EconomEtica

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

N.41 July 2012

Lorenzo Sacconi

ETHICS, ECONOMIC ORGANISATION AND THE SOCIAL CONTRACT

Working papers

ETHICS, ECONOMIC ORGANISATION AND THE SOCIAL CONTRACT

by Lorenzo Sacconi

Abstract

This chapter introduces a notion of social/ethical norm that integrates its description as a self-sustaining regularity of behavior with the normative meanings of the statements by which a norm is formulated in the moral language. This definition is applied to organizational ethics where the main problem – abuse of authority - is identified with the help of a critical reading of the new-institutional economic theory of the firm. Given a game theoretical definition of an institution, it is then shown that only by integrating it with the social contract as shared mode of reasoning the process of convergence to the beliefs system that backs an equilibrium institution may be started. Thus the chapter illustrates the egalitarian social contract as both an impartial justification for organizational constitutions and as an equilibrium selection device. It is shown that equilibrium selection through the social contract solves the problem of legitimization of authority in the organizational relation between a non-controlling stakeholder and the entrepreneur or the management of a firm, holding hierarchical authority over the stakeholder. The result is a fiduciary relation between a stakeholder (the trustor) and the owner, director or manager (the trustees) based on fair distribution of the firm surplus. This is the basis for the explanation of corporate social responsibility, understood as an extended model of organizational governance. It generalizes to all the possible ownership forms of the economic organization, giving credit to the idea that social responsibility is an overarching social norm in the field of organization governance.

1. INTRODUCTION

By an ethical social norm is meant a norm pertaining to ethics in both its normative and descriptive understandings. From a normative perspective, a norm is a rule of behaviour derivable from a principle which is commonly understood as universalisably prescriptive by agents, and whereby they decide to abide by the same rule of behaviour. Descriptively, a norm is a regularity of behaviour which is normally followed by members of a given action domain for most of the time, when they expect that other will also follow it.

Normatively, the social responsible model of organizational governance is the result of the stakeholders' social contract under the moral assumptions of impersonality, impartially, and empathy. Stakeholders agree to select such a governance structure as part of the organization constitution (an economic institution) that they join in order to escape from a mutually destructive 'state of nature'. Solution of the surplus distribution problem is egalitarian (or maximin). This is a consequence of Binmore's reconsideration of Rawls's maximin principle in a game theoretical context (Binmore, 1989, 1991, 2005). This result is replicated in the chapter by reference to a simple repeated game (the trust game), among a non-controlling stakeholder and the firm's management or entrepreneur. In this case, given an already-defined structure of ownership over the physical assets of the firm, the stakeholder grants authority to the entrepreneur only because such authority is exercised in accordance with a principle of fair distribution of the utility surplus.

All this depicts social responsibility owed to the organization's stakeholders as a normative model of multi-fiduciary organization governance. But the distinctive feature of this approach is that the normative social contract theory also works as part of an explanation of how a social norm of organizational governance may be selected, and how it may come to evolve until it establishes itself as one of the main institutions in a given action domain. Institutions are hence defined in game theoretical terms as equilibria supported by belief systems.

The chapter shows that the social contract explains the emergence of a 'social responsibility' norm in the domain of organizational governance, and how the equilibrium selection process leading to such an institution may be started. This is intrinsic in the social contract theory as a theory of agreement on a set of rules which also entails that each agent expects that the same rule is accepted and expected to be accepted by all other participants. But it also means that the fair social contract must fall on an equilibrium point which is a feasible way to implement the agreement reached through the social contract.

This double functioning of the social contract as justification for, and explanation of, an institution's emergence corresponds to the definition given here of ethics and organizational ethics. Descriptively, I accept the social scientist's understanding of ethical norms. But I also show that this cannot work unless the description includes the operation of a normative judgment which is prescriptive and universalisable, and able to override any other reason to act. In fact, in as much as the social contract is a normative principle that entails agreement on a solution for cooperation and coordination problems, it gives rise to a system of beliefs and mutual expectations about reciprocal behaviour which induces convergence to an equilibrium.

2. ETHICS AND SOCIAL NORMS

Economists, as well as social and organization theorists, often understand ethics as a domain of social norms performing pro-social functions that are self-sustaining and thus do not require external enforcement to be complied with. Pro-social functions are typically identified in the facilitation of cooperation and coordination among individuals in both organizational contexts – wherein authority relations are legitimized and hence their acceptance is facilitated by compliance with moral codes – and the markets, where individuals cooperate through imperfect contracts assisted by norms of business ethics. Ethical norms, moreover, are seen as self-sustaining because they induce incentives to conform with the regularity of behavior that gives them a behavioral content. This inducement of motivations to conform is simply explained by the mutual expectation that other members of the same action domain will behave similarly. A notion of reciprocity – either instrumental or intrinsic – is part of this explanation, so that conformity depends on the expectation of reciprocal behavior. When reciprocity is instrumental, conformity is seen as each agent's self-interested best response to the shared expectation that every other agent will also conform.

Normally, those who accept this perspective on ethics do not claim to have considered its normative contents, and they frankly admit that their analysis is entirely descriptive and does not purport to give any *moral justification* in favor of intentional action conforming with a particular norm. At best adhesion to a social norm is *explained* intentionally in terms of enlightened self-interest given common (or more realistically 'shared') knowledge of other player adherence. This is a fully respectable tradition of thought traceable back to David Hume (Hume 1739). Since the work of David Lewis, it has led to the growing contribution of game theoretical models to the understanding of social norms, conventions, and the like (Lewis 1969, Ullman Margalit 1977, Schotter, 1981, Sugden, 1986, Young, 1998, Skyrms, 2004, Bicchieri 2006). However, despite the honest admission that its intent is merely to 'describe' ethics, an explanation of how ethical norms actually work that completely disregards the *normative* content and meaning of these norms seems somewhat paradoxical.

Is it really possible that the cooperative and coordinative functions discharged by ethical norms can be explained without any reference to the content of normative judgments and the meanings of the linguistic utterances by which ethical norms are expressed? Can the effectiveness of these functions be completely indifferent to the normative content of norms, i.e. independent of their content in terms of commands, obligations, rights or permissions – as if the same functions could be performed for whatever obligation, right or duty affirmed, and whatever the reasons justifying them? The self-enforceable nature of ethics itself, which is a basic tenet of this descriptive perspective, is grounded on the idea that mutual expectations of reciprocal compliance – often complemented by self-interest (as in Lewis 1969, Schotter, 1981 or Sugden, 1986, but clearly not in Bicchieri, 2006), but without reference to any intrinsic *normative* reason as to why a norm should be adhered to – are enough to explain norm compliance. Again, the argument must work as if the merit of the purported command, right or duty were completely indifferent. Indifference regards not only the agent's first-person reasons to act but also the formation of his/her expectation concerning why other agents should act in accordance with the norm or why they should expect conformity from each other (which in general will depend on higher level expectations: an agent expects reciprocity and conformity from others precisely because s/he expects that others will also expect conformity from him/herself).

This, in fact, is the distinctive characteristic of Lewis' social conventions of *pure* coordination. No intrinsic reason disposes us to obey one convention amongst others if they perform exactly the function of permitting us to achieve the same goal or objective, or to satisfy the same value through coordination. In this case, the sole reason for conforming with one convention among others is the expectation that others will do the same. Here, however, emerges the deep difference with respect to ethical norms that do not perform a merely *pure* coordinative function but also give us reasons to make a choice among different ways to carry out cooperative endeavors based on terms of value like well-being, fairness, rightness, or the equitable distribution of some cooperative surplus, etc. When reasons to convince other interacting agents to forego their individual incentives to pursue their individually preferred outcomes in order to join a common mutually beneficial plan of action (but not necessarily as good for them as their most preferred alternative) are at stake, and also when reasons are needed for agreeing on one among many possible cooperative plans each characterized by a different distribution of benefits and none of them completely individually irrational (in the sense that they could not be sustained by any equilibrium point), then the content of ethical norms in terms of justice, fairness, equality, righteousness, duties, or otherwise social wellbeing, typically seems to matter (for a definition of ethical norms that modifies the one given by Lewis, including *normative acceptance* see Pettit 1990, see also Donaldson and Dunfee, 1999).

As a matter of fact, also the description of ethics provided thus far seems insufficient once attention is paid to the work of philosophers who have tried to describe the common characteristics of ethical judgments expressed through the use of moral language. One may recall, for example, that from a meta-ethical perspective Richard Hare (Hare 1963, 1981) set the conditions whereby statements, expressed through sentences constructed according to the linguistic rules on utterance formation, may have 'moral meaning'. Ethical statements from this meta-linguistic perspective are 'prescriptive', 'universalizable' and 'overriding'. Note that this too is a *description* of how moral statements work (how they acquire their meaning in accordance with how we regularly use the language), without commitment to any specific ethical intuition or normative theory in particular. The first condition is that any sentence stating an ethical norm is a *prescription,* that is, we understand it as a *guide* to action. Thus, describing the functioning of ethical norms without reference to how we understand their prescriptive meanings seems to be wrong. The second condition – universalizability - is that, in order for any sentence stating a prescription to be understandable (and effective to our action) *as an ethical norm*, it must pass the *universalization* test. This is a minimal requirement of consistency, but it excludes for example the typical contingent explanation of conventions according to which we abide by a norm only because we find ourselves in a particular context wherein a given outcome happens to be 'focal' owing to the contingent reciprocal beliefs of anyone about anyone. The utterance of a norm is understood as ethically prescriptive when it is unconditional on specific contexts and can be reformulated in abstract and universal terms. It can thus be extended from an initial case of application – to which the utterance referred – to all the situations that are invariant with respect to the initial one only in regard to the normatively meaningful characteristics (maybe a certain relation between two or more variables). But these situations may well differ from each other in regard to any other contingent facts, things or individuals described by the relative sentences (for example, the names of persons occurring as variables in a relation that has 'prescriptive' meaning may freely vary). Thus the only way in which the explanation of norm compliance as *exclusively* based on mutual expectations can be rendered consistent with the description of how ethical norms work in language is to state that the norm which we understand as universally prescriptive is 'abide by other persons' expectations about you in whatever context' (i.e. 'do always what others expect from you or believe you will do' – in other words, 'pure conformism'), without any anchorage in the different prescriptive contents (norms) that expectations may have case by case. However, many of us may not understand statements only about what others expect us to do as universally prescriptive.

The last condition – overridingness - imposes some consistency between prescriptive utterance and behavior by requiring that we understand as truly universally prescriptive only norms that we carry out in practice through our actual behavior. This condition may be too strong, because it does not allow any room for 'weakness of the will' – a not-irrelevant psychological phenomenon. If any discrepancy between utterances about values that we understand as universally prescriptive and our choices must be excluded, we are perhaps obliged to discard some of the most intuitively significant moral sentences as void of moral meaning. Nevertheless, this is a condition of realism and genuineness of ethical statements that reminds us that we cannot understand as really universally prescriptive a norm which, after due reflection on all the alternatives, proves psychologically impossible to put consistently into practice (see also Griffin, 1998). This relates to self-enforceability. Since an ethical norm that we truly find meaningful is a motive that overrides any other motive – after the agent's entire set of reasons, incentives and motivations has been considered – it will be revealed as effectively guiding our practical behavior. Such a norm does not need any external enforcement for it to be put into practice: it emerges (so to speak) 'spontaneously' from overall consideration of all the motives for action. Otherwise, in order to say that a moral norm overrides any other reason or incentive to act, the repertoire of motivations that we consider must at least include various sentences whose content is the universally prescriptive meanings of alternative norms.

Moreover, also normative ethics theories are part of the picture. If overriding universally prescriptive norms surface in our intuitive judgments and behavior, normative theories of ethics can be seen as systematic kinds of argumentation wherefrom such judgments are entailed and that are in accordance with an entire class of intuitively normative statements. Alternative normative ethics theories have thus a 'quasi-empirical' content. Each of them accords with a class (not necessarily the same) of sentences with intuitive universalizable prescriptive meanings. At the same time – in so far as it is normative – each theory tries to deny our initial intuitions about the universalizable prescriptive meaning of certain sentences that disagree with it. It is therefore impossible to explain how a behavior conforming with a norm emerges to perform coordination and cooperative functions without reference to a prescription that can be accounted for in terms of some normative ethics theory. Moreover, principles rationalized by a normative theory, besides providing reasons to justify actions, can also be seen as *causal* factors of the initial formation of those systems of mutually consistent (normative) expectations usually invoked to explain why an ethical norm is obeyed (in this chapter, for example, norms will emerge from agents' normative and predictive beliefs in so far as these expectations are shaped by principles of the social contract).

All these considerations suggest that, in order properly to explain how ethical norms perform their pro-social function in the market or organizations, and are 'spontaneously' complied with – even if mutual expectation of conformity may be very important – reference must necessarily be made to the normative reasons to act that are contained in some universalizable ethical judgment usually rationalized by a normative theory. And it will be expressed by some sentence from which a prescription follows which is the normative content of a behavior regularity.

3. ORGANIZATIONAL ETHICS

Ethical norms, as described in the previous section, will be now applied to economic organizations. They have two roles to play here. First by means of explicit normative prescriptions they provide arguments in favor (or against) existing or potentially alternative organizational arrangements, in particular with regard to the allocation among the organization's members of authority and the right to make discretional decisions affecting surplus distribution. In this way, organizational ethics suggests criteria with which to choose governance models for the organization, and principles for their strategic management and decision-making. Ethical theory is taken as the basis for deriving prescriptive and universalizable statements that are then employed to decide whether certain forms of authority allocation can be accepted and hence are morally legitimized forms of organizational authority.

The second role of organizational ethics is to ensure that the legitimate arrangement of organization authority and the governance structure are implemented through a self-enforceable ethical social norm. In other words, it must state whether moral legitimization coincides with enough *de facto* legitimization, sufficient to give rise to a stable enough organizational relationships. This can come about if the organization's moral norm justifying a given arrangement of organizational authority proves to be the basis for a self-enforceable social norm, and if a decision mechanism is identified through which that social norm can be come into being and evolved until its complete establishment.

Accordingly, the main problem of organisational ethics can be retrieved by a critical reading of the new-institutional theory of the firm (Williamson 1975, 1986). In this perspective, the main problem of economic organisation is preventing the opportunism that may occur when contracts are incomplete and the parties to the contract undertake specific investments whose outputs, due to the gaps in the contract, can be expropriated by those who hold control over essential decision variables needed for the success of other agents' investments. This position (usually understood as having control over the firm's physical assets, that can be used in the investment, but which may also consist in controlling some cognitive ability that may be essential to its success) puts its holder in a position to hold up the investor in order to reap as large a part as possible of the surplus engendered by his/her investment. The solution is to allocate authority to the party responsible for the investment, giving him/her control over the decision variable able to condition the investment's success (this usually means giving him/her ownership of the firm's physical assets).

However, as soon as the organisation of the firm increases in complexity, the apparent efficiency of this solution breaks down and becomes less than a second-best optimum. Assume that specific investments are multiple, so that many investments can only obtain a surplus when they are made in the same organization with others, and that some cognitive resources, even if are not necessarily idiosyncratic, are essential for valorisation of other cognitive resources possessed by members of the organisation; and finally assume that these resources or investments are complementary, so that the reduction of one of them may prejudice the value of the others (Sacconi 2000, but see also Aoki 1984, 2010).

The hierarchical governance structure that allocates all the residual decision control to one party is less than second-best in these situations because it is evident that having ownership and residual control legitimates appropriation of the entire surplus generated by the firm. But by predicting that their investment will remain unremunerated with any part of the organisation surplus, and that the importance of a cognitive asset will not be recognized at the level of the governance structure, non-controlling stakeholders will recognize that there is no reason to invest in the use of their idiosyncratic, essential or complementary cognitive assets. This is quite clear in inter-temporal models (see Grossman Hart 1986, Hart and Moore 1991). Once the control structure is settled at time 1, then specific investments that are non-contractible and unforeseen events that open the possibility for essential decisions favourable to the valorisation of investment (but are also non-contractible) occur at time 2. Finally, at time 3 a bargaining session takes place where the distribution of residual decision rights is relevant because they settle the bargaining status quo which conditions negotiation of the surplus division after investments have been made but the decision essential for their valorisation has still to be taken. The status quo entails a threat of exclusion from the firm for those parties who are not protected by residual control but have carried out a specific investment or have invested their cognitive resources in cooperative and complementary modes. The controlling party will threaten non-controlling stakeholders with exclusion if they do not agree that the decision variable has been resolved favourably for investments only if all the surplus will be appropriated by the controlling party. The distribution will thus reflect an imbalance of rights in the governance structure. But it is evident that if the non-controlling stakeholders are provident and risk averse, even if they are not perfectly rational and able to forecast future events, at time 2 they will reduce their specific investments and their effort in employing their complementary cognitive assets in proportion to their expectation of being expropriated. The result cannot but be an inefficient solution. What is important to note here is that the inefficiency must be imputed to the expectation of being unfairly expropriated borne by stakeholders who have legitimate claims to participate in the surplus distribution. Organisational inefficiency reflects expectations of unfair treatment. Abuse of authority results thus the main problem, and organizational ethics comes to the fore. Ethics comes before efficiency.

Note that, in the incomplete contract model with inter-temporal structure and ex post bargaining, inefficiency follows from sub-investments anticipated at the second stage of the model and in some sense from a remarkable amount of cognitive skill and prudency on the part of non-controlling stakeholders. But in the real world it is much more likely that the judgment of being treated unfairly will have a negative effect on the level of surplus. For example, even if this may seem irrational, non-controlling stakeholders (especially consumers or employees) may reject the bargaining proposal backed by the exclusion threat. This may occur precisely because of a fairness judgment, as observed in many experiments on ultimatum games (even if in the case considered the strong party settles the exit option unilaterally while the agreement is left to emerge from bargaining). Second, disloyal behaviour may occur also in a later phase of the game, after the ex post bargaining has been concluded. Disloyal behaviour (especially by employees) may be undertaken by non-controlling stakeholders in the organisation who are advantaged by an asymmetry of information in the implementation stage of the ex post renegotiated bargain. Consider, in fact, that complementarity of cognitive resources typically occurs in teams where the members' productivities are interdependent and inseparable from one another, so that separate monitoring of a single member's output does not yield much information about his/her productivity, which depends on the effort put into complementary resources by other team members. Since productivity is inseparable, a team member can always disclaim his/her responsibility for a reduced output by claiming that other team members are responsible.

Central to organisation ethics, therefore, is the issue of unfairness in surplus distribution due to unilateral authority allocation. Its central concern is to put forward principles of justice acceptable to both controlling and non-controlling stakeholders, and which are able to provide assurance about the principle of justice compliance sufficient to induce them to make optimal effort and an efficient level of investment in their idiosyncratic or complementary assets. Principles of justice in the distribution of surplus also affect the surplus production – as typified by cooperation problems where mutual advantage from joint action and distributive conflict on the surplus distribution are simultaneously present.

A fair balance in the surplus distribution reflects a balance in the stakeholders' rights. A typical solution considered in this chapter is the one related to the intuitive idea of the 'firm social responsibility' owed to the non-controlling stakeholders. It is the responsibility of a corporate board of directors, which is appointed by the owners, to balance the owners' residual right of control and claim/right to the surplus with duties owed to non-controlling stakeholders. These commit the board to allocation and distribution policies that allow non-controlling stakeholders to have a fair share of the cooperative surplus.

This duty corresponds to a constraint on the shareholders' claim to appropriate the entire surplus resulting from the stakeholders' cooperation. It should be clear that this parallels a solution of the organisational authority legitimisation problem discussed at the beginning of this section. Authority is legitimated when stakeholders, typically the non-controlling ones, accept it as the best solution of a cooperation problem among them. When the corporate governance structure is arranged in such a way that non-controlling stakeholders can participate in a fair share of the firm's surplus, they may accept the authority relation as the best solution of their cooperative and coordination problems. They thus agree to enter the authority relation because it satisfies their independent reason to act defined in terms of a claim of justice (Raz 1985, McMahon 1989, Sacconi 2000). In this case authority is provided with moral legitimisation.

4. THE SOCIAL CONTRACT AND THE EMERGENCE OF SOCIAL NORMS IN THE DOMAIN OF ORGANIZATIONAL GOVERNANCE

This and the following two sections explain how the *social contract of the firm* (Sacconi 2000, 2006, 2007, 2010a,b) can be understood as the source of emerging social norms in the domain of organizational governance which satisfies the definition of ethics given in section 2 and able to solve the problem quoted in section 3.

Social norms are in fact nowadays deemed no less important for corporate and organizational governance in general than legal norms. In fact, these two types of norms are complementary (Stout, 2011a). Since the adoption of certain contracts or statutes at the corporate level is to some extent voluntary, social norms may be seen as drivers of the voluntary adoption of one or another legal model (e.g., shareholder vs. stakeholder oriented). Moreover, even if a legal system makes some legal constraints and principles in corporate governance mandatory, it largely depends on social norms whether the legal constraints will be actually followed and whether adherence will spread at societal level. Certain legal institutions of organization governance, such as fiduciary duties, may or may not be established in a given context according to how social norms of trust are shaped at societal level. For example, if bridging social capital and trustworthiness in a given society were very low, assigning the fiduciary duties of autonomous trustees an important role in organizational governance could be pointless (Macey, 2008)

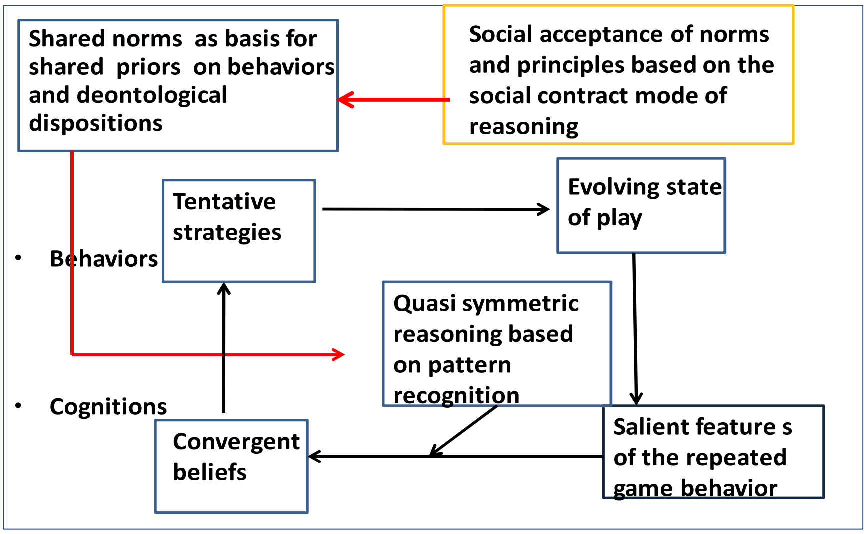
Social norms are even more important for the economic rather than legal analysis of institutions because modern economists understand them as 'conventions' (again see Schotter 1981; Sudgen, 1986, Young 1999). *Conventions* are coordination game equilibria that may endogenously emerge from repeated strategic decisions among players participating in a given domain of interaction. They are stable and self-enforceable once a system of mutually consistent expectations has formed that sustains the common belief that all participants will maintain behavior consistent with the norm. Because of their self-enforceability and incentive compatibility, conventions are the kind of institutions that economists like more, i.e., 'spontaneous orders' (Hayek, 1973; Sugden, 1986).

Hence, the gist of this section is that, once complementarity with the law has been recognized, and assuming that no mandatory laws are obstructing the emergence of a model of organizational governance based on social responsibility, the endogenous beliefs, motivations, and preferences of economic agents such as companies and their stakeholders become the essential forces driving the implementation of the model. In game theoretical terms, the normative model is implementable in equilibrium. This is also the basis for the widely accepted view that ‘corporate social responsibility’ (for an account of this issue see section 6) implementation is mainly a matter of voluntary self-regulation of self-enforceable principles and norms. Thus, its implementation may rest primarily on soft laws, social standards, code of ethics, voluntary adoption of contracts, provisos, and statutes, all of which are self-sustaining norms constraining ‘from within’ the discretion of corporate directors and managers (Wieland, 2003; Sacconi, 2006a).

The best way to integrate social norms into the emergence and stability of organizational governance models is to resort to Aoki’s (2001, 2010) account of institutions. Institutions “are not rules exogenously given by the polity, culture or a meta-game” but “rules created through the strategic interaction of agents, held in the minds of agents and thus self-sustaining” (Aoki, 2001, p. 11). An institution is “a self-sustaining system of shared beliefs about a salient way in which the game is repeatedly played” (Aoki, 2001, p. 11). The content of shared beliefs is “a summary representation (compressed information) of an equilibrium in a repeated game” (Aoki, 2001, p. 11). Thus, the salient feature of the equilibrium played has a symbolic representation inside the agents’ minds and coordinates beliefs that in their turn induce behaviors and their replication over time.

Cognitive components (i.e., beliefs deriving from compressed mental representations of salient aspects of ongoing equilibrium play) and behavioral components (i.e., the iterated play of a given set of equilibrium strategies) are interlocked in a recursive scheme (Aoki, 2010; also see the inner circle of fig. 1). The starting point is cognitive, and it consists in pattern recognition whereby given situations of interaction are framed as games of a certain form wherein players are expected to reason in a given quasi-symmetrical way. At step two, this framing of the situation induces players to entertain quasi-converging beliefs about a certain mode of playing the game. Thus, at step three, on passing from beliefs to the players’ actual behavior, each player adopts a tentative strategy based on the belief that others will also adopt strategies consistent with the aforementioned mode of behavior. Hence, in step four, strategies clash and some of them prove to be more successful and based on a better prediction. By trial and error, therefore, strategies converge towards an equilibrium of the game. This may be construed as an evolutionary result because the mode of playing attracts more and more players through iterated adaptation to the other players’ aggregate behaviors in the long run. At each repetition, however, this evolving

equilibrium is summarily represented in its salient features by a compressed mental model resident in the players mind so the fifth step concluding the circle is again cognitive.
This circle can be recursively iterated so that the ongoing equilibrium mode of playing is repeatedly confirmed by beliefs that translate into equilibrium behaviors, which are represented summarily by mental models, and so on. At some point, this belief system reaches a nearly complete state of 'common knowledge' (Lewis, 1969; Binmore and Brandenburger, 1990) about how players interact. The resulting equilibrium is an institution: a regularity of behavior played in a domain of interaction and stably represented by the shared mental model resident in all the participants' minds. It is essentially equivalent to the notion of social norm as a 'convention.'



(Fig. 1, Aoki's modified diagram representing the recursive process of institution formation)

However, a limitation is apparent in this understanding of institutions, and it concerns the normative meaning of an institution. Institutions in the above game-theoretical definition only ex post tell each player what the best action is**.** Once the players share the knowledge that they have reached an equilibrium state, then playing their best replies is actually a prescription of prudence that confirms the already-established equilibrium. Thus, institutions tell players only how to maintain the existing, already settled, pattern of behavior. They say nothing ex ante about how agents should behave before the mental representation of an equilibrium has settled and a self-replicating equilibrium behavior has crystallized. Institutions only describe regularity of behavior and are devoid of genuine normative meaning and force.

Here the notion of ethics as not only stable social norm but as norm satisficing moral meaning conditions, as explained in section 2, enters the picture. In fact, institutions including organizational governance contain norms (Donaldson, 2012), such as constitutional principles, laws, statutes, ethical codes, standard rules, and shared social values, which are expressed by explicit utterances in the players' language concerning values, rights, and obligations. These statements have a primarily prescriptive meaning, and if individuals attribute them moral meaning, such prescriptions are also universalizable (i.e., extensible to all similar states of affairs) and overriding with respect to alternative prescriptions expressed in the same context (Hare, 1981). Norms thus defined literally have normative meaning independently of the fact that they induce replication of an already-settled collective equilibrium behavior. Thus, a second component of a proper definition of an institution should be the mental representation of the normative meaning of norms.

This makes a great difference. The normative meaning of norms does not depend on knowledge about the ongoing behavior of other players. Instead, norms are able to justify and give first-place reasons for shared acceptance of a mode of behavior addressing all the participants in a given interaction domain before it has been established as an equilibrium point. A norm gives intentional reasons to act independently on the evolutionary benefits of adaptation in the long run because when an individual or a group of agents in a given action domain initiate an institutional change, it cannot stem from the pressure of evolutionary forces, which unfold their attraction only in the long run. Instead, a norm enters the players' shared mental model (Denzau and North, 1994) of how the game should be played, shapes the players' reciprocal disposition to act and their default beliefs about common behaviors, and hence becomes the basis for their first coordination on a specific equilibrium. In other words, it works as the first move in a process of equilibrium selection that activates the recursive process outlined by Aoki (2010). According to a line of theorizing in behavioral game theory, because a norm has been (cognitively) commonly accepted it may affect both dispositions to act (preferences) and expectations (default beliefs about how other players behave), so that the norm becomes a game equilibrium (Grimalda and Sacconi, 2005; Sacconi, 2007, 2011; Sacconi and Faillo, 2010; Sacconi, Faillo, and Ottone, 2011).

This equilibrium selection function of norms is deployed in two contexts: (1) within a well-defined game, where an old equilibrium path (old institution) has been abandoned for whatever reason and a new equilibrium path (new institution) has to be reached; and (2) when the underlying action domain changes because environmental or technological changes have occurred, or some further action opportunity is simply discovered by players, so that achieving a new equilibrium is necessary.

In these contexts, "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" (Aoki, 2001, p. 19). The key point is, therefore, to explain how a norm (basis for a system of normative beliefs) becomes acceptable by agents before the relevant equilibrium behavior is settled through rational best response, evolution, or other behavioral mechanisms such as reciprocity and conformism. What is required is a collective mode of reasoning (cognition) able to explain how a normative mental model arises before any evolutionary pressure has operated in that direction, and on the basis of which a norm may become commonly accepted in a not yet an equilibrium state. Therefore, what is needed is a cognitive mechanism of justification for norms that can operate in a similar way in many different contexts, so as to be able to produce a social norm that adapts to diverse situations.

The best justificatory account for the ex-ante shared acceptance of norms is the social contract model. Contractarian norms result from a voluntary agreement in a hypothetical choice situation that logically comes before any exogenous institution is superimposed on a given action domain, or before any institution has yet emerged. Thus, a norm arises only because of the voluntary agreement and adhesion of agents, even before it is established as an evolutionary equilibrium. To define the agreement, any social contract model sets aside threats, fraud, and manipulation – resources that would render the parties substantially unequal in terms of bargaining power – and considers all the agents as equal in respect to their rational autonomy, so that many of their arbitrary differences are placed under a veil of ignorance. Although a long tradition of different contractarian models could be cited (to exemplify one of the main line of thought in the social contract tradition consider Hobbes, 1651, Buchanan, 1975, Gauthier, 1986, Hampton 1987), the main reference here is to the Kantian model of the social contract developed by John Rawls (Rawls, 1971).

By introducing the social contract as the cognitive mechanism by which a norm may be accepted and become a shared mental model, Aoki's recursive model can be reformulated. The inner circle of fig.1 is retained. What is new (as shown in the upper part of fig. 1) is that the pattern derives from a shared social norm that categorizes the game as the domain of application of some more general principle. From this categorization it follows that some shared idea of the players' disposition to act (preferences) and common beliefs can be applied in the case under examination. In turn, the social norm derives from social contract reasoning (see fig.1) employed by players in order to agree on basic principles and norms when equilibrium institutions are not already established.

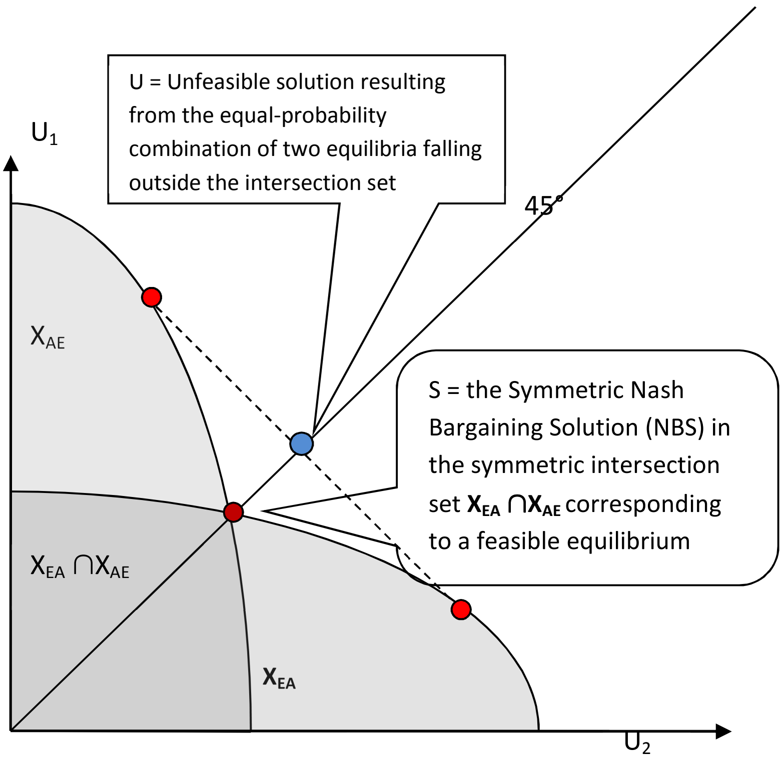
5. THE SOCIAL CONTRACT AS AN EQUILIBRIUM SELECTION DEVICE

5.1 The Rawlsian Maximin Principle Vindicated

This subsection applies Binmore's (2005) game theoretical vindication of the Rawlsian social contract to the organisation stakeholders' interactions (see also Sacconi, 2010b). It gives the basic model of how the social contract (a normative ethics principle) not only justifies an institutional model of organizational governance but also provides a source for its selection as a social norm – i.e. as an institution in revised Aoki's sense. Equilibrium selection has been a growing body of literature over the last three decades in game theory (Harsanyi and Selten 1988, Binmore 1987,1988) from the perspective of how player may interactively solve their uncertainty on how their counterparties are going to play a non-cooperative game with multiple equilibria and hence to calculate the correct and stable best response. What characterizes Binmore's contributions to the social contract theory (Binmore 1989, 1991, 2005) is that an ex ante approach to the selection of repeated play equilibria in (evolutionary) non cooperative games is justified.

Assume that two stakeholders, a poor worker (Eve) and a rich proprietor of means of production and capital (Adam) meet in a 'state of nature' structured as a non-cooperative game. Assume that they repeatedly play the same game resulting in a wide set of feasible outcomes. The 'state of nature' precedes the institution of any legal artifice such as the 'corporation' under which they could form a regulated team. In fig. 2 the convex and compact payoff space X_{EA} corresponds to the outcome set of the state of nature repeated game. Let these outcomes be all equilibria of the repeated game (i.e., when one player chooses his component of one of these strategy combinations the other has no incentive to deviate from it by changing his strategy component).

Then assume that before agents engage in the relevant interaction (e.g., a largely incomplete contract), they want to agree ex ante on the selection of one of these possible equilibrium points/outcomes. This may be seen as agreeing on a social norm singling out to what they should be entitled by playing their roles under a formal organization. Tis distributive norm is a skeletal constitution for the organization that the agents would be prepared to enter. Since the constitution must be fair, impartiality and impersonality of the agreement are required. Taken together, these assumptions are the 'veil of ignorance' hypothesis. In other words, each agent makes his decision "as if" he were ignorant about his true identity, so that in order to reach a deliberation he takes in turn the positions of each possible participant in the game.



(Fig 2. The Binmore-Rawls egalitarian solution)

In this context, *impersonality* means that acceptance of the solution must not depend on personal and social positions. Thus, players should select a solution that cannot be affected by the symmetrical replacement of social roles and personal positions with respect to individual players. Technically, fig. 2 depicts this replacement by the symmetric translation of the initial payoff space X_{EA} with respect to the Cartesian axes representing the utility of player 1 and player 2, respectively. Thus, under the initial payoff space X_{EA}, player 1 will have all the possible payoffs of Eve and player 2 all the possible payoffs of Adam. But under the translated payoff space X_{AE}, roles are reserved and player 1 will then get Adam's possible payoffs and player 2 will get Eve's possible payoffs. Moreover, fig.2 illustrates that each player, when taking the other's perspective, exercises perfect empathetic identification. That is, when player 1, who under X_{EA} was Eve thinks to be Adam under X_{AE}, this player is able to reproduce exactly the same payoffs that player 2 experienced when the player was Adam.

Impartiality means that the players must agree on an outcome under the hypothesis that the reciprocal replacement of positions works in such a way that each stakeholder has an equal probability of finding himself in the position of each of the possible two roles. Equal-probability explains how the solution may not change under the symmetrical translation of the payoff space with respect to the players' utility axes. Take an outcome x_{EA} that by replacing personal positions may realize in two non-coinciding ways (x_{EA} itself and x_{AE}). To make this outcome acceptable requires taking the expected value of an equal probability distribution over the two realization ways: $\frac{1}{2}x_{EA} + \frac{1}{2}x_{AE}$. This would identify a point in the space that is invariant under the players' positions replacement (i.e., an egalitarian solution residing on the bisector).

However, this construction is not meant to be an excessive idealization. Agents retain awareness that the solution must be an equilibrium of the original game. That is, the solution must be a collective behavior that the parties know is self-enforceable and incentive-compatible once they think that they all are playing it. This is a requirement of realism of the agreed solution: agents cannot afford to agree ex ante on a solution if it is not incentive-compatible ex post (beyond 'the veil of ignorance'). The reason is simple. Admit that the impartial solution proves ex post not to be an equilibrium of the original game (does not belong to the original payoff space of the 'state of nature' game). Hence, the player who ex post would be most favored by returning to a solution belonging to the initial equilibrium set would simply deviate to an equilibrium strategy.

Consequently, the stability condition requires that the ex-ante solution (agreed behind the 'veil of ignorance') must correspond to an outcome that under the players' place-permutation would nevertheless belong to the ex post equilibrium set. In other words, the selected outcome must be an equilibrium (say) either if player 1 takes the position of Adam (and player 2 respectively the position of Eve) or in the opposite case when their identification is reversed (player 2 occupies Adam's position, whereas player B takes Eve's position), and all the more so when an equally probable combination of the two identifications is taken.

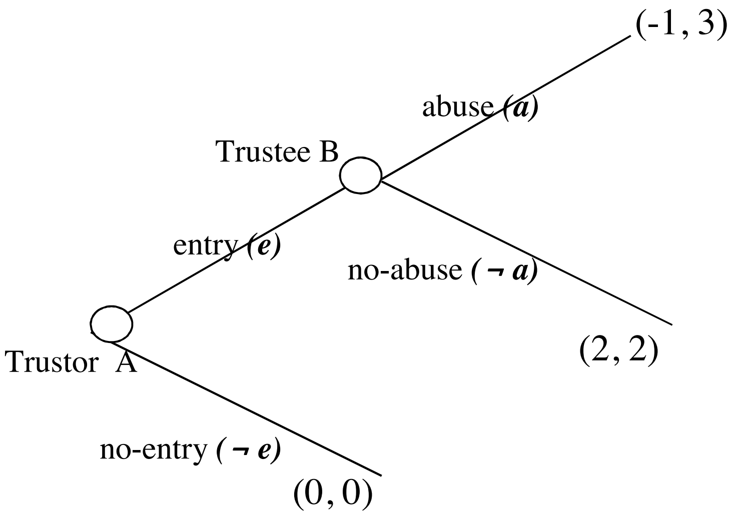
What has been just set is a new feasibility condition. Owing to the state of nature game's assumptions, only equilibria of the original payoff space X_{EA} are feasible. Any further outcome – potentially subject to agreement – would be wishful thinking because no ex post equilibrium would exist that could implement it (see point U in fig.2). Adding the conditions of impersonality and impartiality further restricts feasible outcomes to the symmetric intersection $X_{EA} \cap X_{AE}$ of the two payoff spaces generated by symmetrical translation of the original space, which is a proper subset of the initial outcome (equilibrium) set $X_{EA.}$ as shown in fig.2. This is a symmetrical payoff space wherein any bargaining solution necessarily falls on the bisector, which is the geometrical locus of egalitarian solutions (where parties share the bargaining surplus equally). Note that this result takes for granted an egalitarian status quo preceding the agreement, but this assumption too is a consequence of the veil of ignorance.

In particular, players resort to the Nash bargaining solution (NBS), which is the most widely employed solution for bargaining games (Nash, 1950). It prescribes picking the point of the efficient (north-east) frontier of the payoff space (representing the outcomes set of possible agreements) where the product $\Pi(u_i - d_i)$ of the utilities u_i of players ($i = 1, 2$), net of utility d_i associated with their status quo, is maximal. Assuming that the players bargain according to the typical rationality assumptions of game theory (Harsanyi, 1977), and given that the feasible outcome set is the symmetric intersection sub-space $X_{EA} \cap X_{AE}$, the NBS is by assumption egalitarian and selects the point S of fig. 2.

The striking result of this construction is that the minimal requirement of social justice (impersonality and impartiality) becomes compatible with realism and ex post stability in an interaction where players are free to choose according to their preferences. Freedom of choice and incentive compatibility does not require relinquishing the moral demands of social justice. On the contrary, it entails that the solution must be egalitarian and must coincide with the Rawlsian maximin distribution, even within an originally asymmetrical set of possible outcomes. Thus, given a real-life set of possible outcomes reflecting possible inequality between the participants, the solution falls on the equilibrium that most favors the worst-off player, which in most cases is the egalitarian distribution.

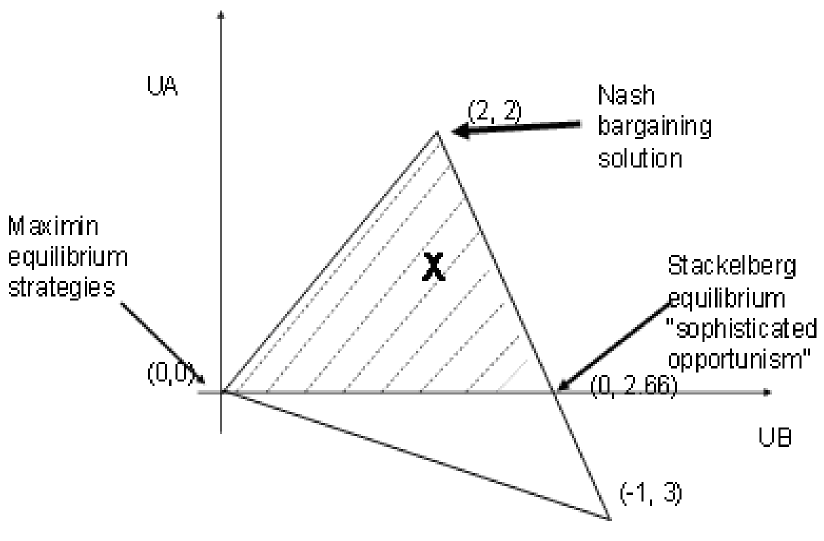
5.2 'Organization Social Responsibility' as norm selected by the social contract in a Trust Game

To give more concreteness to the foregoing exemplification of the organization social contract, consider now a strategic interaction between a non-controlling stakeholder A (i.e., an employee with a specific investment at stake) and a controlling stakeholder B (the entrepreneur or the manager) taking the form of a Trust Game (TG) (see fig. 3). By entering the relationship, the trustor (player A) accepts (trusts) the authority of the trustee (player B). On the contrary, by not entering, he refuses to take a subordinate position in the relationship with B. Moreover by entering A invests idiosyncratically in the relationship. The trustee is an authority who can abuse some discretionary power. Once the trustor has entered, the trustee may choose between abuse and no-abuse. *No-abuse* would maximize the two players' joint payoff and as well the Nash bargaining product in the outcome space, but abusing for a self-interested trustee is strictly dominant. Thus, in the one-shot TG, a self-interested trustee will always abuse, and hence the unique equilibrium solution of the game is no-entrance, abuse.



(Fig . 3 - The stakeholder - entrepreneur/manager/owner Trust Game)

However, the TG game is played repeatedly. On considering repeated strategies and their average outcomes, many possible equilibria exist. These include the original (no-entrance, abuse), the perfectly fair (entrance, no-abuse), plus all the possible pairings of entrance with mixed strategies combining abuse and no-abuse up to a limit probability of ⅔, and ⅓, respectively. Indeed, the entire dashed region of the payoff space in fig. 4 is filled with possible equilibrium points of the repeated TG

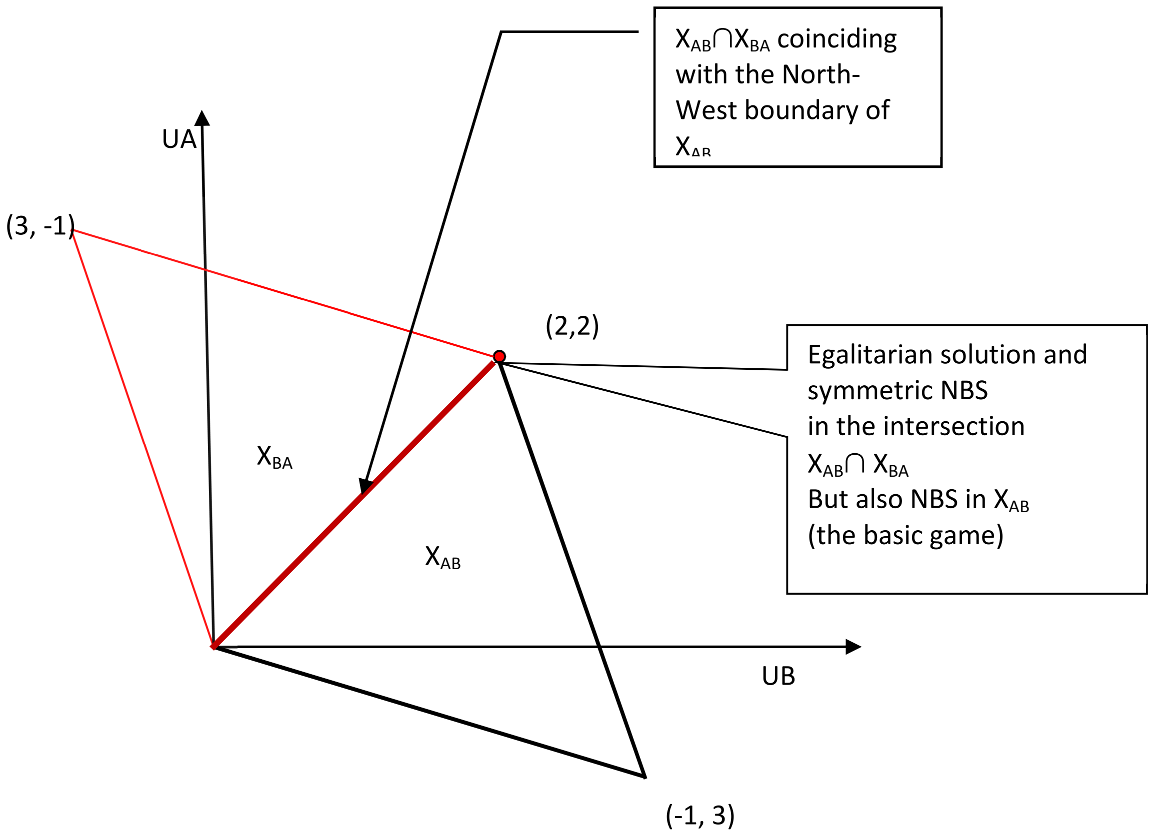


(Fig. 4- Equilibrium set of a repeated TG. Even if there is an egalitarian solution coinciding with the maximum Nash bargaining product, in absence of the social contract equilibrium selection device, the firm would select the Stackelberg equilibrium.)

Given so many equilibria, many possible conventions would emerge from reciprocal coordination. In particular, the trustee has a "conformity problem" with a social norm of fair treatment consisting in the NBS (by which B equally shares the surplus). If the firm is run to the fair reciprocal advantage of both stakeholders, only the equilibrium coinciding with NBS can emerge. By contrast, a model of organizational governance consistent with a purely shareholder-value maximization approach would justify the equilibrium corresponding to the Stackelberg solution.

Application of the Binmore-Rawls theory of equilibrium selection based on the ex-ante social contract is starkly simple in this case (see also Sacconi, 2010b; The idea to study the raise of corporate culture in a repeated Trust Game was originally presented in Kreps (1990); for the use of this game to illustrate business ethics see Sacconi (2000). However both these contributions was focused on the existence of fair equilibria and role of general principles when the game is played under unforeseen contingencies and bounded rationality; and neither of them discussed in depth the equilibrium selection problem.)

Figure.5 illustrates the symmetric translation of the repeated TG payoff space with respect to the player utility axes U_A and U_B, which consists of its rotation around the north-west boundary of the initial space X_{AB}. The symmetrical intersection subset $X_{AB} \cap X_{BA}$ reduces to the rotation axis itself, i.e., no more than a line segment (along the bisector) consisting of all the egalitarian distribution. By simply adding basic strong Pareto Optimality (i.e., agreeing on solutions that permit mutual improvements for all, if available) directly leads to choosing the equilibrium point consistent with the NBS of the original game (2, 2), which is also its egalitarian (and maximin) solution. To say it differently, the intersection set coincides in this case with the bisector where all the egalitarian payoff distributions reside, and by simple application of Pareto Optimality we may choose the best for all egalitarian solution, which is identical with the maximum Nash bargaining product. Nevertheless, once the egalitarian solution was selected, it would be incentive-compatible and stable. This is an abstract representation of an organization ethical norm endogenously emerging from the non-controlling stakeholder vs. entrepreneur/manager interactions aided by the social contract reasoning. Under such an a norm, the trustee behaves as if he owed the trustor (stakeholder) fiduciary duties of fair treatment. We call this norm 'organisation social responsibility' (OSR) owed to the organisation stakeholders.



(fig. 5, the repeated TG with symmetrical payoff space translation, and the egalitarian solution)

6. A MULTI-FIDUCIARY AND MULTI-STAKEHOLDER MODEL OF THE ORGANIZATIONAL GOVERNANCE

When the subject of organizational ethics is studied with reference to today's real world, what is encountered is the ever growing phenomenon of corporate social responsibility (CSR) and its generalization to the social responsibility of productive organizations in general. Firms, business organizations but also cooperatives and nonprofits, are increasingly considered to be subject to commitments, such as discharging socially responsible practices and programs to the benefit of the organization's stakeholders. Management and reporting standard are redefined as being centered on all the stakeholders' interests and not merely on those of the shareholders. Such practices, programs, management and reporting standards are seen as ethical and as involving reasonability of who are in a position of authority toward the organization's stakeholders.

To gain a proper understanding of the global CSR movement requires to question in depth the nature of organisations as social and economic institutions. On the battleground of organizational governance models, CSR denotes a movement that strives to affirms a social norm advocating an extension of the range of obligations owed by firms and productive organisation to all their stakeholders. Thus the CSR movement is the main example of emergence (even if not the only one) of the social norm that in more abstract and skeletal terms we called a norm of organizational social responsibility (OSR) owed to the organisation stakeholders in the previous section.

To be more precise about the basic principle of this norm, CSR - so understood - has been defined as *a model of extended organization governance whereby those who run firms, such as entrepreneurs, directors, managers, have responsibilities that range from fulfillment of their fiduciary duties towards the owners to fulfillment of analogous fiduciary duties towards all the firm's stakeholders* (Sacconi, 2006a, 2006b, 2007, 2010a)

Two terms must be defined for the foregoing definition of to be clearly understood. The first term is fiduciary duties. The assumption here is 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. The trustor therefore delegates decisions to a trustee empowered to choose actions and goals. The trustee may thus use the trustor's resources and select the appropriate course of action. For a fiduciary relationship to arise, the trustor must possess a claim (right) towards the trustee. In other words, the trustee directs actions and uses the resources made over to him/her so that results are obtained that satisfy the trustor's interests. These claims (i.e., the trustor's rights) impose fiduciary duties on the agent who is invested with authority (the trustee) that s/he is obliged to fulfill. The fiduciary relationship applies in a wide variety of instances such as tutor/minor and teacher/pupil relationships. In the corporate domain, the relationship is between the board of a trust and its beneficiaries or between the board of directors of a joint-stock company and its shareholders, and then more generally between management and owners. The term *fiduciary duty* means the duty or responsibility to exercise authority for the good of those who have granted that authority and are therefore subject to it (Flannigan, 1989).

The second term is *stakeholders*. This term denotes individuals or groups with a major stake in the running of the firm and who are able materially to influence it (Freeman 1984, Freeman and Gilbert 1989, Freeman and McVea, 2001, Freeman et al 2010, Donaldson and Preston 1995, Clarkson, 1995). However, from an economist's point of view, most relevant to defining stakeholders is the following distinction between two categories: stakeholders in the strict sense, and stakeholders in the broad sense.

Stakeholders in the strict sense are those who have an interest at stake because they have made specific investments in the firm, such as in the form of human capital, financial capital, social capital or trust, physical or environmental capital, or for the development of dedicated technologies. Such investments may substantially increase the total value generated by the firm and are made specifically in relation to that firm 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 depend largely upon it for satisfaction of their well-being prospects (lock-in effect). By contrast, *stakeholders in the broad sense* are those individuals or groups whose interest is involved because they undergo the 'external effects,' positive or negative, of the transactions performed by the firm, even if they do not directly participate in the transaction, so that they do not contribute to or directly receive value from the firm.

One can thus appreciate the scope of the firm's stakeholder responsibility defined as an extended form of governance: it extends the concept of fiduciary duty from a mono-stakeholder setting where the sole stakeholder relevant to identification of fiduciary duties is the owner of the firm to a multi-stakeholder one in which who run the firm owes fiduciary duties to all its stakeholders. This was exactly the result that was derived in the abstract through the stakeholder /owner (or manager) Trust Game in section 5.2.

This normative model of a firm's responsibility toward its stakeholders can be extended to any form of business firm and productive organization with legal personality, which makes it an institutional 'person' partly insulated from its individual members and constituencies. Hence it is not necessarily tied to the particular cases of publicly-owned companies or family-controlled capitalist corporations. It could be equally applied to a cooperative company where some stakeholders (for example consumers) own the firm and are thus obliged to permit extension of the fiduciary duties of the cooperative's managers to workers, capital lenders, suppliers, local communities, and the like. Or it could be applied to a law firm organized as a limited liability partnership or a professional association, where a few senior partners, who hold residual control rights, occupy the same position that, in the definition of CSR as far as the social responsibility toward stakeholders is concerned, is taken by managers, entrepreneurs or directors.

In all these cases, there is a category of stakeholders who control the organization through ownership constituted by a residual control right and a claim-right on the residual (profit) (Hansmann, 1999). Moreover it appoints the board of directors and managers. Under each of these ownership forms, authority, legitimately assigned to the owner category, may be abused. In fact, organizational authority entails by definition that the authority has at his/her disposal a sphere of discretion wherein ex ante non-contractible decision variables are placed and hence withdrawn from the discipline of an ex ante established detailed rule or contract. These discretionary decisions affect the possibility of non-controlling stakeholders to profit from their specific investments and their cognitive assets employed in a complementary and hence mutually advantageous way. Even though there are numerous ownership forms, the underlying problem – abuse of authority – is therefore recurrent (as stated in sec.3), and cannot be solved as long as the governance structure is focused only on design of the ownership and control rights of a particular stakeholder category. Therefore the organization stakeholder responsibility (OSR) requires the controlling stakeholder (if in a managerial position) or the management appointed by him/her to discharge extended fiduciary duties toward non-controlling stakeholders.

Note that in the public company case the corporate control exerted in general by shareholders is very weak. The corporate governance powers exercised by managers and directors is in many respects autonomous and seems to exercise large part of the actual residual control right. Different schools of thought see this contingency in very different ways. According to supporters of the principal-agent model (for example Jensen and Meckling, 1976, but see also Macey 2008), who affirm 'shareholder primacy', this relative independence of managers carries the risk of managerial opportunism and self-dealing behavior. Hence they suggests curbing it through corporate governance engineering, like the design of incentive contracts (stock options etc.) and contestable control mechanisms like hostile takeovers – whose aim is to provide positive incentives and negative sanctions so as to align the managers' interests with those of the shareholders. Other corporate governance theorists (e.g. Blair and Stout 1999, Stout 2011, but also Elhauge 2005, Gelter 2009) see this relative autonomy as the true manifestation of the nature of a 'corporate actor' which is not constrained to serve solely the interests of its shareholders but must pursue whatever goals the cooperative bargaining solution amongst stakeholders assigns to it (Aoki 1984, Sacconi 2000, 2006a,b). Thus the directors are legitimately entitled to operate as the stewards of the corporation and as trustees of the stakeholders whose wellbeing depends on the company's conduct and that contribute to the company's success. In particular, Blair and Stout (1999) defend the idea that the board of directors plays the role of an impartial hierarchy mediating among the different corporate stakeholders in order to prevent reciprocal opportunism (especially abuse of authority on the part of the controlling shareholders) and favoring their mutually beneficial cooperation. As regards the publicly-owned corporation, the socially responsible model of the economics organization, based on the idea of extended fiduciary duties toward strict and broad sense stakeholders, concurs with this corporate governance doctrine and seeks to provide it with a deep contrarian-ethics foundation and also with the idea that it may be supported by an evolving stable ethical norm based on the stakeholders' 'social contract'.

However, it is noticeable that the idea of an economic organization run by a board of directors or an entrepreneur exercising control rights without being the ultimate proprietor of the firm and owing extended fiduciary duties to all the stakeholders is not restricted to this case alone, and for which the public company is not the only or best example. Another case could be a state-controlled corporation (where the uninformed shareholder is the government), but an even better example is provided by large non-profits – e.g. a private university or a hospital run by a foundation. Although in these cases there is no subject who exercises ownership in its complete sense, because nobody is legitimately identifiable as the 'residual claimant', there are nonetheless managers and directors who are invested with a substantial part of the twofold definition of ownership, i.e. they may exercise residual control over discretionary ex ante non-contractible decisions and hence are endowed with authority (Hansamm, 1996). They are therefore also able to abuse this authority both in the self-dealing sense and by running the organization so as to give an unfair benefit to some stakeholder with which they have privileged relationships. Also this case perfectly enters the domain of application of the multi-stakeholder and multi-fiduciary governance model of the economic organization, once it is taken for granted that 'owner' in the strict sense in the foregoing definition is an empty category. Indeed, these organizations can be taken perhaps as the historical first manifestation of the very idea of social responsibility of the organizational governance toward all the organization's stakeholders (see Aoki 2010).

7. END REMARKS AND CONCLUSIONS

The main result of this chapter may be retold as follows: in general, in a context of constitutional choice concerning the firm's governance structure where two (or more) differently endowed stakeholders confront one another in a 'state of nature', it can be shown that an egalitarian constitution of the firm is agreed (for this general result see also Sacconi 2010b). By 'constitutional choice' is meant the selection of an admissible subset of the stakeholders' 'state of nature' strategies, that if unconstrained would allow them to undertake any opportunistic behavior in their contractual relationships. Under the ethical assumption of the veil of ignorance, however, they reach agreement on a constitution of the firm such that they make a final allocation of payoffs which is identical to the best egalitarian distribution of the cooperative surplus among those feasible for the constitutional choice. In other words, they accept the firms' constitution that allows an equal distribution of the surplus which is as good as possible for both (or all) the player. Egalitarian solutions are monotonically ordered in terms of growing mutual benefits, or symmetric Nash bargaining products, so that the constitution admitting the best feasible egalitarian solution will be chosen.

But what is really essential to the novelty of the result is that the egalitarian constitution is not only the outcome of cooperative bargaining. It is also the result of an equilibrium selection device, which ends up with a (self-sustaining) Nash equilibrium.

Therefore, further to the support given to the multi-stakeholder organizational governance presented in the previous section, there are important consequences of the results set out in this chapter, and especially in section 5, for the fields of new institutional economics, law and economics, and organization design. First, it is received wisdom that when incentive compatibility is required, the optimal design of economic organizations must set aside fairness as an 'idealistic' requirement and seek only some 'second-best' level of efficiency. Moreover, it has been argued that adhering to fairness as the primary value would entail that all the parties involved are doomed to a strictly Pareto inferior position (Kaplow and Shavell 2002) that is, fairness would condemn everybody to faring worse. As we have shown, on the contrary, requiring incentive compatibility in terms of an equilibrium condition entails exactly the opposite. *Equality comes first*, and only on 'second thoughts', within the set of egalitarian solutions (that satisfy the basic ethical requirement of the veil of ignorance), can Pareto dominance be used to choose the best egalitarian solution in terms of general acceptance and mutual benefit (for further analysis of this point see also Sacconi 2010b) .

Secondly, libertarian new-institutional economists and social philosophers, following Hayek, often argue that distributive justice is a 'mirage' which must be eschewed if the tenets of liberty and the endogenous stability of institutions (as if they could emerge as spontaneous orders) are to be maintained. But after our analysis this seems not to be the truth.

If a libertarian is mild – i.e. if s/he accepts the minimum morality contained in the idea of an original position from where the endogenous selection of an equilibrium point is framed by the 'thought-experiment' of agreeing to consider the decision in terms of exchangeable positions among the players – then only an egalitarian solution can follow. Endogeneity of the outcome, understood as a spontaneous order, is fundamental – in addition to the veil of ignorance. And both are consistent with the idea that the decision is taken from a 'state of nature' (original position) perspective. Given these premises, we may only look for the best (Paretian) egalitarian solution among the feasible outcomes.

The emphasis by libertarians on spontaneous orders entails not introducing the hypothesis of a 'deus ex machina' able to enforce whatever artificial common plan to which the participants may agree from the ex-ante perspective. On the contrary, only agreements with which they can comply by dint solely of their individual strategies are regarded as feasible. According to a libertarian, the State cannot be invoked to expand the set of possible agreements beyond those that can emerge from endogenous interaction. Even on admitting all this, however, a choice under the veil of ignorance entails that the solution must be the best (Pareto superior) egalitarian solution compatible with feasibility.

It is the relinquishment of some efficiency, not equality, that the libertarian requires. As far as self-enforceability of institutions is required fairness cannot be escaped from the original position perspective (under the veil of ignorance), while efficiency is constrained by the impossibility of assuming that any whatever constitutional agreement will be enforced. This largely changes the perspective of organization design. If business organizations are seen as orders that must first be self-sustainable and self-regulated – for example through soft laws or self-regulation – before being enacted by mandatory law and becoming part of the (state-enforced) legal system, fairness should be the first requirement of a good design.

8. REFERENCES

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

Aoki, Masahiko. 2001. *Toward a Comparative Institutional Analysis*. Cambridge MA: MIT Press.

Aoki, Masahiko. 2010. *Corporations in Evolving Diversity*. Oxford: Oxford University Press.

Bicchieri, Cristina . (2006), *The Grammar of Society*: *The Nature and Dynamics of Social Norms*, Cambridge U. P. , Cambridge.

Binmore, Ken, (1987), "Modelling Rational Players, part I", *Economics and Philosophy*. 3 , 179-214.

Binmore, Ken, (1988), "Modelling Rational Players, part II", *Economics and Philosophy*. 4 , 9 - 55.

Binmore Ken, (1989), "The social contract: Harsanyi and Rawls", *Economic Journal* , vol. 99, pp. 84-106

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

Binmore, Ken, and Adam Brandenburger. 1990, "Common Knowledge and Game Theory." In Ken Binmore, ed., *Essays in the Foundation of Game Theory*, 105–150. Oxford: Basil Blackwell.

Binmore, Ken. 2005. *Natural Justice*. Oxford: Oxford University Press.

Blair, Margaret M., and Lynn A. Stout. 1999. "A Team Production Theory of Corporate Law." *Virginia Law Review* 85:2, 247–331.

Clarkson Centre for Business Ethics (1991), *Principles of Stakeholder Management*, Joseph Rotman School of Management, University of Toronto, Canada, (reprinted in *Business Ethics Quarterly*, 12, 2, 2002, pp.257-264.)

Clarkson M.C. (1995) "A stakeholder framework for analyzing and evaluating corporate social performance", *Academy of Management review*, 20, 1, pp.92-117.

Denzau Arthur, and Douglass C. North. 1994. "Shared Mental Models: Ideologies and Institutions." *KIKLOS* 47:1, 3–31.

Donaldson T., T.W. Dunfee, (1999), *Ties that Bind; A Social Contract Approach to Business Ethics*, Harvard Business School Press.

Donaldson Thomas and Lee E. Preston. 1995. "The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications" The Academy of Management Review, Vol. 20, No. 1, 65-91.

Donaldson, Thomas. 2012. "The Epistemic Fault Line in Corporate Governance." *Academy of Management Review* 37:2. Forthcoming.

Elhauge, Einer. 2005. "Sacrificing Corporate Profit in the Public Interest." *N.Y.U Law Review*, 80:3, 733–869.

Evan, William M., and R. Edward Freeman. 1993. "A Stakeholder Theory of the Corporation: Kantian Capitalism." In Tom L. Beauchamp and Norman E. Bowie, eds., *Ethical Theory and Business,* 97–106. Englewood Cliffs, NJ: Prentice Hall.

Flannigan, Robert. 1989. "The Fiduciary Obligation." *Oxford Journal of Legal Studie*s 9:3, 285–294.

Freeman, R. Edward, and John McVea. 2001. "A Stakeholder Approach to Strategic Management." In Michael A. Hitt, R. Edward Freeman, and Jeffrey S. Harrison, eds., *The Blackwell Handbook of Strategic Management*, 189–207. Oxford: Blackwell.

Freeman, R. Edward, Jeffrey R. Harrison, Andrew C. Wicks, Bidhan L. Parmar, and Simone De Colle. 2010. *Stakeholder Theory: The State of the Art*. Cambridge: Cambridge University Press.

Freeman, R. Edward. 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

Freeman, R.Edward. and S. Ramakrishna Velamuri. 2006. "A New Approach to CSR: Company Stakeholder Responsibility", in A. Kakabadse and M. Morsing (eds.), Corporate Social Responsibility (CSR): Reconciling Aspiration with Application, New York: Palgrave Macmillan, 9–23

Gauthier, David (1986), *Morals by Agreement*, Oxford, Clarendon Press.

Gelter, Martin. 2009, "The Dark Side of Shareholder Influence: Managerial Autonomy and Stakeholder Orientation in Comparative Corporate Governance." *Harvard International Law Journ*al 50:1, 129–134.

Griffin, J. (1998). *Value Judgment. Improving our ethical beliefs* (p. 192). Oxford: Oxford University Press

Grimalda, Gianluca, and Lorenzo Sacconi. 2005. "The Constitution of the Not-for-profit Organization: Reciprocal Conformity to Morality." *Constitutional Political Economy* 16:3, 249–276.

Grossman, Sanford J., and Oliver Hart. 1986. "The Costs and Benefit of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94:4, 691–719.

Hampton J. (1987*), Hobbes and the Social Contract Tradition*, Cambridge U.P.

Hansmann, Henry. 1996. *The Ownership of the Enterprise*. Cambridge, MA: Harvard University Press.

Hare, Richard M. 1963, *Freedom and reason*, Oxford, Oxford U.P.

Hare, Richard M. 1981. *Moral Thinking*. Oxford: Clarendon Press.

Harsanyi, John C. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98:6, 1119–1158.

Hayek, Fredrick A. 1973. *Law, Legislation and Liberty*. Chicago: University of Chicago Press.

Hobbes, Thomas. 1651, *Leviathan: Or the Matter, Forme, and Power of a Common-Wealth Ecclesiasticall and Civill*, (ed. by Ian Shapiro , Yale University Press; 2010).

Hume David, 1739, *A Treatise on Human Nature*,(ed. by David Fate Norton and Mary, Oxford University Press, 2000)

Michael C., 2001, "Value Maximization, Stakeholder Theory, and the Corporate Objective Function." *Journal of Applied Corporate Finance* 14:3, 8–21.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3:4, 305–360.

Kaplow, L. and S. Shavell (2002), *Fairness versus Welfare*, Cambridge: Harvard University Press.

Kaufman, Allen. (2002), "Managers' Double Fiduciary Duty: To stakeholders and to Freedom", *Business Ethics Quarterly*, 12, 2, pp.189-213

Keeley, Michael . (1988), *A Social-Contract Theory of Organization*, University of Notre Dame

Kreps, David. (1990), 'Corporate Culture and Economic Theory', in Alt J. and K. Shepsle (eds.), *Perspective on Positive Political Economy*, Cambridge: Cambridge University Press.

Lewis, David. 1969. *Convention. A Philosophical Study*. Cambridge, MA: Harvard University Press.

Macey, Jonathan R. 2008. *Corporate Governance*. Princeton NJ: Princeton University Press.

McMahon, Christopher. 1989. "Managerial Authority." *Ethics* 100:1, 33–53.

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18:2, 155–162.

Pettit, Philip. (1990), «Virtus normativa. Rational Choice Perspectives », Ethics, 100, pp.725-755

Posner, Erik A. 2000. *Law and Social Norm*. Cambridge, MA: Harvard University Press.

Rawls, John. 1971. *A Theory of Justice*. Oxford: Oxford University Press.

Raz, Joseph. 1985. "Authority and Justification." *Philosophy and Public Affairs* 14:1, 3–29.

Sacconi, Lorenzo, and Marco 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* 21:2, 171–201.

Sacconi, Lorenzo, Marco Faillo, and Stefania Ottone. 2011. "Contractarian Compliance and the 'Sense of Justice': A Behavioral Conformity Model and Its Experimental Support." *Analyse & Kritik* 33:1, 273–310.

Sacconi, Lorenzo. 2000. *The Social Contract of the Firm: Economics, Ethics and Organization*. Berlin: Springer Verlag.

Sacconi, Lorenzo. 2006a. "CSR as a Model of Extended Corporate Governance, an Explanation Based on the Economic Theory of Social Contract, Reputation and Reciprocal Conformism." In Fabrizio Cafaggi, ed., *Reframing Self-regulation in European Private* Law, 289–346. The Netherlands: Kluwer Law International.

Sacconi, Lorenzo. 2006b. "A Social Contract Account for CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification." *Journal of Business Ethics* 68:3, 259–281.

Sacconi, Lorenzo. 2007. "A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity." *Journal of Business Ethics* 75:1, 77–96.

Sacconi, Lorenzo. 2010a. "A Rawlsian View of CSR and the Game Theory of Its Implementation (Part I): The Multistakeholder Model of Corporate Governance." In Lorenzo Sacconi, Margaret Blair, R. Edward Freeman, and Alessandro Vercelli, eds., *Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*, 157–193. Basingstoke: Palgrave Macmillan.

Sacconi, Lorenzo. 2010b. "A Rawlsian View of CSR and the Game Theory of Its Implementation (Part II): Fairness and Equilibrium." In Lorenzo Sacconi, Margaret Blair, R. Edward Freeman, and Alessandro Vercelli, eds., *Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*. 194–125. Basingstoke: Palgrave Macmillan.

Sacconi, Lorenzo. 2011. "A Rawlsian View of CRS and the Game of Its Implementation (Part III): Conformism and Equilibrium Selection." In Lorenzo Sacconi and Giacomo Degli Antoni, eds., *Social Capital, Corporate Social Responsibility, Economic Behavior and Performance*, 42–79. Basingstoke: Palgrave Macmillan.

Schotter, Andrew. 1981. *The Economic Theory of Social Institutions*. Cambridge: Cambridge University Press.

Skyrms, S. (2004), *The stag-hunt and the evolution of the social structure*, Cambridge U.P.

Stout, Lynn A. 2006. "Social Norms and Other-Regarding Preferences." In John N. Drobak, ed., *Norms and the Law*, 13–35, Cambridge: Cambridge University Press.

Stout, Lynn A. 2011b, "New Thinking On Shareholder Primacy." Working paper , School of Law, UCLA.

Sugden, Robert. 1986. *The Economics of Rights, Co-operation and Welfare*. Oxford: Basil Blackwell.

Ullmann-Margallit, Edna, 1977, *The Emergence of Norms*, Oxford: Clarendon Press.

Vanberg, V. J. (1992), 'Organizations as constitutional systems' *Constitutional political Economy*, Vol. 3, n. 2.

Wieland, Joseph, ed. 2003. *Standards and Audits for Ethics Management Systems – The European Perspective*. Berlin: Springer Verlag.

Willamson, Oliver. 1975. *Market and Hierarchies*. New York: The Free Press.

Williamson, Oliver. 1986. *The Economic Institutions of Capitalism*. New York: The Free Press.

Young Peyton H.(1998). *Individual Strategy and Social Structure , an Evolutionary Theory of Social Institutions,* Princeton: Princeton U.P.