

EconomEtica

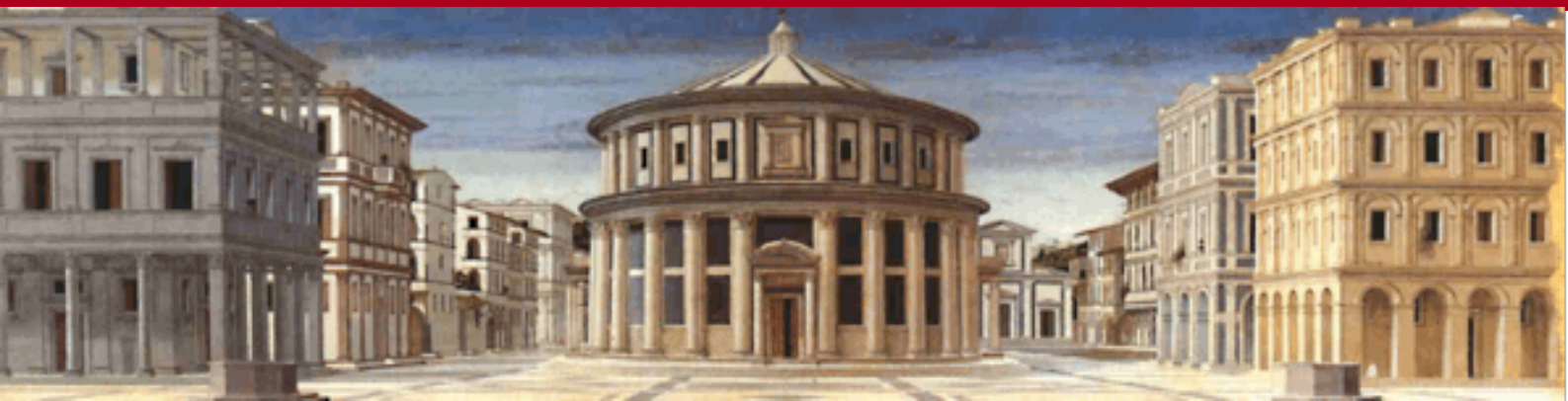
Centro interuniversitario per l'etica economica
e la responsabilità sociale di impresa
promosso dalla Fondazione Italiana Accenture

N.26 March 2011

Lorenzo Sacconi

Multi-stakeholder governance for
effectively sharing social
responsibility 1
(social contracts, deliberative
democracy and endogenous
conformity)

Working papers



Multi-stakeholder governance for effectively sharing social responsibility¹

(social contracts, deliberative democracy and endogenous conformity)

by Lorenzo Sacconi

1. Introduction

The idea of shared social responsibility (SSR) is premised on the definition of a unified methodology of multi-stakeholder governance making that effective the idea of ‘sharing’ social responsibilities. At minimum in fact, shared social responsibility is a matter of allocating responsibility among a number of public, private, individual or organized subjects in relation to social welfare principles and goals that may be shared at European level, and then manifested on regional and local level. Such subjects differ in nature, and their competence and powers have different magnitudes and exist at different levels. Allocating responsibilities, establishing rules for their fulfilment, and enabling cooperation and coordination among such subjects are typical purposes of a multi-stakeholder governance mechanism (for the managerial literature on stakeholder theory see Freeman. 1984, Freeman, and Evan 1990, Donaldson and Preston 1995, and Clarkson 1999).

Hence the main questions that this chapter aims to address are:

- What governance mechanism, mainly on local and regional level, can make the idea of shared social responsibility effective? Is there a model and a methodology of governance based on the sharing of social responsibility among multiple stakeholders that can also be “shared” on European level?
- Which specialized areas and levels (companies, local communities, or multi-agent networks) of ‘shared social responsibility’ can be identified? And what different specifications of the overall model of governance are required in order to account for different configurations of stakeholders, along with their relative power, knowledge and degree of mutual dependence?

¹ To be published in *Towards an Europe of shared social responsibilities. Challenges and strategies*, Trends in Social Cohesion Series, n. 26, 2011, Council of Europe Publishing, Strasbourg

These questions will be addressed in the following pages by developing the basic concepts for an SSR governance methodology. After having worked out a definition of SSR as a *multi-stakeholder governance model on territorial level* based on the analogy with a previous definition of *corporate social responsibility* (Sacconi 2006a,b) and its identification with an endogenously *stable social institution* (Aoki 2001, Sacconi 2010a), the social contract approach to SSR is put forward. Further analysis of two aspects of the social contract model will be suggested in order to provide a contractarian foundation for SSR as multi-stakeholder governance: (i) the idea of a '*local social contracts*', that renders affordable to account for local governance structures in terms of the social contract model (Donaldson and Dunfee 1995, 1999); and (ii) the game theoretical reformulation of the *Rawlsian maximin principle* that uniquely specifies the egalitarian terms of the social contract among stakeholders and understands it not only as satisfaction of a moral requirement but also of a *stability requirement* to be met when an 'original position' perspective is adopted in order to explain SSR as a social institution (Binmore 2005, Sacconi 2010b). Thereafter, the design of the governance methodology that can implement these ideal models of fair bargaining will be introduced. A key role is played by the idea of *deliberative democracy* (Gutman and Thompson 1996, 2005) as the deliberative procedure, constrained by a set of formal and substantial impartiality conditions, to be implemented within SSR governance models. Deliberative democracy makes it possible to constrain governance mechanisms with procedural principles that allow for real life bargaining among stakeholders in order to approximate the model of fair '*local social contracts*'. The implementation of policies through SSR will then be discussed by analyzing the conditions for the emergence of *endogenous motivations and incentives* that allow considering SSR as a self-sustaining norm of behaviour. The basic motivational mechanisms considered will be *reputation* effects (Kreps 1990, Sacconi 2000) and the *preference for reciprocal conformity with a fair norm* or 'the sense of justice' (see Rawls 1971, Grimalda and Sacconi 2005, Sacconi 2007, Sacconi and Faillo 2010, Sacconi 2011). Finally, these endogenous motivations will be studied with reference to specific interaction structures (games) among different stakeholders and hence particular governance models: the concentric model exemplified by CSR (Freeman 1984, Freeman and Evan 1988, Donaldson and Preston 1995, Sacconi 2006b), the egalitarian community model (Ostrom 1990, Ostrom 2009) and the network of heterogeneous stakeholders model (Degli Antoni and Sacconi 2011).

2. ‘Shared social responsibility’ as a social institution of multi-stakeholder governance

There is at present no well-developed view of ‘shared responsibility’ as a governance model. But one can be worked out by exploiting the analogy between shared responsibility and corporate social responsibility (CSR), as the latter has also been defined as *a model of extended corporate governance whereby those who run a firm (entrepreneurs, directors, managers) have responsibilities that range from fulfilment of fiduciary duties towards the owners to fulfilment of analogous – if not identical – fiduciary duties towards all the firm’s stakeholders* (see Sacconi 2006a,b, 2010a).

Since CSR is only one instance, or a special case (see sec. 7), of the wider concept of ‘shared social responsibility’, insofar as the analogy holds it can be exploited in order to define shared social responsibility (SSR) as well. It can thus be understood *as a model of multi-level and multi-stakeholder governance* qualified as (i) a new social institution, involving other formal institutions and organisations of various natures operating on different levels (local, regional, national and continental); (ii) inducing multiple stakeholder reciprocal coordination and cooperation through the building of fiduciary relations; (iii) having reference to interactions and outcomes occurring in a single territory or in a related group of territories; (iv) which is capable of settling impartial agreement on social justice and social cohesion principles, goals and policies; (v) and inducing a significant level of endogenous adhesion also through implementation by the stakeholders themselves. Such interactions and outcomes may comprise some of the following: the provision or depletion of a common or (possibly local) public good, the supply and distribution of a certain (related) set of primary goods or certain social welfare services, some negative/positive externalities of a given market exchange or the inequalities and unfairness generated by abuses of authority or opportunistic behaviour within some (identified) organisations or the management of certain incomplete contractual relationships affecting a definite number of stakeholders (which means that they are those ‘relevant’ for each particular domain).

2.1. The role of fiduciary duties

The main difference between the two definitions (SSR and CSR respectively) is the different emphasis on the role of ‘fiduciary duties’ owed to stakeholders by those who occupy a position of authority. This is not surprising, given that firms in general and the corporation in particular are hierarchical structures of governance that are largely dependent on a previous allocation of ownership and control, which identifies who is in a position of formal authority

(even though this is not sufficient to explain why that authority emerges and is stable, i.e. why it is ‘accepted’), both because he/she directly exercises residual rights of control or delegates control to a board of directors. Due to the fact that it relates first and foremost to a hierarchical structure, corporate social responsibility is mainly defined as a matter of extended fiduciary duties that whoever occupies the position of authority in the company owes to those (non controlling stakeholders) who do not formally hold control and residual decision making rights, but who are nevertheless subject to authority and establish fiduciary relations with the company as part of the pursuit of their stakes in it.

On the contrary, SSR does not formally relate to an analogous pre-existing hierarchical structure and to formal authority positions, and can be seen as a way to extend the democratic method of decision making outside the traditional institutions of democratic representation, by establishing participatory and self-regulated organizations in which responsibility for decisions and implementation can be shared horizontally. However, save for a specific sub-case (see “the concentric model” point (a) sec. 9), SSR is a governance structure where no stakeholder has a clear position of superior authority based on a source of legitimization such as ownership.

The only possible analogy would be with the authority held by constitutional public government institutions (local, regional, national or European) which certainly make up a significant part of any SSR governance mechanism. But this would not be an analogy to be emphasized. In fact, in case this position of authority were the central feature of the governance model under consideration, our emphasis would not be placed on the ‘sharing’ of social responsibilities among public and private institutions, organizations and groups. By contrast, it would be focused more traditionally on the democratic accountability of those in a position of political and administrative formal authority, on the role of officials elected or appointed as public trustees (for example at the local government level), and their fiduciary duties owed to citizens as such (as a constitutional abstraction), but also as stakeholders advancing different kinds of claims. There is no reason to exclude these aspects from the discussion on SSR, but it is fairly clear that this is not at the very heart of a definition of SSR as a model of governance. Quite on the contrary, the definition of SSR implies the idea that some part of that public authority must be ‘shared’ among the representatives of different governmental levels and different types of organisations, or that at most it should be deployed as ‘soft power’ with the goal of building up participatory structures in which different

stakeholder groups are involved, cooperate by “sharing” some decision making right, and hence embrace responsibility for the implementation of policies.

Nevertheless, this distinction must not be pushed too far. When establishing such participatory governance mechanisms, the representatives of some public or private organizations – i.e. that represent various stakeholders and constituencies – will gain some influence over the deliberative process relating to the settlement of the principles, goals and policies of social cohesion. Beyond the natural leading role of public government representatives – which cannot be overlooked, though it is also problematic, since participation by public officials in these multiple stakeholders governance structures runs the risk of being captured by vested interests or stakeholder groups essential to their re-election – there are also well organized stakeholder groups or organizations that are able to take on a leadership role (informal authority) in the governance mechanism of SSR and that are hence capable of exercising greater influence than others. Thus, in one way or another some authority will be *de facto* exerted, which complements the formal authority held by governing bodies delegated by constituencies such as stakeholder associations, public institutions and organizations and informal groups participating on different levels.

Therefore, while in principle the governance structures based on the idea of SSR can be regarded as forms of self-governance within a given policy area and a given territory, as a matter of fact it may be expected that hierarchies of power and authority will emerge, which may be legitimate according to formal procedures internal to the governance mechanism, whilst on the other hand these may also be established *de facto*, with the result that not all the participating stakeholders will be on an equal footing in terms of influence. According to this preliminary descriptive overview (which must be kept distinct from the later normative goal of this chapter), it must be recognized that some stakeholders will be stronger than others, since they have more homogeneous interests, are already concentrated within business and professional associations and organisations, are not dispersed over a large territory and possess valuable resources to be dedicated to coordination and political pressure. Most of them hold merit-based claims (see Sacconi 2011 *infra*) over shares of the value that originates from the implementation of the same policies in which they are involved as implementers. These claims are logically correlated to the control over valuable resources that can be used in productive processes. However these claims are not the most urgent, nor do they have moral priority or the greatest legitimacy – what by contrast can be said of need-based claims

frequently held by weaker stakeholders who are less organized, more dispersed throughout a territory, and by definition less endowed with resources (see again Sacconi 2011, for the definition of different stakeholders' claims and their priority ranking according the social contract model).

Therefore, it can be expected that also in the multistakeholder governance structure, power and influence are not naturally aligned with the urgency, moral legitimacy and priority of stakeholders' claims. It follows that one basic component of the proper design of SSR as a social institution of governance is to include the settlement of fiduciary duties - i.e. obligations and responsibilities restraining and orienting the trustees' discretion to render its exercise ultimately beneficial to their "trustors". Agents in a position of authority or who occupy a privileged position in terms of the availability of resources and claims owe fiduciary duties to stakeholders whose claims are most urgent and retain moral priority, even though they often have fewer resources to be invested in influencing deliberations.

There is no doubt that from a normative point of view the proper design of SSR governance structures should be aimed at rendering the obligations at the basis of shared responsibility reciprocal and even symmetrical, if not identical. This entails empowering each stakeholder so as to enable him/her – as an autonomous agent equally deserving consideration and respect – to participate on equal terms in the deliberation process on the settlement of principles, standard, goals and policies. Moreover empowerment also gives weak stakeholders a role on the policy implementation level, so that relationships of trust must be mutual with reference to the cooperation required in order to achieve goals and implement policies. Nevertheless, such mutuality does not eliminates the need for the players that occupy a powerful position and have a prominent role in the deliberative and implementation process to discharge their fiduciary duties toward all the stakeholders involved.

2.2. Self-regulation

As for CSR, a commonly-held tenet concerning SSR is that it should extend beyond what can be imposed by legislation on local government, private companies, non-profit organizations and associations, as well as individual citizens, with the result 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* (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 a body (i.e. an organization) endowed with its own 'natural' (so to speak

‘unchosen’) internal regularity of functioning whereby its behaviour is completely endogenously directed, with no 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 result of an agreement (explicit or implicit) among individual members of more or less extensive social groups – whereby they establish and adhere to a set of principles or rules expressed (in language, and with a normative content that they understand and which gives them guidance by vetoing some actions and recommending others), but which is *not enforced* by any external authority imposing sanctions, because this is instead performed through the voluntary compliance by individual members of the relevant social group (see Sacconi 2007b, Sacconi 2010a). The self-regulatory nature of shared social responsibility is understood here in accordance with the second view. In particular, this paper endorses the following view of social shared responsibility (SSR) as an effective system of social self-regulation (also in analogy with CSR):

- a) SSR is established by social norms such as a multi-stakeholder charter of rights and obligations, the associated governance standards and management systems, and not merely by discretionary decisions;
- b) these include normative utterances such as general abstract principles and preventive rules of behaviour concerning fiduciary duties, general statements of the principle of fair treatment for each stakeholder, principles of inter-stakeholder justice and fair balancing, and precautionary rules of behaviour in any critical sphere of potentially opportunistic behaviour between any organisation and its stakeholders participating in a territorial network or within nested networks - so that fiduciary duties and related rights may be put into practice by precautionary rules of conduct that pre-empt opportunistic behaviour in typical critical situations;
- c) such norms are agreed upon by the local stakeholders through (voluntary) forms of multi-stakeholder social dialogue (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 local authorities, companies, organizations and associations without external legal enforcement, but rather through reciprocal agreements, the adoption of internal statutes, codes of ethics and standards of behaviour (which typically falls within the domain of soft law), thereby reshaping the organizations’ governance and

participatory structures, their self-organization, training, auditing and control, and which are compatible with voluntariness at the individual organization level;

- e) the previous self-enforcement approach does not prevent self-regulation from being monitored and verified by independent third-party civil society bodies (which do not have conflicts of interest with their mission of providing an impartial overview of public and private organizations voluntarily subjected to self-regulation); this enhances the level of information and knowledge whereby stakeholders define their expectations about the entire system of agent conduct.

Of course, effective SSR self-regulation is a viable option only within an institutional and legal environment that does not obstruct it. Such obstruction occurs, for example, in the field of CSR when overly narrow definitions of the firm's objective-function prescribe the principle of 'shareholder value maximization' as the company's only goal, or the pursuit of the mutual interest of internal members as the only goal of co-operative firms – as occurs today in many company laws at European level (however, consider section 172 of the 2006 "UK Companies Act" for a significant exception).

Laws at the European level should be reshaped in order to *enable*, where necessary, the self-regulation of multi-stakeholder governance systems that are able to allocate and implement shared social responsibilities. However, assuming that company laws or public administration regulations do not obstruct proper social self-regulation, the thrust of my argument is that the endogenous beliefs, motivations and preferences of social agents (local authorities, public servants, companies, non-profit organisations and associations and all their stakeholders including private citizens) are the essential forces driving the implementation of SSR through a model of multi-stakeholder governance.

3. The concept of institution and its normative meaning

Making sense of SSR as an explicit social norm requires a definition of *institution* that is different from the 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 an appropriate concept of institution. 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 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 about one another’s actions henceforth. Boundedly rational players who do not have complete information use this mental compressed representation, in order to form beliefs about how any other player is currently playing the repeated game. And these beliefs are *shared* – in the sense that any two players will make the same prediction about any other player involved – as well as consistent – in the sense that the beliefs whereby any player makes his choice also cohere with his prediction of beliefs whereby other players make their choices. These beliefs replicate the prediction that a particular equilibrium has been established among the many possible, and it is from such beliefs that all players derive their best actions. Since these actions are best responses to beliefs, and since these beliefs correctly summarize current behaviour, these actions are also the best responses to the other players’ actual actions, as represented by beliefs. Thus the derived action profile satisfies the typical Nash equilibrium condition.

An important consequence of Aoki’s view is the following. Even though a statutory law passed by the parliament or another legislative body may expressly specify rights and duties, if there is *no* shared belief that it will be complied with by those who ‘should’ do so, then it will not be considered an *institution*. Instead, the ongoing practice of violating the statutory law could be regarded as the ‘true’ institution of the relevant action domain (see Aoki 2001)

Nevertheless, at first glance, this definition has a major drawback. Institutions thus defined seem to lack any significant normative meaning and force. On the contrary, institutions such

as constitutions or laws, ethical codes, shared social values, organizational codes of conduct and procedures have a primarily prescriptive meaning (in the case of ethics such meaning requires “universalizability” (see Hare 1981)). They are guides to action and not just descriptions of states 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 by contrast appear to give an indication as to 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. As thus defined, an institution tells players only how to maintain the existing pattern of behaviour because it is an equilibrium supporting the existing belief system. An institution such as this seems to have no normative content.

But why would an institution 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? The point is that a necessary component of the belief system defining an institution must not merely replicate the description of behaviour in a given action domain; it must instead prescribe it independently of the description of the ongoing course of action. In other words, it rests on some *a priori* standpoint. Arguably, this is a *necessary* though not sufficient condition for an institution to exist (the condition of sufficiency would be that also the definition of the belief equilibrium were met).

Clearly, this presumes that the belief system supporting an institution also exercises *ex ante* a *justificatory force* capable of achieving the *general acceptance* of some new equilibrium within a given domain where agents find themselves out of equilibrium or where equilibrium has not yet been reached because of some unexpected change in technologies or the social or ecological environment. Once the regularity of behaviour has been generally accepted through an *ex ante* agreement, then (and only then) may it become the ‘salient’ basis for the reciprocal prediction of all of the participants’ actions.

Thus, a second component of the definition of institution – incorporating 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, or values and obligations, which needs to have a prescriptive and universalizable meaning capable of *justifying* its shared acceptance by all participants within a given interaction domain. Because

it is *ex ante* accepted by all players, it enters into their shared mental model (Dezau and North 1994) as to how the game should be played and hence becomes the basis for their coordination according to a specific equilibrium under a given action domain. The key point is then to explain *how* a normative system of beliefs that precedes the evolution of the corresponding equilibrium can *subsequently be accepted by all agents* in the relevant domain.

4. Social contract

To my knowledge, the best justification for norms on the responsible exercise of authority, discretion, autonomy etc. that accounts for *ex ante* shared acceptance is the *social contract model*. Contractarian norms result from a voluntary agreement in a hypothetical situation of original choice which logically comes before any exogenous institution that is super-imposed on a given action domain, or prior to the emergence of any institution (in the equilibrium sense). Thus a norm (and the institution that may encapsulate it) arises and can be maintained only because of the voluntary agreement and acceptance of agents.

With regard to the definition of agreement on a justifiable norm, any social contract model excludes threats, fraud and the manipulation of resources that would render the parties substantially unequal in terms of their bargaining power. In addition to the normative reason for disregarding them, these initial conditions may be viewed as the effect of institutions that already exist in some adjacent domain and which endow some players with more strength than others. The hypothetical choice under the original position is made 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 the rational autonomy, decision-making freedom and intentionality of each participant, which are assumed to be *equal* in weight for all participants in the contract. (This can be disputed on an empirical basis, but *in principle* the idea is to set aside any morally irrelevant difference between participants.) The agreement thus gives equal consideration and respect – i.e. equal treatment – to reasons, interests and decisions put forward by each party to the contract.

It is not only the initial adoption 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 to be a matter of voluntary adhesion. Thus the endogenous nature of institutions with respect to the agents’ voluntary interaction is respected during both stages: an institution is endogenous to the players’ *ex ante* strategic interaction understood as *rational bargaining* among equally situated rational agents, i.e. which can be started only by

the unanimous decision of the players to enter into a voluntary agreement. Moreover, the *ex post* implementation of an institutional arrangement is also regarded as the composition of the autonomous decisions that players make in their strategic interaction when choosing whether or not to comply with the social contract by taking decisions that reflect the entire set of their reasons and motives to act.

Since the social contract is a “thought experiment”, it would impress the players’ minds with a mental model as to how the game should be played and generate an identical ‘salient’ aspect of their interaction, favouring effective coordination over a specific equilibrium point to be achieved by choosing each action. When the shared system of mutually consistent beliefs has been formed for the first time, it provides the basis for a regularity of behaviour (an equilibrium) that also confirms the same beliefs system. The summary information compressed into a mental representation of the regular players’ behaviour throughout repetitions of the game, generated by *ex ante* acceptance of the normative beliefs that a particular equilibrium is to be achieved, can then be understood as an institution.

Thus, the two requirements of (i) the 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, taken jointly, seem to provide the comprehensive definition of SSR as an institution which should guide the building of its proper governance mechanism .

5. Local social contracts

There are many different accounts of the social contract. For example, both Rawls’ and Gauthier’s accounts are compatible with what has been said thus far. However, Rawls’s idea of the original position is basic to the purpose of defining the extent and allocation of shared social responsibility in such a way that these are the normative results of widely accepted moral principles. It is a choice condition requiring unanimous agreement under a ‘veil of ignorance’ as to any details of each participant’s personal identity and social position.

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 this symmetrical exchange of positions, whereby all parties assess the acceptability of any given set of normative statements, agreement is reached that reflects a reasonable impartial combination of all the reasons to act considered in turn. Importantly, the agreement accepted by each of them is necessarily unanimous, as the symmetrical replacement of personal positions is carried out in an identical manner by all of the parties involved, and so they are identically situated in their exercise of institutional assessment.

Thus, it is agreement under the ‘veil of ignorance’ among all the relevant stakeholders that should generate the shared acceptance of a set of social responsibilities among the relevant participants – whether interested parties are individual or organized stakeholders such as local government offices, public service providers, for-profit companies, co-operatives or non-profits associations.

Admittedly, it may seem odd to think of social contracts involving stakeholders as special individuals and groups, and also organizations that (from the legal point of view) are artificial persons with a derivative nature with respect to the overall legal order (perhaps based on the macro social contract over the constitution). However, the idea of the social contract can be considered at both a macro and a micro/local level.²

The macro social contract applies to all of the individual potential participants in the original position, albeit according to different interpretations (for example Rawls 1971, Gauthier 1986, Binmore 2005). Local social contracts apply to individual communities, where communities are self-defined/self-circumscribed groups of persons who interact in the context of shared tasks, aims, or values and who can establish norms of ethical behaviour for themselves (Donaldson and Dunfee 1995, see note 3) (examples of communities are: firms and/or their single offices, cities or part of cities etc.).

The idea of the micro social contract seems to be very useful in order to apply the general concept to a particular institution such as the firm (see Sacconi 2000, 2005, 2006a,b and 2010a) or a territorial system of governance aimed at the implementation of shared social responsibility. Local social contracts establish norms and principles of social justice and wealth distribution at local or micro level with reference to relevant local public goods or commons, a social services provision problem, local externalities of private market

² I follow the approach proposed by Donaldson and Dunfee (1995) who refer to macro and micro social contracts. Donaldson, T., & Dunfee, T. W (1995), Integrative social contracts theory. *Economics and Philosophy*, 11, pp.85-112., see also Dunfee and Donaldson (1999).

interactions or corporate activity, or also the social costs of abuses of authority within corporations or productive organizations operating locally. Such social norms serve as guides for the parties' behaviour so that shared responsibility can be allocated among them.

The roles of the local social contract under a multi-stakeholder governance model are manifold (as will be clear at the end of this section). They must be able to (i) define norms of responsible behaviour among participants that are recognizable as genuine bases for binding obligations (at least in ethical if not legal terms), and hence responsibilities; (ii) identify behaviour that is not beyond the practical capacity for action of individuals or organized agents in that it coincides with behaviour that they are in practice motivated to follow once they have reciprocal expectations that others will also do the same; and (iii) endow participants in the local social contract with the requisite social capital consisting of motivations and beliefs that make them trustworthy for other agents.

Before dealing with points (ii) and (iii), some considerations on point (i) are in order. In fact, one of the possible drawbacks of considering small-scale social norms as the result of local contracts among the stakeholders involved in particular communities is that the resulting norms could be binding at the community level while nevertheless being morally arbitrary. Their stability could not be based on a justification that could be accepted voluntarily by all, but rather on social pressures, threats of ostracism, force exercised against minorities of dissenters, etc.

Local social contracts must therefore not be arbitrary. In fact, they are stipulated within a framework defined by hyper-norms established by the macro social contract (see Donaldson and Dunfee 1995). Hyper-norms imply principles so essential to human existence that they serve as guidance in evaluating lower-level moral norms, with the result that they must be respected by all people regardless of their membership of different local communities. Examples of hyper-norms are those contained in the Declaration of the Rights of Man and of the Citizen or national constitutions.

Nevertheless, the idea of local social contracts concedes that the macro social contract may be deliberately left incomplete (general and abstract principle) so that it can be adapted in line with unforeseen contingencies. This is provided for by allowing local social contracts to specialize general principles, values and norms with respect to particular subsets of interactions and social domains. This renders some norms endogenous within (and emerging from) certain interaction domains.

What is remarkable in this view, however, is that the legal system does not simply delegate absolute decision making rights to actors (for example property rights): it grants space for ‘social contract refinement’ capable of constraining these rights with responsibilities endogenously emerging from small-scale social contracts that can shape small-scale institutions such as corporate governance systems, local governance systems, etc. The incomplete nature of rules and contracts thus does not grant complete discretion to right holders, because such discretion can be constrained by specialized social contracts that, in a world of perfect knowledge, would have been part of the overall set of norms stemming from the macro social contract from the outset (i.e. they would have been endorsed under the ‘original position’).

There are also inherent features of the concept of micro social contracts that protect them against the risk of moral arbitrariness. Authors state that local social norms or micro social contracts must guarantee *genuine consensus* by always leaving participants the option of resorting to ‘voice and exit’. Moreover, their wide acceptance must be mutually acknowledged, and so too the disposition to conform once others are also expected to conform (see again Donaldson and Dunfee, op cit.)

Admittedly however, local social contracts need to be understood in a less generic and lazy manner than has been the case in part of the relevant literature. In order to show that the micro social contract is an appropriate source of genuine (moral) obligations, and hence of shared responsibilities, whilst at the same time being consistent with the agent’s incentives and motivations that entail the stability of the relevant social micro-level institution, we must consider the contractarian methodology in more depth.

6. A proper modelling of micro-level social contracts

Micro-social contracts too can be modelled as social contracts in the proper sense and developed according to a rigorous contractarian methodology (see Binmore’s *Natural justice*, 2005, and also Sacconi 2000, 2006a, 2010b), albeit on a small scale – i.e. restricted to stakeholders interacting within a given domain characterized by relative independence from other social interaction domains. Some examples of semi-independent domains are provided below.

- A small-scale ‘state of nature’ arises in the case of a firm facing a ‘team production problem’ characterized by the possibility of opportunism within the team, and also by

the risk of abuse of authority – which is partially insulated against legal norms by the incomplete nature of contracts and the failure of public regulation due to information asymmetries.

- A local ‘state of nature’ that arises within a community handling the problem of free riding on the costs of maintaining a common resource under the competence of the local community itself.

A remarkable result of a rigorous contractarian methodology is that a local/micro social contract need not be seen as either a mere outcome of a bargaining game without ethical content (because it is affected by arbitrary bargaining and threat power) or as wishful thinking with no correspondence to the effective equilibrium outcome of the real-life interaction among concrete stakeholders involved in a given context. We shall now consider in greater detail (see again Binmore 2005) the basic ideas.

Let us assume that a set of possible states is settled as the feasible outcome of the stakeholders’ interaction (in the relevant domain). Let these outcomes be defined so that they all satisfy the condition of being equilibria – in the sense that when the behaviour corresponding to one of these outcomes is followed by each player in the relevant domain then nobody else in the same interaction domain has any incentive to deviate from the given behaviour.

We then assume that before agents engage in the relevant interaction (for example, by playing through their strategies in an incomplete contract situation, trusting each other in a public good problem, or trusting a company that may abuse its stakeholders), they may wish to agree *ex ante* on how to select one of the possible equilibrium points/outcomes.

Let us also assume that they are looking for an impartially and impersonally acceptable agreement on a single outcome, because what they are seeking is a genuine set of responsibilities that may be ascribed out of a minimal set of (ethical) duties (i.e. they want to determine the equilibrium behaviour of players that is most consistent with some acceptable notion of stakeholders’ duties).

‘Impersonally’ here means that acceptance must not depend on an individual’s own personal position. Thus players must select a solution that remains unchanged under the symmetrical replacement of agents with respect to their social role and personal position. Moreover ‘impartiality’ means that they must agree on an outcome under the hypothesis that the

reciprocal replacement of players' positions assigns equal probability to the chance of finding oneself in the position of each of the possible individuals, when endowed with all of their own social and personal characteristics. This of course is the 'veil of ignorance' assumption.

However, it should be noted that this is not an excessive idealization of the local social contract among stakeholders. Beyond impersonality and impartiality, agents retain their awareness that the solution must be an equilibrium of the original game – that is, a solution that the parties have incentives to self-impose on themselves insofar as they entertain the shared belief that they all are playing the same solution. This is a requirement of realism that requires the stability and incentive compatibility of the agreed solution: you cannot afford to agree *ex ante* on a solution if it is not incentive-compatible *ex post*.

In particular, under conditions of impersonality and the idea of solution invariance to permutations of the players' positions, the stability condition requires that the selected solution must correspond to an equilibrium point coinciding with an outcome which is *ex post* stable under any place-permutation whatsoever with respect to players' social and personal positions. In other words, the selected outcome must be an equilibrium (say) either if player 1 takes the point of view of Adam (and player 2 respectively takes the point of view of Eve) or in the symmetrically opposite case when their identification is reversed (player 2 occupies Adam's position, whereas player B takes Eve's point of view). This means that the solution must be found within the symmetric intersection of the two outcome spaces which are generated from the symmetric translations of the original set of possible outcomes with respect to the players' positions.

A significant result of this construction is that the minimal requirement of social justice (impersonality and impartiality of the solution) becomes compatible with realism and *ex post* stability. But, remarkably, stability under conditions of impersonality does not make it necessary to relinquish the moral demands of social justice. On the contrary, it entails that the solution must be *egalitarian* and must coincide with the *maximin distribution*, even within an asymmetrical set of possible outcomes. Thus, given a real-life set of possible outcomes reflecting possible inequality between the participants, the selected solution falls on the equilibrium that favours the worst-off player most, which in most cases is the egalitarian distribution. It should be noted that this also holds true in situations where inequalities of force, resources or productivity impinge upon possible outcomes, with the result that some of them give one player (for the sake of argument, Adam) advantages that are not affordable to

the second player (Eve) under any outcome in the feasible set (this means that the underlying social situation from which the social contract is selected is not an idyllic ‘already fair’ context, and hence that fairness can only be introduced through the social contract selection).

The basis for this result is a *feasibility condition*: only outcomes that are *feasible* can be taken as candidates for the solution. But owing to the ‘state of nature’ assumption, only equilibria of the original (convex) outcome space are feasible. Any further outcome – potentially subject to agreement – would be *wishful thinking*, because no *ex post* equilibrium would exist that could implement it. Under the conditions of solution impersonality and impartiality, we must restrict feasible outcomes to the intersection set consisting of a subset of the original outcomes resulting from symmetric translations of the outcome space with respect to the players’ positions (see fig. 1). This is necessarily a symmetrical set; and within a symmetric outcome space, any bargaining solution necessarily falls on the bisector which is the geometrical locus of egalitarian solutions where parties agree to share the bargaining surplus equally, taking for granted an egalitarian *status quo* preceding the agreement. It should be noted that this last assumption too is in fact not a precondition concerning the underlying real life game, but rather a consequence of the ‘veil of ignorance’. Given a basic unequal *status quo*, under the veil the proper *status quo* becomes the equal probability mixture of the original *status quo* and its symmetrical translation with respect to the players’ positions which also lies on the bisector.

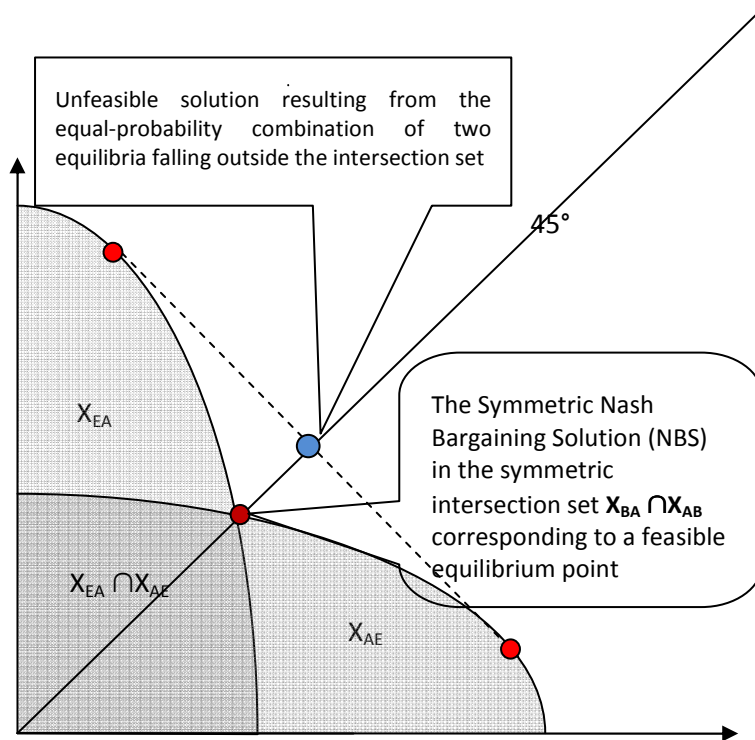


Fig. 1 the Egalitarian feasible solution and the efficient unfeasible solution (see Binmore 2005, Sacconi 2010)

Returning to the idea of shared responsibility, given impartiality and impersonality, the social contract consists of a norm of justice (i.e. a norm of welfare distribution). This is consequently the appropriate basis for *sharing responsibility* among the parties to the contract, because responsibility focuses on the conduct that they are required to carry out in order to implement the corresponding fair outcome. Under a social contract, the parties share the responsibilities for generating a fair distribution of welfare.

Moreover, these responsibilities are ascribable to participants in the agreement because they are compatible with what falls *ex post* within the range of feasible outcomes of their interaction; that is, outcomes that can be produced within their domain of choice. In fact, the above construction ensures that the (local) social contract will be reached at an *ex post* feasible equilibrium point. Participants in the contract are *not* told that they ought to do what they cannot do.

7. Deliberative democracy

It is still necessary to explain what in practice – in a concrete governance environment – will constrain agents to consider only impersonal and impartial agreements when negotiating a real (not theoretical or hypothetical) local social contract on their shared social

responsibilities: that is, solutions the acceptance of which does not depend on any particular personal point of view and that will allocate fair amounts of welfare to all the participants.

There are exceptional institutional situations of deep uncertainty regarding the future and stability of the very identities over time of players which may favour such impartial reasoning also for self-interested political players or interest groups. Typically, these are situations of constitutional choice after revolutions (a good reason for changing constitutions only infrequently, or for submitting acceptance of such changes to a Supreme Court whose members, on account of their institutional role and cultural inheritance, are committed to speaking out of “public reason”).

In laboratory and field experiments, the “veil of ignorance” can be simulated by asking subjects to perform a decision making task concerning agreement on the rules of a further multiplayer task that they will be asked to perform in the future without knowing, at the time the rules are chosen, which role they will occupy in future. Their *ex ante* discussion (where real identities are masked by the use of anonymous computer workstations and where the experimenter makes the credible promise that future roles will be assigned randomly) in practice simulates the thought experiment of decision under “the veil of ignorance” (see Sacconi and Faillo 2010) .

Neurosciences tell us that empathy, i.e. the ability to replicate within one’s own brain the same experience that the subject sees as occurring to a separate agent (assuming that the former has had a similar experience in the past, even if that agent is anonymous and has no relevant affective links with the experimental subject), is not extraordinary or unrealistic, as it seems to be the routine function performed by our ‘mirror neurons’.

However, in defining shared responsibility as the content of governance models, we must primarily resort to the middle-level institutional design of rules, organisations and roles. What concrete set of rules for a deliberative procedure leading stakeholders to a local social contract can successfully approximate the fair outcome of the contractarian model?

This is a matter of deliberative democracy. The main suggestion for the design of a multi-stakeholder governance model conducive to implementing the idea of shared social responsibilities is that *deliberative democracy* must be applied at the local/regional community level, also outside the government and its representative institution (parliaments, municipal and regional councils etc.). The following paragraphs will seek to clarify why this is the case.

First, what is deliberative democracy? It is understood as a deliberative process under which participants are committed to reciprocally offering and accepting impartial reasons to act as justifications, at least in principle acceptable by all, for any policy proposal that they are deliberating and which will be made binding for their reference group/community/nation etc.

Deliberative democracy specifies constraints whereby participants in the deliberative process will converge on an agreement on welfare and social justice principles and policies, as well as minimizing the area of their moral disagreement. Deliberation is constrained by both *substantive* principles (such as basic liberties, and basic and fair equality of opportunities), and *procedural* principles (such as impartiality and reciprocity in the process of reason-exchange, participants' integrity, publicity and accountability for any step in the deliberative process) (see Gutman and Thompson, 1996, 2005).

Even though some authors who pursue this line of thought would perhaps not agree (see again Gutman and Thompson 2005), I understand deliberative democracy as a way to operationalize the idea of an ideal bargaining process under the “veil of ignorance”, starting from a neutralized status quo that would produce a fair agreement (the social contract) and also establish the “overlapping consensus” among the participants' comprehensive ‘conceptions of the good’ – taking it for granted that some moral disagreements cannot be eliminated but only made less disruptive in face of the agreed areas of interests and values.

Why is deliberative democracy so important for the subject of shared responsibility? Since the sharing of responsibility entails the involvement of different kinds of subjects – private or public, organizations or individuals – positioned on different levels (local, regional, national or global), it requires governance mechanisms whose decision making processes are in danger of becoming overly complex, obscure and unaccountable towards the public (as suggested by Claus Offe).

During such processes, bargaining in the literal sense unfettered by any moral constraint occurs, and stronger parties can support their sectional interests by threatening to disrupt the agreement process. Corrupted agreements may be settled because stronger parties may also buy the services of politicians or bureaucrats who, in principle, are expected to play an impartial mediatory role in the process. The permeability of governance mechanisms to corruption, owing to the fact that roles in this case are by definition interlocked because participants from different levels and of differing natures join together to ‘share

responsibilities', jeopardizes both the procedural form of the decision making process as well as the substantive nature of its outcome.

On the one hand, at issue is respect for democratic equality, i.e. citizens' equal rights to participate in the decision making process and not only to have results fairly accounted for, but also the different alternatives assessed during the deliberation phase, as well as the performance of deliberated policies (at least for those that are in their interest). On the other hand, the substantive nature of the agreement reached (from which the allocation of shared responsibility is derived) – i.e. its fairness in terms of distributive effects – is also at stake. However, distrust in the fairness of the decision making procedure and in the substantive nature of its outcome would destroy shared responsibility, for no obligations can be endorsed voluntarily if their allocation is rendered illegitimate.

Shaping the model of multi-stakeholder governance according to deliberative democracy prevents both of these drawbacks. Deliberative democracy is based on the idea that all participants in the deliberative process (for example, local and regional public authorities, companies, NGOs, associations and individual citizens) must introduce into the discussion only impartial reasons to act which could in principle be accepted by all of the participants symmetrically motivated to achieve the goal of reaching an impartial agreement based on mutually acceptable reasons.

Hence a concrete multi-stakeholder bargaining process under which participants bargain under the constraint of providing only impartial justification for their proposals could be taken as an approximation to the ideal of a (small-scale) social contract. During such a process, participants are required to dismiss all proposals that would *not* be justifiable in an impartial manner before their fellow stakeholders, namely those formulated in such a way that either appeals to a party's bargaining power or does not account in any sense for the reasons that counterparts might have to agree (except the fear of a breakdown of the bargaining process).

Moreover, constraining deliberation by the requirement that any decision must be given an acceptable reason in terms of justifications that at least in principle try to account for every fellow citizens' reasons for agreeing, and guaranteeing that the reason-giving process recognizes every fellow participant's right to reciprocate by arguing in turn his/her own reason to act, constitutes not an infringement but an *extension* of the citizens' democratic political equality beyond that exercised by means of traditional political participation, voting etc.

Thus one of the basic promises (but also a challenge) for implementing the idea of shared responsibility is that it is possible to work out the rules of a multi-stakeholder (local) governance system by institutional design, understood as both substantive and procedural constraints on the deliberative process typical of deliberative democracy. Participants endorse such rules before entering the deliberative process, so that the resulting bargaining outcome does not deviate too much from the ideal of a fair social contract. It also incorporates the principle of equal respect for all citizens considered as participants on equal terms in the deliberative process.

According to this view, the following elements characterize a decision making process based on deliberative democracy as applied within multi-stakeholder governance mechanisms directed at the implementation of ‘shared responsibilities’:

- a. it removes force and fraud from deliberation because only impartial arguments intended to persuade other participants can be admitted;
- b. it guarantees equality of opportunities to participate in the deliberation through equal respect for all the participants;
- c. it does not make participation conditional on the possession of particular training, qualifications, formal education or level of sophistication in argumentation; the only requirement is that impartial reasons exchanged must account also for the reason that any counterparty should have to agree to a proposal from his own point of view, which must be stripped of the unreasonable claim that only one’s own interest or point of view should be given positive weight. The ways in which these reasons are given (once the basic impartiality requirement has been satisfied) are not relevant and do not constitute grounds for exclusion;
- d. complete rationality or perfect knowledge and a capability to process information are not required; on the contrary, limited cognitive capabilities and the affective and emotional components of human motivation are considered as obvious. When arbitrators or facilitators of the deliberative process are admitted, they are first committed to reducing the risk that any party may manipulate the other parties’ reasoning by exploiting their cognitive weakness or affective susceptibility;
- e. moreover it recognizes any parties’ claim (positive right) to receive initial information about the future consequences of any policy option, and also to receive training or assistance from technical experts in order to assess these consequences;

- f. it generates as much symmetry and equality as is possible among the participants, for the parties are more equal in ‘reason giving’ than in any other aspect of bargaining in which they would allowed to use the full range of their economic force, threat power etc.;
- g. it induces participants to imagine and assess as many outcomes as possible, including those based on the replacement of positions among the participants; this is accomplished by the frequent posing of questions such as “what would a given decision be in the event that our positions were reversed so that participating stakeholders replaced each another with respect to their economic, social, religious, ideological etc, positions?” This is the main role played by arbitrators and facilitators in the deliberative process;
- h. it explicitly aims at achieving agreement among people who hold different moral conceptions of the ‘good life’ and embeds the criterion of economizing on moral disagreements by maximizing agreements on areas where there is no irreducible dissent. (For example, consider the weakest aspect of Sen’s theory of well-being measurement (Sen 1985), which requires the comparison of functioning and capability vectors. There may be obvious consensus on the absolute importance of some basic functioning and capability vectors. If two basic-functioning vectors present a clear-cut relationship of dominance by one over the other, these partial agreements will be immediately endorsed. With regard to other capability vectors where agreements are not *a priori* obvious because they are at odds with personal life plans or conceptions of the good life, agreement can nevertheless be reached by the mutual exchange of reasons. In these cases, well-being entails social dialog and a deliberative democracy process concerning shared weights for assessing the relative importance of functionings. If agreement is not reached, these functionings will be considered to be equally important for the societal assessment of well-being, and no binding decision can be enforced);
- i. it allows participants to change their minds and to consider deliberations as provisional, i.e. it permits discussion to be reopened for subsequent decisions, even when similar to another already discussed;
- j. disagreement can be argued in public; agreement must be made public and accounted for in public.
- k. the rights to resort to “voice” and “exit” are always guaranteed to all participants during the deliberative process.

Moreover, *accountability* is a corollary of deliberative democracy, with respect to both the deliberation process as such and the implementation phase relating to what has been deliberated. Accountability is owed to all the stakeholders considered as right-holders, who have an entitlement to be informed in order to be better able to deliberate. All the relevant developments of social reporting techniques must be adopted to improve social accountability not only in corporations but also at the level of territorial local governance systems. However, it is basically understood as the systematic, complete, material and relevant reporting on outcomes related to all of the interests at stake. For instance, stakeholders cannot be *chosen* by the reporting subject, they are *out there*, and the reporting subject must recognize them. Information must be provided in a sufficiently succinct manner in order to be *useful for deliberation*. As far as accountability of the deliberative process is concerned, the *alternative* reasons processed during deliberation must be completely reported, included reports of minority positions. Concerning the *ex post* performance of the governance system, a report must account for outcomes clearly related to the commitments undertaken *ex ante*. A complete representation of performance must be reported with respect to each stakeholder concerned and must be done in comparable fashion across stakeholders' conditions. Lastly, the representation of the balance among different interests and values actually struck should be compared with the *ex ante* agreed multi-stakeholder balancing criteria whereby commitments were undertaken in the deliberative phase.

8. How to induce incentives and motivations that make shared responsibility effective

After stakeholders have entered into the local social contract and agreement has been reached, the question arises as to how principles and rules – deliberated in the agreement through extended deliberative democracy – can generate incentives to compliance and implementation, overcoming weakness of will, free riding and opportunism etc. It should be noted that a successful – at least in principle – answer to this question is also relevant for the question as to whether shared responsibility can effectively be ascribed.

From this perspective, the idea is that a shared responsibility governance mechanism should shape the voluntary agreement and its implementation in concrete situations insofar as it is able to:

- i) clarify the mutual, long-run interests of the relevant parties involved;

- ii) create benchmarks in terms of principles and rules of behaviour so that commitments can be undertaken, compliance be assessed, and judgments on trust and reputation be formed (as far as is cognitively possible, and not just for a perfect reasoner);
- iii) affect, through the *basic features* of the same agreement, the emergence of preferences for reciprocity in conformity, and beliefs that support reciprocal compliance with the agreement;
- iv) allow the formation of structural social capital among participants in terms of trust relationships supporting cooperation among agents through large networks.

Essentially, the agreement induces agents (who would not be interested in cooperating without it, and would not trust each other) to acknowledge that their contribution to the creation of social cohesion (or to the production and distribution of specific public goods) conforms to an existing social norm with which they intrinsically desire to conform or they are induced to respect because they care for their reputations with other members who desire in turn to conform. Thus compliance will satisfy the equilibrium property (which was implicit in my initial definition of shared responsibility as a social institution).

The main concept necessary in order to understand compliance is *reciprocity*. In fact, mechanisms that leverage attitudes, motivations and interests related to the fulfilment of commitments and the acceptance of or compliance with obligations are mainly related to some form of reciprocity. When planning to fulfil my part in an agreement or an obligation, if I expect that other parties will reciprocate my behaviour by carrying out a symmetric obligation or complying with the same agreement, then I have a reason and/or an incentive (whatever its motivational force may be) to reciprocate their behaviour, which reinforces my plan.

It is not purely by chance that, from the *ex ante* perspective, reciprocity has already been introduced as the essential feature of deliberative democracy and the local social contract. In suggesting a deliberation, any participant proposes an impartial justification that appeals to the reason that every other party has to agree, in that these other participants are “symmetrically” motivated to seek an agreement based on impartial justifications. No bargaining proposal is advanced in the social contract bargaining process that the proposer would not accept if he/she were to occupy the position of any other participant who reasons symmetrically. Thus symmetry and reciprocity of reasons to participate in the *ex ante* agreement process should anticipate the need for *ex post* reciprocity in the compliance phase.

Reciprocity intervenes in the *ex post* compliance phase through at least two basic mechanisms, alternatively giving instrumental reasons or intrinsic motivations to act within 'social dilemma' situations.

- The simplest responsibility-enhancing mechanism based on reciprocity is *reputation*. In this case, the interest in reciprocity is instrumental, not intrinsic. We desire to reciprocate other agents' behaviour because we may derive an extrinsic benefit from their reciprocating behaviour. Reciprocity is then a mutually beneficial tool. Reputation is the linking concept. Let us assume that compliance with an agreement of mutual advantage is at stake, and that reciprocity for each agent consists in carrying out his/her part in the agreement. If I have the reputation of being someone who conditionally reciprocates compliance, you will reciprocate compliance as well, accepting that cooperation in the long run with someone like myself who is believed to cooperate conditionally on reciprocation is in your best interest (which in general may be assumed to be true). Then my best interest may consist in cooperating in the first move and continuing to reciprocate as long as you have a symmetrical reputation (under the assumption that long-run cooperation is better for me than exploiting an individual opportunity for defection and going through an infinite number or reciprocal conflicts thereafter).

Reputation is the means for inducing trust, and is instrumental for the possibility of experiencing a number of mutually beneficial exchanges. It goes without saying that reputation is of instrumental value only in repeated games (such as the PD) in which a first series of cooperation moves can be reciprocated by a long-run series of responses in kind (long-run cooperation), but also in which a unilateral defection from ongoing mutual cooperative behaviour may be punished in future by reciprocating defection for a sufficient number of times to eliminate the incentive for unilateral opportunism.

There are many conditions that have to be satisfied in order for the reputation model of reciprocity to hold, but the most constraining is its cognitive fragility. Reputation depends on the possibility of forming beliefs about other players' reputable behaviour in terms of their fulfilment of *ex ante* specified and well-known commitments (which in the model are the same as conditional strategies for playing the repeated games). But typically this condition does not hold. Under incompleteness of contracts, for instance, contractual commitments are unspecified in relation to unforeseen states of the world, and this holds in general for commitments understood as long-run strategies (i.e. conditional rules of

behaviour defined for the playing of games that are repeated time and again in future). But if commitments are mute silent with reference to unforeseen states of the world, when these are eventually revealed there is no basis for developing reputation. Hence reciprocity breaks down (see Kreps 1990, Sacconi 2000)

This is where the small-scale model of the social contract enters the picture by making instrumental reciprocity based on reputation possible again. It establishes a set of general and abstract principles that, whilst being to some extent vague, are nevertheless able to engender expectations of the committed player's future behaviour. Since the social contract principles are agreed to under a veil of ignorance, and hence must be of a universal and general nature by construction (insofar as no possibility is contemplated for their fine-tuning in order to fit particular cases and interests), they typically establish pattern recognition devices of a kind that can be deployed in order to decide whether or not a given event belongs within the sphere of obligation of a given agent. Under the shared understanding that the condition for the moral obligation has been satisfied, the fulfilment of a behavioural standard defined *ex ante* may be taken as the basis for assessing the agent's reputation. Hence the model can work again, also in the more realistic context of incomplete knowledge. The key condition however is that commitments are derived as preventive rules of behaviour from the general and abstract principles of the social contract.

However, there are other drawbacks that counsel against assuming the reputation model of reciprocity as a panacea. Essentially, reputations can be of many types and all of them coincide with equilibria. Given a reputation (on a commitment) there is always a reciprocal strategy which is the best response against that reputation, and which is different from other best responses relating to different reputations (on different commitments). These reputations also include bad ones (such as complying only occasionally with the social contract in order to induce stakeholders not to withdraw from the cooperative relationships, whilst for the rest of the time defecting in order to reap as much of the surplus as possible, which I call 'sophisticated abuse reputations'). It is quite obvious that, as far as the decision of the player who can instrumentally profit from the development of his/her own reputation is concerned, and it is assumed that s/he will make the decision according to her/his best self-interest, s/he will prefer to try the selection of the sophisticated abuse reputation and the corresponding iterated strategy. However this

is not the last word about the matter, because it must be considered also how the second player reacts to the first player decision, and this will depend on her beliefs and motivation structure. Mainly it depends on whether the second player's response is only based on her long-run material interest or her preferences are also affected by deontological considerations about reciprocity and fairness (but about this possibility see the following point) .

- A more reliable mechanism inducing reciprocity is the *sense of justice* (Rawls 1971). This is less accepted by rational choice and game theorists because it seems to entail assumptions about individual behaviour which differ from the standard ones, and has sometimes been seen as an idealization of the general assumption of *self-interest*. However, behavioural economics today includes accounts of numerous other-regarding or not strictly self-regarding behaviour. Why therefore should we not take account of a type of behaviour that for centuries has been contrasted with consequentialism, i.e. some form of (weak, i.e.. conditional) deontological behaviour?

The “sense of justice” (i.e. the intrinsic desire to conform with principles or norms of justice established by fair agreement) emerges provided that: (i) a norm is the result of a fair agreement on principles established behind the “veil of ignorance” (or within a deliberative democracy process that replicates the fair agreement by means of the reciprocal exchange of impartial justifications), so that the norm has been reciprocally and impartially accepted; (ii) participants in the agreement have developed a disposition to conform with a fair agreement which is conditional upon reciprocity of conformity; and (iii) participants in the agreement have mutual expectations of reciprocity in conformity. These conditions activate intrinsic preferences to conform with the agreed principles or norms. Theoretical and experimental studies show that the agreement under the “veil of ignorance” may create the endogenous and psychological incentives to respect the shared norms and principles defined in the contract (see Grimalda and Sacconi 2005, Sacconi and Faillo 2010, Sacconi 2011a).

First, the agreement triggers *psychological preferences*: agents have motives to act that are not geared purely towards material advantages because their utility positively depends also on deontological reasons to act – which are measured in line with the degree of conformity displayed by an outcome with a given abstract principle or ideal. An essential feature of psychological preferences is their conditionality on reciprocity of conformity.

Accordingly, an agreement will bring about a preference for conformity provided that each participant entertains the belief that other participants will reciprocally conform with the same norm, i.e. that they will play their own part in implementing the agreed norm. The strength of the motivation to comply depends on the extent to which the agent believes that the counterparties are responsible for conformity. On the other hand, motivational strength is also brought about by the fact that other participants reason symmetrically, so that they too have expectations about the first agent's level of responsibility in fulfilling his/her part of the agreement, which is conditioned by their beliefs concerning the first party's belief, and so on. To sum up, conformist preferences are conditional on mutually expected reciprocity of conformity.

Second, since beliefs and expectations are so important in engendering the sense of justice (or the desire to be just conditional on the expectation that other parties do the same), we may ask from where they originate. The answer given within this approach is that the impartial agreement itself elicits self-fulfilling beliefs. That is, they are beliefs that an *ex ante* impartial agreement on principles of justice will generally be complied with by those who have signed up to it. In other words, if a set of principles has been agreed upon by a fair deliberative process where all the reasons have been considered, the expectation that it will be carried out by consistent behaviour naturally follows. Even if there is no particular self-interested reason to comply with the agreement, so long as we do not have evidence of non-compliance or defection, what we expect to see is that people who have genuinely agreed will, at least to a significant extent, carry out their agreed actions. This is not a valid logical inference from the individual's point of view (because there is no knowledge base from which to infer it validly). Nevertheless, it seems to be a reasonable default reasoning since the only premise which the individual has in his/her mind is the model of an agent who has genuinely agreed and in this vein has expressed the intention to carry out an action subsequently.

9. Areas and forms of shared social responsibility and different specializations of the multi-stakeholder governance model.

This section will consider three institutional models of governance involving networks of stakeholders with different levels of power and influence, but nevertheless all of which making it possible for private citizens, organizations, and public institutions to become involved in the sharing of social responsibility on some issues. They can be regarded as being

arranged along a continuum from the most hierarchical case of the firm (typically understood as a hierarchy) to the purely egalitarian case of a community comprised of nearly equal members without an internal hierarchy (i.e. a village administering a common good such as a natural resource without recourse to a central authority).

Ideally, many forms of networks of agents lie along the continuum from ‘hierarchy’ to ‘egalitarian community’ (no hierarchy). Thus in the middle of the continuum there are various possible network structures connecting non-homogeneous agents, each with different power and influence, but nevertheless to some extent linked with each other (not necessarily all to all others, or with the same intensity) and able to exercise (perhaps indirectly) some level of influence even over the most powerful agent in the network. Such networks may link up local authorities and their communities, powerful players like corporations, and their nearest and most influential stakeholders, but also their weak and in some sense ‘distant’ stakeholders. Three cases can be used to explain how the basic idea of a multi-stakeholder governance mechanism for shared responsibility unfolds along the continuum.

a) *The concentric model.* This model is characterized by the presence of one powerful stakeholder at the centre and many other stakeholders in the surrounding relational network. Typically, this model displays the classic structure of firms as hierarchies, and fits corporate social responsibility as the proper specialization of shared responsibility. Firms’ shareholders (or those who run the firm on their behalf) represent the hierarchical authority which owns the firm and which is entitled to make discretionary decisions on non *ex ante* contractible contingencies. This party is thus safeguarded against the opportunism of others. However, other parties face a risk of abuse of authority, so that they will be *ex ante* discouraged from an optimal level of investment in the firm, while *ex post* they will resort to conflicting or disloyal behaviour in the belief that they are being subjected to an abuse of authority (see Sacconi 2000, 2006a, b). Obviously, this situation may potentially generate social conflicts and reduce social cohesion.

A comprehensive view of SSR should consider this situation and propose a model of socially responsible “corporate governance” (or, more generally, a model which could be used in all of the situations that fall within the category of “concentric models”) capable of increasing social cohesion through the definition of a corporate social contract between the controlling stakeholder and the non-controlling ones, which specifies principles and norms of responsible behaviour capable of remedying the power imbalance and of

preventing opportunistic behaviour. In particular, a power imbalance may be remedied by counterbalancing the residual right of control (allocated to owners and their representative managers) with further fiduciary duties owed by the controlling stakeholder (owners and managers) to the non-controlling counterparts (consisting of ‘strict sense stakeholders’ carrying out firm-specific investments, and ‘broad sense stakeholders’ on whom externalities fall). In fact, the reference idea is that in order for the firm (or more generally each organization characterized by a concentric model) to be a legitimate form of governance, transactions must be grounded on the rational agreement (the social contract) between controlling and non-controlling stakeholders. The agreement stipulates a) that authority be delegated to the stakeholder that is the most efficient in performing governance functions; b) the extended fiduciary duties that this party owes to the non-controlling stakeholders.

Implementation of shared social responsibility in the concentric model is supported by incentives and motivations defined according to the reputation in both a repeated game model, assuming that a CSR set of explicit rules defines the benchmark against which reputation may be assessed, and the conformist preference model (see section 7). Both these models presume that, within the company, a social contract has been agreed upon amongst the stakeholders, involving also owners and the management. The institutional details are largely the same, but the second explanation is much stronger because it also allows for the prevention of sophisticated abuse of stakeholders’ trust (i.e. a company pretending to comply with a code of ethics but which in fact complies in only a minimal number of cases). Social contract-based conformist preferences emphasize the importance of *ex ante* ‘cheap talk’ under the veil of ignorance. The *ex ante* impartial agreement is essential in order to elicit the disposition to conform that gives appropriate weight to the preference for reciprocal conformity, and moreover psychologically affects the emergence of mutual expectations of reciprocal conformity. A governance mechanism for corporate social responsibility should then give as much importance as possible to benchmarking behaviours against a set of self-regulatory standards of social responsibility (to favour reputation formation). But even more emphasis should be given to the cultural and organisational conditions for the formation of genuine impartial *ex ante* agreements (seen as a source of intrinsic and not simply instrumental value) and mutual beliefs about the disposition to conform with the agreement (Sacconi 2007, 2010a, 2011a)

b) *The egalitarian community model.* The second governance model that the SSR methodology should consider is the “social-ecological system” as defined by the 2009 Nobel laureate for economics Elinor Ostrom (1992, 2000, 2009). According to Ostrom, humanly used resources are embedded in complex, social-ecological systems”, they “are composed of multiple subsystems and internal variables within these subsystems at multiple levels analogous to organisms composed of organs, organs of tissues, tissues of cells, cells of proteins, etc.” (see Ostrom 2009). Often, within social-ecological systems, natural resources and commons undergo a process of deterioration due to the difficulty of managing the system’s complexity and the prevalence of free-riding practices and opportunistic behaviour. According to the approach developed by Ostrom however, collective action among nearly equal players in a group without a hierarchical structure, and moreover that is not subject to an external authority, does not necessarily fail in the provision of local public goods and commons. Thus the structure of egalitarian governance does not necessarily need to be replaced with a governance system based on the hierarchical subordination of the community’s members to an external authority, such as a centralized state planning office, or one imposing a system for the management, maintenance and exploitation of the relevant resource, as well as the privatization of the resource and its management under the authority of a private owner. On the contrary, when some contextual and institutional variables are satisfied, a governance of the “social-ecological system” maintaining a regime of common ownership may emerge on community level with the voluntary participation of the individual members of the community, because it is able to overcome the typical free-rider paradox. The following contingent and institutional variables in this situation may be cited:

- *preplay communication* (prior to the actual implementation of individual strategies in the relevant system of interdependent decisions) permitting *agreements* on the settlement of the *system’s rules* that must prescribe reciprocal actions involving cooperation in the management, preservation and usage of the relevant resources to the mutual benefit of participants. ‘Mutual’ does not mean perfectly egalitarian but rather not disproportionately asymmetrical and reasonably fair with respect to the concrete configuration of the contribution and distribution problem;
- agreement on *simple monitoring systems* that can be easily implemented on a symmetrical basis among the community’s members, even if no strong external sanctions are applied as a consequence of reported non-compliance;

- possibility to resort to *endogenous punishment* of the community's members by interrupting cooperation with the defecting members;
- possibility – with reference to the community rules as *benchmarks* – to develop *beliefs and expectations* concerning the members' level of reciprocity in accordance with the agreed and monitored rules.

All of these variables relate to the emergence of social norms associated with the specific common management problem. They operate on the level of the small-scale “social-ecological system” and affect the endowment of cognitive social capital held by the system's members. Users of natural and social resources who share social norms telling them how to behave within the group to which they belong – and in particular endowed with norms of reciprocity – will incur lower transaction costs in reaching agreements and lower monitoring costs.

It is therefore recommended that a governance mechanism based on the idea of shared responsibility and deliberative democracy be developed also in systems of this kind which are considered to be suited for the management of commons and natural resources on community level.

- c) *The heterogeneous players network model.* Finally, let us consider the idea of local social contracts, deliberative democracy and shared responsibility with respect to networks of agents linked not by “hierarchical” but heterogeneous social relations. Agents involved in the network engage in repeated interactions and have different incentives to cooperate or behave opportunistically towards one another. Some agents would wish to defect in their potentially cooperative relations with some of their neighbours (where ‘defect’ means behaving opportunistically or, in other words, trying to gain some advantage from the relationship without considering the loss for others). Other agents however are interested in reciprocal cooperation with all the related agents (i.e. if the agents with which they are associated start to cooperate, they will cooperate as well).

The intuition is as follows. Let us assume that some agents in the network (even only a minor part of them) agree to a small-scale social contract whereby they agree on norms regulating mutually beneficial cooperation and the fair and impartial treatment of all stakeholders. Let us also assume that, for whatever reason, they are endowed with a basic disposition to reciprocate conformity with a fair agreement, and develop the expectation

that other players will also conform with the same agreement. They will then also be endowed with cognitive social capital that makes it easier for them not only to enter face-to-face cooperative relationships but also to support trust and cooperation throughout the entire network. Thus the entire network becomes endowed with structural social capital, with the result that cooperation may become sustainable throughout the network also between pairs of agents who do not as such have sufficient mutual incentive to cooperate. This depends on the fact that agents endowed with cognitive social capital may, even outside their direct interaction, decide to punish subjects who do not respect the cooperation agreement (see Sacconi and Degli Antoni 2009; Degli Antoni and Sacconi 2011).

Let us now imagine a network made up of three agents: A, B and C. A wants to cooperate with B (because cooperation is more advantageous than defection from a material point of view), while A wants to defect with C (because defecting implies a higher material payoff). At the same time, C would like to cooperate with both A and B. B would like to cooperate with both A and C. A may be taken to be a firm which employs immigrant workers (B), whilst C is a local public authority which provides social services. The firm would like to defect with its immigrant workers by attempting to appropriate the entire surplus generated in the relationship with them (for example, it would like to pay very low wages). This is because these immigrants are unskilled workers, are not members of any trade union and may be replaced by the firm very easily. The immigrant workers (B) wish to cooperate both with the firm (they need to work and do not want to lose their jobs by behaving opportunistically) and with the local public authority which provides them with social services. Finally, the local public authority wishes to cooperate (again considering only material incentives) both with the firm and with the immigrant workers. Now let us imagine that the three agents agree to enter into a local social contract under which they agree to cooperate reciprocally. The decision to sign up to a local social contract may be prompted by the fact that the firm (or the local public authority) knows that its consumers (or citizens/voters) will be concerned by such a decision (it should be noted that the local community may also be interested in the level of contribution by the immigrant workers to the local welfare system). The key question is this: what happens after the social contract has been agreed if the firm (which is the only subject which would like to defect given its material incentives) behaves opportunistically vis-à-vis the immigrant workers?

The answer is as follows. If the community (represented by the local public authority) has developed high cognitive social capital consisting in the tendency to reciprocate conformity with the local social contract on the expectation that other parties will also conform, then it will also be ready to sanction the company. This is not due to any material interest but is a response to the psychological payoff associated with the decision to support generalized conformity with the contract itself. However the company, which is interested in protecting its cooperative relationship with the local community (both due to material and psychological payoffs as well as reputation benefits), will react to the incentive of compliance with the agreement by fulfilling its responsibilities towards the immigrant workers. As a whole, the local social contract on shared social responsibility will operate as a support for cooperative relationships in the entire network even for those parties to it (i.e. the link between A and B) that would not be able to support cooperation by itself. The preconditions are in part exogenous (the endowment of cognitive social capital embedded in the cultural heritage of the local community), although they may in part be subjected to institutional design through the proper governance system that makes all of the players converge on a local social contract (which must cover the issues of sharing responsibility for the integration of immigrant workers and the maintenance of the local social welfare system at reasonable cost). This elicits motivations to conform and provides a benchmark for assessing behaviour and hence generating expectations concerning reciprocal conformity. The result may be that the company accepts responsibility for integrating the immigrant workers because this will safeguard its reputation with consumers and the local community. The level of the immigrants' contribution to the community will be raised, thus reducing the basis for racial hostility in the community itself.

10. Conclusion

The main features of the new paradigm of shared social responsibilities and the multi-stakeholder governance systems implementing it can be summarized as follows.

- a) Neither local governments (on municipal or regional level) nor nation states alone – operating through their representative procedures and public policy decision making processes – can satisfactorily undertake to define and implement social welfare policies, mainly on local level. Also a collection of different individuals, social actors and organisations positioned at different levels and endowed with different resources need to

be involved. As such, none of these can have exclusive competence over the entire set of problems because some of them are private companies, or non-profit organisations, informal local communities and also private citizens endowed with different capabilities. Nevertheless, they can share responsibilities for problem solving.

- b) As a consequence, there is a need not only for *government* but also for a model of *governance* in order to enable coordination and cooperation among all these different actors – not only through the formal settlement of the proper allocation of shared social responsibilities among them, but also providing for effective discharge of those responsibilities. In particular, the governance model must enable the achievement of fair settlements and the satisfaction of different and partly conflicting interests that are not all recognized as having the same urgency and priority. This will secure the multi-stakeholder nature of the governance mechanism as well as its basic function in establishing a fair balance among different stakeholders. The fiduciary duties of those occupying positions of power and authority (and the related stakeholders' rights) also derive from the same source.
- c) The strength of this approach is that it should make it possible to identify ways of developing self-enforcing social norms and standards voluntarily agreed among stakeholders, but not inconsistent with the principles that “free and equal persons” would have established under the constitutional social contract on global level (whether national or European). These norms should also activate endogenous motivations and incentives conducive to their self-imposition and effective execution. By virtue of these motivations, stakeholders effectively contribute to the provision of local public goods, the preservation and management of commons, the production of positive externalities and the fair distribution of welfare goods, as well as to the prevention of opportunistic behaviour generating public harms and unfairness in private relations. In other words, as a whole they generate ‘social cohesion’. Thus, typical and apparently insoluble collective action paradoxes and ‘social dilemmas’ (such as free riding) are overcome. Concepts such as ‘local’ or ‘small-scale social contract’ on the one hand, and ‘cognitive social capital’ on the other, correspond to these requirements.
- d) The former (local social contracts) help ensure that social norms or standards whereby shared responsibilities are allocated reflect a genuine consensus unaffected by force, fraud, manipulation or the power of threats. In other words, they reflect the criterion of a

fair agreement consistent with wider principles of social cohesion and social justice, such that they may effectively represent the equilibrium point among different stakeholders' interests and the values that they would accept under impartial and symmetric bargaining conditions.

- e) The latter (cognitive social capital) refers to the development – through the same idea of impartial agreements (local social contracts) – of cognitive and motivational endowments such that effective cooperation is made endogenously possible even though it is not supported by immediate self-interested incentives. They also induce the creation of trust-based relationships capable of supporting cooperation within multi-stakeholder and multi-agent networks whenever cooperation cannot be supported by the mere bilateral benefit that pairs of participants may derive from it.
- f) The shared responsibility approach also entails a danger. If the composition of different interests is achieved by multi-stakeholder governance forums seen as places in which bargaining games are to be played, one could reasonably be worried that the democratic nature of government may be harmed, as well as the equal opportunity for all citizens to participate in the democratic decision making process. Similarly, accountability to citizens – typical of democratic government institutions – also risks being prejudiced. Thus, in these multi-stakeholder governance models, the equality of democratic citizenship risks being jeopardized. This is the reason for making the additional claim that deliberative democracy should become a method applied beyond the limits of representative political institutions for the purpose of shaping multi-stakeholder governance systems on local level as well.
- g) Moreover, deliberative democracy is not simply a preventive measure against a coincidental danger of multi-stakeholder governance. It is a feature inherent in the very model of governance aimed at actualizing the idea of shared responsibility. Since shared responsibilities are largely a matter of voluntary choice and the acceptance of obligations deriving from social norms and standards, stakeholders become responsible mainly in terms of social/ethical (not exogenously imposed) norms. Brute bargaining outcomes – without further qualification – do not qualify as sources of moral values or ethical norms. Individual citizens in particular, who are not formally charged with the fulfilment of a particular public goal, cannot be committed to any social obligation (from which social responsibility stems) without being involved in the appropriate deliberative process

capable of generating the voluntary acceptance of moral commitments. The prototypical feature of deliberative democracy is that all participants in the deliberative process can only introduce impartial reasons to act aimed at justifying any policy proposal to other participants during the *ex ante* discussion (which precedes decisions on binding policy choices). Deliberations are thus made acceptable to all the participants, who are similarly motivated only to advance reasons capable of obtaining general acceptance. Impartial acceptance typically induces fair agreements and ethical standards of behaviour. Since such standards are shared, responsibilities stemming from them are also shared.

- h) *Reciprocity* is the key element in the effectiveness of shared-responsibility governance mechanisms.
- *First*, deliberative democracy commits participants in the deliberative process to reciprocity in the mutual exchange of reasons intended to justify different deliberative proposals. This is a reciprocity-based standard of acceptance like ‘I cannot claim you should accept a policy proposal that I would not accept had I replaced you in the deliberative process by taking precisely your (social and personal) position and viewpoint’. Thus deliberative democracy saves the multi-stakeholder governance model from being reduced to a mere bargaining game played by self-interested players without further qualifications.
 - *Second*, reciprocity is a precondition for the creation of the motivational base that induces voluntary participation in the implementation of shared social norms. In fact, reciprocity in the agreement constitutes the basis for a desire to reciprocate the behaviour of other agents who also conform with the agreed norms, provided that the counterparties are also believed to reciprocate, whereas reciprocity in the agreement also psychologically elicits the expectation that others will conform.
 - *Third*, this effect of reciprocity (supporting social norm compliance) also spreads through large social networks in which some agents may also entertain relationships that do not support effective cooperation between members – for example, large multinational corporations and weak employees in delocalized plants. Prevention of breaches of trust can in these cases be based on further links also present in the network, for example between the strong agent (the company) and other stakeholders (the local community), whose mutual relationships are shaped by impartial micro social contracts. These relationships then expand the preference for reciprocity and the desire to reciprocate

punishment for unilateral breaches of social norms also beyond their direct relationships, so that they are extended to cover further parts of the social network where the weak stakeholders are located.

REFERENCES

- Aoki, M. (2001), *Toward a Comparative Institutional Analysis*, Cambridge, MA: MIT Press.
- Aoki, M. (2007), 'Endogenizing Institutions and Institutional Change,' *Journal of Institutional Economics*, **3**, pp. 1–39.
- Bacharach M. (2006), *Beyond Individual Choice. Teams and Frames in Game theory*, edited by R.Sugden and N.Gold, Princeton U.P., Princeton and Oxford.
- Berle A., G. Means (1932), *The Modern Corporation and Private Property*, Mac Millan, New York.
- Binmore, K. (2005), *Natural Justice*, Oxford, Oxford University Press.
- Clarkson, M. (1999), *Principles of Stakeholder Management*, Toronto: Clarkson Center for Business Ethics.
- Degli Antoni G. and L. Sacconi, (2011), "Modeling cognitive social capital and CSR as preconditions for sustainable networks of cooperative relations" in L. Sacconi and G. Degli Antoni, *Social Capital, Corporate Social Responsibility, Economic Behaviour and Performance*, (eds), Palgrave Macmillan, London.
- Dezau, A. and D. North (1994), 'Shared mental models: Ideologies and institutions', *KIKLOS*, **47** (1), pp. 1–31.
- Donaldson T.. and T.W. Dunfee, (1999), *Ties that Bind; A Social Contract Approach to Business Ethics*, Harvard Business School Press.
- Donaldson T., L. Preston (1995), "The stakeholder Theory of the Corporation: Concepts, Evidence, and Implications", *Academy of Management Review*, 20, 1, pp.65-91
- Donaldson T. and Dunfee, T.W. (1995), "Integrative social contracts theory", *Economics and Philosophy*, Vol. 11, pp. 85-112
- Dunfee T.W. and T. Donaldson,(1995), "Contractarian Business Ethics", *Business Ethics Quarterly*, April, vol.5, n.2.
- Freeman R.E. (1984), *Strategic Management: A Stakeholder Approach*, Pitman, Boston.
- Freeman R.E. and J. McVea (2002) *A stakeholder approach to Strategic management*, Working paper n.01-02 , Darden Graduate School of Business Administration
- Freeman R.E. and P.Evans (1988), "Stakeholder Management and the Modern Corporation: Kantian Capitalism", in Beuchamp T. and N.Bowie (eds.) *Ethical Theory and Business*, 3rd ed., Englewood Cliffs N.J. (Prentice Hall).
- Gauthier D. (1986), *Morals by Agreement*, Clarendon Press, Oxford..

- 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.
- Gutman A. and D. Thompson (1996), *Democracy and disagreement*, Cambridge Mass, Harvard U.P.
- Gutman A. and D. Thompson (2005), *Why deliberative democracy*, Princeton, Princeton U.P.
- Kaufman, A. (2002), 'Managers' dual fiduciary duty: To stakeholders and to freedom', *Business Ethics Quarterly*, vol. 89, pp.189-214
- Kreps D. (1990), "Corporate Culture and Economic Theory" J.Alt and K.Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge U.P., Cambridge.
- Ostrom, E. (1990), *Governing the Commons*, Cambridge University Press, New York.
- Ostrom E. (1999), "Social capital: a fad or a fundamental concept?" In *Social Capital: A Multifaceted Perspective*, P. Dasgupta and I. Seraeldin (eds.), 172-214. Washington, DC: The World Bank
- Ostrom E. (2000), "Collective Action and the Evolution of Social Norms" *Journal of Economic Perspectives*, Volume 14, Number 3-Summer 2000-Pages 13 7-158
- Ostrom, E. (2009), A General Framework for Analyzing Sustainability of Social-Ecological Systems, *Science*, 325, 419
- Ostrom, E., Walker, J., Gardner, R., (1992) "Covenants with and without a sword: self-governance is possible", *American Political Science Review*, 86, 404–417
- Rawls J. (1971), *A Theory of Justice*, Oxford U.P., Oxford.
- Rawls J. (1993), *Political Liberalism*, Columbia U.P. New York.
- Sacconi L. (2000), *The Social Contract of the Firm. Economics, Ethics and Organisation*, Springer Verlag, Berlin.
- Sacconi L. (2005), (a cura di), *Guida critica alla responsabilità sociale e al governo di impresa*, Bancaria editrice, Roma
- Sacconi, L. (2006a), '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. (2006b) "CSR as a model of extended corporate governance, an explanation based on the economic theories of social contract, reputation and reciprocal conformism", in F. Cafaggi (ed.), *Reframing self-regulation in European private Law*, Kluwer Law International, London. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=514522
- Sacconi, L. (2007), '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. (2010a), "A Rawlsian view of CSR and the Game Theory of its Implementation (Part I): The Multistakeholder Model of Corporate Governance", in L. Sacconi, M. Blair, E. Freeman and A. Vercelli (eds.) *Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*, Palgrave Macmillan, Basingstoke.

- 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 (eds.) *Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*, Palgrave Macmillan, Basingstoke
- Sacconi L. (2011a), "A Rawlsian View of CSR and the Game Theory of its Implementation (Part III): Conformism and Equilibrium Selection" in *Social Capital, Corporate Social Responsibility, Economic Behavior and Performance*, edited by L. Sacconi and G. Degli Antoni, Palgrave Macmillan, Basingstoke.
- Sacconi L. (2011b), "From individual responsibility to 'shared' social responsibility: concepts for a new paradigm", this book.
- Sacconi L., Degli Antoni G. (2009), "A theoretical analysis of the relationship between social capital and corporate social responsibility: concepts and definitions", in Sacchetti S. and Sugden R. (eds.), *Knowledge in the Development of Economies. Institutional Choices under Globalisation*, Edward Elgar Publishing Ltd, London.
- Sacconi L., Faillo M. (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, N. 2 / June](#), pp.171-201
- Sen A.(1985), *Commodities and Capabilities*, North Holland.
- Sunstein C. (2000), (ed.) *Behavioral Law & Economics*, Cambridge U.P., Cambridge.