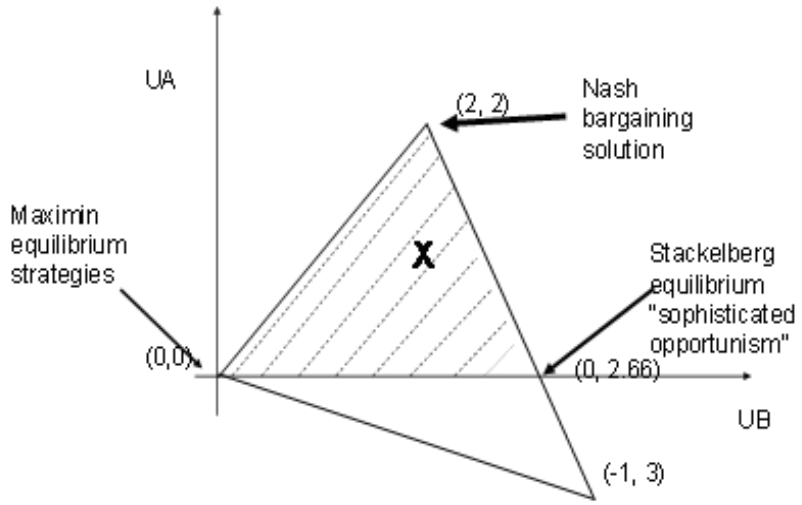

Fig. 2 Equilibrium set X of the repeated TG

Before going a step further, however, note that we have already obtained an important result – even if it is an admittedly partial one. It follows naturally from what has been said about the *constructive role* of explicitly agreed CSR social norms (and the related *multiplicity* problem) that effective self-regulation should not be confused with the standard economic view that if CSR is to emerge as an equilibrium behavior from endogenous incentives, its driving force must simply be 'enlightened self-interest in the long run'. According to this view, a self-interested entrepreneur who owns the firm, and cares only for his/her own self-interest in the long run (or, if s/he does not own the firm personally, cares for the self-interest of all the company shareholders) would adopt behavior that spontaneously satisfies the company stakeholders' interests with no need to single out a principle of fairness, either to agree on any social contract or to state explicitly any charter on the firm's fiduciary duties to stakeholders. Self-interest in the long run – or more concretely, maximizing total shareholder value in the long run - would naturally guarantee that the treatment of corporate stakeholders will fulfill their interests and claims, thus making any explicit statement of extended fiduciary duties superfluous. As a consequence, the only goal that should be specified as the proper constraint on managerial and entrepreneurial discretion in the management of the firm is the coherent pursuit of shareholder-value in the long run. The stakeholders' legitimate interests would be satisfied simply as a side-effect of this main goal, because they are related to it

through a means-end relation. Hence whilst stakeholders are to be taken into account by the corporate strategy in the domain of *means*, only shareholders are recognized as sources for corporate *ends*.[4] This view, of course, does not recognize any need for a norm that explicitly states a principle of fair balancing amongst stakeholders, even if it may be understood as not externally enforced but as self-imposed through self organization by those in an authority position in the firm.

From what we already know, however, this self-interest-in-the-long-run view is clearly untenable. First of all, without the explicit statement of a CSR norm - based at least hypothetically on agreement by the company stakeholders reached under ideal conditions of impartial bargaining - a long-run self-interested corporate strategy simulating the discharge of fiduciary duties owed to stakeholders may simply not exist (or be something that the firm cannot be aware of at all). This is implied by the case just discussed of unspecified game form. Under incompleteness of contracts, and without the protection of a constitution charter or a code of ethics stating general abstract principles and prophylactic rules of behavior about stakeholders' fair treatment, no conditional commitment is defined with respect to unforeseen states of the world. Thus the firm cannot accumulate reputation due to its expected behavior in these states.

Moreover, if such a behavior in the long run could be worked out as something of which the firm might be aware (and this will happen when a CSR norm is given), nevertheless other behaviors in the long run could also be worked out by the company, such that they provide very limited and minimal satisfaction of the stakeholders' claims for fair treatment. These further behaviors would not only be preferable to the firm's owners; they would also command a certain acquiescence by the stakeholders – which could be made indifferent between the prospects of giving in to these firm's opportunistic strategies or refraining from entering any relationship with it. We must conclude that the simple self-interest in the long-run view, translated into shareholder value in the long-run doctrine, would imply a large amount of violation of stakeholders' legitimate claims and abuse of ownership-based authority.

By contrast, the self-regulatory view defended here requires the establishment of explicit norms arrived at by social dialogue and multi-stakeholder agreements, and taking the form of CSR governance codes or management standards voluntarily accepted by firms because they contain and specify the terms of the ideal and fair social contract between the firm and its stakeholders. They are explicitly formulated in language (written or oral) and their utterances

state the extended fiduciary duties and obligations that the firm owes its stakeholders. At the same time they are voluntarily adhered to. And, as far as enforcement is concerned, they are not imposed by external legal sanctions but instead through endogenous social and economic sanctions and incentives. In this sense they are self-enforceable explicit norms put into practice essentially by means of endogenous economic and social forces such as reputation effects and conformity. As a matter of fact, such a norm will correspond to just one equilibrium among the many possible (see again *fig. 2*; it is quite obvious that a norm of fair treatment will require play of the repeated strategy equilibrium with average discounted payoffs (2, 2)). Part II will show that the social contract on an explicitly expressed CSR standard and norm also performs a normative role by providing an ex ante guide for the solution of the equilibrium selection problem.

---

[1] At first glance, one might object to the idea that many stakeholders, in both the 'strict' and 'broad' senses, do not have relations with a firm such that they formally delegate authority to those who run it (for example, they do not vote). The consequence is that the fiduciary duties as defined earlier do not apply to them. In the model of the social contract as a hypothetical explanation of the origin of the firm, however – see section 5.2 – all the stakeholders participate in the 'firm's second social contract'. The consequence is that their trust constitutes the authority of the firm's owner and manager. This also explains how the latter's authority may be accepted by these subjects. Moreover, the hypothetical social contract is typically used to explain how authority – that is, legitimate power – may come about at both the political and organizational levels; see, for example, Green (1990), Raz (1985) and Watt (1982). For a discussion of managerial authority, see MacMahon (1989) and Sacconi (1991).

[2] However, consider debates on the business judgment rule in relation to its consistency with 'team production theory' as inherent in the American tradition of company law (Blair and Stout, 1999; Meese, 2002), but also see the recent UK company law reform – especially the introduction of the directors' obligation to run the company "*in the way he considers, in good faith, would be most likely to promote the success of the company for the benefit of its members as a whole, and in doing so have regard*"…. for the interest of stakeholders other than the "members" of the company (employees, customers, suppliers, communities and others), for the impact on the environment, and the company reputation conditioned by these relationships, which moreover states that when these further purposes are to be considered, beyond the interest of shareholders, the meaning of '*promoting the success of the company in the interest of its members*' must be understood as if it included the pursuance of also these further purposes and interests. (The 2006 UK company law reform, Art. 172). Such an enlargement of the purposes that directors must pursue as the definition of the company success concept effectively opens the way to effective CSR self-regulation.

[3] Aoki pays much attention to institutions of different level ('generic, substantive and operational') and their mutual complementarities (Aoki 2007a, 2002). On the contrary, my view of CSR as a corporate governance

institution emerging form the firm's social contract is a 'state of nature' explanation such that other institutional levels do not significantly affect the interaction among stakeholders, and between the stakeholders and the firm (see also Sacconi 2000, 2006a,b, 2009). Admittedly, there are benefits and costs in both the modelling strategies. I maintain that there is an advantage in being able of considering what would happen in case the law in general made room for the firm's social contract among all its concerned stakeholders seen as an endogenous institution making process, including both the ex ante settlement of a set of explicit norms and the solution of the ex post compliance and equilibrium selection problem. Nevertheless, in order to model the stakeholders' social contract on the firm's control and accountability structure as a governance institution, there is no need to consider it as a completely isolated object lost in a institutional vacuum. It is enough to borrow the idea of "morally free zone" - as it was re-elaborated by Dunfee and Donaldson (1995) in quite a different way with respect to the original version given by David Gauthier (1986). 'Small scale social contracts' at industry, local or sectional levels are explicitly allowed by hyper-norms that are the object of the 'general social contract'. The general social contract leaves intentionally room to them due to the parties' awareness of bounded moral knowledge and rationality. However, by contrast also with Dunfee and Donaldson's view, the small scale social contract of the firm is here explicitly modeled as the result of an ex ante bargaining between stakeholders under the 'veil of ignorance' (see also part II), and not just as an ex post equilibrium institution. Whereas the equilibrium condition was also true of the local norms' definition in Dunfee and Donaldson's ISCT, seeing them as 'approved social convention', that theory was unable to provide a proper social contract model for the emergence of local norms – i.e. to explain them in terms of an impartial agreement among the firm's stakeholders on constitutional general principles and preventives rules of behavior. This is provided by the Rawlsian view of CSR.

[4] This is probably the opinion of Jensen when he says "*Indeed, it is a basic principle of enlightened value maximization that we cannot maximize the long-term market value of an organization if we ignore or mistreat any important constituency. We cannot create value without good relations with customers, employees, financial backers, suppliers, regulators, and communities. But having said that, we can now use the value criterion for choosing among those competing interests. I say "competing" interests because no constituency can be given full satisfaction if the firm is to flourish and survive*." (Jensen 2001). See also Sternberg (1999).

References

Alchian, A. and H. Demsetz (1972), 'Production, Information Costs and Economic Organization' *American Economic Review*, 62.p.777-95.

Aoki, M. (1984), *The Cooperative Game Theory of the Firm*, Cambridge: Cambridge University Press.

Aoki, M. (2001), *Toward a Comparative Institutional Analysis*, Cambridge, MA: MIT Press.

Aoki, M. (2007a), 'Three-Level Approach to the Rules of the Societal Game: Generic, Substantive and Operational' paper presented at the conference on 'Changing Institutions (in developed countries): Economics, Politics and Welfare' Paris, May 24–25, 2007.

Aoki, M. (2007b), 'Endogenizing Institutions and Institutional Change', *Journal of Institutional Economics*, 3, pp. 1–39.

Binmore, K. (1987), 'Modeling rational players', *Economics and Philosophy*, 1 (3), pp. 9–55 and 2 (4), pp. 179–214.

Binmore, K. (1991), 'Game theory and the social contract' in R. Selten (ed.), *Game Equilibrium Models II, Methods, Morals, Markets*, Berlin: Springer Verlag.

Binmore, K. (1994), *Game theory and the Social Contract (Vol. I): Playing Fair*, Cambridge MA: MIT Press.
Binmore, K. (1998), *Game theory and the Social Contract (Vol II):Just playing*, Cambridge MA: MIT Press.
Binmore, K. (2005), *Natural Justice*, Oxford: Oxford University Press.
Blair, M. and L. Stout (1999), 'A Team Production Theory of Corporate Law', *Virginia Law Review*, 85 (2), pp. 248-328.
Blair, M. and L. Stout (2006), 'Specific Investment: Explaining Anomalies in Corporate Law', *Journal of Corporation Law*, 31, pp. 719–44.
Clarkson, M. (1999), *Principles of Stakeholder Management*, Toronto: Clarkson Center for Business Ethics.
Coleman, J. (1992), *Risks and Wrongs*, Cambridge, MA: Cambridge University Press.
Donaldson, T. and L.E. Preston (1995), 'Stakeholder theory and the Corporation: concepts evidence and implication', *Academy of Management Review*, 20 (1), pp. 65–91.
Dezau A. and D. North (1994) , 'Shared mental models: Ideologies and institutions', *KIKLOS*, 47, pp.1-31.
Dunfee, T. and T. Donaldson (1995), 'Contractarian Business Ethics', *Business Ethics Quarterly*, 5, pp. 167–72.
Faillo, M. and L. Sacconi (2007), 'Norm Compliance: The contribution of Behavioral Economics models', in A. Innocenti and P. Sbriglia (eds), *Games, Rationality and Behavior*, London: Palgrave Macmillan.
Flannigan, R. (1989), 'The Fiduciary Obligation', *Oxford Journal of Legal Studies*, 9, pp. 285–94.
Freeman, R.E. (1984), *Strategic Management: A Stakeholder Approach*, Boston: Pitman.
Freeman, R.E. and P. Evans (1989), 'Stakeholder Management and the Modern Corporation: Kantian Capitalism', in T.L. Beauchamp and N. Bowie (eds), *Ethical Theory and Business*, 3rd ed., Englewood Cliffs, N.J.: Prentice Hall.
Freeman, R.E. and J. McVea (2002), 'A stakeholder approach to Strategic management', working paper No. 01-02, Darden Graduate School of Business Administration.
Freeman, R.E and S. Ramakrishna Velamuri (2006), 'A New approach to CSR Company Stakeholder Responsibility', in A. Kakabadse and M. Morsing (eds), *Corporate social responsibility reconciling aspiration and application*, London: Palgrave Macmillan. Pag 4
Fudenberg, D. and D. Levine (1986), 'Limit games and limit equilibria', *Journal of Economic Theory*, Elsevier, vol. 38(2), pages 261-279;
Fudenberg, D. and D. Levine (1989), 'Reputation and equilibrium selection in games with a patient player', *Econometrica*, 57, pp. 759–78.
Fudenberg, D. and J. Tirole (1991), *Game Theory*, MIT Press, Cambridge Mass.
Fudenberg, D. (1991), 'Explaining cooperation and commitment in repeated games', in J.J. Laffont (ed.), *Advances in Economic Theory*, 6th World Congress, Cambridge: Cambridge University Press.
Gauthier, D. (1986), *Morals by Agreement*, Oxford: Clarendon Press.
Geanakoplos J., Pearce D. and Stacchetti E. (1989), 'Psycological Games and Sequential Rationality', in *Games and Economic Behavior*, vol. 1, pp. 60-79.
Green, L. (1990), *The Authority of the State*, Oxford: Clarendon Press.
Grimalda, G. and L. Sacconi (2005), 'The constitution of the not-for-profit organisation: reciprocal conformity to morality', *Constitutional Political Economy*, 16 (3), pp. 249–76.
Grossman, S. and O. Hart (1986), 'The Costs and Benefit of Ownership: A Theory of Vertical and Lateral Integration', *Journal of Political Economy*, 94, pp. 691–719.
Hansmann, H. (1996), *The Ownership of the Enterprise*, Cambridge, MA: Harvard University Press.
Hare, R. M. (1981), *Moral Thinking*, Oxford: Clarendon Press.
Harsanyi, J.C. (1977), *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge, MA: Cambridge University Press.
Harsanyi, J.C and R. Selten, (1988), *A General Theory of Equilibrium Selection*, Cambridge, MA: MIT Press.
Hart, O. and J. Moore (1990), 'Property Rights and the Nature of the Firm', *Journal of Political Economy*, 98, pp. 1119–58.
Hart, O. (1995), *Firms, Contracts and Financial Structure*, Oxford: Clarendon press.
Hobbes, T. (1994), *Leviathan*, English edition With Selected Variants from the Latin Edition of 1668, Edwin Curley (editor) , Indianapolis: Hackett Publishing Company Inc.
Jensen, M.C. (2001), 'Value Maximization, Stakeholder Theory, and the Corporate Objective Function', *Journal of Applied Corporate Finance*, 14 (3), pp. 8–21.
Kalai, E. and M. Smordinski (1975), 'Other Solution to Nash's Bargaining Problem', *Econometrica*, 43 (3), pp. 880–95.
Kreps, D. (1990), 'Corporate Culture and Economic Theory', in J. Alt and K. Shepsle (eds), *Perspectives on Positive Political Economy*, Cambridge: Cambridge University Press.
Kreps, D. (1990), *Games and Economic Modelling*, Oxford: Oxford University Press
Kreps, D. (1992), 'Static Choice in the Presence of Unforeseen Contingencies' in P. Dasgupta, Rae D., Hart O. and Maskin E. (eds), Economic Analysis of markets and Games, MIT Press, Cabridge, Mass.

Lewis, D. (1969), *Convention. A Philosophical Study*, Cambridge, MA: Harvard University Press.
Maskin, E. and J. Tirole (1999), 'Unforenseen Contingencies and Incomplete Contracts' *Review of Economic Studies*, 66, pp. 83-114.
McMahon, C. (1989), 'Managerial Authority', *Ethics*, 100, pp. 33–53.
Meese, A.L. (2002), 'The Team Production Theory of Corporate Law: A critical Assessment', *William and Mary Law Review*, 43, 1629–39.
Nash, J. (1950), 'The Bargaining Problem', *Econometrica*, 18, pp. 155–62.
Posner, E.A. (2000), *Law and Social Norms*, Cambridge, MA: Harvard University Press.
Rabin, M. (1993), 'Incorporating Fairness into Game Theory', *American Economic Review*, vol. 83 (5), pp. 1821-1302.
Rajan, R. and L. Zingales (1998), 'Power in a Theory of the Firm' *Quarterly Journal of Economics*, CXIII.
Rajan, R. and L. Zingales (2000), 'The Governance of the New Enterprise', in X. Vives (ed.) *Corporate Governance, Theoretical and Empirical Perspective*, Cambridge: Cambridge University Press.
Rawls, J. (1971), *A Theory of Justice*, Oxford: Oxford University Press.
Rawls, J. (1993), *Political Liberalism*, New York: Columbia University Press. Togli
Raz, J. (1985), 'Authority and Justification', *Philosophy and Public Affairs*, 14 (1), pp. 3–29.
Raz, J. (1999), *Engaging Reason: On the Theory of Value and Action*, Oxford: Oxford University Press.
Sacconi, L. (1991), *Etica degli affari, individui, imprese e mercati nella prospettiva dell'etica razionale*, Milano: Il Saggiatore.
Sacconi , L, (1997), *Economia, etica, organizzazione,* Bari: Laterza.
Sacconi, L. (2000), *The Social Contract of the Firm. Economics, Ethics and Organisation*, Berlin: Springer Verlag.
Sacconi, L., S. De Colle and E. Baldin (2003), 'The Q-RES Project: The Quality of Social and Ethical Responsibility of Corporations', in J. Wieland (ed.), *Standards and Audits for Ethics Management Systems, The European Perspective*, Berlin: Springer Verlag, pp. 60–117.
Sacconi, L. (2006a), 'CSR as a model of extended corporate governance, an explanation based on the economic theory of social contract, reputation and reciprocal conformism' in F.Cafaggi (ed.), *Reframing self-regulation in European private Law,* Kluwer Law International, London,
.Sacconi, L. (2006b), 'A Social Contract Account For CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification', *Journal of Business Ethics*, Volume 68, Number 3 / October, 2006, pp.259-281
Sacconi, L. (2007a), 'A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity', *Journal of Business Ethics*, Volume 75, Number 1 / September, 2007, pp. 77–96.
Sacconi, L. (2007b), 'Incomplete Contracts and Corporate Ethics: A Game Theoretical Model under Fuzzy Information', in F. Cafaggi, A. Nicita and U. Pagano (eds), *Legal Orderings and economic institutions*, London: Routledge.
Sacconi L. (2008), 'CSR as Contractarian Model of Multi-Stakeholder Corporate Governance and the Game-Theory of its Implementation, University of Trento - Department of Economics Working paper N.18
Sacconi L. (2009), 'Corporate Social Responsibility: Implementing a Contractarian Model of Multi-stakeholder Corporate Governance trough Game Theory' in J.P. Touffut and R. Solow (ed.), *Does Company Ownership Matter?*, Centre for economic Studies Series, Edward Elgar Publishing Ltd., London.
Sacconi L. (2010b),"A Rawlsian view of CSR and the Game Theory of its Implementation (Part II): Fairness and Equilibrium", in L. Sacconi, M. Blair, E. Freeman and A. Vercelli (ed.) '*Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*', (*infra*)
Sacconi L. (2010c), 'A Rawlsian View of CRS and the Game of its Implementation (Part III): Conformism and Equilibrium Selection' in L. Sacconi and G. Degli Antoni (ed.), *Social Capital, Corporate Social Responsibility, Economic Behavior and Performance*' Palgrave McMillan, London (Infra)
Sacconi, L. and M. Faillo (2010); "Conformity, Reciprocity and the Sense of Justice. How Social Contract-based Preferences and Beliefs Explain Norm Compliance: the Experimental Evidence" Constitutional Political Economy, Volume 21, Number 2 / June, 2010, pp.171-201
Sternberg, E. (1999), 'The stakeholder Concept: a Mistake Doctrine, Foundation for Business Responsibility', Issue Paper No. 4, November 1999, online available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=263144.
Stout, L. (2006), 'Social Norms and Other-Regarding Preferences', in J.N. Drobak (ed.), *Norms and the Law*, Cambridge: Cambridge University Press.

Tirole, J. (1999), 'Incomplete Contracts: Where do We Stand?' *Econometrica,* vol. 69(4), pp. 741-781.
Tirole, J. (2001), 'Corporate Governance', *Econometrica*, 69 (1), pp. 1–35. sez 3
Trebilcock, M. (1993), 'The Corporate Stakeholder Conference', *University of Toronto Law Journal*, 62 (3), pp. 297–793.
Watt, E.D. (1982), Authority, London: Croom Helm.
Wieland, J. (ed.) (2003), *Standards and Audits for Ethics Management Systems, The European Perspective*, Berlin: Springer Verlag.
Willamson, O. (1975), *Market and Hierarchies*, New York: The Free Press.
Willamson, O. (1986), *The Economic Institutions of Capitalism*, New York: The Free Press.
Zimmerman, H.J. (1991), *Fuzzy Set Theory and Its Applications*, 2nd revised ed., Dordrecht-Boston: Kluwer Academic Press.