EconomEtica

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

N.23 August 2010

Lorenzo Sacconi

A Rawlsian View of CSR and the Game Theory of its Implementation (Part II): Fairness and Equilibrium

Working papers

A Rawlsian View of CSR and the Game Theory of its Implementation (Part II): Fairness and Equilibrium[1]

Lorenzo Sacconi

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

1 Introduction

This is the second part of an comprehensive essay on the Rawlsian view of corporate social responsibility (CSR thereafter) understood as an extended model of corporate governance and objective function, based on the extension of fiduciary duties owed to the sole owner of the firm to all the company stakeholder (for this definition see part I, Sacconi, 2010a) *infra*). As in the first part, CSR is also understood as a self sustaining institution – i.e. as a self sustaining system of descriptive and normative beliefs consistent with the equilibrium behaviors performed repeatedly by agents in the domain of action of corporate governance (firms and their stakeholders). But equilibria are multiple in the game representing the strategic interaction among the firm and its stakeholders - modeled as a repeated trust game or some similar 'social dilemma game' (Ostrom, 1990). Thus asserting that CSR satisfies the Nash equilibrium condition as an institution is not enough. There is also an equilibrium selection problem. This the place where the Rawlsian social contract (Ralws, 1971, 1993) enters again the picture by performing its main role as normative equilibrium selection device from the ex ante perspective: that is, the ex ante impartial selection of a unique equilibrium amongst the many possible in the repeated trust game involving the firms and its stakeholders. Note that this was its second role previously suggested (see sec. 5 part I, and left to this part where it is treated at length), as distinguished from the role of shaping the players' expectations so that in the ex pot perspective they are able to predict the agreed solution as the

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

result of a cognitive process of beliefs convergence to the equilibrium, which is focused on in part III, (see Sacconi, 2010c and Sacconi 2008).

To this end (in section 2) I shall discuss at length the rehabilitation of the Rawlsian maximin principle provided by Ken Binmore's game theoretical reformulation of the social contract. (Binmore, 1984, 1989, 1991, 1994, 1998, 2005). Contrary to the belief that Rawls' view was utopian, it is shown that the maximin principle provides the best account of the social contract under the assumption that in a 'state of nature' any agreement on principles for institutions must be self-sustainable. In other words, to be self-sustainable and incentive-compatible, the agreement must be egalitarian, or in the best interest of the worst-off player.

Such an unconventional result has overarching implications also for the constitutional contract on the firm's governance and control structures. This is a theory to make sense of the idea of extended fiduciary duties put forward in previous works (Sacconi, 1997, 2000, 2006a,b, 2007). Its main point was that the stakeholders' constitutional agreement (seen as the rational solution of an original bargaining game) will complement the efficient control structure with further social responsibilities toward non-controlling stakeholders, enabling them to participate in the surplus created by joint production through a redress rule against the abuse of authority (sec. 3). However, when a constitutional bargaining situation is considered such that the only feasible constitutions are allocations of exclusive property and control rights, a strong imbalance of bargaining power is inevitable, so that asymmetry in the final surplus distribution will reflect the asymmetry of decision rights. Then, an outcome corresponding to the arrangement of rights (ownership and control rights plus redress rights with the attached fiduciary duties) that immunizes non-controlling stakeholders against abuse of authority, and gives them an opportunity to participate in the surplus created by joint production, may not belong in the equilibrium space of the constitutional choice game (sec. 4). This means that the outcome of such a redress mechanism cannot be obtained in equilibrium (violating the self-sustainability condition) .

The idea is that each constitution corresponds to a set of feasible (equilibrium) outcomes, and each of them comprises a post-constitutional bargaining solution within its feasible set of outcomes. Different constitutions - as they allocate rights of control to one player or another - will have post-constitutional bargaining solutions differently favorable to one or another player, but not equally favorable to all. Agreement at the constitutional stage selects the allocation of exclusive rights of ownership and control endowed with the most efficient post-constitutional solution in terms of incentives for the accomplishment of specific investments and in terms of wealth maximization. Players who forgo control in order to make agreement on the most efficient control structure possible, then need to be redressed through fiduciary duties. Implementation of such duties is an outcome coinciding with an equitable compromise (a linear combination) of the post-constitutional rational solutions preferred by different stakeholders as they relate to different allocations of rights, some in favor of one stakeholder, some in favor of another. But when the assumption is made that the only feasible outcomes (corresponding to equilibria) are those belonging to the outcome set of constitutions asymmetrically allocating ownership and control rights, then the quite obvious possibility arises that the symmetric outcome of an equitable redress mechanism does not correspond to any feasible outcome.

Many scholars of corporate governance accustomed to accepting second-best solutions would then be ready to give up fairness and extended fiduciary duties in order to achieve nothing more than the most efficient constitution of the firm. Remarkably enough, application of the Rawls-Binmore theory to the social contract on corporate governance structures yields quite the opposite suggestion (see sec. 5). In order to be consistent with the requirement of self sustainability, the impartial agreement must select the constitution with the best egalitarian solution among all the alternative feasible constitutions. That is to say, a constitutional arrangement must be chosen such that, within its feasible outcome set, the solution that maximizes the position of the worst-off stakeholder is traceable accepted because this is the best egalitarian solution with respect to all the egalitarian solutions available under alternative constitutions. Pareto dominance, as a principle of unanimous agreement, is therefore to be applied only to the comparison of feasible egalitarian solutions under alternative constitutions. The social contract will select the constitution with the relatively most Pareto-efficient egalitarian solution. What is most important here is that this result follows straightforwardly from the requirement that the social contract should select an outcome belonging to the set of (impartial) equilibria, i.e. a self-sustaining institution.

Moreover, the Rawlsian theory of corporate governance refutes much of the traditional wisdom in the domain of corporate governance as it has been viewed by both new institutional economics and law & economics (sec. 6). Quite unconventionally again, fairness precedes both efficiency and welfare maximization (contrary to Kaplow and Shavell), and it also precedes aggregate transaction costs minimization (against Hansmann 1988, 1996). Even libertarians like Hayek's followers - who typically believe that rules of behavior should spontaneously emerge from endogenous motivations respecting free choice – will have to concede that under the simple ethical constraint of impartiality egalitarianism is a natural consequence of the self sustainability of institutions in the domain of corporate governance.

2 Normative selection of an equilibrium: Binmore vindicates Rawls

By ‘normative role’ I mean the function of a contractarian fairness principle in giving impartial reasons for singling out a unique equilibrium solution amongst the many possible. Note that the normative principle is here used to choose an equilibrium point within the equilibrium set of the game to be played afterwards in the implementation phase. The perspective is still that of an *ex ante* impartial choice, but it now concerns equilibria, that is, game solutions that are self-enforceable.

In order to accomplish this endeavor a social contract theory is needed as an ex ante equilibrium selection tool. Ken Binmore has provided such a theory as a game theoretical reinterpretation of John Rawls’ famous maximin principle of justice (Binmore, 2005)[1].

2.1 The game of life

The social contract on constitutional principles takes place against the background of a *state of nature* called the “game of life” (Binmore 2005). Assume that there are two players for simplicity; and then that it is a repeated game, for example a repeated asymmetrical prisoner’s dilemma (PD) or something similar to it (for example a repeated Trust Game, whereby the second player has an advantage over the first because she may abuse her trust, whereas she can only protect herself by refraining from any cooperation). Its payoff set is a convex-compact space resulting from attaching the players’average discounted payoff to each repeated game strategy profile mixing both players’ cooperation and cheating in whatever proportion along the repetitions of the stage games. To exemplify, the payoff space represents outcomes of profiles whereby both players completely cooperate, they both never cooperate, they choose cooperating and cheating with the same frequency, as well as profiles whereby one party adopts cooperation more frequently (in whatever proportion) than the other and vice versa. As a whole, the payoff space (in terms of average discounted payoffs) amounts to the set of all the convex combination in whatever proportion of the stage game pure payoff vectors. According to the folk theorem, the equilibrium set of this game again in terms of average discounted payoffs is represented by an extensive region of the convex compact payoff space (see Fudenberg and Tirole, 1991)[2]. On the south-west side of the payoff space (possibly at the utility axes’ origin), in correspondence to the profile “never cooperate throughout all the repetitions", there is the worst possible equilibrium point for both the players. The payoff space's region to the north-east of this point is made up of points corresponding to equilibrium strategy profiles affording the players any non-negative surplus over the worst possible equilibrium result. In this perspective, the social contract works as a way to single out principles able to select just one amongst the many equilibrium profiles of the repeated game, affording some mutual advantage to both the players.

To keep things simple, let us again assume that there are only two players. The repeated game is played by player 1 in the role of Adam, A for short, and player 2, in the role of Eve, E for short. Adam is systematically in an advantage position over Eve because of some natural or historical brute fact (natural power, brute force). Hence the repeated game equilibrium set is Z_{AE} (from the name of the players - Adam and Eve; see *fig.1*), which is an asymmetric space. This means that within the equilibrium set Z_{AE} of the repeated game there are equilibrium pairs advantaging A over E or E over A in the relative sense; but the in absolute sense the equilibrium pairs preferred by player A give him a much higher payoffs than those given to player E by the equilibrium pairs she prefers. The best chances of profiting from the game are quite different for the two players. In other words, there are many outcomes in which Adam gets a much higher payoff than Eve, whereas symmetrical outcomes, giving Eve a similar higher payoff, are not possible.

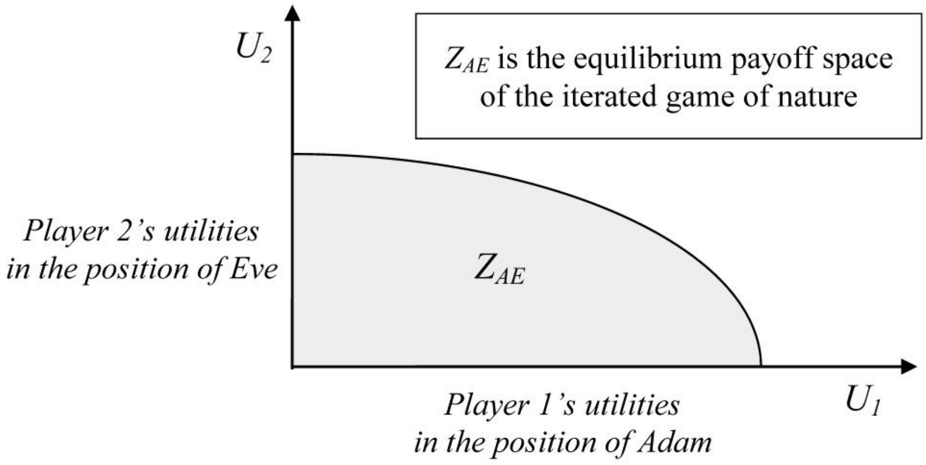
U_2
Z_{AE} is the equilibrium payoff space of the iterated game of nature
Player 2's utilities in the position of Eve
Z_{AE}
Player 1's utilities in the position of Adam
U_1

Figure 1 - The repeated game equilibrium set Z_{AE}

The *game of life* is repeated in the long run. As it is repeated, some details may occasionally change as new generations of players join. Thus, there is a chance that a player 1 is sometimes called upon to play in the position of Eve, while a player 2 is called upon to play in the position of Adam. Evolutionary games typically select players at random from given populations (viz. players from population 1 and players from population 2) to play any role in each repetition of a given. The situation is such that throughout the evolutionary history of humankind or societies, players that usually play as weak stakeholders may also sometimes (even though with small probability) occupy the role of the owner of a firm and vice versa. Consider that player 1's progeny consists of many more players taking the role of Adam with respect to Eve but, due to a mutation at some point in time, Mother Nature has selected for a while only player 1's sons to play the role of Eve. By chance, these Eves may play against player 2's heirs, who are Adams. Hence player 1 and player 2 have undergone a permutation of their roles across these game and they may retain memories of this position exchange through their evolutionary history. This is the evolutionary basis for the capacity to assume the other's perspective and develop empathetic preferences. Put in neuroscience language, player A's "mirror neurons" fire when A sees poor E getting such a modest payoff x that it as if it was player A himself who had received that same payoff x.

2.2 The game of morals

All this is simply preparatory (i.e. gives an evolutionary basis) for introduction of the social contract as an *ex ante* generally acceptable and stable equilibrium selection mechanism. Following the Rawlsian idea of a hypothetical "original position", Binmore calls the relevant choice situation "the game of morals", which re-elaborates the game of life from an impersonal, empathetic and impartial perspective (Binmore, 2005). It is a hypothetical choice situation whereby each player consider the entire set of possible equilibrium outcomes of the repeated game as if he/she were able to occupy each role (Adam or Eve) under each outcome and to receive each possible role-related payoff from each outcome. Consequently, neither of the players identifies with his/her role, and each of them (player 1 or 2) takes it for granted that there is an equal chance of occupying the positions of both A or E interchangeably. These are the typical assumptions made when the original position is seen as a choice under the "veil of ignorance". However there are distinct hypotheses that must be introduced step by step.

2.3 Impersonality and inter-**changeability of the players' posit**ions

First of all, impersonality is the capacity to consider not just one's own narrow personal point of view and to assume every possible personal perspective when assessing the outcome space – i.e. both players 1 and 2 view the decision problem from the personal perspectives of both Adam and Eve. This requirement is captured by the geometrical construction of a payoff space translation with respect to the Cartesian axes representing player 1 and 2's utilities (payoffs) respectively. Given the initial payoff space Z_{AE} , the translation generates a new payoff space Z_{EA}. For each "physical" outcome of the original game (represented by a point in Z_{AE}) this translation generates an outcome (a point in Z_{EA}) with the players 1's and 2's social and personal positions (A and E respectively) symmetrically replaced. So that player 2 (ex-E, now in the role of A') obtains exactly the outcome that was got by player 1 in the role of A "before the translation", whereas player 1 (ex-A, now in the role of E') gets exactly the outcome that were got by player 2 when s/he was in the position of E. Hence, for every equilibrium point in the original outcome set Z_{AE}, whatever the equilibrium outcome afforded to player 1 in the initial representation, the same outcome will be afforded to player 2 under the translated outcome set Z_{EA}, and vice versa (see Binmore, 2005).

2.4 Empathetic preferences and interpersonal utility comparisons

However, a point must be raised here. Player 1 and 2 are just labels for individual players, but a complete description of a player's preference can only be given when s/he takes a particular social role and personal position as Adam or Eve. In assuming the role/position of Eve, player 1 (normally Adam) tests his psychological capacity for empathetic identification with the preferences held by player 2, who usually plays in the role/position of Eve.

Consider first what is not an exercise of empathy (but autism – as Binmore suggests, see Binmore, 2005). Although player 2, now in the role A', receives the consequences of player 1 when he was A, she is incapable of evaluating them in terms of the same preference as player 1's in the role of A, and to compare these preferences and their utility measure with the preferences he had in the role of E. On the contrary, she keeps the preferences and utility measure she had when she was in the role of E. Hence the translated Z_{EA} need not be a symmetrical image of Z_{AE}.

However, this is not the proper manner to construct the original position, which is designed to enable the players to exercise their capacity for empathetic identification. What is required of player 1, while he is E', is to understand what it means for player 2 to be in the E role with her own preferences, and vice versa. Under empathetic preference, player 1 (respectively, player

2), when he (resp. she) takes the position E' (resp. A') experiences being in this position with the preference that another player had when she (he) was in position E (resp. A). They thus carry out interpersonal comparisons of utility, which means that player 1, both in the role of A or E', uses the same utility unit to represent and compare his *empathetic* preferences with his *personal* preference between the two positions (see Harsanyi 1977). The capacity for empathetic preference is a distinctive trait that makes human psychology what it is. Binmore assumes (and I follow him) that biological evolution has equipped us not only with a capacity – maybe our "mirror neurons" - for empathetic introspection and simulation but also with the competence to represent different individuals' preferences in a fairly similar manner, that is, by means of fairly similar utility units (Binmore 2005).

What we have now are two spaces X_{AE} and X_{EA}, one the *symmetrical* image of the other (see *fig. 2*).

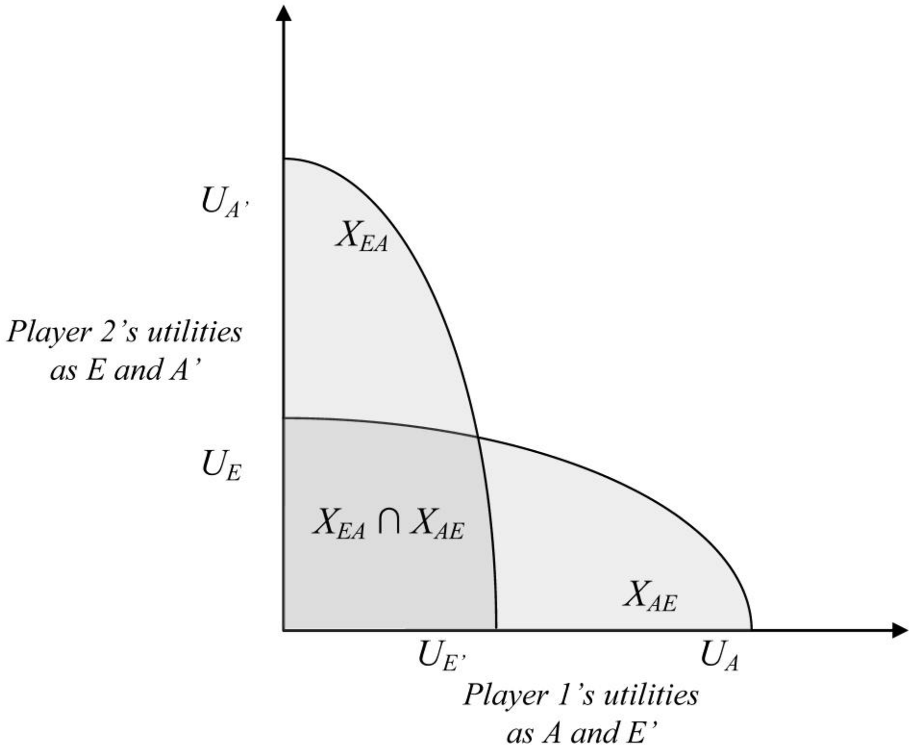
$U_{A'}$
X_{EA}
Player 2's utilities as E and A'
U_E
$X_{EA} \cap X_{AE}$
X_{AE}
$U_{E'}$
U_A
Player 1's utilities as A and E'

Figure 2 - Symmetric translation of the payoff space X_{AE} with respect to the individual utility axes, so that the utility function U_A is replaced by $U_A' = U_E$ and vice versa

Space X_{EA} results from the *symmetrical* translation of all points of the first space into (symmetrical) points of the second. Recall that in the game with payoff space X_{AE} player 1 is A (with payoff measured on the horizontal axis), and player 2 is E (with payoff measured on the vertical axis). Under the translation , player 1 (ex A) becomes E' (with utilities identical to E) and player 2 (ex E) becomes A' (with utilities identical to A). Owing to the symmetry of the translation, for each outcome x in X_{AE}, where the two players get payoff x_A, x_E, respectively for player 1 and 2, we may find within the space X_{EA} a point $x' = (x'_{E'}, x'_{A'})$ where payoffs are simply exchanged between the players 1 and 2, i.e. such that player 1 gets $x'_{E'} = x_E$, and player 2 gets $x'_{A'} = x_A$. Hence, exactly what was got by player 1 (as A) now belongs to player 2 (as A'), while the payoff got before by player 2 (as E) is now obtained by player 1 (as E').

2.5 Impartiality and solution invariance

This construction allows each player to put himself into the shoes (A or E roles) of the other player and vice versa. But now that the players are *impersonal* - i.e. they properly (empathetically) consider the decision problem from every personal point of view, but do not identify themselves with whatever personal perspective - what is required is that they give an *impartial* solution to the problem; a solution that is not biased to the advantage of either player, and does not put any personal role in a position of differential advantage with respect to others. A natural consequence for the equilibrium selection problem is that the solution must have some *invariance* under the position replacement, so that the player can continue to recognize and choose it in both positions. Impartiality thus simply implies that the solution must be invariant under this pay-off space translation, because the solution has to be accepted by each player under both the roles s/he will occupy, i.e. it cannot be contingent on a particular role-position s/he occupies. This seems to mean that each player must get from the solution the same "acceptable" payoff whatever the role (A or E) he takes, i.e. whatever the party's position he takes in the game. Thus an *impartial* solution is an equilibrium point that allows each player to achieve a payoff which is invariant, whatever the role the player happens to occupy. By contrast, a solution (given a particular representation of the game pay-off space) is said to *depend* on the *particular personal and strategic position* that players hold in the game if implementing the corresponding equilibrium yields payoffs that the players could not obtain if the same equilibrium point were implemented under the symmetric translation of the pay-off space – that is, under the symmetric replacement of the players with respect to each outcome. Translation invariance must be satisfied in order for the equilibrium point selected to be normatively considered *the solution*.

It is fairly clear that this property is satisfied if the initial payoff space X_{AE} is restricted to the bisector of the Cartesian plan, i.e. if the outcome space is constrained to satisfy the condition that any outcome is mapped onto itself by a symmetric translation of the outcome space with respect the Cartesian axes. But of course this is very far from being the general case (consider however section 5 where this case is relevant). In general a payoff space, whether symmetrical or otherwise, will contain many outcomes that under a payoff space symmetric translation will be mapped onto another point in the Cartesian plan by inverting individual payoffs in the payoff vector. In other words, invariance would require a solution to be located on the bisector, which seems at a first glance to be a very restrictive condition with respect to payoff spaces in general.

To be sure, symmetric and asymmetric payoffs spaces are not on an equal footing in this respect. A symmetrical outcome space can be simply assumed to have a symmetrical solution. When an outcome space is perfectly symmetrical, there is no reason to imagine that there are major differences between the players. Nor there is any need to impose explicit impersonality and impartiality between players who are completely equal in any respect: they will directly jump to the egalitarian solution, which is typically on the bisector where any symmetric translation of the outcome space will result in outcome invariance (this was also John Nash's intuition, see Nash (1950)).

But now assume that the equilibrium space is asymmetrical, as X_{AE} in fact is. Why not admit that, without an explicit requirement of impartiality and impersonality, unequal self-interested players would produce by their bargaining process whatever result other than a perfectly equal one? Thus, assume that any player would ex ante accept (under a given representation of the pay-off space) any equilibrium point but an egalitarian one as the solution. Under the pay-off space translation X_{EA} this equilibrium point translates into a *different* point outside the original pay-off space. Once the player positions have been exchanged, the pay-off space translation identifies a point corresponding to the same equilibrium, but this point (a pay-off vector) does not afford each player the same pay-off as before (simply because it replaces the pay-off of the previously ‘fortunate’ player with that of the previously ‘unfortunate’ one, and vice versa). Thus the solution cannot be invariant.

2.6 Veil of ignorance, and equally-probable mixtures

The invariance condition in the case of a large space with numerous asymmetric outcomes is regained by introducing another step in the construction of the "original position", i.e. by imposing (following Harsanyi and not Rawls on this point) the probabilistic interpretation of the "veil of ignorance". The veil of ignorance according to this version (see Binmore, 2005) consists of complete (probabilistic) uncertainty about player 1 and 2's roles (A or E) in the game, i.e. complete uncertainty about which of the two asymmetric spaces X_{AE} and X_{EA} will actually take place. This amounts to saying that each space has probability ½ to represent the actual outcome space of the game. If the players were required to choose a joint strategy that produces the outcome x in the outcome space X_{AE}, they would consider that this choice will achieve the outcome x only with probability ½, whereas it may also achieve by probability ½ the symmetric outcome x' where the players' positions are mutually exchanged.

The probabilistic version of the veil of ignorance implies that when a player chooses in the original position s/he must always account for the expected value of any decision. For any selection of a particular equilibrium point, this amounts to always considering the equally-probable mixture of the payoffs s/he gets under that particular outcome and its symmetric translation. We are thus back to the 45° bisector, where all the expected values of equally probable mixtures of symmetric outcomes belonging to spaces X_{AE} and X_{EA} do in fact lie.

This is what gives invariance to the solution also in the case of an initially asymmetric payoff space: when a player considers as the candidate solution an equilibrium point ***s*** in X_{AE}, s/he must also account for its translation ***s***' into X_{EA}, and in fact s/he takes as the actual candidate solution payoff the mid-point on the straight line representing the linear combination of the two outcomes ***s*** and ***s***'. What matters for this choice is the expected value of the equally-probable combination of his/her payoff for the equilibrium ***s*** in X_{AE} and his/her payoff for its symmetric translation ***s***' in X_{EA} .

2.7 Feasibility

Decision making under the veil of ignorance raises the further question as to whether equally probable combinations of symmetric outcomes are themselves *feasible* terms of agreement. The question is whether is it feasible to agree on a jointly randomized pair of strategy combinations that generates two outcomes with the same probability, in such a way that one may consider at least ex ante the expected value as the utility that one will actually receive from selecting the joint strategy combinations. This makes sense only if one is confident that, whatever outcome may be selected by the random device attached to the pair of strategy combinations (or outcomes), it will be put into practice. Put differently, whatever outcome is selected, it will be automatically enforced. The opposite hypothesis is that when the time at last arrives that the agreement must be implemented by a random choice of the actual outcome, if the selected outcome does not satisfy a player, the latter can renegotiate it. Typically, player 1, when by chance an outcome is selected in which he is E', may ask to renegotiate the outcome selected in order to have a new chance of occupying the luckiest role of A as an outcome is selected. After all, in the game of life he *de facto* plays in the role of A (see Binmore 2005) .

The question would be simply solved if the mid-point of the probabilistic mixture was an equilibrium point on its own. If in correspondence to this mid-point there is an equilibrium point formed of strategies (pure or mixed) that in practice the players may adopt in the ex post game, then that equilibrium can be selected in order to generate an impartial solution. I would say that this is not beyond any doubt, for player could maintain doubts about the obedience of other real-life players to an action dictated by the random mechanism. However, there is no incentive in this case to defect from the outcome selected by the random mechanism. The case is different if the 'mid-point' results from the convex combination (joint randomization) of two points each alternatively belonging to one of the two basic pay-off spaces, but it actually falls outside both the basic spaces and their intersection. Certainly, such mid-points of equally-probable mixtures falling outside both the space X_{AE} and X_{EA} cannot be equilibria in the "game of life".

2.8 The Deus ex machina hypothesis

Here a basic methodological decision must be made. Joint randomization is an admissible operation within the context of cooperative games, where joint strategies (plans of action) can be always randomized by an interpersonally valid random mechanism without fear that individual players will act according to separate mixed strategies in practice. But cooperative games assume that an exogenous mechanism will enforce whatever agreement on any jointly randomized outcomes: this amounts to what can be called a *Dues ex machina* hypothesis.

At the methodological level, however, the modeller must decide whether or not it is appropriate to assume – or whether or not the players actually believe in – the existence of God as an external enforcer for whichever agreement to which the players subscribe in the "original position". If God exists, then the outcome space will expand significantly because it will also include all the linear combinations of any pair of points in X_{AE} and X_{EA}, i.e. the bargaining game in the original position will become the convex hull of all the points in the union of X_{AE} and X_{EA} – which is necessarily a symmetric space of expected payoff (see *fig. 3*).

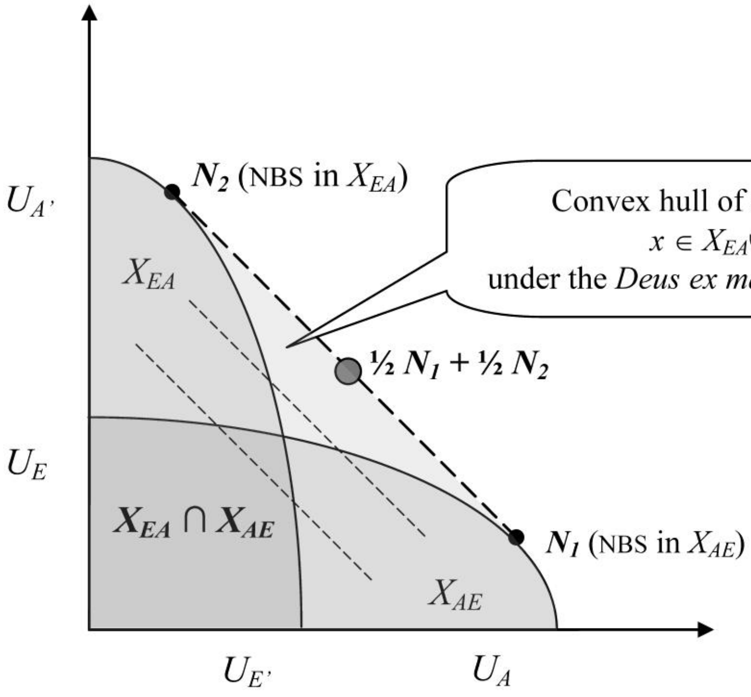

Figure 3 - Veil of ignorance and convexity

In this case there is an open choice among a wide variety of principles. For example, the utilitarian solution seems reasonable because it suggests taking as the solution the point in each space where the utility sum is maximised, and then considering their mean value. We thus do not have to concern ourselves with what the players will do when the veil of ignorance is removed and hence face the situation where one player is reduced to extreme poverty in order to maximize the utility sum.

We are looking for contractarian principles. Assume that under each representation of the payoff space players agree by rational bargaining on the relevant Nash bargaining solution. Hence, the equally probable combination of the two Nash bargaining solutions (NBS), each belonging to space X_{AE} or X_{EA} respectively, seems to be the obvious candidate. This means that player 2 will take it for granted that s/he will be afforded the payoff resulting at the mid-point along the straight line joining his/her payoffs at the two NBS, N1 and N2, each belonging to the relevant payoff space X_{AE} or X_{EA} respectively. What s/he gets in fact is his/her expected payoff at the point ½ N1+ ½ N2, a point that requires the presence of a *Deus ex machina* to be implemented because it does not belong either to X_{AE} or to X_{EA}.

Nevertheless, believing that God will always be ready to play the role of an external enforcer is not the most appropriate hypothesis for a decision in the original position. The idea of a "state of nature" would be pointless in this case. In fact it means maintaining that only agreements corresponding to equilibrium points of the underlying non-cooperative game of life can be expected to be implemented, because they are self sustaining and doesn't require any previous authority to impose them. In other words, the game considered here is non-cooperative. Thus one is *not* allowed to generate from the original outcome space and its symmetric translation the convex hull of all their components (see Binmore, 1987).

It follows that both the equally probable combinations of the Utilitarian and the Nash bargaining solutions are ruled out because they do not belong to the payoff space intersection $X_{AE} \cap X_{EA.}$

To explain, assume that a random mechanism is agreed upon, and it randomly selects the payoff distribution corresponding to N2 where player 1 is in the role of E'. Since, in the actual game of life player 1 is in fact occupying A's role, he can to decline to comply with the randomly selected solution N2 because it is not enforced by itself. Thus, in the event that the players agreed on the NBS equally probable combination under the veil of ignorance, this would simply amount to player 1 getting his Adam's payoff for N1 with probability one, because his alternative N2 payoff (Eve's payoff) cannot be enforced if it is selected. If the coin was to fall on the side that would dictate the payoff of A' to player 2, player 1 would simply refuse to comply by asserting that his actual role in the *game of life* is playing as A. Why, then, should player 2 enter the original position. It seems cheap talk without any relevance to the players' actual behavior. Summing up, there is no scope for agreeing under the veil of ignorance on outcomes that cannot be enforced.

2.9 No Deus ex machina

Contrary to the conventional wisdom, this does not require giving up either the original position or the veil of ignorance. Binmore suggests retaining symmetric payoff translations (impersonality), empathetic preferences and equally-probable mixtures (impartiality), but to skip the hypothesis that God is ready to serve as an external (*dues ex machina*) enforcer, thus adding the requirement of *self-sustainability* (Binmore 2005). This consists of restricting the selection of the acceptable solution only to within the *intersection* of the original outcome space and its symmetric translation i.e. $X_{AE} \cap X_{EA}$. Any selection within this set, in fact, does not create the feasibility problem just considered because any point in the intersection set corresponds to an equilibrium point that is always existent as long as it belongs to both the original and the translated outcome sets, viz. an equilibrium outcome that would always materialize if either X_{AE} or X_{EA} were actually the case.

Thus one way to satisfy the condition of solution invariance under the symmetric replacement of players with respect to the payoff space follows quite naturally. As before, the veil of ignorance entails considering as admissible only equally-probable mixtures of each player's pay-offs derivable from an equilibrium point and its symmetric translation. Necessarily, the solution will be a point on the 45° straight line (the bisector) connecting the origin of the intersection space $X_{AE} \cap X_{EA}$ to its north-east frontier, where all the admitted equally-probable mixtures lie (see fig. 4). Each outcome resident on the bisector is invariant under the symmetric translation of the outcome space. But each of such "mid-points" also necessarily identifies one equilibrium that the players can ex post achieve by a feasible pure or mixed strategy as long as it belongs to the intersection set $X_{AE} \cap X_{EA}$.

Moreover, consider that the space $X_{AE} \cap X_{EA}$ is also a symmetric space on its own. It is, in fact, the collection of all those pairs of symmetrical points - like x and y generated one from the other by a symmetrical payoff space translation - which are at the same time elements of both the spaces X_{AE} and $X_{EA.}$ separately. Thus $X_{AE} \cap X_{EA}$ coincides with the symmetric sub-set of each space X_{EA} and $X_{EA..}$

Given symmetry of the payoff space, bargaining theory becomes extraordinary simple. The bargaining solution must be taken on the 45° bisector deriving from the origin at the point where it intersects with the north-east boundary of the payoff space. Being on a straight line deriving from the status quo and pointing north-east simply means that the solution provides mutual gains to both the players with respect to the status quo. Being on the bisector means that mutual gains are equal. This depends on the symmetry of the payoff space. Given any agreement on which a player may insist, there is a symmetric agreement in the same outcome space, with the same payoffs exchanged between the players, on which the other party may insist as well. The reasons for insisting on each side are equally strong (under whichever definition) and would be perfectly balanced. It is then reasonable to expect rational bargaining to lead to an agreement located at the midpoint of the linear combination joining any symmetric pair of possible agreements. Lastly, that the solution is at the intersection point with the north-east boundary simply implies Pareto optimality - which means that equal mutual gains must be as high as possible.

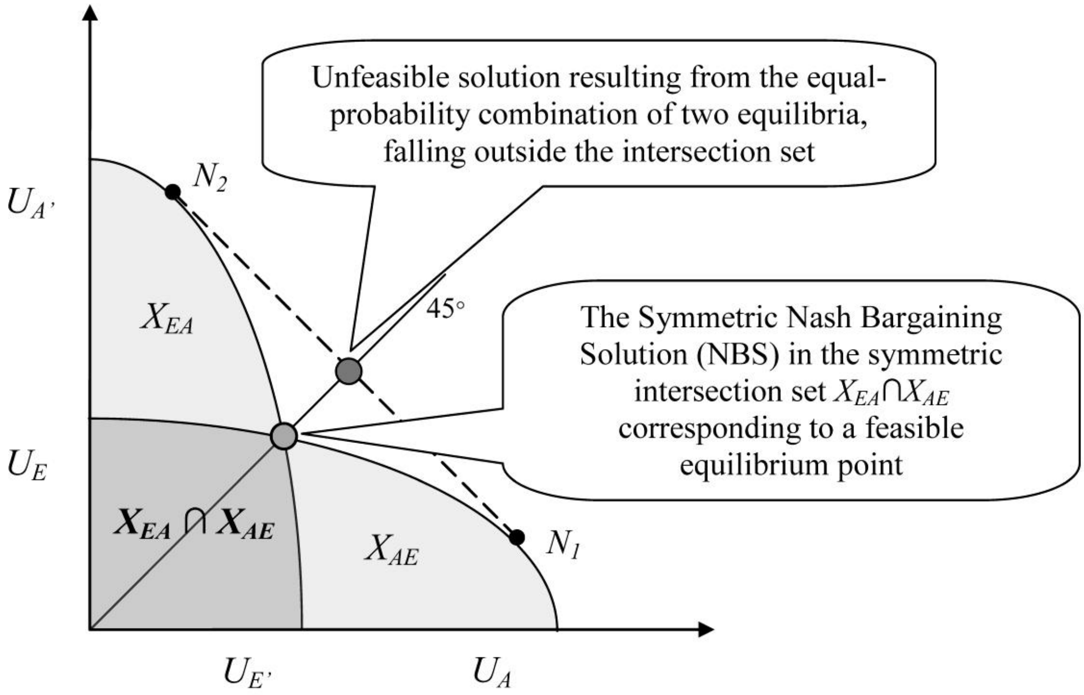
Unfeasible solution resulting from the equal-probability combination of two equilibria, falling outside the intersection set
$U_{A'}$
N_2
X_{EA}
45°
The Symmetric Nash Bargaining Solution (NBS) in the symmetric intersection set $X_{EA} \cap X_{AE}$ corresponding to a feasible equilibrium point
U_E
$\mathbf{X_{EA} \cap X_{AE}}$
X_{AE}
N_1
$U_{E'}$
U_A

Figure 4 - Egalitarian feasible solution and efficient unfeasible solution

All these qualifications seems very natural for the selection of a single equilibrium point within the intersection set $X_{AE} \cap X_{EA}$ given its symmetry. The result is the Nash Bargaining Solution (NBS) for the special case of a symmetrical payoff space, which is also the same as the *egalitarian solution*: the surplus over the status quo point is distributed to players in (feasible) maximal equal shares. Since, in our construction, we have assumed interpersonal utility comparability, this means that the players get substantially the same amount of welfare or the same level of needs satisfaction over the status quo.

2.10 General validity of the egalitarian solution

However, our starting point was not a symmetric payoff space. Hence the decision to restrict the solution to the symmetric intersection set $X_{AE} \cap X_{EA}$ must rest on some reasons direct or indirect in favour of *egalitarianism*. To appreciate this, consider that egalitarianism requires that if it is wanted to reach an agreement under the "original position", the agreed solution must be such that the players' payoffs are invariant to the symmetrical permutation of the players' positions and roles. The solution is a point in the payoff space such that the individual payoff allotted to each player must remain perfectly unchanged under the symmetric translation of the payoff space with respect to the players' utility-Cartesian axes.

This invariance condition is much stronger than the simple requirement that the *solution concept* (and its corresponding maximum value, i.e. the maximal value resulting from aggregation from whatever social welfare function) be invariant under the mutual replacement of players with respect to their roles and positions. In this second case, whereas the value of the solution function would remain unchanged (for example, the outcome where the Nash bargaining product is maximal is invariant to any independent affine utility transformation of the payoff space and hence also to its symmetrical translation from X_{AE} to X_{EA}), the payoff allocated to each player would vary according to the translation. Hence, in general, players would not preserve the same payoffs that they had before the replacement.

By contrast, the egalitarian solution amounts to saying that the anonymity of social roles does not justify any inequality of distribution. "Who gets what" cannot depend on who gets the social role of Adam or Eve, no matter that the assignment of social roles is anonymous, and both player 1 and player 2 think it equally possible to be in A or E's roles. Egalitarianism seems to rest on a more basic idea of equality among people, which is antecedent to the differences (utility function, strategy set, etc.) associated with their A or E social roles. It seems to reflect a basic feature of the original position where all these difference are weighted out. Only perfect equality is acceptable in the original position because if all the positions must be mutually interchanged, nobody is able to claim a payoff that others could not also claim. And in case the claims each player would make from any different standpoints were mutually incompatible, they should be compromised by an equally probable mixture of the two.

However, the *egalitarian distribution* does not necessarily follow directly from the *equality* of participants in the original position. The main argument in its favor is indirect. Stability, which is not an ethical assumption, is sufficient here. In fact, in order to make such agreement *credible*, it may be constrained to belong to the symmetric subset of the two equally possible spaces of claimable outcomes. Owing to the symmetry of this space the solution is necessarily egalitarian. But what requires a symmetric payoff space, which in turn implies egalitarianism, is the ex post feasibility and stability of outcomes. Hence *stability* plus *impersonality* (symmetric inter-changeability) and *impartiality* (equally probable mixtures) leads to the egalitarian solution.

2.11 Rawls vindicated also to not Kantians

By this route Binmore vindicates Rawls and his proposal of the maximin principle as a choice rule in the original position also when it is seen in the apparently alien context of a game-theoretic social contract (Binmore, 1991, 1998, 2005). In fact Eve's payoffs, those allotted to the disadvantaged player, are maximized within both the payoff spaces X_{AE} and X_{EA}. When players 1 and 2, through their position permutation, take Eve's role under the alternative label of E and E' respectively, they both have their payoffs maximized.

It should be noted, however, that *egalitarian* and *maximin solution* are based neither on a direct intuition in favour of such payoffs distributions nor on an extreme form of risk aversion (as Rawls himself seemed to think). According to Binmore, they depend on the requirement of the *ex post stability* of any agreement reached in the original position when joined with the genuine ethical requirements of symmetrical place permutation of players, veil of ignorance and the capacity for empathetic preferences (Binmore 2005).

In essence, an agreement in the original position must be taken *seriously*. Each player – the disadvantaged one in particular – is thus entitled to decline an agreement that renders the impersonality and impartiality of the solution purely illusory due to its ex post instability. Solution invariance under the exchange of the players' position with respect to the payoff space, and equally probable mixtures of symmetric outcomes, are hypotheses that any credible agreement in the original position must satisfy *effectively*, not fictitiously. But this would not be possible if the agreement fell outside the intersection set wherein all agreements can be implemented in equilibrium. Hence, the disadvantaged player has veto power over such an illusory agreement. This point resembles the one that Rawls made by stating that in the original position - due to the recognized moral arbitrariness of inequality in general - the disadvantaged party has *veto power* over all the inequalities that do not maximize his/her benefit as well. Here, alternatively, s/he has the capacity to veto every agreement that cannot be trusted as fair because its implementation will necessarily turn out to be biased in favour of the advantaged player.

3 Constitutional contract over the control structure of the firm

What does this Rawlsian social contract theory tell us about the selection of a CSR model of corporate governance and a firm control structure? In order to give an answer I need to return to the theory of constitutional contract on control structures of the firm, which was at the basis of my previous definition of the normative multi-stakeholder model of corporate governance (see Sacconi, 1991, 1997, 2000, 2006a,b, 2007a, 2008). It is a contractarian theory of an ex ante choice concerning the control structure of the firm seen as the firm's "constitution" (see also Vanberg, 1992). The model rests on the analogy between social contract theories used to justify on one hand the legal ordering by constitutional contract (Buchanan, 1975; Brock, 1979) and the mutually advantageous moral rules of a society 'by agreement' (Gauthier 1986), and on the other hand the economic theory of the efficient control structure of the firm based on the idea of contractual incompleteness (Williamson, 1975; Grossman and Hart, 1986; Hart and Moore, 1990; Hart, 1995).

3.1 A multistage decision model.

As far as the latter is concerned, this model is a multi-step decision model with timing, involving the potential members of a productive coalition S. At time t = 1 the allocation of rights is decided, and this determines the control structure exerted over the productive coalition S. At this step, however, not only are the ownership structure and the related residual rights of control allocated but also any other right and responsibility owed to non-controlling stakeholder such that they give them any level of protection against the "absolute power" of those in the position to make residual decisions (here there is a departure from the standard incomplete contract model) .

At time t = 2 the right-holding individuals (both owners and non-owners) take specific investment decisions with a view to the completion of subsequent transactions. Such investment decisions cannot be required in the ex ante contract because they cannot be ex ante described in a formal contract.

At time t = 3 events may occur which are also unforeseen by the initial contract. These events reveal the possibility of further decisions that may be essential to the value of investments already undertaken. For example, these decision are essential for implementing some technical innovation that the foregoing investment has made possible. Such decisions may physically pertain to one player or another. However, "ex ante" rights allocate control over these decisions in an indirect way. A party in the position of an authority in the firm may order those parties who do not formally control the firm but are in the physical condition to implement decisions, to execute actions chosen by the first party. In this way, an investment - when introduced at time 2 – is exploited so as to derive surplus value from it.

At time t = 4 a new bargaining game begins, defined for each allocation of rights, given whatever investment decisions were taken at time 2. Time 4 bargaining concerns decisions revealed as possible at step 3, according to control rights and responsibility. How time 4 bargaining is resolved depends on the allocation of rights at time 1. Thus, according to the firm's constitution, ex post bargaining will be in favour of one or other of the participants, in the sense that these will be able to appropriate shares of the corporate surplus depending on how may rights (ownership, control, protection, verification, accountability etc.) they have acquired at step 1.

Here the analogy with constitutional economy theory emerges: in fact, the overall collective decision problem is modelled as a compounded two-step bargaining game: an ex ante constitutional bargaining game G_C on the "constitution" and an ex post "post-constitutional" bargaining game G_I on the collective agreement concerning the surplus's distribution amongst the coalition S members. First, the constitutional bargaining game G_C is carried out (at time t = 1), when what is at stake is a "constitution": i.e. a subset of the logically possible strategies open to each player at time 1 is singled out. This set will constrain the bargaining strategy set open to each player at the post-constitutional stage. Because it is a restriction on the initial set of strategies, and defines a subset of strategies available to each player, it can be understood as a "constitution", that is, a delimitation of the natural liberties of each player that institutes the correlated set of rights and responsibilities held by all the other players. The not obvious point here is that the first agreement concerns not just a single joint strategy profile, but a set of possible joint strategies. Accordingly, the G_C game is a game that does not single out a joint strategy but an entire set (subset) of joint strategies that could constitute the possible actions and agreements allowed by the given constitution. Second, a subsequent bargaining game is played (at time t = 4) within the limits of the given constitution, and wherein the players make a choice among the available joint strategies allowed by the agreement reached at the constitutional step.

The constitutional economics aspect of the model introduces an ex ante social contract on the allotment of rights at step 1 as a bargaining game; whereas bargaining was admitted by the incomplete contract model only at step 4 (where also the constitutional economics model posits the post-constitutional bargaining) so that the ex ante decision remained quite unspecified – a somewhat mysterious collective decision based on the intent to minimize transaction coasts.

However, the analogy with the incomplete contract model explains why the constitutional contract is a two-stage decision. The social contract is incomplete: it cannot provide for whatever particular decision in detail. On the contrary, it only provides for the ex ante assignation of decision rights. In the second stage, therefore, decision rights influence the post-constitutional division of the surplus by means of post-constitutional bargaining, after investments have been undertaken and also after new decision opportunities have been revealed.

Nevertheless, as in much of the incomplete contract literature, here the simplifying assumption will be made that a resolution in terms of surplus division can be assigned to each constitution at the first stage, so to speak. Given each constitution, players can forecast the single post-constitutional solution for that constitution in terms of post-constitutional bargaining: a fact that the player can assess by looking onward from the first stage in order to decide the constitution on which s/he wants to agree. Put simply, at the first step the game is split into numerous sub-games, each defined in terms of a given subset of the basic strategy space. Then a solution is computed for each sub-game. Hence the overall range of the sub-game solutions is assessed and the different ex post solutions are compared at the constitutional stage (ex ante decision) in order to give a basis for the constitutional choice in terms of each constitution's outcome. This is a strong simplification indeed, because it should be admitted that, owing to proper contract incompleteness, the realisation of the possible available amounts of surplus (and hence the payoff value related to each concrete joint strategy) must be learnt only after specific investments have been made, and after the revelation of unforeseen events that allow surpluses to be made out of investments. These facts, because they cannot be included in the contract, would be unforeseen at the first stage, and hence would not allow the onward assessment of alternative constitutions in terms of their final payoffs distribution.

This would require modelling the constitutional contract as a choice with vague payoff variables (maybe fuzzy payoffs) – which is also consistent with our solution of the constructive/cognitive problem in part I of this essay (see Sacconi 2010a, *infra*). – i.e. the specification of the vague game form of the underling trust game played by stakeholders and firms under unforeseen contingencies. In fact, in that unforeseen events are defined as fuzzy sets, understood as application domains (sets) for principles of behaviour (corresponding to strategies) contingent on unforeseen states, the players' payoffs attached to joint strategies can be modelled in a similar way. Because these payoffs are functions of unforeseen events, they could become vague variables. For simplicity, however, I set this point aside for the moment by assuming that, even if in a vague way, players have a fairly good understanding of the payoff space of the constitutional choice game as a set of outcomes each associated (vaguely to a certain degree) with (many) possible constitutions (subset of the initial strategy space) (Kreps, 1990; Zimmermann, 1991; Sacconi, 2000, 2007).

3.2 The "state of nature" game

Having assumed that the constitutional choice is about rights and restrictions on the admissible sets of free actions and their outcomes, where do these actions and outcomes come from? The answer is (in part) from the "state of nature". Many of the possible constitutional outcomes, based on the use of some action capabilities by players, are state-of-nature outcomes virtually already possible in the case that these action were adopted. They are not *all* state-of-nature possible outcomes simply because, in the constitutional phase, we can devise many intermediate cooperation modes that we did not appreciate in the rough picture of our actions opportunity in the state of nature(for example, the opportunity to randomize between two possible agreements). Nevertheless, most of these outcomes and strategy profiles were already possible in the state of nature.

Thus before the constitutional game is played, we must consider the state-of nature-game G_N. This is a generic game with a finite number of players (at least two) and any finite number of pure strategies, which is a generalised form of PD or social dilemma. In this game, players have any degree of liberty allowing them to cooperate or act favourably toward each other, or to defect from any degree of cooperation, cheating and using offensive or defensive action one against the other. The salient aspect of this game is that players (without any constraint or obstruction, external or internal, physical, legal or motivational) are able to resort to any level of "natural" liberty. At the same time, the only equilibrium point in this game played as a *one shot-game* is a combination of pure strategies d* that represents an extremely poor and mutually unprofitable state of interaction in which they do not restrain in any significant way activities aimed at appropriating other natural endowments. Not only are they unable to cooperate, but the logic of choice induces them to adopt actions able to steal any benefits from the counterparty if s/he is ready to act kindly toward them. As a matter of fact, this is a

Hobbesian “state of nature”, with an unique equilibrium solution wherein the conduct of players’ reciprocal business relations render their lives “solitary, poor, nasty, brutish, and short". It has to be understood as a market interaction characterized by any sort of contract failure and incompleteness leading to very high transaction costs which makes almost impossible to attain in equilibrium mutually advantageous exchanges.

The outcome space P_N of the state-of-nature game G_N is shown in *fig. 5*. This includes a large number of discrete outcomes because it represents many possible levels of mutual or unilateral cooperation and defection, friendly or aggressive attitudes in the conduct of many business activities by the two players. What matters in this representation is that the unique equilibrium point is interior to the payoff space, which is pushed toward the origin (in order to avoid the extreme but not completely unreasonable possibility that they may also get negative payoffs in the one-shot version of this game) but (as in Hobbes’ state of nature) is equally bad for everyone. Formally, the unique equilibrium d* is Pareto-suboptimal.

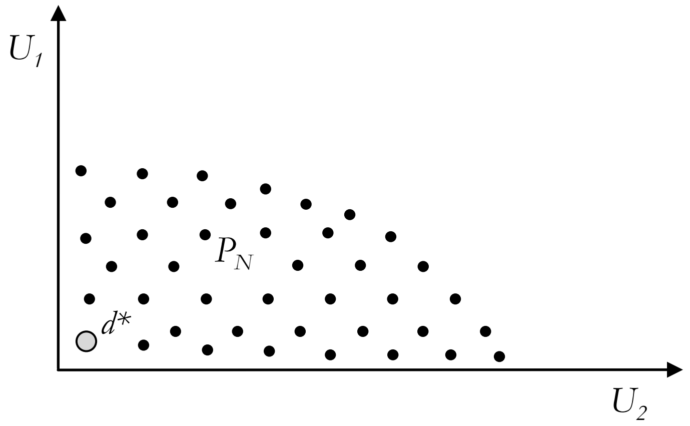

Figure .1 - The “state of nature” game

3.3 The “all possible constitutions” game

Let us move from this payoff space to the constitutional choice-game G_C payoff space. Firstly, the G_C outcome space P1 consists of the symmetrical ‘state of nature’ equilibrium d*, taken as the *status quo* where the game would remain if the players were incapable of reaching any agreement, plus the other 'state of nature' possible outcomes and all their (convex) combinations as outcomes of possible enforceable agreement. This means that agreements on constitutions can generate all the outcomes that were previously only "virtually" possible, and also all their convex combinations that were not allowed in the state of nature. In fact, the state of nature is a *non-cooperative* game, whilst the G_C is a *cooperative* bargaining game. Given any pairs of pure joint strategies (each corresponding to a profile of individual pure strategies), a cooperative game admits joint randomisations on such pairs that generate jointly randomized joint strategies or (to put it differently) mixed joint strategies as additional possible agreements of the bargaining game. Such jointly mixed joint strategies are effective in this game because any joint strategy (pure or mixed) can be enforced. That is, given agreements on two pure strategy combinations, a randomizing mechanism may dictate which of the two will be implemented without fear of individual defection from the selected combination. This defines the outcome space of G_C as, at least, the convex hull of the state-of-nature game outcomes.

A legitimate question is how the cooperative game G_C could ever emerge from the non-cooperative G_N. The answer is that G_C is a "thought experiment" that players may conduct at any time when, in order to devise a *justifiable* escape to the sub-optimality of G_N solution d*, they are willing to suppose that a solution can be given by agreement - i.e. by admitting that they are able to subscribe to whatever agreement without the fear that any player (him/herself included) may fail to comply with it. Hence, in moving forward form the state of nature game G_N to the constitutional choice game G_C it is not necessary to assume that the underlying real world situation is substantially changed. Simply, we assume that players may *frame* it as *different* games. Firstly, as a non-cooperative game G_N. Secondly, as a cooperative bargaining game G_C generated form the same physical action set and possible outcome set as G_N but with a major framing difference: the assumption that "whatever agreement is reached by players can be automatically enforced". This can be understood as taking a different perspective or point of view on the game, starting from the question "what constitution would we *fairly* agree granted that our agreements were enforceable?", which entails a completely different but internally consistent frame of the game with respect to the case of G_N.

However, this different framing of the situation allows to enlarge the outcomes space even further. Because the players are considering "all the possible" cooperative agreements, their imaginations must not be limited by their real-life power relations. They can decide to subscribe to whatever terms of agreement. This introduces a second step in the definition of the outcome space of the constitutional choice game – i.e. assuming that the G_C game outcome space is in general symmetrical and convex for whatever configuration of the outcome space of the basic state-of-nature game G_N. As far as symmetry is concerned, we proceed as follows. Players considering all the logically possible agreement, given a basic state-of-nature outcomes set, can account not just for all the probabilistic mixtures of possible agreements P but also for those resulting from a symmetric translation P1’ of the outcomes space with respect to the Cartesian utility axes, i.e. from the idea that they can also agree to exchange each other’s positions with respect to any possible agreement directly accounted for by outcomes of the basic game. Recall that G_C derives from G_N as a “thought experiment” intended to devise a *justifiable* agreement enabling the players to escape from the suboptimal equilibrium d* of G_N The need for justification (or impartial justification) is what entails that the G_C outcome space accounts for not just the convex combinations of the basic game possible outcomes, but also for the symmetric translation of these outcomes with respect the Cartesian axes representing the players’ utility payoffs. Once all these possibilities have been taken in account, also all the linear combinations among all the resulting symmetrical points are allowed, so that the space is also convex as in standard cooperative bargaining game theory. What results is a convex symmetrical outcome space P resulting from the more basic outcome space P_1 (see *fig. 6*). Note that because the status quo d* was already on the bisector, it remains unvaried under the payoff space translation (otherwise we would have taken as the relevant *status quo* the convex combination of the original one and its symmetrical translation).

Note that because the status quo d* was already on the bisector, it remains unvaried under the payoff space translation (otherwise we would have taken as the relevant *status quo* the convex combination of the original one and its symmetrical translation). As we already know, the distinctive feature of the constitutional choice game is that it seeks a solution understood as an *optimal* (in a sense to be clarified) *subset* of the possible agreements in G_C. Players simply choose a subset I_i of the joint strategies set I admissible in G_C. Each subset of the G_C strategies space is a limitation on the players’ choice freedom. Thus, the choice of any subset coincides with the choice of a ‘constitution’. Each subset (constitution) in turn defines a cooperative sub-game G_i whose outcome space P_i is a subset of the outcome space P of G_C . These sub-games may be understood as post-constitutional coalition games in which the players negotiate on how much they obtain from cooperation according their “constitutional rights”.

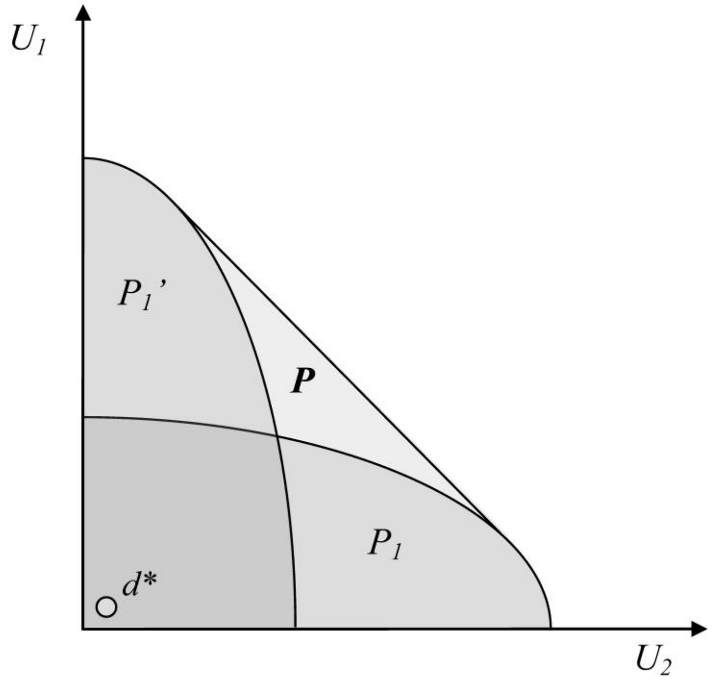

Figure 6 - The G_C payoff space

As we already know, the distinctive feature of the constitutional choice game is that it seeks a solution understood as an *optimal* (in a sense to be clarified) *subset* of the possible agreements in G_C. Players simply choose a subset I_i of the joint strategies set I admissible in G_C. Each subset of the G_C strategies space is a limitation on the players' choice freedom. Thus, the choice of any subset coincides with the choice of a 'constitution'. Each subset (constitution) in turn defines a cooperative sub-game G_i whose outcome space P_i is a subset of the outcome space P of G_C . These sub-games may be understood as post-constitutional coalition games in which the players negotiate on how much they obtain from cooperation according their "constitutional rights". Hence, each post-constitutional sub game G_i is constrained by the constitution (its set of possible strategies) chosen in G_C. Formally, the outcome space P of the constitutional choice game G_C is the union of all its possible subsets P_1…..P_n (see *fig. 7* for a case where seven payoff subspaces of P are represented), and the decision problem in G_C concerns the selection of the "best" subset of P (Nash, 1950).

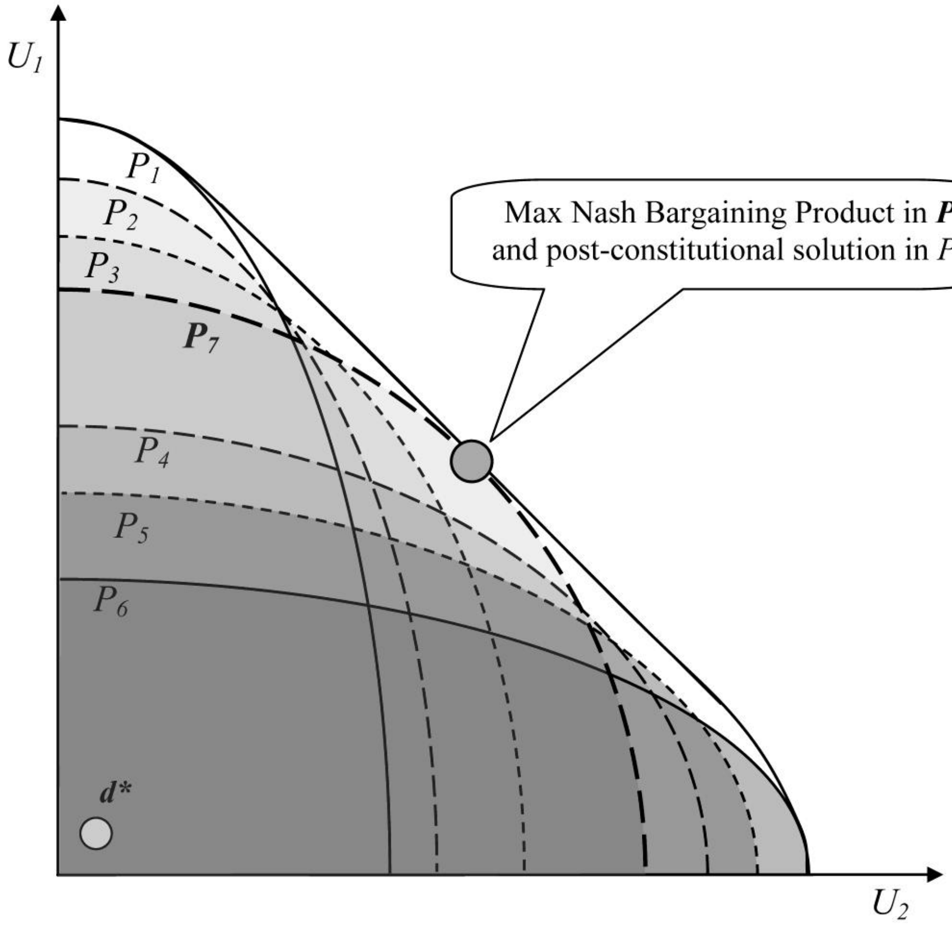

Figure 7 - Possible pay-off spaces of post constitutional sub-games

3.4 A backward-induction solution of constitutional choice as a sequential game

How must the best constitutions be identified? Recall that even if the constitution is selected as a set of joint strategies, nevertheless, for each sub game constrained by a specific strategy set, we assumed that from the constitutional point of view players may learn the unique bargaining solutions of the post-constitutional games. They thus use this information to select the best constitution. Every outcome sub-set reduces to the unique outcome coinciding with the sub-game solution relative to that particular sub-set, and these solutions are compared in terms of the relevant constitutional property.

As a whole, this amounts to saying that players take part in a sequential game in two steps so that the constitutional contract can be worked out by backward induction. Given the complete description of all the possible sub-games, players start to solve the game from its second step, i.e. by solving each post-constitutional game G_i defined for each possible constitution (each possible subset of the outcome space). Given each sub-game hypothetically, the players calculate the payoff assigned to them by the *Shapley value*, which is the relevant solution concept for n person cooperative coalition games

$$V_i = \sum_{S} [(s-1)!(n-s)!/ n!] [v(S) – v(S-\{i\})]$$

(note that in two-player bargaining in which the coalition structure reduces to the “solo-coalitions” and the “total-coalition” of two players , this reduces to the Nash bargaining game taking the “solo-coalition” as the status quo d*). For each sub game G_i there is thus a well-defined solution σ_i of the coalition problem such that $\sigma_i \geq d^*$. Then, moving backwards, the players solve the first-stage constitutional choice game. Because the G_C solution is a social contract, it must be the unanimous choice of a unique constitution by all the members of S. If this agreement is not reached, players are doomed to play the unprofitable ‘state of nature’ game with solution d*. Since G_C is a typical cooperative bargaining game, the most accredited solution is the Nash bargaining solution (N.B.S), which follows from different sets of very general rationality postulates (Nash, 1950; Harsanyi–Zeuthen, 1977)

$$\underset{\sigma i \in \mathsf{I}}{\text{Max}}\ \Pi_i(U_i – d^*_i)$$

In G_C the N.B.S must be found within the symmetrical outcome space P generated by the power-set I of all the *logically possible* subsets of the strategies set I of G_C itself. All the points in this space are understood as being solutions for possible post-constitutional games. What is remarkable is that this payoff space P is the same as the payoff space P assigned to G_C when seen as a bargaining game directly played on possible agreement concerning specific joint strategies included in the set I. The N.B.S hence selects a constitution such that the relevant post-constitutional game will distribute equal parts of the cooperative surplus calculated with respect to the entire G_C outcome space P (= P). In other words, the constitution chosen in G_C will have a post-constitutional solution coinciding with the maximization of the Nash bargaining product also relative to P. In our example (where for simplicity we exemplify only seven subsets of P), the selected constitution is identified by the space P7, wherein the Nash bargaining solution coincides with the N.B.S valid for the “all encompassing” space P.

3.5 Distributive justice interpretation

The sequential bargaining game solution can be given an intuitive ethical interpretation not only because of the symmetrical shape of the bargaining game, but also on the basis of the correspondence between each of the two concepts of solution that I have employed and the intuitive principle of justice appropriate to the respective bargaining phase in question. The solution to each post-constitutional game according to the *Shapley value* can be interpreted as an application of the principle of *remuneration on the basis of relative contribution*. The Shapley value is in fact the linear combination (weighted with equal probability assigned to all the coalitions with the same number of members) of the marginal contributions that an individual can make to all the coalitions. On the other hand, the Nash bargaining solution – provided that the units of measure for the individual utilities are assumed to be interpersonally calibrated (which is not required for simple calculation of the Nash bargaining solution) – can be interpreted as an equivalent solution to the distribution proportional to relative needs, that is, proportional to the relative intensity of marginal utility variations comparison for the players at the point where the solution falls. In fact the ratio in which the shares of the surplus are distributed under the Nash bargaining solution is proportional to the ratio between the marginal variations in the players' utilities $\partial U1/\partial U2 = -a1/a2$. Thus, once the utility units have been interpersonally calibrated, so that each unit expresses the same magnitude of preference for both the players, the ratio between their marginal variation measures the players' relative needs at the solution point (see Brock, 1979; Sacconi 1991, 2000, 2006b).

The twofold distributive justice characterisation of the bargaining solutions matches the different nature of the problems of collective choice modelled by the post-constitutional games, on the one hand, and the constitutional choice game G_C on the other. Before the parties play a post-constitutional sub-game, they undertake their specific investments bearing in mind the guarantees offered by the constitution in regard to their possibilities of reaping the benefits of cooperation. They then calculate the effect of their participation in each possible sub-coalition of S, and finally contract with S the part due to them for concluding an agreement which will enable S to pursue its best joint strategy, associated with which is a super-additive production function (or characteristic function). The solution of each sub-game distributes benefits to which the players have already contributed through their investment decisions and through their decision to join the coalition S. Therefore appropriate at this point is the distribution criterion based on *relative contribution* or, put otherwise, *relative merits*. Instead, in the case of the constitutional bargaining game G_C, none of the parties subscribing to the agreement has yet contributed anything, so that the merit or relative contribution criterion does not seem to be a valid criterion of distributive justice in this case. Chosen in G_C is the constitution on the basis of which the investment decisions will be taken. What the various players will be willing to contribute depends on which constitution is chosen. These rights-for-incentives, however, must be incorporated into an agreement among participants in the constitutional bargaining phase which considers only what is relevant from their current point of view. In the absence of any relevance of merit, in this case only *needs* can matter for the players' agreement. Hence an appropriate criterion for the solution will refer to the *relative needs* of the parties in regard to what will subsequently enable them to contribute to joint production.

3.6 Dealing with exclusive property rights

Thus far *every logically* possible constitution for the productive organisation has been considered as equally feasible. This case can be called *Utopian*, because any constitutional design can be devised out by the players' imaginations, without any constraint in terms of "institutional feasibility". This amounts to saying that, for example, property rights may be allotted amongst players as if they were a continuous variable based on some qualitative object or property (i.e. control over a good or an action) indefinitely divisible amongst them, so that rights can be distributed in whatever proportion among all the players. Non-separable discrete objects are completely excluded in this case.

However, more realistic is the hypothesis that only certain kinds of restrictions (constitutions) on the set of all the possible joint strategies of G_C are "institutionally feasible". Specifically, only "exclusive" allocations of property rights on all the physical assets of the firm may be institutionally feasible. For example, control structures could allow the assignation of authority (residual decision rights) to some party or another, but not any intermediate or equal degrees of authority to all parties - understood as whatever splitting of the same decision right on the very same asset. (Note, however, that this does not imply that other rights combinations are impossible, for example ones complementing a residual decision right held by a party with a responsibility or an accountability duty owed to those who do not hold that right). If these indivisibilities are admitted, the N.B.S. relative to the all-inclusive payoff space of G_C may not coincide with the solution of any of the institutionally feasible sub-games, since the choice must fall within the set of *institutionally feasible* solutions, which will not coincide with the entire payoff space P.

A reasonable interpretation is that "realism" constrains desirable normative properties such as ideal social efficiency and fairness. (In fact it is a standard assumption in transaction cost economics that governance and authority costs entail that any whatever governance structure is second-best. Moreover, we know that this occurs because of abuse of authority and unfairness under each exercise of ownership as an exclusive right). Thus feasible sub-games are assumed to have outcome spaces that coincide with only a few of the proper subsets of the all-encompassing outcome space P. The resulting candidate set of constitutions (deriving from the post-constitutional solutions of feasible sub games) is defined as a set of second-best solutions with respect to the outcome space P.

Consider a two-players case (see *fig. 8*). There is one feasible constitution G1 (which assigns ownership to player 1) with payoff space P1, whose solution is more efficient than that of the alternative feasible constitution G2 with payoff space P2 (which assigns ownership to player 2). Since these constitutions give complete control to one player or another, but not to both, it is natural that such constitutions should also assign a significant advantage to owners in terms of the surplus shares that they may appropriate. Assume that there are not other institutionally feasible constitutions of the control structure in terms of property rights allotment. Both the feasible constitutions have second-best solutions with respect to the all-encompassing space P. Efficiency is here understood as proximity to the Pareto frontier, i.e. how large the aggregate surplus is under the two ownership allocations. In ex post efficiency terms, ownership should be given to player 1 (which entails availability of a Kaldor-Hicks side payment that would allow a shift from one solution to the other but not vice versa). However, under the G1 game we may predict a significant level of abuse of authority by player 1 as s/he appropriates an unjustly large share of the surplus. Why should player 2 agree to such a control structure?

The only way to legitimize such an inequality into the distribution of property rights by ex ante agreement is for player 1 to render it acceptable from the ex ante perspective also to player 1, who will be disadvantaged under such a control structure. Player 1 must then take account of player 2's claims and compensate him/her for the prospective abuse of authority and injustice that s/he will suffer under player 1's control. The agreed control structure must then provide for player 1 a constitutional commitment to implement a utility side payment drawn from the surplus that s/he will appropriate under his/her control of the firm's assets and transferred to player 2: the utility side transfer will continue until player 2's fair claim of redress has been satisfied so that the most efficient control form is accepted by unanimous agreement.

But what is the fair and efficient amount of the side utility transfer form 1 to 2?

The problem is that at first glance we do not have a Pareto convex frontier along which the players can move until they reach a mutually acceptable bargain. But we can provide it by construction as follows. There are two payoff spaces, each relative to an institutionally feasible constitution (set of strategies).

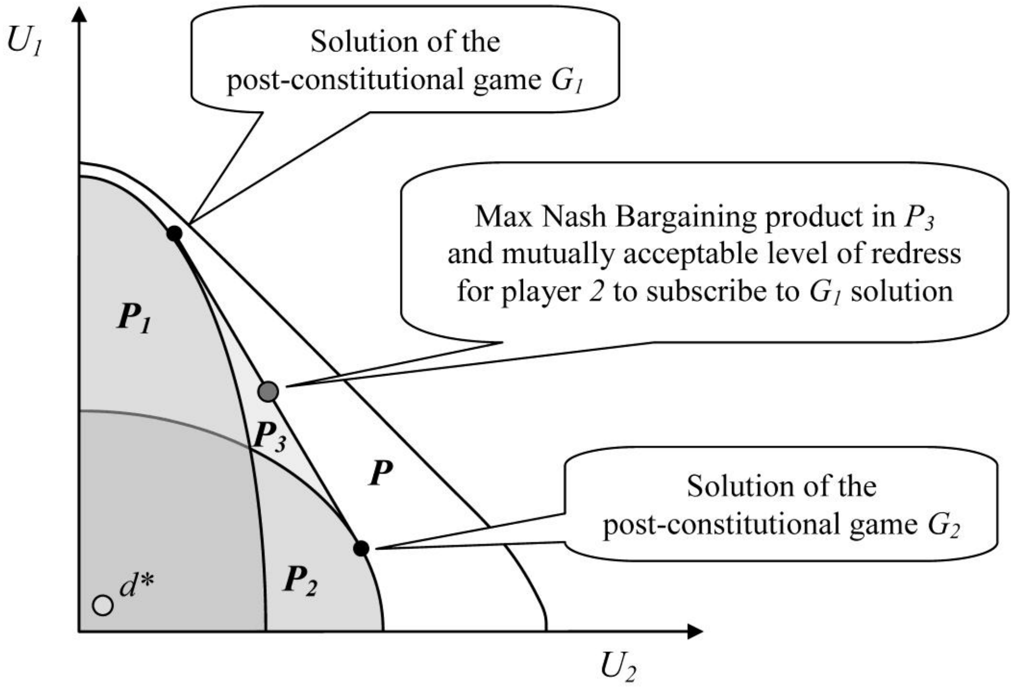
U_1
Solution of the post-constitutional game G_1
Max Nash Bargaining product in P_3 and mutually acceptable level of redress for player 2 to subscribe to G_1 solution
P_1
P_3
P
Solution of the post-constitutional game G_2
d^*
P_2
U_2

Figure 8 - Constitutions pay-off spaces under feasibility

The constrained constitutional imaginations of the players can be simply used to allow any convex combination of each pair of possible agreements, where one agreement in each pair belongs to a different feasible strategy set respectively. In particular, we focus on all the convex combinations of the two post-constitutional sub-game solutions and interpret such convex combinations as random mechanisms implementing each of the two solutions with given probabilities. The set of all these convex combinations defines the relevant north-east frontier of the payoff space P3 worked out by taking the convex hull of outcomes belonging to spaces P1 and P2 associated with the feasible constitutions. The rational utility side transfer is identified by the point where NBS is maximised along the north-east frontier of the outcome space P3. In order to allow the acceptance of the solution reachable under sub-game G1, player 1 must then ex ante commit him/herself to transferring to player 2 an ex post side-payment such that the surplus shares will be equal to those that will maximise the NBS calculated with respect to P3 (see again *fig. 8*), which is the same as allowing an appropriate random mechanism to make the choice between the two relevant sub-game solutions.

Thus, even in the context of this reduced set of feasible constitutions, we can identify a unique solution for the firm's constitution: the most efficient control structure plus the mutually acceptable (from the constitutional perspective) level of redress for the disadvantaged party.

3.7 Institutional feasibility

Institutional feasibility, as I have implicitly understood it in the previous subsection, is a twofold condition:

a) Institutional feasibility means "a consistent manner to introduce constraints on the complete players' natural capabilities to act" (held by some or all of them), and thus to assign different players' rights and responsibilities. Here "consistent" must be understood not in a pure mathematical sense but in terms of compatibility with our best knowledge about norms, institutions and legal orders as matter of facts and values.

For example assigning ownership - residual right of control - to all the interested stakeholders in the same measure, or giving each of them the same right, could be *inconsistent* with *facts* about the non-divisibility of assets or rights over some assets. By contrast, allotting control rights so that one stakeholder is given the right to take residual decisions, while another stakeholder is given protection against some extreme form of that decision, could be "consistent". So that the latter is given the following rights: (i) to ask the first stakeholder to account for his decision and (ii) to be redressed under certain conditions. The "impossibility of social choice" (Arrow, 1951) is an example of inconsistency-related to certain mechanisms of collective choice that presupposes certain decision rights of the society's members plus ethical and structural assumptions concerning the mechanism that represents some facts and values about social choice. More in general, institutional consistency requires us to have discovered an institutional arrangement consistently describable in our normative language and which can prescribe the allocation of decision rights and responsibility among the players that does not clash with our best knowledge of the subject matter. One might say that the highly fine-tuned and continuous allotments of decision rights entailed formally by taking as the basis for the constitutional choice all the logically possible subsets of the payoff space P is not institutionally feasible because we still have not designed in practice a plausible legal order able to allot legal rights in this continuous and fine-tuned mode. Thus, whereas in the mathematical model we may think of infinite subsets of the outcome space P, and we can think of moving from one subset to another by a continuous marginal change in the distribution of rights, on the contrary, within the language of institutions, we may only face a description of discrete objects permitting only rough divisions into discrete "pieces of rights" held on such objects. Some rights can be indivisible and not sharable, whereas they can be counteracted by different rights, also indivisible but consistently able to curtail the first right abuse. Even if this second institutional structure may be consistent, there is no reason to say that it does not entail a loss in terms of ideal efficiency and fairness. Indeed, the perfect divisibility of property rights would give a perfect modulation of investment incentives to all the players in proportion to the importance of these investments for social surplus production, whereas the feasible arrangement may be less fine-tuned to this purpose. Moreover, it is fairly obvious that institutional feasibility, by requiring the assignation of authority to one party and submission to the authority of another party, has unequal payoff distributions.

b) Institutional feasibility entails a sufficient level of effectiveness, i.e. a control and governance structure which can be intended as a protection of some rights or interests is feasible only if it can be put into practice effectively.

This condition has various interpretations. The most obvious one is to equate effectiveness with self-interested incentive compatibility in the pure game-theoretical sense. Thus the agreed solution should be required to correspond to a pre-existing equilibrium point in the underlying game (the state of nature) which implements the agreement. However, in our case - where the state of nature is seen as a one-shot game - this interpretation cannot work, because only the status quo d* corresponds to a pre-existing equilibrium point of the "state of nature" game. A possible way to introduce this type of effectiveness would be to assume that G_N is an infinitely repeated game, so that each one-shot game outcome may be reached in equilibrium as the average payoff of an appropriate combination of repeated strategies.

Nevertheless, the use of this strict notion of incentive compatibility is not necessary in order to account for institutional effectiveness. As an alternative explanation, consider only those constitutions which define allotments of decision rights such that a bargaining sub-game within these agreed constitutional constraints is supported by motivations sufficiently strong to induce players to stay within the limits of that agreement. In other words, effectiveness comes about if the constitution distributes rights and action opportunities in such a way that players in the corresponding sub-game will reach agreements that are effective causal factors in inducing intrinsic motivations to implement that same agreement. The difference, of course, is in the role that constitutions as such may play in generating incentives and motivations that are effective in the implementation phase There is no need to make a choice between these two interpretations at this stage (however this line of thoug will be undertaken in Part II of this essay, see Sacconi 2010c).

Thus far we can maintain that effectiveness is a constraint on the "all possible constitutions" set P, so that only proper subsets are feasible (which entails that the effective constitutions outcome spaces are proper subsets of the all-encompassing space P, and because these subsets will not include the north-east boundary of space P, in general they are quite obviously *second best* in terms of efficiency). However, it is not obvious what this means in term of fairness.

4 Difficulties in the constitutional contract of the firm

Constitutions are not simply logically possible but also institutionally feasible if their design is "consistent", and some mechanism (able to carry out their constitutional agreement) exists. The mechanism may be of any nature, internal or external, legal, social, moral or psychological. Simply, there must be positive inducements or negative sanctions (internal or external, material or psychological) able to induce individuals to comply with the agreement, which may operate through the legal system, the social acceptance mechanism, or through internal motivations like moral sentiments, the sense of moral obligation, or the belief that God will condemn us to hell.

That assumption was implicitly made when the idea of an ex ante grand social contract on the constitution of the firms was introduced, and which was admitted to be about all the logically possible institutional arrangements of the control structure and other legal rights. Then, by dealing with exclusive property rights alone, I have simply constrained this hypothesis to hold only for a subset of the logically possible institutions, i.e. for the subsets in which property rights are exclusively assigned to one or another stakeholder. This intentionally makes the problem of designing a multi-stakeholder control structure of the firm more realistic and serious, because we cannot now rely on an all-encompassing institutional structure in which every stakeholder is granted an equal proportion of control rights. Hence we need to define the redress duties or responsibilities owed to those stakeholders that cannot share rights of control.

In the context of the theory of the firm, this line of reasoning could be pursued without too many difficulties, because some parts of the institutional system can be presumed to be already enacted before the social contract of the firm occurs. Hence it is admissible that at least some institutional arrangements that are deliberate through the social contract of the firm may also be externally enforced by some other mechanism (social or legal) which pre-exists the firm itself. Nevertheless, I do not want to rely too much on these presumptions, because the basic thesis of this essay is that the CSR model of corporate governance is self-enforceable, and hence can rest primarily on endogenous forces.

The question must then be asked of "how self-sustaining is a solution that, given two feasible arrangements of property rights, defines a side-payment from the owner to the non-owner in order to redress the abuse of authority that will take place under each feasible institutional structure of control?" Recall that the exact dimension of this side-payment was identified through the construction of a small-scale constitutional choice problem, i.e. the convex combination of the two sets of outcomes admitted by the outcome space of the two institutionally feasible sub-games, and by the straight line joining their NBS. In other words, this implies resolving the problem of collective choice within the linear combination of the two bargaining solutions, one for each sub-game.

But we must now address a problem: this linear combination does not necessarily satisfy the same assumptions that we made for the two institutionally feasible sub-games. Hence its agreed solution on the north-east frontier of the convex combination of their payoffs spaces does not need to be feasible. How can we deal with this difficulty? And must a proper escape from the feasibility problem compromise the request for fairness and accordance with intuitive principles of justice in the constitutional choice on control structures? Of course, any successful attempt to solve this difficulty will contribute essentially to the very basis of the idea that CSR is a governance system not externally imposed by the law but implementable as a self-enforceable social norm incorporating the normative requirements of contractarian ethics. To be sure of the relevance of these questions, let us look at the institutionally feasible solution more carefully, with the aid of some geometry (see *Fig. 9*).

Fig. 9 shows a line segment joining points S_1 and S_2 and that represents the linear combination of the two bargaining solutions relative to subspaces P_1 and P_2 respectively. Along this line segment, there are all the possible probabilistic combinations of S_1 and S_2. Also represented are all the possible utility side-payments which, given solution S_2 - the more efficient one and nearest to the north-east frontier - may be agreed to redress player 1's loss for agreeing to give up control over the firm. The utility transfer in L is calculated as the constitutional agreement within P_3, i.e. a subset of the all-encompassing payoff space P, which is constructed as the convex hull of the sub-game spaces P_1 and P_2 representing institutionally feasible sub-games. The status quo is assumed to be at the origin. Hence, L is the NBS of P_3, and thus is also proportional to relative needs contingent to this subspace P_3. This last property may be seen by considering that the slope of the line segment joining S_1 and S_2 is the same, with inverse sign, as the dashed line joining the origin (status quo) and L, where it is incident on S_1S_2, which in fact is the frontier of the convex (compact) space P_3.

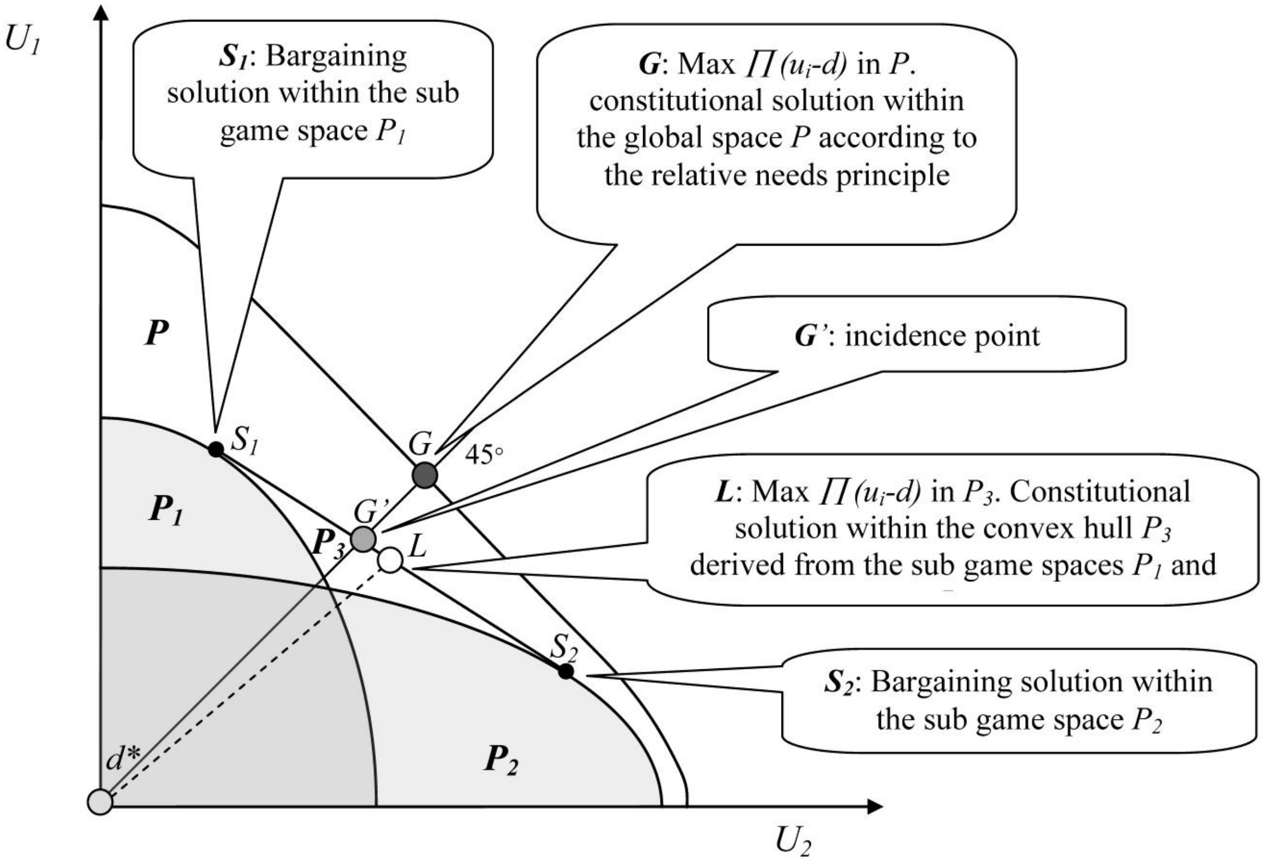
U_1
S_1: Bargaining solution within the sub game space P_1
G: Max $\prod(u_i\text{-}d)$ in P. constitutional solution within the global space P according to the relative needs principle
P
G': incidence point
S_1
G
45°
L: Max $\prod(u_i\text{-}d)$ in P_3. Constitutional solution within the convex hull P_3 derived from the sub game spaces P_1 and
P_1
G'
P_3
L
S_2
S_2: Bargaining solution within the sub game space P_2
d^*
P_2
U_2

Figure 9 - Alternative bargaining solutions, feasible and unfeasible

Two points are raised by this case:

i) *Instability of the equitable institutional arrangement*:

The institutional mechanism granting that player 1 will agree ex ante to enter a control structure that legitimizes player 2's control, and also allowing him to profit considerably from control, is the utility-side payment represented by L on S_1S_2. But whereas P_1 and P_2 are assumed to be institutionally feasible sub-game payoff spaces, i.e. to have bargaining solutions that are enforced by some mechanism or motivation, the same does not hold for any points in P_3 lying outside the union of P_1 and P_2. Combining points like S_1 and S_2 does not ensure that the resulting linear combination lies inside the institutionally feasible set of solutions. The linear combination may give rise to outcomes that are not enforceable; and this is exactly the case when, as for L, the point representing the optimal redress lies outside the P_1 and P_2 union. What will make point L feasible? Notice that L is an ex ante social contract on the institutional structure of the firm which would induce the players to give their ex ante consent to entering the institutional arrangement of the firm. Nevertheless, it does not necessarily coincide with any solution of the ex post implementation problem, and is therefore unstable. On anticipating such instability, player 2 would not effectively endorse such an agreement. But then on what should they reach an agreement?

ii) Divorce between local and global justice.

Global justice is represented by point G in *fig. 9*, where the NBS relative to space P is located. Here the institutional structure is arranged so that it reflects a measure of relative needs with respect to the all-encompassing space of possible institutions P such that it is uniquely reflected by the NBS's distribution of payoffs. This space is properly understood to be symmetrical in so far as any logically possible allocation and distribution of control rights is taken into consideration. In fact, the dashed line segment from the origin to G has the same slope (with inverse sign) as the tangent to the north-east boundary of P at the incidence point G. Because point G lies outside any institutionally feasible sub-game payoff space such as P_1 and P_2, we recognize that this solution is merely utopian. Nevertheless, the line segment joining the status quo to G represents the distributive proportion that would incorporate the relative needs principle with respect to the "global" payoff space P. The point G' at which this line segment crosses the north-east boundary of P_3 (incidence point) is hence a natural candidate for the agreement according to the constitutional choice principles, the one that mostly approximated the global justice solution (call it *constrained global justice*). Here payoffs are allotted so that the relative needs principle is satisfied not so much with respect the contingent subspace P_3 as with respect to the set of possible institutional alternative P in general. This would be a natural requirement derived from the general theory of constitutional choice: select the sub-game with a payoff space such that its bargaining solution is the one closest to the point where NBS is maximized on the all-encompassing payoff space P. In other words select a sub-game such that its own bargaining solution lies on the line segment joining the status quo to G, as near as possible to G (that is as mutually advantageous as possible). If there are no such sub-games, take as an acceptable level of redress to the disadvantaged party the point within the convex combination of feasible subspaces that lies on the line segment joining the status quo to the global justice point G. By contrast, L is a *local justice* solution: it allots payoffs in such a way that the relative needs principle is respected only with reference to the contingent subset of institutionally feasible sub-games.

Which of the two should prevail? Intuition helps only when we consider extreme cases. Let us therefore concentrate on the case where local justice diverges from global justice owing to the asymmetrical shape of all the institutional feasible subsets and hence also to their convex combination. *Fig. 10* illustrates this case: P is symmetric, but both its institutionally feasible subsets are rather asymmetrically placed in the region where player 1 always fares somewhat better than 2 (incidence point). In a sense, this means that only property rights assignations to player 1 are allowed - which gives player 1 a plain advantage - even if these regimes may be more or less favourable also to player 2 (i.e. they leave player 2 unprotected at different levels against player 1's discretion). Within this subset of institutions, the sub-game corresponding to the outcome space P_1 has a solution nearest to the Pareto frontier of P. This means that there are Kaldor- Hicks side payments that allow reaching the solution P_1 form the solution P_2 but not vice versa. Moreover, there is an arrangement in which player 1 partially redresses the imbalance in the payoff distribution generated by the most extreme form of ownership in favour of player 1 by a utility side-payment in favour of player 2, calculated as the bargaining solution within the bargaining subset P_3 derived from the convex combination of P_1 and P_2. Nevertheless, this seems to be a caricature of the redress principle: the best feasible case for player 2 - the solution under P_2 - has already asymmetrically shifted in favour of player 1. Indeed, drawing the convex combination of spaces P_1 and P_2 simply induces a compromise between two solutions both to the advantage of player 1; and any whatever linear combination of these solutions will shift the final result even more toward player 1's advantage than will taking the solution directly in P_2. So why should player 2 not insist on the less efficient but nevertheless feasible solution in P_2?

Global justice here seems to prevail over the alternative. Following the straight line joining the status quo to the global justice solution G in P, the north-east boundary of P_2 is crossed in G'. Because P is a perfectly symmetric payoff space, this happens along the 45° straight-line. Hence the solution G' is egalitarian and also proportional to relative needs in a global sense.

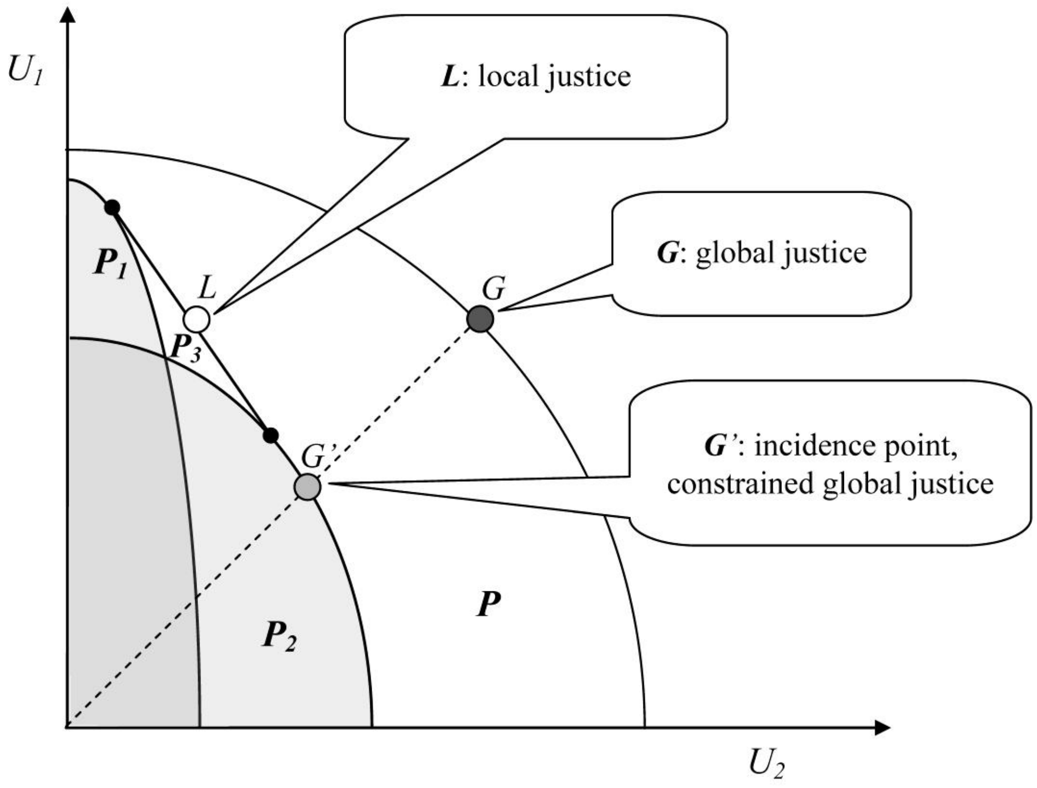

Figure 10 - Global justice and local justice

By contrast, the locally fair solution L, located on the north-east boundary of P_3, seems excessively to reflect the arbitrary fact that only institutions that favour player 1 are feasible. Apparent realism would mistakenly suggest abandoning global justice for local justice, but this is not the case. G' lies on the boundary of the payoff space of a sub-game pertaining to a feasible institution, while this is not the case of L, which lies outside any feasible payoff space. Hence proper realism would suggest proceeding the other way round, and admitting an allocation of control rights compatible with selecting the approximation to utopia G'. Thus both the ethical intuition of distributive justice and the requirement of ex post stability seem to suggest a reformulation of the "non utopian" version of the firm's constitutional contract. Rawlsian contractarian theory, as already illustrated, provides this reformulation.

5 The Rawlsian theory of corporate governance and control

As already discussed, for whatever (repeated) game, based on a constituent *social dilemma* game, however endowed with an asymmetrical equilibrium (convex) outcome space, the Rawls-Binmore social contract always selects a non-cooperative Nash equilibrium coinciding with an application of the Rawlsian maximin principle of welfare distribution. It is computed as the egalitarian solution within the symmetrical intersection set generated by the original (equilibrium) outcome space and its symmetrical translation with respect to the Cartesian axes, i.e. the Nash Bargaining Solution (NBS) computed with respect to this symmetrical payoff subspace.

5.1 Egalitarianism and constitutional choice amongst different control and governance structures

From this general result let us return to the constitutional choice of a governance and control structure of the firm. Consider two different institutionally *feasible* subsets derived form the all-inclusive set of the possible governance and control structures. I interpret this hypothesis as stating that, by proper design of the related corporate constitutions, we find two outcome spaces - subsets of the all-inclusive outcome space - corresponding to non-cooperative Nash equilibria sets (in the sense of the Rawls – Binmore theory). Given that such equilibria can only derive from the outcome space of an underlying non-cooperative game, it follows that we are necessarily considering constitutions whose outcomes belong as proper subsets to the equilibrium set of the "state of nature" game played as a *repeated* game. In other words, by proper design we are able to select outcome spaces that are different subsets of the basic outcome space P_N of *fig.5* (according to the *folk theorem*, the region lying between the status quo d* and the north-east frontier of the convex and compact envelope of outcomes depicted in P_N is the equilibrium set of the repeated basic game G_N).

Taking such two outcome sets as the starting point, the "veil of ignorance" hypothesis is introduced with respect to each of them - i.e. the hypothesis that players consider *each* feasible constitution from an impartial standpoint by allowing the mutual replacement of the roles (and utility function) that they play under each constitution. Not only is the basic outcome space symmetrically translated, but also *each* feasible subset - candidate for the outcome space of an acceptable constitutionally sub game - must be impartially considered. This means that a symmetrical translation with respect to the Cartesian axes is taken for *every* candidate outcome space, and an acceptable solution is accounted for in terms of candidate solutions that are invariant under the symmetric translation of the respective outcome spaces. Hence, what we relinquish are not impartiality and empathy but only the possibility to take for granted the feasibility of every convex combination of feasible outcome spaces. This is a requirement of realism that reminds us that the implementation of whatever constitution we could devise by institutional imagination is constrained by feasibility. Proposition I logically follows.

PROPOSITION I:

> Given any pair of feasible convex outcome sub-spaces P1 and P2, relative to a pair of constitutions and their respective post constitutional cooperative games, if the "veil of ignorance" hypothesis is introduced, but the "Dues ex machina" hypothesis is rejected, then the Constitutional Choice selects a constitution corresponding to the bargaining sub-game endowed with a feasible outcome sub-space P* such that the *egalitarian solution* in P* dominates any other egalitarian solution belonging to the alternative feasible sub-space.

More specifically, given any two feasible convex outcome sub-spaces P_1 and P_2 and their symmetric translations $P_{1'}$ and $P_{2'}$, *no matter how other characteristics of the relevant spaces are established*,

$$\sigma_2{*}>\sigma_1{*} \qquad \text{if and only if} \quad P_1 \cap P_1' \subset P_2 \cap P_2'$$

where σ* is the egalitarian solution within the respective outcome space P_i and the order relation > should be understood as *strictly superior unanimous acceptance* (strong Pareto dominance). Thus inclusiveness of the symmetric intersection is the only property relevant to the constitutional choice of sub games (see *fig. 11* for an example).

From a purely formal standpoint, this proposition is fairly trivial. Recall the relation > between points s and s', representing players' payoff pairs on the Cartesian plane, is *strong Pareto dominance* (i.e. if s'>s then in s' both players' payoffs are greater than in s). If we take two payoff spaces S and S', both symmetric and convex, such that $S \subset S'$ (S is a proper subset of S'), and two points $\sigma \in S$ and $\sigma' \in S'$ respectively equal to the *loci* where the bisector of the Cartesian plane intersects the north-east frontiers of S and S' (i.e. they are the *egalitarian solutions* relative to spaces S and S' under the condition that $\sigma \in S'$ but $\sigma' \notin S$), then the relation $\sigma' > \sigma$ holds *necessarily* for these points. In fact, all points taken along the bisector are strictly increasing toward north-east as a function of the players' pairs of (identical) increasing payoffs. Since the two egalitarian solutions σ and σ' coincide with two of those points - not identical given $\sigma' \notin S$ - they are also ordered in the same way.

In other words, if two symmetrical payoff spaces S and S' are defined so that $S \subset S'$ and each point $s' \in S'$ is a function of the same increasing monotonic – symmetry and convexity preserving - transformation of a pair of players' payoffs corresponding to a point $s \in S$, then also the egalitarian solution point $\sigma' \in S'$, which lies on the bisector, will be a monotonic increasing transformation of the egalitarian solution point $\sigma \in S$, which also lies on the bisector - that is $\sigma' > \sigma$.

Of course the intersection between any generic convex space and the space generated by its symmetrical translation with respect to the Cartesian axes is also a symmetric space. Thus, when many intersection sets are generated by this operation from generic convex spaces, an entire collection of symmetric spaces results so that they are related to each other by set theoretic inclusion. It follows that Pareto-dominance among egalitarian solutions, each belonging to a different payoff space, is monotonically related to how much inclusive these symmetric intersection sets are.

From a substantive point of view, however, it is important that Pareto-dominance *only* between egalitarian solutions should be considered as the decisive condition for the unanimous choice of constitutions, *no matter how other characteristics of the payoff spaces are settled*. From this perspective, the proposition states that the level of unanimous acceptance of a constitution (and hence its outcome) dominates the level of acceptance of another constitution only if its egalitarian solution is Pareto-superior to the egalitarian solution of the alternative, no matter what the same Pareto dominance relation states about other points in the respective payoff spaces. From sections 2 and 3 we know that this restriction of unanimous acceptance to egalitarian solutions rests on a concern for *impartial feasibility*, i.e. an individual rationality criterion (equilibrium) under the hypothesis of impartiality (veil of ignorance), rather than for maximizing some welfare aggregate. We choose then the most efficient (in the Paretian sense) point within the collection of egalitarian solutions, which are monotonically ordered according to the inclusiveness of the respective intersection sets, since this restriction guarantees satisfaction of an ex post stability condition granted that the decision must be ex ante impartially acceptable under the "veil of ignorance" .

To illustrate proposition I, consider *fig. 11*. The "all-encompassing" outcome space P represents all the logically possible ways to cooperate on choice of a constitution. It is assumed that no equilibrium points exist that are able to implement all outcomes in P, and in particular there is no such equilibrium corresponding to the utopian solution U in P, i.e. its symmetric NBS. Thus our attention is restricted to two subspaces P_1 and P_2, which are *feasible* subsets of P. These subsets are construed so that they can be also understood as proper subsets of the convex equilibrium space P_N of the "state of nature game" played as a *repeated* game.

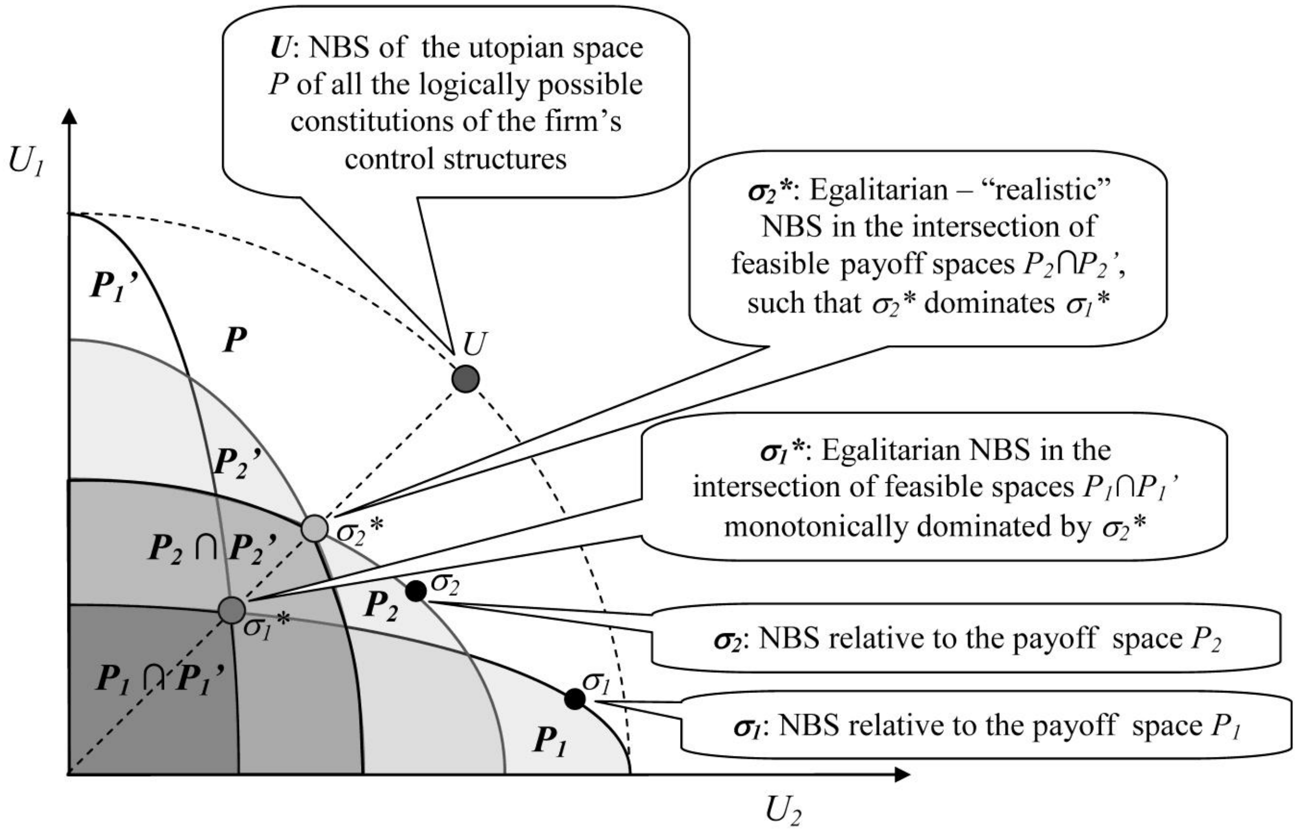
U: NBS of the utopian space *P* of all the logically possible constitutions of the firm's control structures
U_1
σ_2*: Egalitarian – "realistic" NBS in the intersection of feasible payoff spaces $P_2 \cap P_2'$, such that σ_2* dominates σ_1*
P_1'
P
U
P_2'
σ_1*: Egalitarian NBS in the intersection of feasible spaces $P_1 \cap P_1'$ monotonically dominated by σ_2*
σ_2*
$P_2 \cap P_2'$
σ_2
P_2
σ_1*
σ_2: NBS relative to the payoff space P_2
$P_1 \cap P_1'$
σ_1
σ_1: NBS relative to the payoff space P_1
P_1
U_2

Figure 11 - The Pareto-dominant egalitarian solution dominates local justice

Because they are related to the asymmetrical space P_N embodying natural inequalities between the two players, both spaces P_1 and P_2 are asymmetrical and give some advantage to player 2, but at different levels. In comparison with P_2, P_1 is a more asymmetric outcome space with a cooperative solution σ_1 of the post-constitutional cooperative game quite near to the north-east frontier of P. In terms of NBS or other welfare measures, this entails that this post-constitutional game would produce a larger amount of aggregate utility as solution - i.e. compared with P and P_2, the solution σ_1 of P_1 is *second-best* in term of efficiency (again taking the utopian solution of P as the first best), even though the aggregate value is quite unfairly distributed. P_2 on the contrary entails a cooperative solution σ_2 of the cooperative post-constitutional sub-game which is *third-best* in terms of efficiency. However, because its solution σ_2 lies nearer to the bisector joining the origin with the egalitarian solution U, it would distribute payoffs in fairer shares. Recall that according Rawls-Binmore theory a constitution needs to be found by impartially acceptable choice. In other words, i.e. a constitution must be chosen with an invariant solution under the symmetric replacement of the players' roles, which at the same time must be ex post stable (equilibrium). Picking solution σ_1 or σ_2 as such is thus ruled out. But feasibility also debars us from any arbitrary operation on the convex combination of spaces P_1 and P_2. So what properties does constitutional choice impose on the final payoffs in terms of ex post distribution? And which outcome space corresponds to the selected constitution?

For each feasible outcome space, *fig. 11* also shows the respective symmetrical translation P_1' and P_2'. Assuming that no convex combination of P_1 and P_1', and P_2 and P_2' can be generated, we must focus on the respective intersection sets $P_1 \cap P_1$' and $P_2 \cap P_2$', where it is clear that the former is a proper subset of the latter. Both intersection sets are symmetrical spaces, and have symmetrical NBS equal to the egalitarian solutions σ_1* and σ_2* belonging to P_1 and P_2 respectively and lying on the bisector. Both these solution are impartial because they are invariant under the players' role replacement. But they are also *feasible*, given that all the points included in these intersection sets are equilibrium points of the underlying "state of nature" game, so that any convex combination of outcomes falling *within* a symmetric intersection set would be implementable in equilibrium. Any agreement within each of these sets would not be ruled out by unfeasibility if one player's role were interchanged with the other, since the resulting agreement would nevertheless be an equilibrium. However, the symmetrical intersection set $P_2 \cap P_2$' strictly includes $P_1 \cap P_1$', so that the egalitarian solution within P_2 strictly Pareto-dominates the egalitarian solution relative to P_1.

Fig. 11 shows why. The more asymmetric a payoff space and the more unequal its post-constitutional NBS with respect to the available alternative, the less inclusive is its intersection set, and the less unanimously acceptable (in term of constrained Pareto dominance) its egalitarian solution.

Summing up, constitutional choice falls on the constitution with outcome space P_2, which would have a post-constitutional bargaining solution σ_2 (as far as the pure exercise of ownership and control rights is considered). But in order to make such a constitution impartially acceptable and at the same time to preserve its *feasibility* of, the constitutional choice requires an *ex post* redistribution with respect to the solution σ_2 belonging to P_2 such that the egalitarian solution σ_2* in P_2 is *de facto* implemented. Thus *egalitarian redress* of the disadvantaged stakeholder is the main constitutional constraint on implementation of the constitution of ownership and control rights denoted by P_2. It entails maximizing the benefit of player 2, who even under this less unfair constitution still occupies the role of the disadvantaged player. Note that because the dominant egalitarian solution is an equilibrium of the underlying game, reaching an agreement on the redistributive mechanism is not "wishful thinking". No constitutional agreement may be acceptable without the ex ante acceptance of such an egalitarian condition, and the selected egalitarian solution – admitted that it coordinates expectations also in the post-constitutional game – is also *ex post* stable as it is a Nash equilibrium.

5.2 Global justice overrides local justice

*T*hus far we have been concerned only with the *instability of the equitable institutional arrangement* problem. Let us now turn to the second problem: the divorce between global and local justice in the choice of the firm's constitution. The Rawlsian theory of corporate governance solves this problem because neither global justice nor local justice as such simply succeeds; but considerations from global justice make it possible to derive an approximation to global justice that always overrides local justice. In fact, the egalitarian solution is always on the bisector where also the global justice solution lies, and given any two different feasible payoff subspaces, and the symmetrical intersection sets that they generate with their symmetric translation, their egalitarian solutions always stand in a relation of monotonic dominance of one over the other. Thus the Pareto-dominant egalitarian solution provides the best feasible approximation to global justice. No room remains for considerations of local justice, which are rebutted simply by the unfeasibility of the collateral utility transfer mechanism.

To see why, for the moment discard the strict concern for adherence of the feasible payoff sub-spaces to the underlying state-of-nature equilibrium space, and allow constitutions to be *feasible* in a less constrained sense, so that effectiveness may be granted by hypothesis to whatever subset of the all-encompassing space P. In this light we can reconsider the cases of *fig. 9* and *fig. 10* (see *fig. 12* and *fig. 13* respectively)

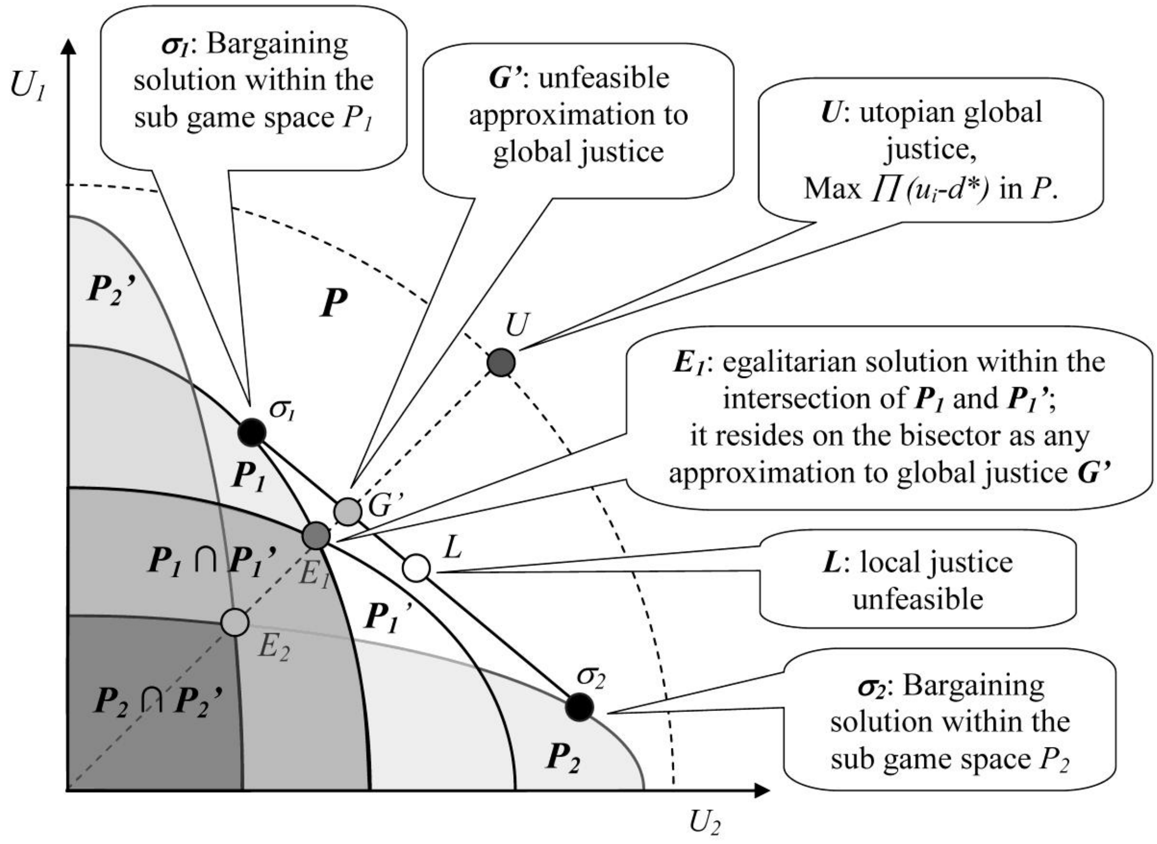
U_1
σ_1: Bargaining solution within the sub game space P_1
G': unfeasible approximation to global justice
U: utopian global justice, Max $\Pi(u_i\text{-}d^*)$ in P.
P_2'
P
U
E_1: egalitarian solution within the intersection of P_1 and P_1'; it resides on the bisector as any approximation to global justice G'
σ_1
P_1
G'
L
L: local justice unfeasible
$P_1 \cap P_1$'
E_1
P_1'
E_2
σ_2
σ_2: Bargaining solution within the sub game space P_2
$P_2 \cap P_2$'
P_2
U_2

Figure 12

In *Fig. 12*, P_1 and P_2 are two outcome spaces corresponding to institutionally feasible constitutions such that either player 1 or player 2 is alternatively advantaged (by alternative assignments of exclusive property rights). Note that this presumes that feasible institutions do not coincide with state-of-nature equilibria, or – put differently – players are able to generate other equilibria or stable configurations of play through their institutional imaginations and artifice. This figure also considers the spaces P_1' and P_2' resulting respectively from the symmetric translation of space P_1 and P_2 with respect to the Cartesian axes. The intersection between space P_1 and its translation P_1' entirely includes the intersection between space P_2 and its translation P_2'. Thus its egalitarian solution E_1 dominates the second E_2. It is noticeable that what was said to be a local justice solution L is no longer affordable because it is infeasible. What about the egalitarian solution G' previously called "approximation to global justice" because it was resident on the bisector where also the utopian solution U lies? Even though it is Pareto-dominant over the alternatives, it is nonetheless ruled out because it is unaffordable due to unfeasibility. However, the Rawls-Binmore solution E_1 provides a new second-best approximation to global justice which is compatible with feasibility.

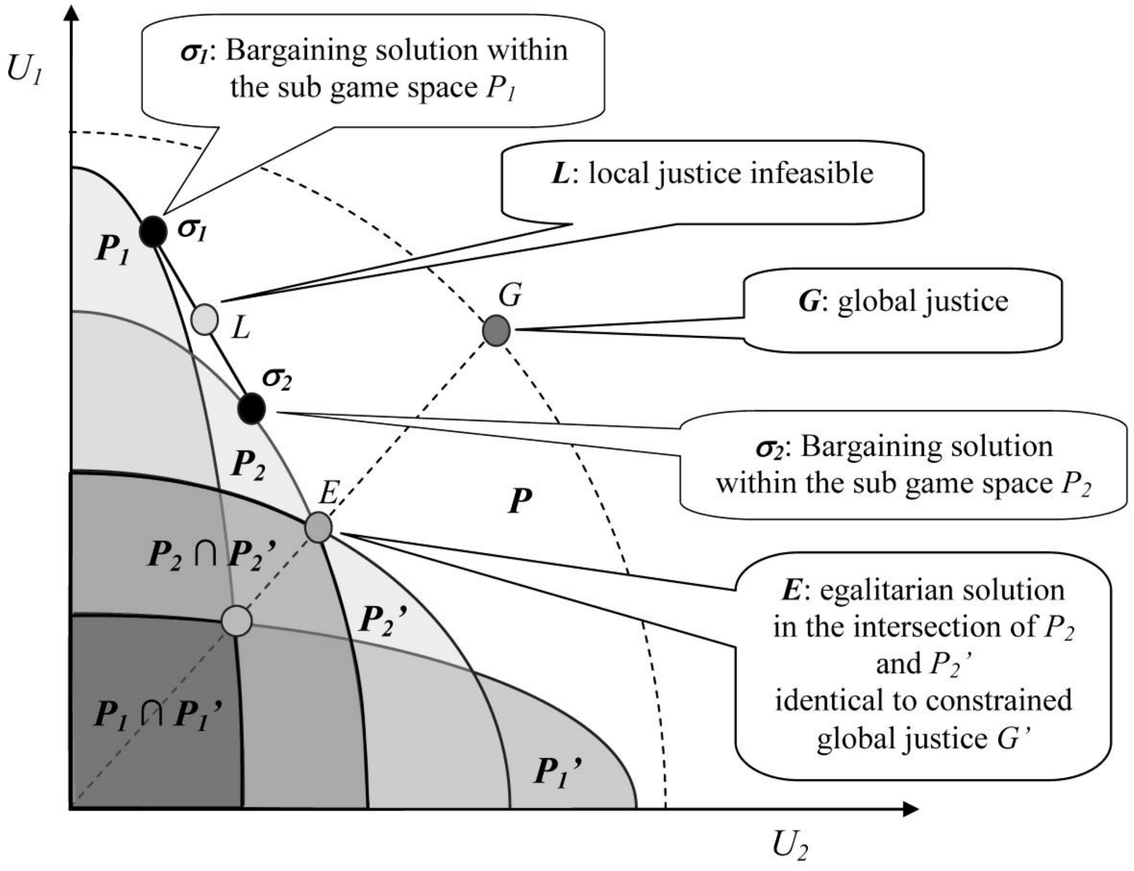
U_1
σ_1: Bargaining solution within the sub game space P_1
L: local justice infeasible
G: global justice
σ_2: Bargaining solution within the sub game space P_2
E: egalitarian solution in the intersection of P_2 and P_2'
identical to constrained global justice G'
P_1
σ_1
L
G
σ_2
P_2
E
P
$P_2 \cap P_2$'
P_2'
$P_1 \cap P_1$'
P_1'
U_2

Figure 13

The case of *fig. 13* is somewhat clearer in terms of its implications for the global VS/ local justice problem. We started with two feasible outcome spaces P_1 and P_2 both benefitting player 1 at different levels. This case can be regarded as one where ownership is always allotted to player 1, granting some degree of abuse of authority to player 1. But under the constitution corresponding to the outcome space P_2, player 1's residual right of control is moderately constrained. All this can be seen by looking at the respective post-constitutional bargaining solution annexed to the two constitutions (S_1 and S_2). In order to redress such unfairness of the feasible solutions, the local justice collateral utility transfer L and the constrained global justice solution G' (directly belonging to the feasible space P_2) have been proposed. The latter coincides exactly with the egalitarian solution E selected by the Binmore–Rawls social contract, because it was already the egalitarian solution selected by the incidence point of the bisector on P_2 frontier, which is the most symmetrical payoff space among those considered here. By introducing into *fig. 13* also the symmetrical translations of spaces P_1 and P_2, accounting for considerations of impartiality and "veil of ignorance", the intersection set $P_2 \cap P_2$' results more comprehensive than $P_1 \cap P_1$'; hence its egalitarian solution is dominant. Again, the local justice solution is unaffordable because it does not belong to any feasible payoffs space. I do not have to deal with its anti-intuitivism from the distributive justice point of view (it redresses player 2 less than does solution S_2). Feasibility already rules out it from the outset.

6 Challenging received wisdoms

Some corollaries are required to illustrate the relevance of the main proposition given in the previous section to the economics of institutions and in particular to the selection of the firm's governance and control structures. They concern two typical positions playing important roles in the literature on institutions design: the aggregate welfare maximizer and the libertarian one.

6.1 Fairness VS. Welfare?

Consider two feasible outcome spaces P_1 and P_2 such that P_1 includes both the maximal utilitarian solution and the best solution in terms of Kaldor- Hicks efficiency. Nevertheless, P_2, with its symmetric translation P_2', generates an intersection set which strictly includes the intersection of P_1 and its own symmetric translation P_1'. Then, any rational constitutional choice must prefer the constitution of the firm corresponding to the outcome space P_2 - no matter what the efficiency properties of P_1.

Assume that the Utilitarian and Kaldor-Hicks solutions do not coincide with the egalitarian solution of any relevant outcome space P_i as such. We are thus in a situation such as depicted by *fig. 13*, where the quite unequal NBS solution S_1 in P_1 is the also the one that satisfies both the foregoing welfarist conditions. Since a constitutional choice must be reached under the "veil of ignorance", a natural way to preserve this solution would be to take the equally probable lottery between this solution treated as a point belonging to the original space P_1 and its realization under the symmetric translation in space P_1'.

But without the *dues ex machina* assumption, a convex combination of these symmetric Utilitarian or Kaldor-Hicks solutions does not generate any feasible outcome. On the other hand, the feasible intersection of $P_1 \cap P_1$' is Pareto-dominated by $P_2 \cap P_2$', so that P_1 cannot be constitutionally chosen. An efficiency criterion (Pareto dominance) is then decisive for the unanimous acceptance of a constitution in so far as it is *restricted* to comparison between egalitarian solutions. Hence, *equity constraints efficiency.* It follow that

COROLLARY 1: Equity comes before efficiency.

Often the quest for social efficiency does not extend to requiring satisfaction of the demanding standard of utilitarianism. Many *law & economics* analysts are sufficiently content with wealth maximisation taken as a proxy for the more demanding utilitarian requirements. But wealth maximisation as a solution concept performs no better than the former two in the context of constitutional choice (for example in *fig. 13* the space P_1's solution S_1 also maximises the payoffs sum understood in simple monetary terms). Joint feasibility and impartiality rules out wealth maximisation. Even if it may sound iconoclastic to the standard theorizing in law and economics, the following proposition naturally obtains.

PROPOSITION II:

> In order to select an institutional form of corporate governance under the constraint of being ex post stable – i.e. implementable by an equilibrium point – do not bother with welfare maximization or its proxy, wealth maximization. Instead, look for the best "egalitarian solution", in the qualified sense of being the best monotonic Nash bargaining symmetric solution among those related to the symmetric intersection sets resulting from symmetrical translations of the outcome equilibrium sets annexed to feasible constitutions.

Students of corporate governance may be struck by this result, which contradicts many of the subject's basic credos – as they have been extensively elaborated by, for example, Kaplow and Shavell [3]. Let us quote them extensively:

"Our argument for basing the evaluation of legal rules entirely on welfare economics, giving no weight to notions of fairness, derives from the fundamental characteristic of fairness-based assessment: such assessment does not depend exclusively on the effects of legal rules on individual's well-being. As a consequence, satisfying notion of fairness can make individual worse-off, that is , reduce social welfare. Furthermore, individuals will be made worse off overall whenever consideration of fairness leads to the choice of a regime different from which would be adopted under welfare economics because by definition the two approaches conflict when a regime with greater overall well-being is rejected on grounds of fairness" (p. 52). This thesis is particularly compelling because also in important and simple situations, i.e.

"symmetric contexts – those in which all individuals are identically situated – it is always the case that everyone will be worse off when a notion of fairness leads to the choice of different legal rule from that chosen under welfare economics" (p.52).

The violation of strong Pareto optimality (choosing a rule under which everyone is worse off) is particularly unacceptable in such a symmetric context. In order to avoid such a risk, the conclusion is that no institutional regime should be chosen primarily on the basis of fairness; or better, fairness as an independent criterion with respect to aggregate welfare maximization must have no role to play in the choice of institutions.

On the contrary, given the previous analysis, it may be shown that

(i) In the simplest symmetrical cases, egalitarianism and strong Pareto optimality always go hand in hand;

(ii) In most cases where only asymmetric payoff spaces are feasible, but individuals are symmetrically situated by imposition of the "veil of ignorance" (the typical case of symmetric situation also for Kaplow and Shavell) it is very reasonable to put maximization of aggregate welfare completely aside in order to maintain egalitarianism, without any contradiction of "general acceptance" understood as a strong Pareto condition;

(iii) Even in the special case where the legal regimes under assessment correspond to a feasible payoffs space that renders egalitarianism Pareto-dominated, egalitarianism has reasonable priority over welfare maximization as the criterion for identifying the payoffs allocation that should be generated in order to make such a regime acceptable. It constrains Pareto improvements reasonably acceptable by all players to be consistent with the *least* deviation from perfect egalitarianism; moreover, it reasonably debars players from reaching solutions of welfare maximization that would be naturally acceptable if no weight were given to fairness.

Before arguing in favor of these propositions, let us recall that Kaplow and Shavell define a fairness principle as an assessment criterion not consequentialist and not entirely based on personal well-being measures, i.e. not entirely reducible to an assessment of the individuals' subjective welfare perceptions annexed to consequences that happen to each individual under such a legal rule.[4] Thus a fairness principle is an assessment criterion Z(x) where x is a legal regime, or rather a state of affairs described in terms of individual actions regulated by the relevant regime, but not necessarily (and only) their consequences.[5] Thus Z is not reducible to a description of personal well-being levels or utilities and their aggregation (summation or multiplication or whatsoever) because it evaluates x in terms of other characteristics - for example, fairness, rights or duties. Egalitarianism falls within this assessment category: it accounts for the state x in terms of a *ratio* between agents' payoffs, which admittedly presupposes a description of personal utilities but says more. It states how equal is the *proportion* between players' payoffs, whatever they are in absolute terms. It is a *relation* not reducible to a measure of how well individuals fare as distinct persons or as an aggregate.

Be warned that Kaplow and Shavell's argument is tricky. Fairness considerations are accommodated by welfare maximization because individuals possibly develop a *taste* for fairness.[6] Thus fairness becomes an object of preference exactly like any other consequence or good whereby it can be accounted through the personal subjective well-being that individuals attach to this taste. No doubt, the formal treatment of preferences can be extended to make room for fairness principles as motives to act and represent them through utility functions (for a proper enlargement of the motives to act represented by utility functions see part III of this essay). But calling a *taste* the motivational importance that we give to adhering to principles is quite at odds with intuition. In fact, there is no reason to reduce preferences – i.e. binary relations expressing whatsoever *betterness* judgment consistent with behavior (see Broome, 1999) - to the idiosyncratic case of *tastes.*

It is also noticeable that this immunization move entails that Kaplow and Shavell's theory is virtually devoid of any empirical content (and perhaps paradoxical). Assume that most people are convinced of the view that Kaplow and Shavell wish to confute. Nearly everybody prefers to assess legal regimes by fairness principles not completely dependent on individual wellbeing - for example, by using equality as a choice criterion. Since they prefer to perform assessments of this kind, Kaplow and Shavell would say that the people have a taste for fairness, and hence that people's welfare is maximized by assessing legal regimes on the basis of a criterion that gives no essential relevance to welfare maximization. Given such a social preference, Kaplow and Shavell would conclude that legal regimes are chosen solely on the basis of considerations of personal well-being and welfare maximization, even though the actual assessment of legal rules accommodated by their own theory rests on fairness principles which do not primarily refer to personal well-being. Could one say that such a theory is useful in any sense? Defining a different social choice rule consistent with the fact that individual utilities are functions (also) of fairness principles - appropriately understood as measures of the motivational strength of individuals' adhesions to fairness principles - would be more useful than collapsing everything into generic welfare maximization.

However, let us set aside these comments and take Kaplow and Shavell's thesis at its best. How would it work in our context of constitutional choice on intuitional regimes of corporate governance and control? It is clearly irrelevant in the simplest case where only constitutions represented by symmetric payoff space are feasible. Such constitutions are increasingly ordered in terms of Pareto dominance by inclusiveness of their payoff spaces; and the acceptability of their egalitarian solutions monotonically depends on the inclusiveness ordering defined on spaces. In this case, there is no divorce between egalitarianism and efficiency. Given the perfect equality of players, no reasonable bargaining theory may ask players to accept any solution except the symmetrical one. At the same time, the intuition that the solution must fall on the bisector is simply completed by the requirement that it also resides on the payoff frontier. As this is true under any initial symmetric feasible payoff space, it is also true under any symmetrical translation of the payoff space which cannot destroy the original symmetry of the situation. In fact, impersonality and the veil of ignorance, operationalized through symmetric translation of the payoff space, map the space onto itself, generating a perfectly identical payoff space. Players are perfectly identically situated and see the solution in exactly the same way under both the players' roles. Solution invariance under symmetric translation of the payoff space (which is the egalitarian requirement derived from impersonality and impartiality) is naturally satisfied by keeping to the symmetric solution that already proved intuitive given the initial payoff space representation. Even though egalitarianism is defined in term of the payoffs *ratio* (*1/1*), not a specific allocation of any welfare amount, it is not inconsistent, but rather perfectly compatible, with 'general acceptance' as Pareto dominance because it requires taking the intersection of the bisector with the north-east boundary of the payoff space as uniquely defined solution.

However, Kaplow and Shavell's thesis seems rather relevant to cases where the only outcome spaces corresponding to feasible constitutions are asymmetrical and reflect inequalities among players. Players can then be identically situated with respect to the decision problem precisely because of the symmetrical translation of the payoff space that allows mutual replacement of their personal and position-relative points of view, and the introduction the veil of ignorance in order to seek a solution which is impartial and independent from any personal perspective. Owing to feasibility and the *No Deus ex machina* assumption, identically situated players must choose the solution from within the intersection set and pick it up on the bisector. Thus, in the case of two possible feasible constitutions, no matter what their further efficiency properties, the one with highest egalitarian solution must be chosen - because it is identified by a monotonic function of symmetrical intersections sets inclusiveness. No doubt, this solution will not generally satisfy most of the usual welfare maximization concepts, such as utilitarianism, or the largest Nash bargaining product with respect to alternative feasible constitutions. Moreover, such welfarist solutions could be easily reached from the egalitarian solution through Kaldor-Hicks utility side-transfers that testify to the social efficiency of these further solutions.

Nevertheless, there are very good reasons for not accepting these solutions instead of the best egalitarian one. These reasons are *feasibility* together with the "*veil of ignorance*" and awareness that there is *no Deus ex machina* able to enforce any agreement that players may reach in the constitutional choice context. Impartiality and impersonality (underlying the veil of ignorance) are independent of personal well-being and they constrain the solution to be on the bisector. Feasibility together with the *No dues ex machina* hypothesis requires that such a solution must be reached within the intersection set. Quitting this outcome set in order to reach the welfare maximizing solution would simply mean that one party can impose looking at the solution solely from his/her point of view, because s/he is effectively the stronger player in the actual game of life. Conversely, looking at the solution from the perspective of the symmetrically translated payoff space would be considered pure wishful thinking. But the egalitarian solution within the intersection set is also feasible, i.e. it corresponds to an equilibrium under both the payoff spaces representations. Its implementation is incentive compatible whatever personal role is taken by players. This impartial realism overrides the claim of the fortunate player to profit unilaterally from his strongest position. For an example see *fig. 13*, where S_1 in P_1 is both the utilitarian solution and the highest value of the Nash bargaining product among any feasible spaces; but nevertheless the chosen constitution is P_2 because its egalitarian solution is better. What about acceptability in terms of making all players worse off or better off? No solution Pareto-dominates the alternative; hence there is no room for asserting that egalitarianism worsens each player's position. It is true that a Kaldor-Hicks utility transfer could improve player 2's position if he agreed to switch from the egalitarian solution to S_1. But why should s/he accept this change rather than any other one more sensitive to fairness considerations?

In order to clarify this point, consider the third case illustrated in *fig. 14*, which is also the most problematic from the egalitarian point of view. The feasible payoff space P_1 is so asymmetric that by considering its translation P_1', the intersection set is a very narrow region of the plan and the egalitarian solution in $P_1 \cap P_1$' proves to be Pareto-dominated by S_1, where both the maximal utilitarian solution and the maximum Nash bargaining product reside, with respect to any other feasible outcome. This seems to be a case where keeping to fairness makes every players worse off, which - according to Kaplow and Shavell - is unacceptable. In fact, player 1 could try to convince player 2 to relinquish egalitarianism with the reasonable argument that there is a mutual advantage in switching to S_1. To be sure, this would entail also relinquishing adhesion to principles of impersonality and impartiality, because accepting S1 means selecting the bargaining solution rationally reachable by playing the post-constitutional bargaining game related to space P_1 as a separate game, without any pretence of choosing a solution under a veil of ignorance. But in the end, why defend impersonality and impartiality if these principles condemn everybody to having the worse?

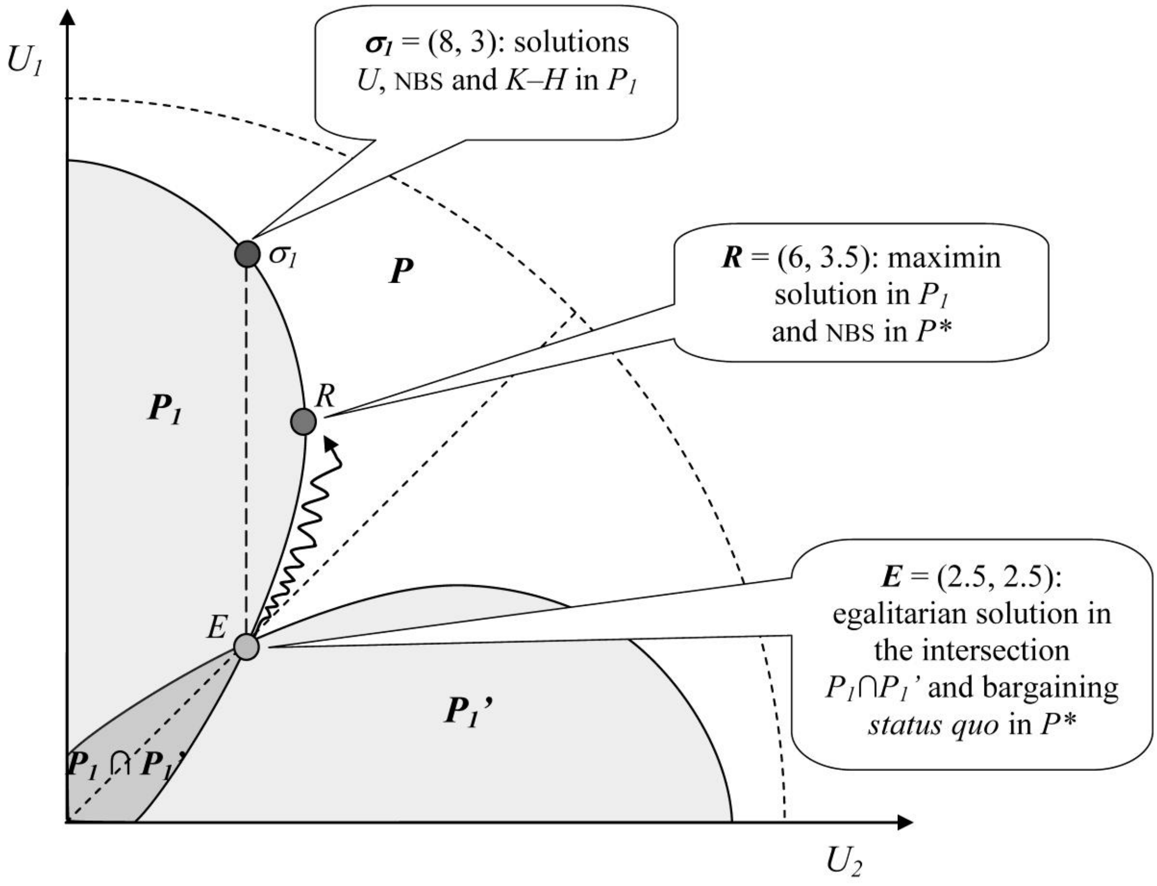
U_1
U_2
σ_1 = (8, 3): solutions U, NBS and K–H in P_1
σ_1
P
R = (6, 3.5): maximin solution in P_1 and NBS in P^*
R
P_1
E = (2.5, 2.5): egalitarian solution in the intersection $P_1 \cap P_1$' and bargaining *status quo* in P^*
E
P_1'
$P_1 \cap P_1$'

Figure 14

But this is not the case. On the contrary, giving egalitarianism priority over welfare maximization is perfectly reasonable because it allows mutually acceptable Pareto improvements with respect to the egalitarian solution itself. Egalitarian solutions constrain Pareto efficiency in so far as egalitarianism is taken to be the proper starting point from which acceptable Pareto improvements are calculated. This solution is the maximin point R on the north-east frontier of the space P_1, where player 2's payoffs (the disadvantaged player) are improved as much as possible, no matter what the marginal payoff improvement of player 1 (who for each player 2's improvement obviously fares better than player 2 him/herself). According to this solution, Pareto improvements with respect to E are achieved by moving along the frontier of P_1, and they end as soon as no better improvement in player 2 payoff is possible. This solution dominates E, but it makes sense only because E is taken to be the appropriate status quo from which the Pareto improvements process is started.

Assume that E is initially accepted owing to impersonality and impartiality seen as independent (from personal well-being) conditions, under the additional assumptions of feasibility and *No Dues ex machina*. Then, player 1 proposes to player 2 a switch from E to S_1 for reasons of mutual advantage. Player 2 can reply that it is also unfair not to consider the alternative Pareto-dominant solution S_1', that would advantage her rather than player 1 if the symmetrical translation P_1' were assumed as the payoff space from which to select the solution. Thus she suggests that some compromise between the two solutions S_1 and S_1' should be agreed upon in order to improve over E. However, player 1 may insist that seeking a solution in P_1' is pointless: space P_1' is only a virtual, conjectural payoff space admitted for convenience of the veil of ignorance exercise, but only P_1 is the relevant payoff space of the game players will actually play. Agreeing on S_1 prevents mere cheap talk because it entails reaching an equilibrium point that will be executed in the implementation stage. By contrast, if an agreed random mechanism were to select the corresponding solution S_1', player 1 could simply veto its implementation. Since all this is common knowledge, it can be also anticipated by both the players at the stage where they are to select a proper constitution by the social contract. In other words, S_1' is outside the feasible agreement set that they can reach at this stage because player 1's *actual* concession limit does not extend to include S_1'.

Note that all these are arguments of rational bargaining. Hence, by similar argument, player 2 can recall that the solution E, being itself an equilibrium point lying within both spaces P_1 and P_1', is the status quo of a bargaining game seen as a second thought in the constitutional choice. In fact, E has been accepted at least as a first step in the selection of the solution; so it is the outcome that will be effectively implemented if the players do not agree on any further improvement on E. By sticking to E, player 1 can effectively veto any unacceptable change to the constitutional solution. What results is a new bargaining problem which takes E as the *status quo* that delimits the set of possible agreements as those included within the players' concession limits on the Pareto frontier of P_1.

A peculiarity of the new bargaining problem is that the *status quo* point E defines as the relevant bargaining set the outcome subspace P*. In P*, the players' incentives to reach an agreement are different. Whilst player 2 is restricted to claiming only her minimal acceptable payoff fixed at E (e.g. *2.5*), on the other hand a very large surplus appropriable by player 1 is created (e.g. *8.5*). Any movement from that position in order to improve player 2's payoff entails a trade-off (a conflict) between player 1 and player 2. By contrast, restricting player 1 to claiming only her minimal acceptable payoff set at E (*2.5*) is of no value to player 2. Moving from this position along the payoff frontier in order to improve player 1's payoffs on the status quo is also in the best interest of player 2. She fares better and better by also raising player 1's payoff until player 2's maximum possible payoff in P* is reached at R = (*6, 3.5*). This means that player 2 is a much more profitable bargaining partner for player 1 than the other way round, because there is much less bargaining attrition in reaching player 2's most desirable agreements - which are also desirable to player 1 - than player 1' most desirable agreements. In other words, player 1 is much readier to satisfy player 2's claims to improve her payoff than player 2 is in regard to player 1, since in order to satisfy player 1's most desired claim, s/he needs to forgo any possible improvements, whereas player 1 does not face any payoff renunciation by satisfying player 2's highest claim. This clearly reflects upon the Nash bargaining solution relative to the bargaining sub-problem (E, P*) because it coincides with the maximin point R, where the disadvantaged player 2's payoff is maximized.

Consider again the numerical example of *fig. 14*. Payoffs at S_1 are (*8, 3*) for player 1 and 2 respectively. Both the utilitarian solution (*11*) and the Nash bargaining product (24) are maximal at S_1 with respect to the entire P_1 space. But now impose E as the *status quo* of a new bargaining problem with the subspace P* as the appropriate bargaining set. Players' payoffs at E are (*2.5, 2.5*). Then at the maximin point R = (*5, 3.5*) the Nash bargaining product is greater than at S_1:

$(6 - 2.5) \text{ x } (3.5 - 2.5) = 3.5 \; > \; (8 - 2.5) \text{ x } (3 - 2.5) = 2.75$

Thus the players would accept the point R as the constitutional choice of the final payoff allocation that must be carried out by selecting the constitution corresponding to P_1, which entails a redress (from *3* to *3.5*) of player 2 with respect to the solution S_1 reachable in the relevant post-constitutional bargaining game. This shift of the bargaining solution is entirely caused by taking the egalitarian solution E as the appropriate *status quo* of the second bargaining step in constitutional choice, an assumption due to impersonality and impartiality considerations that are independent of personal well-being. True, this induces setting aside welfare maximization solutions belonging to P_1. However, it does not contradict Pareto-dominance at all, because the solution R Pareto-improves on E; or rather, it is the only acceptable Pareto improvement attainable by rational bargaining from E.

Summing up, fairness precedes efficiency in that it establishes the relevant *status quo* from which the proper Pareto improvement can be calculated. Moreover, it constrains such improvements to converge to the maximin solution R, so that no Pareto-efficient improvement is admitted whenever there exists another that would reduce the distance from perfect egalitarianism more (indeed R is the point belonging to the Pareto frontier of P_1 nearest to the bisector).

6.2 Just minimizing transaction costs?

Much closer to the corporate governance literature is Hansmann's theory of "ownership of the firm", which is based on the principle that a single stakeholder class should be given property and control over the firm when this regime minimizes the aggregate value of transaction costs resulting from the summation of governance costs held by the controlling party and the aggregate contract costs held by all the remaining (non-controlling) stakeholders (see Hansmann, 1988, 1996). This is also an aggregate efficiency or wealth maximization criterion seen as a proxy for the utilitarian solution. Hence it is set aside by Rawlsian theory as a solution for the constitutional choice of corporate governance institutions.

Let us assume that each post-constitutional game played under its relevant constitution generates aggregate costs allocations according to Hansmann's formula, and that one particular ownership regime minimizes them. Player 1 could bear the minimal governance cost with respect to any other player, and also his governance costs could be smaller than his contract cost, so that giving him control over the firm would certainly reduce overall costs with respect to a situation of "no corporate ownership and control" - admitted that it does not increase other players' contract costs too much. This can also minimize the overall costs if player 1's contract costs, replaced by his minimal governance costs, are higher than other players' contract costs. Nevertheless, this solution could also not be Pareto-dominant with respect to a more costly institutional alternative if player 1's ownership and control regime were more abusive in terms of player 2's contract costs rather than player 2's control regime in terms of player 1's contract costs (induced by player 2's abuse). This may hold even though, by substituting her "natural" contract costs with her governance costs, player 2 could only gain a small improvement in terms of efficiency. For example, assume that in a "state of nature" of no ownership and control over the productive organization where business relations are only subject to incomplete contracts, players 1 and 2 bear contract costs (7, 7) respectively. Giving ownership and control to player 1 would replace his contract costs with the minimal governance cost 1, but owing to his abuse of authority such a control structure would only slightly reduce player 2's contract costs to 6. On the other hand, giving ownership to player 2 would give more protection to player 1 by reducing his contract costs to 5, but it would inefficiently replace player 2's contract costs with her high governance costs set at 4. Overall, transaction costs under player 1's control score 7 and are minimal, whereas the "state of nature" badly scores 14 and player 2's control scores 9. Nonetheless, there is no reason for player 2 to agree to give control to player 1 rather than claiming control for herself, as long as her cost amount to 4 by controlling and to 6 by not controlling.

The natural response would be to resort to a Kaldor-Hicks efficient side-payment that would immunize player 2 under player 1's control against the effect of his authority abuse, so that her contract costs are kept below 4. But of course in our context the question arises of whether or not this side-payment may fall within a feasible outcome set. Giving so much authority to party 1 under the non-credible promise that he will repay player 2 in the future for his authority abuse may not correspond to any feasible (equilibrium) solution in the ex post perspective.

According to Rawlsian theory, in this situation it may be necessary to chose a different governance structure; for example, by giving control to player 2 if this structure may have a better egalitarian effect on the payoffs allocation. This happens if this better (in the Paretian sense) egalitarian allocation (i) is an equilibrium point resident within the intersection set of the payoff space corresponding to the less efficient governance structure (player 2 control) and its symmetrical translation, and (ii) it can be reached from the cost allocation of the post-constitutional game (e.g. the cost allocation (5,4)) by moving within the equilibrium set of the game. In fact, whereas the first side-payment could be unfeasible, this redress mechanism in favor of player 1 corresponds to an equilibrium point and is therefore perfectly implementable.

6.3 Really is social justice a mirage?

There are other commonplace tenets in the field of the economics of institutions that the Rawlsian theory calls into question. Most of the new-institutional theorising on the governance and control structures of the firm (and other institutions) is based on the implicit postulate that institution design cannot go further than prescribing outcomes interpretable to a certain extent as *spontaneous order*s, or at least as corresponding to outcomes that could be achieved by a spontaneous order. Hayek would certainly see commercial law and corporate governance codes, institutions and principles as sets of norms resulting as spontaneous orders from evolution (see also Vanberg's idea of corporations as constitutional contracts, Vanberg 1992).

Only spontaneous orders are self-enforcing norms, i.e. they do not require the intervention of an external *Deux ex machina* that would heavily constrain individual freedom. This responds to a demand for stability. But this statement points out a concern for freedom of choice. It is the same, but in milder form, as the requirement that any institutional design must be "incentive compatible" – incentives are only relevant to decision makers who are at a certain level free to choose.

Often, this is not just a descriptive belief concerning the fact that economic agents are more or less free and hence able to circumvent any strict regulation that does not provide for an equilibrium property. It is also a normative presumption that freedom of choice must be respected. Now take this normative value as granted and understand it as the central concern of the libertarian standpoint. Our theory has unexpected implications for mild libertarians as well.

COROLLARY 2: Mild libertarians cannot but be egalitarians.

A mild libertarian would not reject the contention that individual agents must enter the "original position under the veil of ignorance". Granted the priority of freedom and spontaneous order, s/he would take the veil of ignorance standpoint at least in order to make an impartial assessment of possible spontaneous order outcomes and to voluntarily agree on such an outcome that is also invariant under the symmetrical permutation of players' roles.

However, constraining the libertarian position with a concern for impartiality, plus the concern for ex post stability (no *Deus ex machina*), has dramatic consequences for the libertarian point of view. Freedom requires spontaneous order (equilibrium), but constraining it by impartiality entails that the only admissible subset of spontaneous orders is the symmetric intersection of the equilibrium set with its symmetric translation. Thus only governance and control structures providing for an egalitarian payoffs distribution (at least in term of redress) are acceptable. Once the "spontaneous order" outcome space has been restricted to the symmetrical subset resulting from the intersection of the original space and its symmetrical translation, the egalitarian solution is the only one acceptable through the players' free agreement.

Libertarians such as Hayek (Hayek, 1973) and Nozick (Nozick, 1974) have militated strongly against any redistributive notion of social justice. But far from ostracizing the "mirage of social justice", even in the small-scale society constituted by the stakeholders of a firm, a moderate impartial libertarian *cannot but be egalitarian* in the selection of the firm's governance structure.

7 Unique ex ante equilibrium selection in the repeated Trust Game and end remarks

Let us return to the problem of the ex ante justification of a particular equilibrium as raised in part I of this essay. The "game of life" played by the firm and its stakeholders was then represented as a repeated Trust Game (TG) where the entire positive region of the payoff space is constituted by Nash equilibria. In this second part, I have been concerned with a generalization of this case by taking the constituent game played by the firm (Adam) and the stakeholder (Eve) as a generic social dilemma resembling an asymmetric prisoners' dilemma (PD) with an enlarged set of pure strategies. The basic difference is that, in the TG, only one side (the firm) can profit form abusing the other player's trustworthy behavior, whereas the only profitable payoff for the stakeholder is reaching the symmetrical cooperation outcome (2,2) when – as usually assumed – it exists. In a typical PD representation of the stakeholder/firm interaction, the two parties would have symmetric abilities to cheat one another. The asymmetric PD-like social dilemma here assumed was midway between the two. Eve (the stakeholder) is allowed some defection opportunity from the contract, even though non-cooperative resources with which to take advantage of the other side's cooperation are in general more profitable to the stronger player Adam (the firm) – what in fact represents in our situation the "game of life" imbalance of power, and also captures the effects of abuse of authority in the stakeholder /firm interaction. But we can now come back to the trust game, which was assumed to be the simplest and most typical formal representation of the implementation problem related to a CSR social norm based on the social contract of the firm, because this problem is addressed through the firm and its stakeholders' strategic interaction.

It is remarkable that Rawlsian theory gives a particularly simple and compelling solution to the ex ante equilibrium selection problem when the repeated TG is considered. The requirement of selecting a solution within the intersection of the basic outcome space X_{AE} (see *fig. 15*) and its symmetric translation is sufficient for singling out a unique solution, once the obvious Pareto dominance condition has been granted, which cannot but be the egalitarian Nash bargaining solution of the original game. In order to achieve this result, we need not concern ourselves with the complex construction of equally probable linear combinations between outcomes resident in a payoff space and its translated version - which is typical of the probabilistic interpretation of the veil of ignorance.

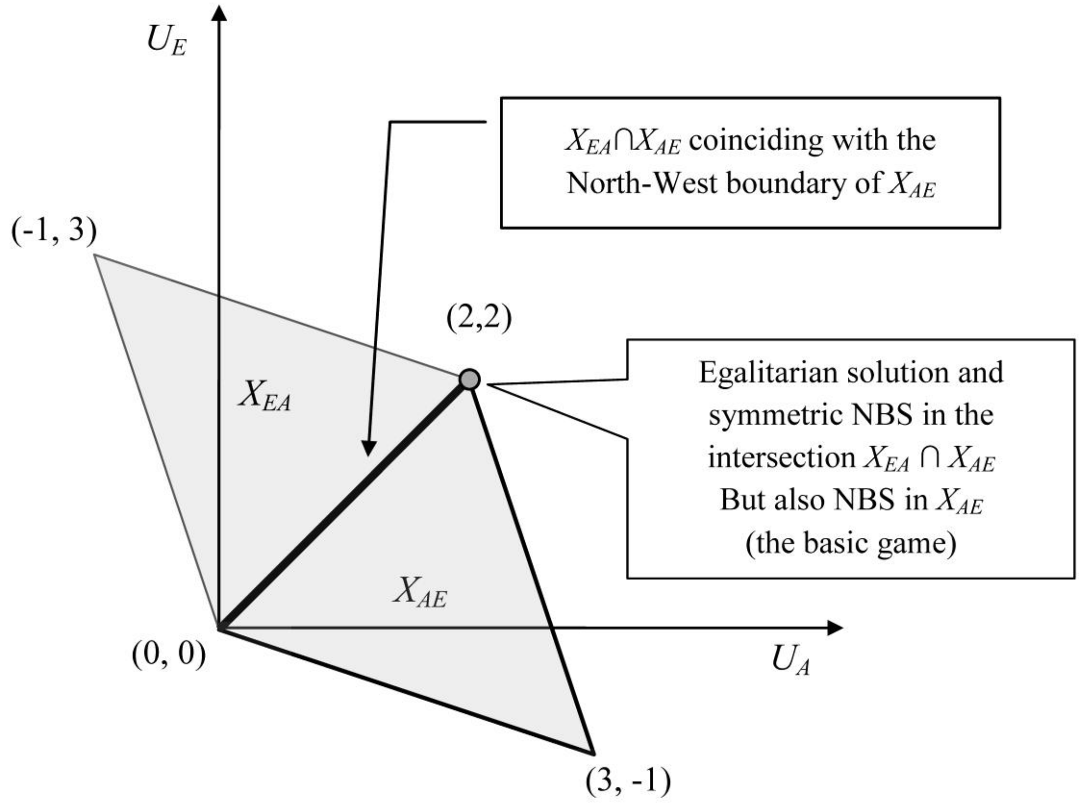
U_E
$X_{EA} \cap X_{AE}$ coinciding with the North-West boundary of X_{AE}
(-1, 3)
(2,2)
X_{EA}
Egalitarian solution and symmetric NBS in the intersection $X_{EA} \cap X_{AE}$ But also NBS in X_{AE} (the basic game)
X_{AE}
(0, 0)
U_A
(3, -1)

Figure 15

Only relevant assumption are *impersonality* (the capacity to permute the individual players' points of view) plus *feasibility* (to stay within the intersection set generated through impersonality), so that the solution must reside within the intersection set generated by rotation of the payoff space X_{AE} around its north-west boundary. But the intersection set is quite peculiar in this case. It exactly coincides with the north-west boundary itself of the payoff space, which lies on the bisector. Because it is reduced to a segment of the 45° line, the solution cannot but be the only point on this line segment belonging to the Pareto frontier, i.e. the symmetric Nash bargaining solution (2,2).

Thus applying the "veil of ignorance reasoning" without a *Deus ex machina* provides a reason for selecting the intuitively fair outcome (2,2) of the Trust Game.

Note that the key point in arriving at this conclusion is simply that an impartial exercise of choice (replacement invariance) must select an equilibrium point *within* the intersection set; that is, an equilibrium point that necessarily exists and is therefore implementable by each player whatever the position he or she occupies in the ex post perspective. A stability condition (the solution must lie in the set of those points that correspond to ex post implementable equilibria) linked with the weak fairness condition of invariance to players' replacement is sufficient to derive the egalitarian solution. Thus, the social contract as an explicit normative method of impartial reasoning helps resolve the multiplicity problem from the ex ante perspective in an extremely simple way in the repeated Trust Game.

However this result should not be overemphasized as far as the equilibrium selection problem is concerned. What would effectively solve the multiplicity problem is an equilibrium selection theory able to predict the *ex post* game equilibrium solution so that it is consistent with the *ex ante* solution identified. In other words, selection is *ex post* effective only if it gives reasons to act that fit the *ex post* reasoning context. *Ex post*, only common knowledge of the solution – that is, a system of mutually consistent expectations converging on the prediction of a uniquely determined equilibrium point – conveys to each player the appropriate reason to act, because choosing an equilibrium strategy amongst many others requires having a clear prediction of other players' behavior and beliefs. However, from that in the *ex ante* perspective a solution is invariant to the players' position replacement, there is no logical reason to conclude that that solution will be effectively implemented. The reason that justifies a particular decision in the *ex post* game is knowledge of what the players will effectively do. Moreover, this knowledge about the other players' decisions must be consistent with their being symmetrically able to predict the others' behavior and to choose their best response to those predictions. Therefore, it is not the impartial selection of a desirable *ex ante* solution, but the knowledge of other players' *de facto* behaviors that provides the proper reason for acting in the *ex post* context. Moreover, there is no logical implication from what is fair *ex ante* selection (even if it falls on an equilibrium point) as to what other players will actually do. Maybe they will act in accordance with the principle, maybe not. The fair *ex ante* agreement, or impartial choice, does not gives us common knowledge of the ex post behavior of players. If, however, one does not know how other players will behave, one has no reason to play a given strategy, even though the fair solution is part of an equilibrium point.

This is not to say that the ex ante agreement on an impartial solution does not provide any cue to believe that players will act according to the same principle in the ex post interaction. But this is simply a matter of fact, or of cognitive psychology, not a matter of logic. Common knowledge, on the contrary, is a matter of epistemic logic: this means recursive group knowledge of what everybody knows to be true (a *truism*).[7] It is the case that a given equilibrium is *commonly known* to be played only if each player has many layers of knowledge about every other player's action, beliefs, beliefs about beliefs, and so on, that are consistent and justify the prediction that this equilibrium will be played. This state of knowledge can be approximated by a theory of belief formation that at last leads us to a stable prediction of any other player's equilibrium choice and belief (see Sacconi 2010c). Ex ante selection, on the contrary, does not predict how one will actually decide; it only answers the question of what equilibrium *should* be chosen, because it is invariant under the individuals' position replacement. The step from an answer to the question of which equilibrium is *fair* to an answer to the question of how players will *actually* behave is a *default inference* that some player may in fact make; but this is just a possibility. Thus, from the perspective of the ex post game, there is still much to do before the multiplicity problem is solved.

References

Arrow, K. (1951), *Social Choice and Individual Values*, New York, Wiley (trad. It., Scelte sociali e valori individuali, Etas libri, Milano, 1977).

Binmore K. (1984) *'Game theory and the social contract'* ST/ICERD discussion paper (84/108), LSE , London.

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

Binmore K. (1989) 'Social contract I: Harsanyi and Rawls', *The Economic Journal,* Vol. 99, No. 395, pp. 84-102.

Binmore, K. and A. Brandenburger (1990), 'Common Konwledge and Game Theory', in K. Binmore, *Essays on the Foundations of Game Theory*, Oxford: Basil Blackwell.

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

Binmore, K. (1994), *Game theory and the Social Contract (Vol. I): Playing Fair*, Cambridge MA: MIT Press.

Binmore, K. (1998), *Game theory and the Social Contract (Vol II):Just playing*, Cambridge MA: MIT Press.

Binmore, K. (2005), *Natural Justice*, Oxford: Oxford University Press.

Buchanan, J. (1975), *The Limits of Liberty*, The University of Chicago Press, Chicago.

Brock, H. (1979), 'A Game Theoretical Account of Social Justice', *Theory and Decision*, 11, pp. 239–65.

Broome, J. (1999), *Ethics out of Economics*, Cambridge: Cambridge University Press

Fagin, R. Halpern J.Y., Moses Y. and Vardi M. Y. (1996), 'Common Knowledge: How you have it, now you don't' in Intelligent Systems: A Semiotic Perspective, Proc. 1996 Int. Multidisciplinary Conf., VOL. 1, pp. 177-183.

Fundenberg, D. and J. Tirole (1991), *Game Theory*, MIT Press, Cambridge Mass.

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

Grossman, S. and O. Hart (1986), 'The Costs and Benefit of Ownership: A Theory of Vertical and Lateral Integration', *Journal of Political Economy*, 94, pp. 691–719.

Hansmann, H. (1988),'Ownership of the firm', *Journal of Law Economics and Organization*, 4, 2, pp. 267-304.

Hansmann, H. (1996), *The Ownership of the Enterprise*, Cambridge, MA: Harvard University Press. Mettere nel suo paragrafo

Harsanyi, J.C. (1977), *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge, MA: Cambridge University Press.

Harsanyi, J.C and R. Selten, (1988), *A General Theory of Equilibrium Selection*, Cambridge, MA: MIT Press.

Hart, O. (1995), *Firms, Contracts and Financial Structure*, Oxford: Clarendon press.

Hart, O. and J. Moore (1990), 'Property Rights and the Nature of the Firm', *Journal of Political Economy*, 98, pp. 1119–58.

Hayek, F. A. (1973), *Law, Legislation and Liberty*, Chigago: University of Chicago Press.

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

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

Kreps, D. (1990), *Games and Economic Modeling*, Oxford: Oxford University Press.

Nash, J. (1950), 'The Bargaining Problem', *Econometrica*, 18, pp. 155–62.

Nozick, R. (1974), *Anarchy, State and Utopia*, New York: Basic Books.

Ostrom, E. (1990), *Governing the Commons*, Cambridge University Press, New York.

Rawls, J. (1971), *A Theory of Justice*, Oxford: Oxford University Press.

Rawls, J. (1993), *Political Liberalism*, New York: Columbia University Press.

Sacconi, L. (1991), *Etica degli affari, individui, imprese e mercati nella prospettiva dell'etica razionale*, Milano: Il Saggiatore.

Sacconi , L, (1997), *Economia, etica, organizzazione,* Bari: Laterza.

Sacconi, L. (2000), *The Social Contract of the Firm. Economics, Ethics and Organisation*, Berlin: Springer Verlag.

Sacconi, L. (2006a), 'CSR as a model of extended corporate governance, an explanation based on the economic theory of social contract, reputation and reciprocal conformism' in F.Cafaggi (ed.), *Reframing self-regulation in European private Law,* Kluwer Law International, London.

Sacconi, L. (2006b), 'A Social Contract Account For CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification', *Journal of Business Ethics*, Volume 68, Number 3 / October, 2006, pp.259-281

Sacconi, L. (2007a), 'A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity', *Journal of Business Ethics*, Volume 75, Number 1 / September, 2007, pp. 77–96.

Sacconi, L. (2007), '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. (2008b) 'CSR as Contractarian Model of Multi-Stakeholder Corporate Governance and the Game-Theory of its Implementation, University of Trento - Department of Economics Working paper N.18

Sacconi L. (2009), 'Corporate Social Responsibility: Implementing a Contractarian Model of Multi-stakeholder Corporate Governance trough Game Theory' in J.P. Touffut and R. Solow (ed.), *Does Company Ownership Matter?*, Centre for economic Studies Series, Edward Elgar Publishing Ltd., London.

Sacconi L. (2010a),"A Rawlsian view of CSR and the Game Theory of its Implementation (Part I): The Multistakeholder Model of Corporate Governance", in L. Sacconi, M. Blair, E. Freeman and A. Vercelli (ed.) '*Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*'

Sacconi L. (2010c), 'A Rawlsian View of CRS and the Game of its Implementation (Part III): Conformism and Equilibrium Selection' in L. Sacconi and G. Degli Antoni (ed.), *Social Capital, Corporate Social Responsibility, Economic Behavior and Performance*' Palgrave MacMillan London (publication date August 2010).

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

Willamson, O. (1975), *Market and Hierarchies*, New York: The Free Press.

Zimmerman, H.J. (1991), *Fuzzy Set Theory and Its Applications*, 2nd revised ed., Dordrecht-Boston: Kluwer Academic Press.

Notes

[1] This section presents my own account of Binmore's theory. Because it has evolved over time (Binmore 1984, 1989, 1994, 1998, 2005), I do not claim that my treatment is entirely consistent with all the theory's statements, especially with its multifaceted attempt to give biological and evolutionary foundations to the Rawlsian social contract. But it is the best way for me to make sense of it, and to put it at the basis of my own revision of the theory of constitutional choice on corporate governance structures. Even if reference could be made to many of Binmore's papers and books, and especially to his first paper 'Game Theory and the Social Contrac't (1984), I will confine my references in this section mainly to the last one (Binmore 2005).

[2] For an example, in the case of the repeated trust game see fig. 2 part I.

[3] For a detailed exposition of how the dogmas of the overriding ness of welfare maximization and efficiency over fairness permeate all the economics of institutions, see Kaplow and Shavell (2002).

[4] see Kaplow and Shavell (2002).

[5] see Kaplow and Shavell (2002).

[6] see op. cit. pag. 78.

[7] The ex post rationality of the Nash equilibrium – implied by the notion of common knowledge – was already clear in Lewis (1968), who also suggested that an agreement could give an empirical explanation of how a state of common knowledge could emerge. He, however, focused on the different cognitive phenomena of salience. On the game theoretic definition of common knowledge, see Binmore and Brandeburger (1990) and Kreps (1990); on the epistemic logic of common knowledge, see Fagin, Halpern, Moses and Vardi (1996).

On the selction of Nash equilibria based on common knowledge of the unique solution see Harsany and Selten (1988).