



UNIVERSITA' DEGLI STUDI DI TRENTO - DIPARTIMENTO DI ECONOMIA

A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity

Lorenzo Sacconi

Discussion Paper No. 11, 2004

The Discussion Paper series provides a means for circulating preliminary research results by staff of or visitors to the Department. Its purpose is to stimulate discussion prior to the publication of papers.

Requests for copies of Discussion Papers and address changes should be sent to:

Dott. Stefano Comino
Dipartimento di Economia
Università degli Studi
Via Inama 5
38100 TRENTO ITALY

A SOCIAL CONTRACT ACCOUNT FOR CSR AS EXTENDED MODEL OF CORPORATE GOVERNANCE (PART II): COMPLIANCE, REPUTATION AND RECIPROCITY

by

Lorenzo Sacconi

Department of Economics, University of Trento

and

EconomEtica, interuniversity centre of research , Bicocca University, Milano

lorenzo.sacconi@economia.unitn.it

Abstract

This paper is the second of a two parts essay aimed at giving a contractarian foundation to the concept of Corporate Social Responsibility (CSR), meant as an extended model of corporate governance of the firm. The first part has been focused on rational bargaining and justification. Here, on the contrary, I am focused over the implementation and compliance domain. First this domain is defined in terms of the relevant game theoretical model, which is required to represent ex post implementation of an ex ante agreed contract and is able to account for the self-enforceability of institutions resulting from the social contract. The model of reputation games, with reference to the basic game of trust, is introduced in order to make sense of self-regulation as a way to implement the social contract over the multi-fiduciary model of corporate governance. This allows understanding why self-regulation, meant as the mere resort to a long-run strategy in a repeated game of trust, fails in making sure the reputation of the firm when the game is surrounded by incomplete contracts and unforeseen contingencies. Hence, two basic problems for the functioning of the reputation mechanism are faced: the cognitive fragility problem, and the motivational problem in turn. As far as cognitive fragilities of reputation (resulting from impact of unforeseen contingencies and bounded rationality) are concerned, the paper elaborates upon the idea of proper self-regulation, based on explicit norms, and develops the logic and the structure that these self-regulatory norms must satisfy if they have to serve as a gap-filling tool for the remedy of cognitive limitations in the reputation mechanism. This coincides with an ethical decision procedure employing fuzzy logic and default reasoning. Then, a motivation problem emerges from the possibility of sophisticated abuse on the part of the firm, employing equilibrium mixed strategies in the repeated trust game. Here an entirely new application of the theory of conformist preferences is developed as an approach to the problem of the motivational role played by business ethics norms. This section answers the “real life” question raised about stakeholder’ activism and why they do not acquiesce to sophisticated abuse by the firm. The result derives from straightforward calculation of stakeholders’ overall utilities and equilibrium strategies under the hypotheses of conformism-and-reciprocity-based preferences and overall utility functions.

Key words: Corporate social responsibility, Social contract, Self-regulation, Ethical norms and Management Standards, Reputation games, Unforeseen contingencies, Fuzzy logic and Default reasoning, Reciprocity and fairness, Conformist preferences.

1. The compliance problem

In part I of this essay I have provided a contractarian justification for the CSR model of corporate governance (see the related paper “A social contract account for CSR as model of corporate governance: rational bargaining and justification”, *supra*). Here I move from justification to implementation of the model. The question asked is whether the social contract is also able to induce endogenous incentives and motivations, which might be strong enough to drive individual behaviour to carry out the normative model of extended fiduciary duties. In other words, I am wondering whether the institutional model of firm governed in the interests of all its stakeholders may be self-enforcing in the sense that it does not ask, in order to be complied with, to be enforced by an external authority - the authority of the law – or at most requires only a mild external imposition, which can be considered as residual.

Reasons for asking this question are both conceptual and practical. From the theoretical standpoint, it should be recognized that self-enforceability is an inherent feature of any sound contractarian mode of theorizing. Contractarianism is a double-faced approach, normative and explanatory at the same time. In so far as it affords an explanation of how a social or economic institution (in our case the multi-fiduciary firm) could have emerged from a rational agreement endorsed in an original pre-institutional situation of strategic interaction, it also provides a justification of the same institution, thanks the normative force of the very idea of an impartial unanimous agreement amongst rational and free individuals – respecting equally their rationality. Nevertheless, at the same time it gives a hypothetical deductive explanation of the institution deriving it, as an endogenous artefact, through the logic of collective bargaining from the “state of nature”. Decisions are taken by the interacting individuals in a pre-institutional situation from which the institution emerges as a theorem in the very logic of their collective bargaining. Hence the institution (and the social contract itself) must basically rest on these individual decisions, without resorting to any further social or legal institution.

However, no institution can be explained as an endogenous artefact, unless it is not only borne out of an agreement but it can also satisfy some condition of *ex post* stability, i.e. a condition of effective enduring compliance with the agreement by those who have struck the contract *ex ante*. The social contract were simply “cheap talk”, no agreement in state of nature would be a satisfactory explanation of enduring institutions at all, for nobody will comply with it, so that no effective behaviour could be expected in correspondence to the contractarian normative model. To say all that in terms of non cooperative game theory, the social contract or, better, the corresponding institution, must fulfil an equilibrium condition *ex post*, which implies non existence of incentive to deviate from it once the individuals know that the institution has been set out through an agreement, and each player knows that this is known by every other player. Given that any significant game in which equilibria are existent, also usually admits a multiplicity of equilibrium points, I can say that the bargaining model of social contract may be employed as an equilibrium selection mechanism within the set of equilibria of the underlying game (that is the set of institutions that satisfy the equilibrium condition).¹ Then the result of this selective mechanism must necessarily coincide with one of the feasible equilibrium points of a non cooperative game appropriately defined to represent the situation in which – after the agreement - the individual agents will have a choice over either conforming or not to the requirements put forward by the institution selected *via* social contracting.

This opens the route however to a lasting debate over the choice of an appropriate game model for representing the “original” interaction situation, such that the social contract may be shown a self-enforcing agreement. An ongoing dissent amongst rational-choice-contractarian theorists focuses around which game should be used to represent the “state of nature” in which the social contract is used as device to select a social beneficial solution of interaction coinciding with a norm or an institution. It should be clear that if the traditional representation of the “state of nature” as a Prisoner’s Dilemma Game (PD) traceable back to Hobbes and renewed by Buchanan (1975) and Gauthier (Gauthier 1969, 1986) is employed,

¹ This is exactly the intuition underlying Binmore’s approach (see Binmore, 1994, 1997), see also note 1, Part I.

then the social contract cannot succeed in giving rise to a stable arrangement of social interaction. In fact a PD game as such has only one equilibrium solution, moreover in dominant strategies, which coincides to the defection from any cooperative agreement, the social contract included. On the other hand, if we employ games like “stag hunt”, “assurance game” or some other game of coordination, as the representation of the situation of interdependent choice the players face in the “state of nature”, we have *at least* two equilibria (in pure strategies) in the game, but possibly many others (Hampton 1987, Sacconi 1995, Skyrms 2004).² As far as “stag hunt” is concerned the equilibria in pure strategies are one corresponding to the stable persistence of a state of “war of all against all”, whereas the second corresponds to a state of mutually beneficial cooperation that could be rationally chosen by unanimous agreement over a set of rules for the players’ behaviour. The problem would reduce to the apparently more tractable question of how to select the social optimum in a situation of uncertainty generated by the fact that the social optimal equilibrium is also risk-dominated by the inefficient “state of war” one. Hence, only if the players have mutually consistent expectations compatible with the hypothesis that they will play the socially optimal equilibrium, they will in fact play their strategy belonging to that equilibrium. In this case the social contract, understood as a way to generate a system of mutual beliefs, could work quite well as equilibrium selection mechanism. But once these uncertainties were resolved, there would be no further incentive to deviate from the requirements dictated by the social contract.

Basically a similar story can be told also for any other representation of the state of nature game as a multiple equilibrium repeated game, that essentially, when as seen as a whole, result in a mixed interest coordination game (Schotter, 1981, Binmore 1994, 1997, Sacconi 2000). At the end what we would get to is a characterisation of the norm or institution selected *via* social contract in terms of a social norm or a convention in the sense of Hume and Lewis (1969), i.e. a coordination equilibrium in a (possibly mixed) motives situation of social interaction, supported by a system of mutually consistent expectations. An agreement amongst players in this situation solves the coordination problem by fixing their mutual expectations and behaviours over one of the possible equilibrium points and generating the system of common knowledge required by the very definition of a convention³. Less demanding definitions of conventions have been given, for example skipping the notion of common knowledge and putting them in an evolutionary framework (Sugden 1986), but it is

² To be sure, Jean Hampton (1987) suggested that the hobbesian tradition must be reinterpreted in order to make justice to its basic program. She explained that there is a change of the strategic interaction’s situation in the transition from the “state of nature”, seen like a PD, to the mixed motives coordination game where the players have to decide over the selection of one form of the social contract among the many possible, and at the same time about social contract compliance. I worked out a more formalised model for this change of perspective in Sacconi (1995). Concerning “Stag hunt”, it has the following matrix-form (see Skyrms 2004):

	hunt stag	hunt rabbit
hunt stag	5, 5	0, 4
hunt rabbit	4, 0	4, 4

³ A regularity R in the behaviour of a population P whose members are playing repeatedly a situation of strategic interaction S characterised as a coordination problem is a convention if it is common knowledge that i) everybody plays according to R, ii) everybody knows that all the participants play according to R and iii) everybody prefers to play according to R given what he knows, for the situation of reference S is a coordination game (see Lewis 1969)

quite clear that some related notion of social norm is required if the stability and enduring acceptance of business ethics norms and institutions have to be captured – see for example the notion of “genuine consensus” based on the re-elaboration of “social norm” adapted by Donaldson and Dunfee (1995) from Alan Pettit (1990).

Nevertheless, when one attempts to model the situation in which a business norm and institution (like property rights, rules of contracts or the firm, with its hierarchical relationships and allocation of decision rights) has to be carried out by players that retain some freedom of choice and a margin of discretion in the decision of their effective strategy (which can conform or not to the requirements put forward by a given institutional arrangement or legal norm), the PD-like logic regains all its intuitive force – at least in the one shot game perspective. Take for example business ethics norms like due care in abiding by each steps to a single long term contract, which cannot be split in a chain of many linked short term contractual phases – i.e. the supply of a single very complex and inseparable item of high technology to a client that will not be able to buy a second piece in the foreseeable future - or, being in the role of a CEO, respecting the accountability obligation owed to non controlling shareholders or stakeholders, when there is a chance to hide the relevant information for a long time due to enduring collusion with auditors and monitoring agencies, or to acknowledge the valuable contribution of a young specialised employee who, due to his investment in specific human capital, is locked-in into the company for long time ahead. It seems very natural to think that defecting from the obligation to comply with these normative requirements can be individually advantageous, as far as the other party complies, whereas it may seem obvious that if the other player does not comply it would be individually destructive to comply unilaterally.

To be sure, this argument is construed out of a very pessimistic picture of human motivation, in which the players reason only from the standpoint of their immediate self-interest. I agree indeed with the program to enlarge the notion of preference in order to include also other motivations in the picture of a rational Self who makes choices given a comprehensive consideration of all her motives for preference (as it will be seen in section 4).⁴ Nevertheless, at least at this starting point the pessimistic picture can be maintained in order to set a simple benchmark – as in fact mere selfish utility maximisation is a very clear and coherent but simplistic representation of human behaviour - against which we may test the more complex hypotheses (cognitive or motivational) needed to escape the very dark perspective of the prisoner’s dilemma situation. It may be useful in order to introduce as few hypotheses as they are truly necessary to account for more complex social phenomena – as for example voluntary adherence to other-regarding rules of behaviour (I undertake these tasks in sec. 3 and sec. 4, where the cognitive and motivational roles of explicit ethical norms are considered in turn).

In particular, starting from the pessimistic picture seems not so unrealistic in our case. Who in the firm holds the widest discretion are those in the position of an authority, because they hold residual right of control over the not ex ante contractible decision variables. Abuse of authority is inherent in this situation and it is also the starting point of our reasoning about the need for a social contract foundation of extended fiduciary duties (see sec. 3, part I, *supra*).

⁴ Among the many studies suggesting to go beyond the mere selfish representation of human preferences let remind at least Bernheim (1994), Rabin (1993), Charness and Rabin (2002), Sugden (1998), Frey (1997), Falck and Fishbaker (2000), Fehr and Schmidt (2001), Falk, Fehr and Fishbaker (2003).

The model admits that those in position of authority in the firm will retain discretion and decision rights associated to property rights, but at the same time they will also be requested to manage the firm in accord with the fundamental social contract from which they derive their authority by endorsing extended fiduciary duties owed to all the firm's stakeholders (see sec. 6, part I, *supra*). Hence, they have the opportunity to defect from these obligations because, even though a normative model asks them to constrain their discretion, the effective implementation of these obligations rests largely on their voluntary obedience to the constraints, or to say it differently, these constraints may operate only through the internal channels of their motivations, incentives and deliberations. It seems that a grain of the DP-like logic is thus constitutive of the implementation phase in the social contract theory - in the sense that we face an asymmetrical Prisoner's Dilemma situation where a strong party at least is in the position to abuse other players whereas the latter - i.e. the non controlling stakeholders - can only protect themselves by preventing the first from abusing (this is the 'trust game' I will discuss at long in the next section).

Summing up, in order to accomplish the contractarian program by a proper modelling of the compliance game, the difficult task is to account for the basic incentives to defect from behaviours required by the normative model, and at the same time to make room for a solution such that in equilibrium the parties will conform to the normative model's requisite behaviours. A natural place in game theory to answer this question is the field of repeated games of reputation.⁵ In the following section I will discuss how strong or weak the idea of reputation is when it confronts the problem of compliance with an institutional framework of fiduciary duties in a world of incomplete contracts.

From the practical standpoint however - at the cost of giving up contractarianism as self-contained theory -, one could say that I might be dispensed from an equilibrium solution of the compliance problem, given that whatever we justify by means of the normative model we could also make enforceable by the law (originated externally with respect to the interacting players in the social contract model). This is not the case however. Welfare State regulations, labour-market laws and environmental regulations, establish a general legal framework, but they cannot regulate every detail of firms' decisions. They may lay down compulsory conditions, but in many settings their application requires interpretation of a 'grey' zone; or else the conditions for verifying compliance with them may not be observable. Moreover, even when management decisions closely affect the stakeholders, the law cannot regulate them in every respect: the decision whether or not to restructure or downsize a firm is always a business decision, notwithstanding the requirements of the law with regard to the protection of third parties or employees. Regulations aimed to dictate how such decisions should be made in every circumstance would inevitably be inefficient (as it has proved to be in any command economy). Nor do contracts provide a solution. Indeed, the fact that contracts are generally incomplete has prompted the present discussion of CSR. On my hypothesis, the social contract amongst the stakeholders is the 'hypothetical contract' which works as a basis for default or 'gap-filling' rules to complete incomplete contracts (Coleman 1992). The obligation to fulfil explicit but incomplete contracts hence does not guarantee compliance

⁵ Another way is following Gauthier's project to introduce a theory of rational choice over the dispositions to choose, but see my discussion of this approach in sec. 1, point B of part I.

with the obligations deriving from the hypothetical contract, exactly because the explicit contract is mute on the matters that the hypothetical contract helps to clarify.

This carries us to consider the role of self-regulation. We may distinguish between two approaches to the CSR self-regulation. The first I call the *discretionary approach* and it is addressed in sec. 2. I deem it *unfit* in so far as it pertains to the mere sphere of consequentialist expected utility maximising choice over predicable (or at least foreseeable) action and consequences. It does not imply any explicit principle or rule constraining behaviours in general, but at most a commitment over a given strategy (self-regulation in the *proper sense* is addressed in section 3). Its basic tenet is that there is no reason to add any further specification or constraint apart from the *enlightened* self-interest of those who run the firm for their own advantage. Respect for the stakeholders' claims will come about through free choice, or through the firm's free exercise of discretion. Enlightened self-interest would thus be an endogenous force able to induce self-discipline because it induces to account for personal interest in the *long-run*. By virtue of the long-run, the firm, as it pursues the simple goal of profit maximisation, would be induced to respect the fiduciary relation with the stakeholders and take due consideration of their well-being. On this view, self-regulation is nothing more than *the consistent adherence to a strategy* whereby the firm does not behave in a manner such to abuse the trust that stakeholders have placed in it. The firm does not self-impose any formal system of rules or adopt any explicit management system required to abide by standards or norms, even if voluntary: this is self-discipline *without* explicit rules.

It would be too easy to discredit this thesis by considering in its favour the only argument based on the obviously unrealistic hypothesis of perfect competition and the 'invisible hand' of the market. If competition were perfect, no firms would exist in the sense of being alternative institutions of governance to minimize transaction costs. It is therefore obvious that the argument does not pertain to the ideal world in which the 'invisible hand' operates. I shall instead take this thesis at its best, although even in this case it fails.⁶

At best, the thesis maintains that enlightened self-interest leads to respecting trust of stakeholders, and therefore not to abuse them, in that the firm recognizes the importance of safeguarding and enhancing its reputation, which depends on non-abuse of the stakeholders. Reputation is one of the most valuable, albeit intangible, of the firm's assets. It is reputation that induces the stakeholders to trust the firm and consequently to cooperate with it, so that transactions come about at low costs of control or bargaining. Moreover, reputation activates

⁶ Let me add that I'm here much fairer to the critics than they are to the very idea of multi-stakeholder firm. For example Michael Jensen (2001) resorts to the idea of discretionary self-regulation based on long run enlightened self-interest exactly under the hypothesis of "quasi-perfect" competition that ensues when the State perfectly provides for market failures like public goods and externalities, hence leaving the owners of firms free to act upon their selfish motivations i.e. maximising shareholders value even though spread over the long run. It would be too easy remembering that in the transaction costs approach the firm itself is an imperfect solution to market failures due to incompleteness of contacts – as we saw in section 3 part I. It is with the imperfections of this institutional solution that we are dealing in the search for a multi-fiduciary model of corporate governance (which of course implies that we cannot be satisfied with the solution of providing control to capital investors capable to hold risk for their investments). Either these imperfections are inexistent, thus we are left to deal only with the traditional public goods failures by means of State regulations, or these costs or imperfections cannot be solved by a free market approach, even under the State providing for externalities and public goods, simply because their existence implies an additional strong imperfection in the functioning of the market.

some sort of self-fulfilling virtuous circle which leads to spontaneous compliance with social norms – that is, makes them self-enforcing. In fact, compliance with the norms *creates* reputation; reputation *induces* a cooperative response from the stakeholders; those who abide by the norms are thus *offered* a benefit; and this benefit acts as an *incentive* for complying with the norms. Unfortunately, however, reputation does not support the discretionary approach to self-regulation, which, without imposing any constraint or explicit rule upon the firm, would allow it to decide unilaterally what actions to undertake in the best interest of both the firm itself and its stakeholders. To understand this, we must delve at least a little way into the theory of games of reputation.⁷

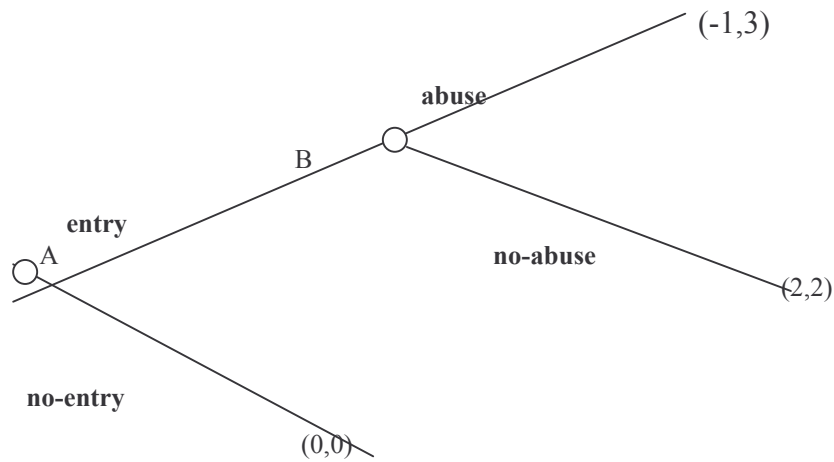
2. The reputation mechanism and its fragilities

I will illustrate reputation mechanism by means of a simple interactive situation (called *the trust game*, see Fig. 1) representing a transaction based on the fiduciary relation between a stakeholder A and the firm B.

The stakeholder must decide whether or not to place its trust in the firm, entering or not into an exchange relation with it (assume that if it does so, necessarily he is making a specific investment). The firm then decides between abusing and not abusing. If, after the stakeholder has entered, the firm does not abuse its trust, there will be a reasonably good outcome for both. However, if the stakeholder places trust in the firm, the latter has an interest in abusing that trust, because in the current game this is the more remunerative option. Consequently, the stakeholder will not grant its trust and the transaction will not take place.

The underlying idea of a game of reputation is that there is an alternative solution which permits the transaction between the two parties to take place if the basic game is infinitely repeated, and if an incentive is thus created for the firm to protect its reputation (Fudenberg and Levine 1989). We thus have an infinitely repeated game (which expresses the idea of long-run) whose stage-game is the trust game already defined. The players in the game are on the one hand an infinite series of stakeholders, called A_i (where i denotes the order of entry into the game), each of them lasting only for the stage-game in which they decide whether to enter or not enter (and are therefore short-run players), and on the other a firm (B), the long-run player, which last throughout all repetitions of the game.

⁷ Basic reference are Kreps, Milgrom, Robert and Wilson (1982), Kreps and Wilson (1982), for a general presentation of the subject see Fudenberg and Tirole (1991) ch. 9.



(Figure 1. The trust game)

Information is crucial for the players: each A_i is uncertain about the *type* of B – in that B may be a *type* that never abuses trust or a *type* that always abuses it, or even a *type* that abuses with a certain probability and does not abuse with the residual probability. Let us assume that the various stakeholders believe that there exist only *pure types* which always abuse or do not abuse trust (and at most only two types of minimum deviations from these pure types, which represent the possibility of a *mistake* by the pure types). Hence the *types* can be understood as commitments to the stereotyped use of a given basic game strategy (unless a mistake occurs). For every player A_i , all these types of B have some *a priori* positive probability (and, in particular, the type that never abuses - which for simplicity I may call ‘honest’- is assigned positive, though very low, probability). At each stage game the current A_i player changes his beliefs (the probability assigned to types) according to what he has learned from the previous stage game. In each of these stages, the conditioned probabilities of the types are updated on the basis of the evidence furnished by the manner in which B has played the previous game.

Player B’s reputation is the probability assigned by each player A_i in the current stage to the various types of player B. Player B’s reputation of being a certain type increases as evidence is gathered which confirms that *type* (the probability of a *type* increases with the observation of those actions whose likelihood is positive, given the type), but it diminishes dramatically if a single observation is made that falsifies the type (if, in any stage, an action by B whose likelihood given its type is 0 is observed, the conditioned probability of that type is nullified as well). Player B, on the other hand, is perfectly rational and informed, so that his strategic reasoning scheme also includes awareness of the limitedly informed reasoning performed by players A_i , and in particular it enables player B to predict about the mechanism by which they learn and update beliefs about types. The players’ interests are such that each stakeholder A_i maximizes its benefits in the current game (i.e. is short-sighted) while B is interested in long-run benefit. B may therefore be more or less far-sighted on the basis of a discount rate on future utilities which, in every period, increasingly reduces (even though at a marginally decreasing rate) payoffs associated to further outcomes of the repeated game.

These factors influence calculation of the players’ rational choices. On the basis of its calculation of expected utility, each A_i chooses between **entry** and **no-entry** in light of the

current conditioned probability of the types of B. Obviously, in the first stages-games, the probability of types is such that the first players A_i will usually not place their trust in B, because the expected utility for **entry** is less than that of the alternative. Sooner or later, however, some A_i will decide to trust B if they have observed a series of **non-abuse**, as a result of which the conditioned probability of the honest type has increased sufficiently to give the **entry** choice an expected utility greater than **no-entry**. There is always a calculable number of learning periods (in which learning must never contradict the hypothesis that B is an honest type) necessary for the probability of the type in question to reach the critical threshold p^* above which a short-run player A_i for the first time will rationally decide to trust B.

Analysing player B's choices requires consideration of the equilibrium strategies of the iterated game. At a first sight, B may opt for the equilibrium strategy of each stage-game, namely **abuse**, which is certainly the best response to the choices made by the players A_i in the first periods. This strategy by player B gives rise to an equilibrium profile in the iterated game which outcome consists of an infinite series of (**no-entry**, **abuse**). However, player B has a different strategy available, which consists in exploitation of its knowledge of the mechanism by which the beliefs of the various A_i are updated. It may choose to simulate the behaviour of the 'honest' type until the stage occurs in which the conditioned probability of this type reaches the critical level p^* at which the first A_i will enter. At this point, B calculates whether to play the **no-abuse** action, and consequently induce the players A_i to enter again, or to profit from the first opportunity to defect by choosing **abuse**, thereby gaining a unilateral advantage from the (-1, 3) outcome on the first time but thereafter condemning itself to an infinite series of (0, 0) outcomes. If B is not impatient, and therefore if the discount rate of future utilities does not excessively reduce the value of the future prospects of cooperation, infinite outcomes of future cooperation (which begin once the first A_i has entered) are able to off-set the cost of the initial series of null outcomes (in which no A_i enters but B does not abuse), and to thwart the incentive to take advantage of an individual stakeholder as it enters. One possible B's rational strategy therefore is to sustain its reputation and to induce the sequence of stakeholders to trust it. The best response to this strategy by stakeholders A_i , from the time that the first of them has placed its trust in B, is to continue to be trustful until they observe a period in which B abuses. This gives rise to a game equilibrium profile in which a series of cooperative outcomes (**entry**, **non-abuse**) are observed from a certain point onwards (Fudenberg and Levine 1989).⁸ Hence, the long-run search for reputation induces the firm to behave as if it wants to fulfil its fiduciary duties towards the stakeholders.

However, it is essential to understand the conditions under which this result holds:

- (a) *signalling the types*: the firm must be able to signal the possibility that it is an honest type which does not abuse trust;
- (b) *quasi-simultaneity*: the firm and the stakeholder observe the result of each game simultaneously, for if the stakeholder acts first, the firm would have no reason to reveal its choice had the stakeholder not entered, so that there is no basis for *learning*;

⁸ For an simple illustration of this model and result see Sacconi (2000).

- (c) *observability of the results*: at the end of each stage-game, the stakeholder must be able to observe the outcome of the firm's choice without ambiguity, and it must be able to determine without ambiguity whether the firm has behaved according to a *type*. Because types can also be viewed as commitments (to a certain game action), the essential condition is that at the end of each stage-game each stakeholder (the current one) should be able to observe that 'what had to be done has been done';
- (d) *shared knowledge among stakeholders*: each stakeholder must be able to transmit what it has learnt in a given period to the stakeholder that comes next: that is, all the stakeholders in succession must have the same judgement on the firm's fulfilment of its commitments;
- (e) *absence of optimal mixed strategies types*: were the firm able to calculate the probability of *non-abuse* at which the stakeholder would be indifferent between entering and not entering, then it would abide by its *type* that refrain from abusing only at that minimum indispensable level of probability. This type would obviously not induce an equilibrium in which the firm does not abuse, but instead one in which it abuses with the maximum possible probability compatible with maintaining stakeholder indifference between entry and non-entry. Hence mixed equilibrium strategies must be excluded for the virtuous effect of reputation has to emerge.

In general, these conditions as a whole *are not* spontaneously fulfilled in situations relevant to the purposes of CSR, the consequence being that discretionary self-discipline (based on simple enlightened self-interest) normally fails. Putting aside the last assumption that will be traced back in section 4, the main reason for discretionary self-discipline fails is the *cognitive fragility* of reputation. This is evinced by conditions (a), (b), (c) and (d) above - all of which refer to the knowledge that the players must possess if the model is to hold true. Accumulating reputation may be prohibitively difficult if in order to show that a commitment has been maintained it is needed enabling *each* stakeholder to observe that *concrete* actions have been undertaken, or that the *concrete* results have been obtained, so that they match their description established *ex ante* in a commitment announced by the firm (a possible *type*).

Typically, CSR is involved in *incomplete contract* situations where a contract does not contain clauses covering unforeseen contingencies, so that there is no concrete benchmark against which claims of renegotiation can be assessed when unforeseen events occur. Moreover consider *unobservable quality*, such that the customer may not be able to verify a commitment over a quality level of a good or service on the basis of the information available to her by inspection or experience; or *organizational authority* where the 'boss' takes genuinely discretionary decisions with regard to tasks to be ordered the employees. Lastly, in *collusion situations* information about illicit agreements is reserved, so those not present when a bribe is negotiated are unable to determine whether a breach to a commitment of not bribing has taken place. These are all settings in which information or knowledge about the firm's action is incomplete or highly asymmetric. Either commitments have not been defined in relation to unforeseen events, and therefore cannot be verified, or their fulfilment is not observable. The problem is that incomplete information makes it impossible to determine whether 'what had to be done has been done': either it was not established *ex ante*, so that there is nothing to verify, or it is impossible to observe results by which it can at least be inferred whether the commitment has been respected (since the result coincides with at least

one of the possible results contemplated *ex ante*). Activation of the reputation mechanism is obstructed by a cognitive gap.⁹

3. Filling the gap in the reputation mechanism: a cognitive role for explicit ethical norms

If reputation may not be effective in supporting the firm in carrying out its commitments, may we nevertheless continue to say that extended fiduciary duties can be self-enforcing? Reputation is one of the most obvious social mechanisms backing endogenous compliance to social norms. Its failure puts at risk the same idea of a voluntary approach to CSR. This would imply that either the multi-fiduciary governance, even though morally justified, cannot be implemented, or that to be put in practice it needs a large reform of company law, which should make the board directly accountable and legally subjected to the control of all stakeholders, skipping any special status until now reserved to the owner.

Consider however that this is not the design of extended governance derived from the social contract model. What I have derived from that model was in fact that the stakeholders step forward from the first social contract to the second, where they legitimise one party to take over ownership and control over the firm, granted that his discretionary power be constrained by compliance with the extended fiduciary duties derivable from the first social contract. But this must operate as internal constraint, that is from within the owner's or his delegate manager's exercise of decision rights. This means that there must be some endogenous force compelling the board to comply not only with the owner claims (which may be even imperfectly enforced by residual rights of control), but also to duties owed to the non-controlling stakeholders. Thus only the existence of internal motivation or incentives is compatible with the design of the firm as an institution satisfying the requisites of the constitutional contract devised in part I.

Fortunately, there is an answer to this worry. It consists of substituting self-regulation in the *proper sense* - namely understood as a set of voluntary but nevertheless *explicit* and *standardised* ethical norms which are addressed to fill the cognitive gap - for the discretionary approach based on mere farsighted self-interest considered so far. Rather than responding to a concern over the enforcement of CSR norms by an external authority, it concentrates on the need to create the cognitive and informational bases that enable the social mechanism of reputation - with its endogenous rewards and punishments - to function properly. This comes about through the voluntarily-taken decision to accept *explicit norms with an appropriate structure*, established by the firm in the light of a multi-stakeholder social dialogue, such to induce impartial acceptability. For this reason, self-regulation is a voluntary but *not* discretionary approach. Voluntariness resides in the decision to endorse an *explicitly announced standard* for the firm's management system and governance, which is *ex ante* shared among the firm and its stakeholders. This standard sets out general principles, whose contents are such to elicit stakeholder consensus, as well as explicit commitments to compliance with principles and rules which are to be known *ex ante* by stakeholders. It is clear that stakeholders' consensus can be more easily obtained if the standard relative to the

⁹ Kreps made this point first (see Kreps 1990a).

strategic management system, intended to ensure CSR, is established by the firm through explicit dialogue with the stakeholders.

However, explicit statement and dialogue does not detract from the voluntary nature of the standards; nor does it preclude that compliance may then be obtained via the self-enforcement of the fiduciary duties established. How this occurs can once again be explained by referring to the reputation mechanism. The standard, and the procedures ensuring compliance with it, are announced *ex ante*; and it is *on these* – not in relation to particular (unforeseen) events or to particular (unobservable) actions or outcomes – that firm and stakeholders pass homogeneous judgement on *ex post* compliance with them. These duties assert – in the proper form – what is to be expected of the firm in *unexpected* situations, too, or in ones where the results of actions *are not observable*. Once the gap is filled, it is thus possible to reactivate the reputation reward and punishment mechanism, which generates endogenous incentives to comply with the standard itself. Everything rotates around the gap-filling function performed by the CSR management standard, whereby the firm's fiduciary duties towards its stakeholders are made explicit and announced. In order to work properly, three conditions must be met¹⁰:

(i) General and abstract principles

Principles define the *vision* of the social contract that each firm proposes to its stakeholders (which must therefore be completely identified). These principles must hence offer fair terms of treatment acceptable to each stakeholder. Their form is abstract and general, so that they apply to a wide variety of events, including those which cannot be predicted or described beforehand. Consequently, their application does not require an *ex ante detailed description* of any possible situation; all that is necessary is *ex post* recognition of the presence of certain abstract features which reflect a pattern established at the outset.¹¹

(ii) Precautionary rules of behaviour

Definition of principles allows identification of areas of potential opportunism where interactions between stakeholders and firm put those principles at risk (take this term in intuitive sense, without its probabilistic specification). Given each of these risky areas, precautionary rules of behaviour can be established which assure the relevant stakeholder that a particular form of opportunism has been avoided. The distinctive feature of these rules is that their implementation is not conditional on the actual occurrence of concrete foreseen situations. Operationally, they are applied when the belonging of the occurring situation into the domain in which a principle is breached exceeds a pre-announced threshold. Hence, their conditions of implementation can be established *ex ante* by the firm, and on these the stakeholder may legitimately form expectations about the firm's behaviour. Their application

¹⁰ This theory of reputation under unforeseen contingencies and incomplete contract is fully developed in Sacconi (2000, 2004a). For a design of a CSR management standard corresponding to these requirements see the Q-RES model in Sacconi, DeColle, Baldin (2003), another example is Clarkson Centre for Business Ethics (2002), *Principles of Stakeholder Management*.

¹¹ I'm here elaborating on a suggestion derived by David Kreps' theory of "corporate culture" (Kreps 1990a), however my modelling of principles is completely different than the one attempted by Kreps that used a focal points approach, a concept that I deem not very useful in this context (see Sacconi 2000 ch.4).

constitutes the evidence that no principle has been intentionally breached, and consequently that the firm's reputation is well-deserved.

(iii) Communication and dialogue with the stakeholders

Stakeholders base their assessments on the match amongst principles/rules announced *ex ante*, levels of membership into the principles domain exhibited by any event that has occurred, and the behaviour adopted. Dialog and communication allow developing a common understanding between the firm and its stakeholders concerning the former's principles and commitments and the latter's expectations. Dialog provides for a symmetric interpretation amongst the parties about critical situations so that no serious divergence on interpretation may emerge about contingencies where rules of conduct must be implemented..

Remarks

Concerning the first, general principles identify moral properties associated to abstract and universalizable characteristics, which are not necessarily bound to a complete description of every concrete contingency that might occur in all the states. Hence, in order to find out the characteristic identified by a principle, we do not need to look after a complete and detailed description of all the features of the possible states of the world. An abstract and general principle corresponds to a domain of application (a set of states) and membership to that domain is a matter of degree. Foreseen states of the world will belong or not to it in a clear-cut way, for they will exhibit or not the properties that are assumed to be associated to the given ethical principle. Unforeseen contingencies will define instead a vague domain of application of the principle, namely the set of states of the world such that their holding the descriptive characteristics associated to the principle is a *matter of vagueness*. Although their belonging can be a matter of vagueness, nevertheless, unforeseen contingencies will always belong to some degree to the domain of application of a general principle. It is exactly the abstractness of principles that makes possible their application to *every* situation, even if these are *ex ante* unforeseen, whereas concrete rules, contingent upon detailed state description, would be simply *mute*.

Vagueness is the price to be paid for being able "to say something" also about the unforeseen states of the world. But it is worth to be paid. The resulting situation is in fact completely different from the one occurring when expectations concerning the unexpected state are completely undetermined. In the case of general/abstract principles, we can try to manage vagueness by a measure of membership of each state into the domain of application of any principle. I suggest that a fuzzy membership function, taking its values in the real interval $[0,1]$, can be defined for each unforeseen state that occurs (or reveals to be possible *ex post*), which implies that the domain of application of a principle must be understood as a *fuzzy set*¹².

Whenever our language contained only concrete descriptions of actions, contingent upon *ex ante* known states, the occurrence of unforeseen states would make impossible to say neither what would be required by commitments at these states, nor what the meaning of action and outcomes *ex post* observed would be. Quite differently, we can use general and abstract

¹² Fuzzy set have been introduced by Zadeh (1965), for a partial synthesis the burgeoning literature on *fuzzy set* see Zimmerman (1991), for simple definitions useful to my application see Sacconi (2000).

properties to which unforeseen states adapt at least imperfectly. In this case, through a default inference, we may conclude that an unforeseen state with a degree of membership to the principle at least equal to α (for $1 > \alpha > 0$) is “normally” an exemplar of the principle¹³. Thus I conclude by the same mode of reasoning, even if I have not a complete proof that this is the case, that a rule of behaviour conforming to the principle must be carried out. The logic at work here is *default reasoning* - as based on a *fuzzy semantic* (i.e. a theory of degrees of truth at possible states of the world). Even if it is clearly fallible, default reasoning is the better we can reasonably do in order to face unforeseen contingencies that testify how limited human rationality is (Reiter 1980, Ginsberg 1987)¹⁴.

The second part of the cognitive device here has just entered the scene. Once it is established the membership threshold above which it is by default accepted that a state fulfils the properties required to be recognized as a member of the domain, a rule of conduct, conditional *simply* on satisfaction of *that threshold*, can be *ex ante* stated, without any request that we have *ex ante* any detailed knowledge of the relevant states of the world. Hence, if an unforeseen state belongs to the domain of a principle at least to degree α , then we conclude by default that such a state is a case in which “normally” a given rule of conduct does apply. What really matters is that the couple principles/rules of conduct allows to specify *ex ante* the *conditions* under which a certain procedure must be carried out *ex post*, *without request* to give *ex ante* any detailed concrete description of these states of affairs (as they may be completely unforeseen). Such conditions in fact consist essentially in that a situation, whatever it may be (even if *ex ante* unforeseen), must belong *at least to a certain degree* to the domain of application of a principle. Thus it is possible *ex ante* to *undertake* commitments and generate expectations on future behaviours, without any detailed knowledge or forecast of future states of the world, which may be left at least in part unforeseen.¹⁵

No doubt, such preventive rules of conduct will not induce utility maximization in every state. It is implicit in non-monotonicity of default reasoning that from vague information we are allowed to conclude that “normally” cases “such and such” are to be managed according to a certain procedure, but also that in the presence of additional information those conclusions may have to be revised.¹⁶

The main point however is that principles and rules of conduct allow at least provisory “completion of the contract”. This completion is allowed by the specification of what the stakeholder may expect in terms of commitments over certain *ex ante conditions*, by which we intend *not* so much the occurrence of future (unforeseen) states *per se*, but just *the fulfilment a threshold* of membership into the fuzzy-set domain, defined for a general principle. *Ex post* execution of rules of conduct will then give a reliable base in order to decide if “what had to be done has been done”. This allows resorting even under unforeseen

¹³ Notice that when degrees are 0 or 1 we are in a situation of no vagueness concerning the truth of statements about whether a case belongs to the domain of a principle or not. But I assume that this is true only for foreseen states of affairs, which are *ex ante* described within exactly the same language and state of knowledge in which we also *ex ante* put forward the principles.

¹⁴ For the application of default reasoning in the contexts of contract theory and reputation theory see Sacconi (2000, 2004a)

¹⁵ See Sacconi 2000 chap.8 and Sacconi (2004a).

¹⁶ On *Non-monotonic logic* see Ginsberg (1987).

contingencies to the mechanism of reputation effects. The “honest” type of the firm has to be replaced by the type who conforms to a rule of conduct when an ethical principle in the CSR management standard asks to do that, that is in all the contexts where the *ex ante* announced conditions on principles and rules are fulfilled.

Summing up, quite paradoxically, the criticism raised by Jensen (2001) that asking for multiple objectives and multiple obligations, CSR would make too complex managerial decisions, given decision-makers endowed with only bounded rationality, retorts on itself. The multi-fiduciary model of governance in fact is implemented through a voluntary but explicit CSR set of principles and rules of conduct that are able to overcome the problem of undertaking commitments in a world of bounded rationality and unforeseen events. A principle-and-rule-following model of decision, conditional on requisites specified in order to face vagueness and unforeseen constancies, fits the idea of bounded rationality. Those who conceded that the multi-stakeholder approach would over-complicate decision processes, have conceded too much to the adversaries (for this “excess” in concession, see Phillips, Freeman, Wicks 2003).

4. The motivational role of explicit ethical norms

4.1 *The problem of refined abuse*

Compliance with CSR voluntary but explicit norms (codes of ethics, management system standards etc.) can be effective once they work as parameters against which to assess a firm’s reputation. Yet a firm which has acquired a reputation as a relatively mild abuser of its stakeholders’ trust can nonetheless acquire their cooperation if the incomplete fulfilment of its duties generates to the stakeholder an expected payoff not less than that promised by his alternative choice of withdrawal from all transactions. In this case, the firm would fulfil its duties only to the minimum extent necessary to dissuade the stakeholder from exiting the relation (although it is quite unfair). Moreover if this strategy is available to the firm and it is able to support the corresponding reputation, it would be an equilibrium strategy and one that the firm would certainly try to select, given that under this equilibrium the stakeholder will acquiesce by relinquish to the firm the largest part of the surplus and getting practically nil for himself. Call this firm’s behaviour a strategy of *refined abuse* ¹⁷

¹⁷ This trace us back to condition (e) at the end of section 2 (‘absence of optimal mixed strategies’) as a possible obstruction of the use of reputation to support self-regulation. Note that in the game illustrated in Fig. 1, the payoff to player A from the non-entry strategy is zero, so that B’s optimal strategy to persuade A to enter is the offer of an infinitesimally positive payoff by the appropriate probability mixture of **abuse** and **no-abuse** strategies. Consider, however, that the dependence of reputation on the ‘general principles/precautionary procedures’ pairing, where the latter are activated by default on the basis of satisfaction of a sufficient relevance threshold for the principle, would suffice in itself to render the firm unable to calculate the optimal mixed strategy with the minimum probability of conformity necessary to prevent the stakeholder’s exit and enabling the firm to act opportunistically in all other cases. In fact, the firm is unable to foreseen *ex ante all* the possible states of the world in which it should implement a given action. Consequently, it is unable to calculate beforehand the probability with which it should abuse or not abuse trust optimally (given that the context of incomplete information would cause it to commit gross errors). As seen in sec. 3, rationality asks the firm to implement a more rigid decision-making procedure available to it, which is to apply a code of precautionary rules of behaviour whenever the relevant ethical principles are involved to a sufficient extent (see Sacconi 2004a).

This introduces the need considering the motivational force attached to explicit but self-imposed ethical norms (or CSR management standards). Given that these entail ethical principles and duties of conduct – and are therefore couched in the language of *deontology* – they enable resorting to stakeholder’s motivations, which extend beyond the mere material advantage deriving from transactions with the firm. Many stakeholders, in fact, have motives to act that are not purely self-interested or geared to material advantages (consequences). These stakeholders also place importance on the firm’s fulfilment of duties of CSR, deriving from the social contract, especially if the firm enunciates these duties in codes of ethics and communicates them externally. Hence, any deviation from the CSR standard, or from the firm’s ethical commitments, may undergo harsher punishment than would be the case if simple material interest were concerned.

In fact, assume the firm is a single player with a low level of intrinsic desire to conform to CSR fiduciary duties, although it has an ethical code and a standardized CSR management system that may constitute the term of reference for a behaviour ideally conforming to those duties. Unlike the firm, however, assume the stakeholders have strong conformist orientation, in that if the firm has adopted a standard and a code of ethics which affirms an ideal of fairness – and if they expect the firm to comply with the principles – then the stakeholders associate high intrinsic utility with the fact that the firm is behaving consistently with the code of ethics and principles of mutual advantage and fairness. By contrast, if the stakeholders behave cooperatively and expect the firm to know that they do that, any deviation by the firm from its principles will affect them not only materially but also because it contradicts their conformist principles. (This seems to be the case of the ethical investors or responsible consumers which expect that the respect for human rights and concern over environmental impact currently required in their countries, and on which companies are usually accepted as corporate good citizens at the national level, should be shown by those same companies wherever they operate.)

This prompts the question whether these hypotheses are such to prevent the firm’s refined abuse strategy. By hypothesis, the firm will adopt this strategy only if it is able to induce the stakeholder to enter (see again the trust game illustrated in Figure 1). However, it can be shown that if the stakeholder is conformist, it will refuse to enter, with the consequence that the firm’s refined abuse strategy does not induce equilibrium anymore.

4.2 *Conformist preferences and reciprocity*

In this section I summarize briefly some results of previous works (Sacconi 2004b, Grimalda and Sacconi 2002, 2004, Sacconi and Grimalda 2004), before applying them to the problem raised at the end of the previous section. Let me assume that stakeholders have not only self-interested motives to prefer but also ideological motives to prefer and that their accepted ideology coincides with the social contract of the firm - i.e. the guiding principle of an extended corporate governance system. In Grimalda and Sacconi (2002, 2004) we suggest that these two classes of motives can be accounted for by two types of preferences of the self and by their relative mathematical representation in the corresponding utility function.

To begin with, it must be recognized that strategic interaction generates states of affairs, which can be differently described according to their characteristics. A first description of states understands them as *consequences*. Consequences may be described as attributed just to

the acting Self - what happens to the decision maker in any state. This description is the basis of Self-interest: the Self has preference over consequence which are Self-referred. On the contrary, consequences may be attributed to every person (extended consequences) in so far as they can be understood as what happens to whichever individual. This makes a case for some sort of impartial consequentialist ethics like utilitarianism, or altruism etc. In general, if a player defines his preferences only over states *described as consequences*, then he has *consequentialist personal preferences*.

Let me come to the second type of preferences, what I call *conformist personal preferences*. States description is no less important here, but they are now described as sets of interdependent actions characterised in terms of whether they conform or not to a given abstract principle or ideal. This ideal can be captured by means of a function of individual first type utilities attached to states, which measures fairness of welfare distribution within each states of the world. Fix a pattern of behaviours (a vector of strategies), and define it as perfectly *deontological* if it fully conforms to the abstract principle of fairness - that is it maximises the function just defined. I call such a state the *ideal*. Then a degree of conformity to the ideal appended to each strategy choice can be identified by seeing whether the *ideal* comes about through the choice of each player given what he believes about other parties' choice. For each strategy combination each player's conformist preferences will follow by a measure of expected reciprocal conformity, which is based on:

- (a) an index of how much the player himself contributes to carry out the ideal by conforming or deviating from it, given what he believes about the other player's choice;
- (b) an index of how much the other player contributes to carry out the ideal by conforming or deviating from it (as seen through the eyes of first player), given what second player believes (and first player believes that second player believes) that first player will do.

To be sure, there is a relation between the two kinds of preferences - consequentialist and conformist. In fact, in order to define fairness, I look at distributions of payoffs, derived from *first type of preference* - i.e. material utilities. But this does not reduce second type preferences to the first one. First type utilities are no more than *rough materials* for the criterion defining second type preferences. What matters to the second description are not consequences or material payoffs as such, but a distributive property defined over payoffs, which is expressed by the function representing a principle of fairness.

Thus characterisation of second type preferences accords more with deontology than consequentialism. The more an expected state of affairs conforms to the ideal, the more it is preferred by a player (due to a measure of expected reciprocal conformity, which is at basis of each player's conformist preference over states). Moreover, there is no reason to link deontology to a belief that there is some objective source of value that would have an ontological reality "out there" (independent on the decision maker's affections). In fact, while conformist preferences depends on degrees of deontology, nonetheless deontology itself may be understood, as I do, simply as individual compliance with a fair distribution principle *that players could have rationally agreed in an ex ante hypothetical bargaining situation* .

Once these concepts are translated in a formal model, a player's overall utility function that combines two parts (which we assume to be separable), i.e. the representations of consequentialist and conformist preferences, can be defined. First, a form of the fairness-

function T must be specified, which represents formally the ideal. It has to be a mapping from the set of states (and first order utilities attached to them) to a fairness ordering ranging over states. A characterisation in contractarian terms of the ideal principle T is given by the Nash bargaining solution, i.e. *Nash social welfare function* $T(\sigma) = \prod_{i=1}^N (U_i - d_i)$ where d_i represents the reservation utility that agents can get when the process of bargaining breaks down. In this case d_i coincides with covering the costs of each player's specific investments, which means that fair bargaining over the surplus may start only if parties are assured that they will end up at least with reimbursement of the cost they must bear in order to participate in cooperation.

Then let define the two personal indexes of conformity, which will be compounded in a measure of mutual expected conformity and will enter the utility function of players.¹⁸ In this construction I take the point of view of player i (any other player j 's perspective is symmetrical).

A) Player i personal index of conformity:

It is player i degree of deviation from the ideal principle T (which varies from 0 to -1), due to player i choice, given her expectation on the player j behaviour. It is normalised by the magnitude of the difference between player i full conformity and no conformity at all conditional on player j choice

$$f_i(\sigma_i, b_i^1) = \frac{T(\sigma_i, b_i^1) - T^{MAX}(b_i^1)}{T^{MAX}(b_i^1) - T^{MIN}(b_i^1)} \quad (1)$$

where b_i^1 is the belief of player i over player j action, $T^{MAX}(b_i^1)$ is the maximum attainable by the function T given i 's belief, $T^{MIN}(b_i^1)$ is the minimum attainable by the function T given i 's belief, $T(\sigma_i, b_i^1)$ is the effective level attained by T when the player adopts strategy σ_i , given his belief on the other player's behaviour.

B) Estimation of the second player index of conformity to the ideal:

It is player j degree of deviation from the ideal principle T (which also varies from 0 to -1), as seen through player i beliefs - as well normalised by the magnitude of the difference between player j full conformity and no conformity at all, given what j believes (and player i believes that he believes) about player i choice

$$\tilde{f}_j(b_i^1, b_i^2) = \frac{T(b_i^1, b_i^2) - T^{MAX}(b_i^2)}{T^{MAX}(b_i^2) - T^{MIN}(b_i^2)} \quad (2)$$

where b_i^1 is player i *first order* belief over player j action (i.e. formally identical to a strategy of player j), b_i^2 is player i *second order* belief over the belief of player j about the action adopted by player i (i.e. formally identical to a player i strategy predicted by player j). These indexes are employed to build up the following ideal component of the utility function

¹⁸ In Grimalda and Sacconi (2002, 2004) we elaborate on Rabin (1993) in order to define our own model of reciprocity. As Rabin, therefore we are in the context of *Psychological games* as defined by Genakoplos et al (1989)

$$\lambda_i \left[1 + \tilde{f}_j(b_i^2, b_i^1) \right] \left[1 + f_i(\sigma_i, b_i^1) \right]$$

where the weight λ_i - which may be any positive real number - is an exogenous psychological parameter that expresses, previously than any consideration of reciprocity, how important deontology is within the motivational system of player i . As a whole the formula works as follow: if player i perfectly conforms to the ideal, given her expectation, while the player j is also expected to perfectly conform, then the two individual indexes take values zero, so that the resulting utility value due to conformism is $(1)(1) \lambda_i$. Thus the maximum conformist utility value equals λ_i . On the contrary if a player does not entirely conform, while not expecting the other player also entirely to conform, then the two indexes take negative values (possibly -1). Thus the utility calculation for conformist reasons reduces to $(1-x)(1-y)$ (possibly both equal to zero) multiplied the weight λ_i and gives less than λ_i (possibly zero) as conformist utility value.

The overall utility function V_i is the linear combination of the two components

$$V_i(\sigma_i, b_i^1, b_i^2) = U_i(\sigma_i, b_i^1) + \lambda_i \left[1 + \tilde{f}_j(b_i^2, b_i^1) \right] \left[1 + f_i(\sigma_i, b_i^1) \right] \quad (3)$$

It suggests that if a player predicts reciprocal conformism (as it enters the utility function), as far as weight λ_i is high, it is then possible that the overall utility of a strategy choice reverses the effect of player i simple consequentialist preferences represented by $U_i(\sigma_i, b_i)$. For example it may push the player to select strategies that he would never choose when he rest on material utility only. That is enough for the theory of conformist preferences.

4.3 How conformist preferences prevent refined abuse

In this section I provide a proof of the following proposition:

If we assume that at least stakeholders have (inter alia) conformist preferences, then the firm's opportunistic resort to a mixed strategy of minimal compliance with a code of ethics will not be anymore a rational strategy to the firm itself.

To do that, let first come back to the infinitely repeated game of reputation between a long-run firm B and one short-run stakeholder at time A_i , whose stage-game is the game of trust (see again fig. 2). To begin with, consider player's B type that adopts a mixed strategy $(2/3 \mathbf{a}, 1/3 \mathbf{no-a})$, i.e. B may try to develop the reputation of being this type by playing the two pure strategies with the attached probability throughout all the repetitions of the game (let assume that these mixed strategy types can be understood and learned by the A_i players). This type, once the stakeholder believes it with probability one, renders each player A_i indifferent among all his pure or mixed strategies, since in every case he gets no more than 0. Typically player A_i 's best response is remaining indifferent between her pure strategies and using the mixed strategy $(1/2 \mathbf{e}; 1/2 \mathbf{no-e})$, which gives as well A_i the expected payoff 0, and allows player B an expected payoff 1.33.

Of course, this is not an equilibrium of the stage-game, for against any probability of entrance by players A_i player B's best response is to play abuse. However we are considering the repeated game, assuming that mixed-strategy-types of player B can accumulate reputation by playing repeatedly their characteristic strategy. This suggests the following repeated game equilibrium: B could even improve his stage-game payoff over the Stackelberg pure

strategies payoff 2, by simply resorting to a commitment on the mixed strategy $(2/3-\varepsilon a, 1/3+\varepsilon no-a)$ (with ε as small as possible.) In fact, if player B is able to accumulate reputation that he is this type, player A_i necessarily enters in order to get the expected positive payoff 3ε (which, due to the infinitesimal ε , is practically nil), and gives player B a stage game expected payoff $2.66-\varepsilon$. Notice that player B best response is then not to deviate to stable defection (**abuse**), for this would change his reputation till a player A_i will consider only possible a nearly complete abusive type, so that no more A_i players would enter thereafter. This suggests that there is a mixed strategy equilibrium in which, also by not complying entirely, player B (the firm) can get a reputation such that stakeholders enter and acquiesce to a firm appropriating large part of the surplus. But this implies that the firm for the most of time would not carry out its fiduciary duties toward the stakeholders.

This pessimistic result obtains when no role is played by conformist preferences. Let now the more complex representation of stakeholders' system of preferences enter the picture. I will be as parsimonious as possible in introducing motivations that make easier compliance with fiduciary duties on the part of the firm. Thus, as before, I assume that only stakeholders (for example employees or consumers, or investors) stick to the ideal of socially responsible firm as defined by a code of ethics or a voluntary CSR management system, expressing the ideal model of a social contract amongst the firm and its stakeholders. This presupposes that the firm has at least deliberated and signalled a commitment to such a code, but *not necessarily* that it has developed that kind of attitude that I call conformist preference toward the reciprocal compliance with it. Hence, put at 0 the λ_B parameter in the manager's or entrepreneur's utility function by which I capture the weight of conformity within the utility function of player B, the "firm". On the contrary stakeholder A (I skip the individual index for simplicity of notation) has a positive weight λ_A and holds an *overall utility function* combining both consequentialist (self-interested) motives and conformist motives to act. Hence, the stakeholder's ideal of socially responsible governance of the firm may be calculated by means of the Nash social welfare function. Due to the very simplistic representation of the basic trust game, this welfare function is defined only over the firm's payoffs and the payoffs of the sole stakeholder taking part in the current stage-game, without considering that at any time there would be more stakeholders involved or that what the firm decides at one time may be relevant to other stakeholders at later times. I am confident that these more complex interactions would only reinforce the result of this section, as they would imply that more stakeholders' utility functions should be accounted for that include conformist preferences.

The relevant first order and second order player A's beliefs in this exercises are

$$\begin{aligned} b_A^1 &= (2/3a, 1/3 no-a), & \text{in short } (2/3, 1/3) \\ b_A^2 &= e, & b_A^2 = no-e \end{aligned}$$

In other words I first define the overall utility function of player A for a situation in which she believes that player B will abuse with probability $2/3$ and not abuse with probability $1/3$, while she has the second order belief that players B predicts that she (player A) will enter. True, this is not the effective mixed strategy equilibrium of the repeated game without conformism, for this would be the case only for a infinitesimal ε added to the probability of "no-abuse" strategy. However we may disregard the small ε as it practically does not change

the two players' payoffs, so that we may proceed as if the effective mixed strategy equilibrium were $(e, (2/3, 1/3))$. Hereafter, I will contrast this utility value with the *alternative* utility value for the case that player A believes that player B will play the "equilibrium" mixed strategy $(2/3, 1/3)$, but she will not play the entry strategy, so that her second order belief is that herself *does not enter* and the firm predicts that she doesn't enter (formally $b_A^2 = no-e$). In order to calculate player A's utility for the two alternatives (e and $no-e$), we must consider

- *strategy e* (case A*): player A's conformity index when she chooses e given that she believes player B is choosing $(2/3, 1/3)$, combined with (case B*) player B's conformity index based on player A's 2nd order beliefs, i.e. the index annexed to B choosing $(2/3, 1/3)$ (and player A believing it) given that he believes that she chooses e (and A believes that B believes it).
- *strategy no-e* (case A**): player A's conformity index when she chooses $no-e$ given that she believes that B plays $(2/3, 1/3)$, combined with (case B**) player B's conformity index based on player A's 2nd order beliefs, i.e. the conformity index resulting from B choosing $(2/3, 1/3)$ (and A believing it) given that he believes that she chooses $no-e$ (and player A believes that B believes it).

Let me first computing according to formula (1) the player A's conformity indexes of for the two cases considered

Case A* (player A's strategy e , given beliefs $(2/3, 1/3)$)

$$\frac{T(e, (2/3, 1/3)) - T^{MAX}(2/3, 1/3)}{T^{MAX}(2/3, 1/3) - T^{MIN}(2/3, 1/3)} = 0$$

Case A** (player A's strategy $no-e$, given beliefs $(2/3, 1/3)$)

$$\frac{T(no-e, (2/3, 1/3)) - T^{MAX}(2/3, 1/3)}{T^{MAX}(2/3, 1/3) - T^{MIN}(2/3, 1/3)} = 0$$

In fact *Nash social welfare function* in case A* gives $T(e, (1/3, 2/3)) = 0 \times 2.66$, and it is $T(no-e, (2/3, 1/3)) = 0 \times 0$ in case A**. At the same time, $T^{MAX}(2/3, 1/3)$ and $T^{MIN}(2/3, 1/3)$ are both 0 because player A's expected payoff is always 0 under player B strategy $(2/3, 1/3)$. The meaning of these 0-levels of the conformity index is better understood according to their interpretation as degrees of deviation from complete compliance with the ideal, conditioned on the other player expected choice. In both A* and A** cases the expected player B mixed strategy $(2/3, 1/3)$ nullifies any effort that payer A could make to enhance the level of ideal attainment. Whatever choice player A tries, in fact, the level of T is always 0. Thus A has not responsibility in any deviation from the maximum feasible level of T, given B's choice. Formally in fact the T level is at its maximum (which is also its minimum in this case) given player B opportunistic strategy that always reduces to 0 (or practically nil) player A's payoff. Then consider player B's conformity indexes in the two cases considered, i.e. when B uses

the mixed strategy against e or $no-e$. This index according to formula (2) takes the following values

Case B* (strategy $(2/3, 1/3)$ used by player B given his belief that A chooses e)

$$\frac{T((2/3, 1/3), e) - T^{MAX}(no-a, e)}{T^{MAX}(no-a, e) - T^{MIN}(a, e)} = \frac{4}{7} = -0.57$$

$T^{MAX} = 2 \times 2$ is in fact the maximum value of *Nash product* that ensues when B does not abuse given that A enters, while $T^{MIN} = -1 \times 3$ is its minimum value ensuing when B abuses given that A enters. The ideal's value if B plays the mixed strategy when A enters is nevertheless zero due to $T = 0 \times 2.66$. Notice that the index asks contrasting the T value brought about by player B opportunistic mixed strategy, in the case player A acquiesces, with the T^{MAX} value that player B could attain given the same player A choice. But this implies a relevant deviation from maximal conformity conditional on A behaviour, which can be *imputed entirely to player B* decision to play his mixed strategy in place of his $no-a$ strategy. In this case player B *relevantly* does not conform to the ideal and this results in the negative value taken by his conformity index.

Case B** (strategy $(2/3, 1/3)$ used by player B when he believes that A chooses $no-e$, and player A believes that B believes it)

$$\frac{T((2/3, 1/3), no-e) - T^{MAX}(no-a, no-e)}{T^{MAX}(no-a, no-e) - T^{MIN}(a, no-e)} = 0$$

In fact, for whichever player B choice given the expected $no-e$ by players A, T^{MAX} as well as T^{MIN} are both zero, and this is also true to the single payoff got by player A under $(2/3, 1/3)$. Given his belief $no-e$, players B cannot significantly deviate from the ideal, for also by playing his mixed strategy he is not accountable for a deviation from the maximal ideal's value given no entry by player A. Comparing B* and B**, it can be said that the *intention* of exploiting player A attitude to acquiesce *implies a significant responsibility* of B for a deviation (non conformity) to the ideal only conditional on the expectation that in effect player A will acquiesce, for it is exactly in this case that he does not reciprocate player A's conformity.

Finally, according to formula (3), I can calculate player A's overall utility values for the two alternative strategies e and $no-e$ respectively, given that player A predicts player B will use strategy $(2/3, 1/3)$, under the assumption that player B predicts player A choice, i.e. he believes (and she believes that he believes) that player A uses either strategy e or strategy $non-e$ respectively. The material payoff to player A if she plays "enter" under the equilibrium mixed strategy is 0 (accounting also for an additional infinitesimal probability ϵ attached to $no-a$, the payoff nevertheless is practically nil), whereas her conformist utility is based on the indexes A* and B*. Thus player A overall utility for strategy e is

$$V_A(e, b_A^1, b_A^2) = 0 + \lambda_A(1 + (-0.57))(1 + 0) = 0.43\lambda_A$$

On the other hand player A conformist utility for strategy *no-e* is given by the indexes A** and B**, whereas her material payoff is again 0. Thus A's overall utility for strategy *no-e* is

$$V_A(\mathbf{no-e}, b_A^1, b_A^2) = 0 + \lambda_A(1+0)(1+0) = \lambda_A$$

A conclusion follows straightforwardly: player A with conformist preferences does not accept to play the (psychological) mixed equilibrium strategy of the repeated trust game. As far as player B's conformity index annexed to the equilibrium mixed strategy is negative and, *ceteris paribus*, the weight λ_A the stakeholder attaches to conformity is positive, the logic of (psychological) strategic choice under conformist preferences reverses the result of standard strategic calculation in a repeated trust game with mixed strategies.¹⁹ Consider that, granted λ_A is positive, this result typically follows from the opportunistic nature of player B's mixed strategy type. In fact he endeavours to minimize (till to nil) stakeholders' payoffs, whereas he is convincing nevertheless them to enter in order enabling the firm to appropriate a part of the surplus larger than the "cooperative" payoff (2) in each stage-game. Of course were B ready to concede more to A, for example by rising the additional probability ϵ of "no abuse" to a substantial level, then A's firmness in rejecting "entry" would be lessened. However straightforward calculation shows that, *ceteris paribus*, A stops rejecting "entry" only when $T((2/3-\epsilon), 1/3+\epsilon), e$ takes value 4, which means that probability of "no abuse" equates 1 - i.e. player B mixed strategy type degenerates to his "honest" type.

5. Conclusion

In part II of this essay I have dealt with compliance and implementation of the social contract over the fiduciary duties' institutional framework of the socially responsible corporation. Endogenous incentives and motivations to comply with the CSR corporate governance structure have been searched for in the domain of reputation games, according to the idea that the firm may benefit from reputation when it behaves respectfully to the fiduciary duties owed to its stakeholders. However resort to the standard models of reputation has been shown to be obstructed: first, cognitive limitations jeopardize the undertaking and specification of commitments with respect to unforeseen contingencies. Second, firm's sophisticated opportunism may take advantage of the other's self-interest, inducing the stakeholder to acquiesce to abuse on the part of the firm in place of inducing the firm itself to conform to the social contract.

As far as the first problem is concerned, I provided an answer based on the cognitive role of explicit norms of ethics and CSR management and governance standards. These fill the gaps thanks to their logical structure, requiring general and abstract principles of ethics, precautionary rules of behaviour and dialog with the stakeholder in order to get to common understanding about the belonging of the occurring states of affairs in the domain of application of the couple principle/rules. It is remarkable that the use of explicit norms as decision device, able to overcome cognitive limitations in managing reputation, introduces a

¹⁹ Equilibrium strategies here must be understood as Nash psychological equilibrium strategies, the main difference being that the conditions for a non cooperative equilibrium are satisfied under the hypothesis that beliefs are inserted directly within the scope of the payoff function of every player.

stark change in the economic agents' logic of decision, as fuzzy logic and default reasoning are put at the basis of what elsewhere is called procedural rationality.

This is strictly connected to the underlying contractarian nature of business ethics principles and norms. They are fruitful essentially for their abstractness and generality. Albeit they are vague, their virtue is ability to extend from one context to another. Such principles can typically be understood as chosen in a hypothetical decision position, in which any decision makers by a counterfactual mode of reasoning detaches himself from details and concrete personal characteristics of any particular situation, in order to get to an agreement on universalisable principles acceptable by all and applicable to very wide classes of situations, even if ex ante unforeseen. Moreover the principle that is used to discipline the choice of the firm in the trust game under unforeseen contingencies is exactly the contractarian principle of fair bargaining we arrived at in part I, i.e. the Nash bargaining solution of an ideal situation of contracting among the firm and its stakeholders. This principle, applied within the domain where the occurring situation fulfils the relevant general and abstract proprieties required, commands the firm not to abuse the stakeholder and letting him to participate in the surplus (see Sacconi 2000). The bulk of this theory is a model of reasoning that makes possible undertaking principle-based-commitments with reference to unforeseen contingencies.

By the same contractarian principle of fairness I also give an answer to the second drawback of reputation theory, which leads us to the motivational role of contractarian business ethics norms. In fact the contractarian principle of fairness (the Nash bargaining solution of an ideal bargaining game) is a building block of the model of conformist preferences, which allows formalising the prompt to act out of the desire to conform to an ideal, granted that the same ideal is agreed upon by other players and also these players are expected to reciprocate conformity to the same ideal.

This new model of preference is employed in giving the following remarkable result: if stakeholders share the ideal of a multi-stakeholder firm (to say the contractarian principles of a corporate code of ethics), and they earn ideal utility from reciprocal expected conformity to that code - as far as it is assumed that the firm just announces a code of ethics requiring that Nash welfare function be maximised - *then* a mixed strategy of the repeated trust game inducing stakeholders to forbear exploitation cannot be any more an equilibrium.

This depends on the very nature of conformist preferences as they follow from reciprocity in conformity. If one examines the significance (or the intention) of the stakeholder's decision whether or not to enter, given the hypothesis that the firm wants to adopt the sophisticated abuse mixed strategy, the stakeholder is entirely unable to alter the situation's proximity to the social contract ideal, because the value resulting from her choice in terms of approximation is always close to zero (whether it decides to submit to the firm's strategy or whether it eschews the transaction). However, when she considers the firm's behaviour, given that the latter expects that in equilibrium the former will want to enter, the significance of the firm's choice is a marked deviation from the social contract ideal, because the intention is to expropriate the stakeholder precisely when it follows a cooperative behaviour. This, therefore, negatively affects the conformist preferences of the stakeholder, which will be doubly dissatisfied with the firm's behaviour, for her material remuneration is zero and at the same time the firm's action in itself increases the distance from the principle. The inclusion of a negative conformity index in the overall utility function entails that the stakeholder will not

enter the relation with the firm. Indeed he will punish the firm more harshly than the damage that the firm's calculated abuse would warrant in the absence of conformist preferences. Put otherwise, the firm must increase material payments to the stakeholder in order to off-set the negative impact of its incomplete compliance with the principle on the stakeholder's preferences. But this obviously also entails a choice of behaviour increasingly less distant from fairness, which therefore also reduces the stakeholder's negative evaluation of the firm's compliance

References

- Bernheim, B. (1994), "A Theory of Conformity", *Journal of Political Economy*, Vol. 102, N. 5, pp.841-877.
- Binmore K. (1994), *Playing Fair*, Cambridge Mass., MIT Press.
- Binmore K. (1997), *Just Playing*, Cambridge Mass., MIT Press
- Buchanan J. (1975), *The Limits of Liberty*, The Univ. of Chicago Press,
- Camerer C. and E.Fehr (2002), *Measuring social norms and preferences using experimental games: A guide for social scientists*, Institute for Empirical Research in Economics, University of Zurich, WP N.1424-0459
- Charness, G. and Rabin, M. (2002), "Understanding Social Preferences with Simple Tests The Quarterly Journal of Economics, August 2002, pp.818-869
- Clarkson Centre for Business Ethics (2002), *Principles of Stakeholder Management, Business Ethics Quarterly*, 12, 2, , pp.257-264.
- Coleman J. (1992), *Risks and Wrongs*, Cambridge , Cambridge University Press.
- Donaldson T. and T.W. Dunfee (1995), "Integrative Social Contracts Theory". *Economics and Philosophy*, 11, pp.85-112
- Donaldson T. and T.W. Dunfee, (1999), *Ties that Bind; A Social Contract Approach to Business Ethics*, Harvard Business School Press.
- Falk A. and E. Fehr , U. Fischbacher (2003), "On the Nature of Fair Behaviour", *Economic Inquiry*, Vol. 41, N.1, pp.20-26
- Falk A. and Fischbacher, U. (2000), "*A Theory of Reciprocity*", Institute for Empirical Research in Economics, University of Zurich, WP N. 6
- Fehr E. and A.Falk (2002), "Psychological Foundation of Incentives", Centre for economic studies & ifo Institute for economic research, working paper N.714
- Fehr, E. and Schmidt, K. (2001): "*Theories of Fairness and Reciprocity – Evidence and Economic Applications*", Institute for Empirical Research in Economics, University of Zurich, WP N. 75
- Frey B. (1997), *Not Just for the Money*, Brookfield, Edward Elgar.
- Fudenberg D. and J.Tirole (1991) , *Game Theory*, Cambridge Mass., MIT Press
- Fudenberg D. and D.Levine (1989), "Reputation and Equilibrium Selection in Games with a Patient Player", *Econometrica*, 57, pp.759-778.
- Fudenberg D. (1991), "Explaining Cooperation and Commitment in Repeated Games", in J.J.Laffont (ed.) *Advances in Economic Theory, 6th World Congress*, Cambridge, Cambridge University Press.
- Gauthier D. (1986), *Morals by Agreement*, Oxford, Clarendon Press.
- Gauthier D. (1969), *The Logic of Leviathan*, Oxford, Clarendon Press
- Geanakoplos, J., Pearce, D. and Stacchetti E. (1989), "Psychological Games and Sequential Rationality", *Games and Economic Behavior*, Vol. 1, pp. 60-79
- Ginsberg M.L. (1987): *Reading in Nonmonotonic Reasoning*, Los Altos California , Morgan Kaufmann Publisher Inc.
- Grimalda G. and L.Sacconi (2002), *The Constitution of the Non profit Enterprise: Ideals, Conformity and Reciprocity*, LIUC paper n.110, Catellanza, Varese.
- Grimalda G. and L.Sacconi (2004), "The Constitution of the Not-For-Profit Organisation: Reciprocal Conformity to Morality", *Constitutional Political Economy*, (forthcoming 2005)
- Hampton J. (1987), *Hobbes and the Social Contract Tradition*, Cambridge U.P.

- Hobbes T. (1651), *Leviathan*, London
- Hume D.(1739), *Treatise on Human Nature*, ed. by D.Fate Norton and Mary Norton, Oxford University Press, 2000
- Jensen M.C. (2001), "Value Maximization, Stakeholder Theory, and the Corporate Objective Function" *Journal of Applied Corporate Finance*, Vol. 14, No 3, Fall.
- Kreps D. (1990a), "Corporate Culture and Economic Theory" J.Alt and K.Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge, Cambridge University Press.
- Kreps D. (1990b), "Static Choice in the Presence of Unforeseen Contingencies", in P.Dasgupta, D.Rae, O.Hart, E.Maskin (eds.), *Economic Analysis of Markets and Games*, Cambridge Mass. , The MIT Press
- Kreps, D. (1998), "Bounded Rationality", in *The New Palgrave Dictionary of Economics and Law*
- Kreps D. and R.Wilson (1982), "Reputation and Imperfect Information", *Journal of Economic Theory*, 27 pp.257-279.
- Kreps D., Milgrom P., Roberts J. and R.Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma", *Journal of Economic Theory*, 27 pp.245-252 .
- Lewis D. (1969), *Convention, A Philosophical Study*, Cambridge, Mass. Harvard University Press.
- McDermott D. and J.Doyle (1980), "Nonmonotonic Logic I", *Artificial Intelligence*, 13, pp.41-72.
- Pettit P. (1990), «Virtus normativa. Rational Choice Perspectives », *Ethics*, 100, pp.725-755.
- Phillips R., Freeman E. and Wicks A.C. (2003). "[What Stakeholder Theory is Not](#)" *Business Ethics quarterly*, vol. 13, 4, pp. 479-502
- Posner E. A. (2000), *Law and Social Norms*, Cambridge Mass, Harvard UP
- Rabin, M. (1993): "Incorporating Fairness into Game Theory", *American Economic Review*, Vol. 83, N. 5, pp. 1281-1302.
- Reiter R. (1980), "A Logic for Default Reasoning", *Artificial Intelligence*, 13, pp.81-132.
- Sacconi L. , S. DeColle and E.Baldin (2003), "The Q-RES Project: The Quality of Social and Ethical Responsibility of Corporations" in Wieland J. (ed.), *Standards and Audits for Ethics Management Systems, The European Perspective* , Springer Verlag, Berlin, pp.60-117
- Sacconi L. (1991), *Etica degli affari, individui, imprese e mercati nella prospettiva dell'etica razionale*, Milano , Il Saggiatore
- Sacconi L. (1995) "Equilibrio e giustizia (parte I): la stabilità del contratto sociale; (Parte II): la selezione del contratto sociale" *Il giornale degli economisti*, n.s. LII, 10/12, '93, pp.479-528, pp.529-575
- Sacconi L. (2000), *The Social Contract of the Firm. Economics, Ethics and Organisation*, Springer Verlag, Berlin.
- Sacconi L. (2004a), "Incomplete Contracts and Corporate Ethics: A Game Theoretical Model under Fuzzy Information", in F.Cafaggi, A. Nicita and U. Pagano (eds.), *Legal Orderings and economic institutions*, London, Routledge.
- Sacconi L. (2004b); "The efficiency of the non profit enterprise: constitutional ideology, conformist preferences and reputation", in B. Hodgson (ed.) *The Invisible hand and the common good*, Berlin, Springer Verlag.
- Sacconi L. and Grimalda G. (2004), "Ideals, conformism and reciprocity: A model of Individual Choice with Conformist Motivations, and an Application to the Not-for-Profit Case" in L.Bruni P.L.Porta (eds) *The paradoxes of happiness in economics (proceedings of the international conference)*, Brookfield, Edward Elgar, (forthcoming 2005)
- Schotter A. (1981), *The Economic Theory of Social Institutions*, Cambridge, Cambridge U.P.
- Skyrms S. (2004), *The stag-hunt and the evolution of the social structure*, Cambridge, Cambridge U.P.
- Sugden R. (1986), *The Economics of Rights, Co-operation and Welfare*, London, Basil Blackwell
- Sugden, R. (1998), "Normative expectations: the simultaneous evolution of institutions and norms", in Ben-Ner, A. and Putterman, L. (eds): *Economics, Values, and Organization*, Cambridge, Cambridge University Press, pp. 73-100.
- Zadeh L.A. (1965): "Fuzzy Sets", *Information and Control*, 8, pp.338-353.
- Zimmerman H.J. (1991): *Fuzzy Set Theory and Its Applications*, 2nd revised ed., Dordrecht-Boston (Kluwer Academic Press).

Elenco dei papers del Dipartimento di Economia

2000.1 *A two-sector model of the effects of wage compression on unemployment and industry distribution of employment*, by Luigi Bonatti

2000.2 *From Kuwait to Kosovo: What have we learned? Reflections on globalization and peace*, by Roberto Tamborini

- 2000.3 *Metodo e valutazione in economia. Dall'apriorismo a Friedman*, by Matteo Motterlini
- 2000.4 *Under tertiarisation and unemployment.* by Maurizio Pugno
- 2001.1 *Growth and Monetary Rules in a Model with Competitive Labor Markets,* by Luigi Bonatti.
- 2001.2 *Profit Versus Non-Profit Firms in the Service Sector: an Analysis of the Employment and Welfare Implications,* by Luigi Bonatti, Carlo Borzaga and Luigi Mittone.
- 2001.3 *Statistical Economic Approach to Mixed Stock-Flows Dynamic Models in Macroeconomics,* by Bernardo Maggi and Giuseppe Espa.
- 2001.4 *The monetary transmission mechanism in Italy: The credit channel and a missing ring,* by Riccardo Fiorentini and Roberto Tamborini.
- 2001.5 *Vat evasion: an experimental approach,* by Luigi Mittone
- 2001.6 *Decomposability and Modularity of Economic Interactions,* by Luigi Marengo, Corrado Pasquali and Marco Valente.
- 2001.7 *Unbalanced Growth and Women's Homework,* by Maurizio Pugno
- 2002.1 *The Underground Economy and the Underdevelopment Trap,* by Maria Rosaria Carillo and Maurizio Pugno.
- 2002.2 *Interregional Income Redistribution and Convergence in a Model with Perfect Capital Mobility and Unionized Labor Markets,* by Luigi Bonatti.
- 2002.3 *Firms' bankruptcy and turnover in a macroeconomy,* by Marco Bee, Giuseppe Espa and Roberto Tamborini.
- 2002.4 *One "monetary giant" with many "fiscal dwarfs": the efficiency of macroeconomic stabilization policies in the European Monetary Union,* by Roberto Tamborini.
- 2002.5 *The Boom that never was? Latin American Loans in London 1822-1825,* by Giorgio Fodor.
- 2002.6 *L'economia senza banditore di Axel Leijonhufvud: le 'forze oscure del tempo e dell'ignoranza' e la complessità del coordinamento,* by Elisabetta De Antoni.
- 2002.7 *Why is Trade between the European Union and the Transition Economies Vertical?,* by Hubert Gabrisch and Maria Luigia Segnana.
- 2003.1 *The service paradox and endogenous economic growth,* by Maurizio Pugno.
- 2003.2 *Mappe di probabilità di sito archeologico: un passo avanti,* di Giuseppe Espa, Roberto Benedetti, Anna De Meo e Salvatore Espa.

(Probability maps of archaeological site location: one step beyond, by Giuseppe Espa, Roberto Benedetti, Anna De Meo and Salvatore Espa).

2003.3 *The Long Swings in Economic Understanding*, by Axel Leijonhufvud.

2003.4 *Dinamica strutturale e occupazione nei servizi*, di Giulia Felice.

2003.5 *The Desirable Organizational Structure for Evolutionary Firms in Static Landscapes*, by Nicolás Garrido.

2003.6 *The Financial Markets and Wealth Effects on Consumption An Experimental Analysis*, by Matteo Ploner.

2003.7 *Essays on Computable Economics, Methodology and the Philosophy of Science*, by Kumaraswamy Velupillai.

2003.8 *Economics and the Complexity Vision: Chimerical Partners or Elysian Adventurers?*, by Kumaraswamy Velupillai.

2003.9 *Contratto d'area cooperativo contro il rischio sistemico di produzione in agricoltura*, di Luciano Pilati e Vasco Boatto.

2003.10 *Il contratto della docenza universitaria. Un problema multi-tasking*, di Roberto Tamborini.

2004.1 *Razionalità e motivazioni affettive: nuove idee dalla neurobiologia e psichiatria per la teoria economica?* di Maurizio Pugno.

(*Rationality and affective motivations: new ideas from neurobiology and psychiatry for economic theory?* by Maurizio Pugno.

2004.2 *The economic consequences of Mr. G. W. Bush's foreign policy. Can th US afford it?* by Roberto Tamborini

2004.3 *Fighting Poverty as a Worldwide Goal* by Rubens Ricupero

2004.4 *Commodity Prices and Debt Sustainability* by Christopher L. Gilbert and Alexandra Tabova

2004.5 *A Primer on the Tools and Concepts of Computable Economics* by K. Vela Velupillai

2004.6 *The Unreasonable Ineffectiveness of Mathematics in Economics* by Vela K. Velupillai

2004.7 *Hicksian Visions and Vignettes on (Non Linear) Trade Cucle Theories* by Vela K. Velupillai.

2004.8 *Trade, inequality and pro-poor growth: Two perspectives, one message?* by Gabriella Berlofffa and Maria Luigia Segnana.

2004.9 *Worker involvement in entrepreneurial nonprofit organizations. Toward a new assessment of workers? Perceived satisfaction and fairness* by Carlo Borzaga and Ermanno Tortia.

2004.10 *A Social Contract Account for CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification* by Lorenzo Sacconi

2004.11 *A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity* by Lorenzo Sacconi

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO