

# INCOMPLETE CONTRACTS AND CORPORATE ETHICS: A GAME THEORETICAL MODEL UNDER FUZZY INFORMATION(\*)

Lorenzo Sacconi<sup>1</sup>

## 1. Introduction and Motivation

Developing a corporate code of ethics amounts to something like playing the role of a “constitution designer” on small scale. It is an experiment of rational decision about the general and abstract norms that have to work as constitutional constraints on a “corporate actor” - i.e. a firm seen as a micro-constitutional order (Coleman 1990, Vanberg 1992). According to empirical surveys in 1980 only 8% of American largest companies (the *Fortune 1000* list of industrial and service corporations) had a code of ethics. Afterward the phenomenon has been very fast growing in the US. According to surveys performed in 1985/86 and 1990/92 on the *Fortune's* first 1000 US firms, those endowed with a code of ethics were respectively 77% and 93%. Although it is less extensive, nevertheless the phenomenon is clearly recognisable also in Europe. A research performed in 1988 – the CEOs of the 600 largest industrial companies in RFT, GB and France was asked to answer a questionnaire and the level of acceptance was around 30% - concludes that at that time code of ethics was present in 51% and 41% of the RFT and GB companies respectively. The very fast development of the field of business ethics in Europe in the last decade allows us to say that these levels has certainly grown in the meantime. Best structured codes of ethics both in US and Europe clearly reflect the idea of corporate responsibility towards all the firm's stakeholders, and are organised into separate chapters defining the corporate fiduciary duties towards shareholders, customers, employees, suppliers, government agencies, competitors, local communities, political representatives etc.

Corporations allocate to their corporate governance structures authority over a large part of the transactions they carry out, both regulated by (incomplete) labour contracts or by (incomplete) arm's length contract with suppliers, customers, partners, capital lending organisations. These contracts are incomplete *per se*, but as the occurrence of unforeseen contingencies is anticipated, they are completed by residual rights of control allocating discretion upon ex ante non-contractible decisions to one party in the contract. Ownership and control structures respond to a need of minimising some transaction costs. But they also admit the overreaching risk of abuse of authority. This is true in particular when many parties involved in transactions make specific investments and face sunk costs. As a consequence corporations tend to be surrounded by the

fear of abusing their power and by distrust on the part of those stakeholders that in principle might interact with them in the perspective of mutual gain. If this distrust does not end up with the collapse of these economic institutions, which on the contrary are overwhelming in the contemporary economies, it must be due to norms and institutions *other than* residual rights of control, which constrain their abuse. Both empirical investigation and theoretical deduction from the theory of firm suggest that corporations need systems of self-regulatory norms of behaviour like codes of ethics exactly because of the ‘abuse of authority’ problem (Kreps 1990, Sacconi 2000).

Let me state three main requirements for a code of ethics may play the role of a rational constraint on abuse of authority:

- a) It must reflect a reasonable and acceptable agreement amongst the corporate stakeholders about how the surplus produced by their joint co-operation will be shared;
- b) It must work as a self-enforcing social norm, as it should not be meant as a legal expropriation of the residual rights of control seen as a (second best) efficient economic institution;
- c) It must answer the question how who holds authority in the firm may undertake commitments over events and situations that cannot be *ex ante* contractible or describable within contracts or detailed regulation. Such situations make contracts incomplete in the very sense that parties *a priori* do not even have an idea of the possible states that should be used in order to condition obligation or concrete commitment. Under such situations commitments tend to be empty, simply because they are mute regarding those states that were *ex ante* unknown.

The first two points are quite naturally amenable to a game theoretical analysis. Cooperative bargaining games provide the natural theoretical tool for understanding the social contract amongst the stakeholders of the firm, who are essential to the joint production of a surplus (Harsanyi 1977, Brock 1979, Gauthier 1986, Binmore 1991, Sacconi 1999, 2000). Moreover, non cooperative game theory provides games of reputation as the natural way of modelling codes of ethics as social norms that are be put in practice only because they induce endogenous incentive to compliance, as they satisfy the conditions for the existence of a Nash equilibrium (Binmore 1991, Sacconi 2000).

However, the third point – how undertaking significant commitments under unforeseen contingencies – constitutes a serious challenge to current game theoretical modelling (Kreps 1990). Our understanding of incomplete knowledge – the label under which this point should be analysed – must go further than treatments where unforeseen contingencies are seen as a sort of statistical uncertainty over a qualified subset of states amenable to standard Bayesian modelling (Tirole 1999, Felli, Anderlini 2000, Al-Najjar, Anderlini, Felli 2000). What is needed are models of rational decision under situation such that (i) players are aware of the possibility of generic unforeseen contingencies (because of the limitation of their linguistic resources), but (ii) they are also *ex ante* unable to figure out each of them. Thus the model of player’s reasoning must not assume that the complete and exhaustive set of the possible states of the world is already resident in the back of his mind.

This paper suggests a first, very rough and tentative, modelling of unforeseen contingencies in terms of fuzzy sets theory (Zadeh 1965, Zimmerman 1991). Our incomplete knowledge about unforeseen contingencies is captured by defining some events as fuzzy sub-sets of the set of unforeseen states of the world - which are states that admit *ex post* only an incomplete and vague description in terms of the concepts

formulated in our *ex ante* language. Such events correspond to terms of the language the players were able to use before coming to learn about the unforeseen states. By these terms we define the domain of general abstract norms, i.e. abstract descriptive characteristics that have to be satisfied in order a moral principle may be applied. Due to their generality, abstractness and universalisability, these terms admit to be at least in part satisfied by many situation, even though these don't share many details that we *ex ante* used to describe as belonging to their domain. The const of their all-encompassing nature is vagueness. Membership into the set defining the domain of a general, abstract, principle of ethics is a matter of degree and this opens the route to defining domains of general norms as fuzzy sets of unforeseen states of the world.

This is only a first step however. Then a model is needed of the reasoning the players are able to perform given this incomplete (vague) base of knowledge, in order to jump from vague premises to what it has to be done in any unforeseen contingency. A second tentative modelling of players' inferences in terms of the logic of default reasoning (Reiter 1980, Ginsberg 1987) is then suggested. The intuition is that, given incomplete knowledge that does not allow refuting the "normal course" of things, we extend our belief that the "normal course" does in fact hold also to the unforeseen situations. As a consequence I can define the commitments that a player conforming to the code is expected to carry out. This allows replicating within the new context of unforeseen contingencies some well-known results in the theory of reputation games (Fudenberg and Levine 1989, 1991)

## 2. The Hierarchical Transaction as a Game

In order to explain how a set of general and abstract principles and norms - call it a code of ethics - may complete the holes of an incomplete contract (Coleman 1992), we need to define the context of strategic interaction under which a transaction takes place between two or more parties (the players). Because the contract is incomplete, some decision variables contingent upon unforeseen events are *ex-ante* uncontractible. Thus, them are left under control of some party in the *ex-post* perspective - that is when the acceptance of the contract is over and some important decisions within the relationship will have already been taken. Such party will exploit them in the resulting *ex-post* bargaining in order to renegotiate the incomplete contract. Because this may generate transaction costs and inefficiency, an institutional design of the residual right of control over these *ex ante* un-contractible decisions variable is in order. The resulting situation is one where some transaction occurs under an incomplete contract cum residual rights of control (authority) over the *ex ante* un-contractible decisions allocated to one of the parties in the contract. This is what we call a hierarchical transaction (Williamson 1986, Grossman, Hart 1986, Hart, Moore 1988, Kreps 1990). A Game theoretical model of the hierarchical transaction is our starting point.

### 2.1. A Hierarchical Transaction

Consider two parties who may undergo to a contract in order to carry out a hierarchical transaction. Party A may enter or not into a relationship of dependency with a second player B. This means that, when A has entered the relationship with B, he will in fact accept to perform some *ex-ante* unspecified task that B will

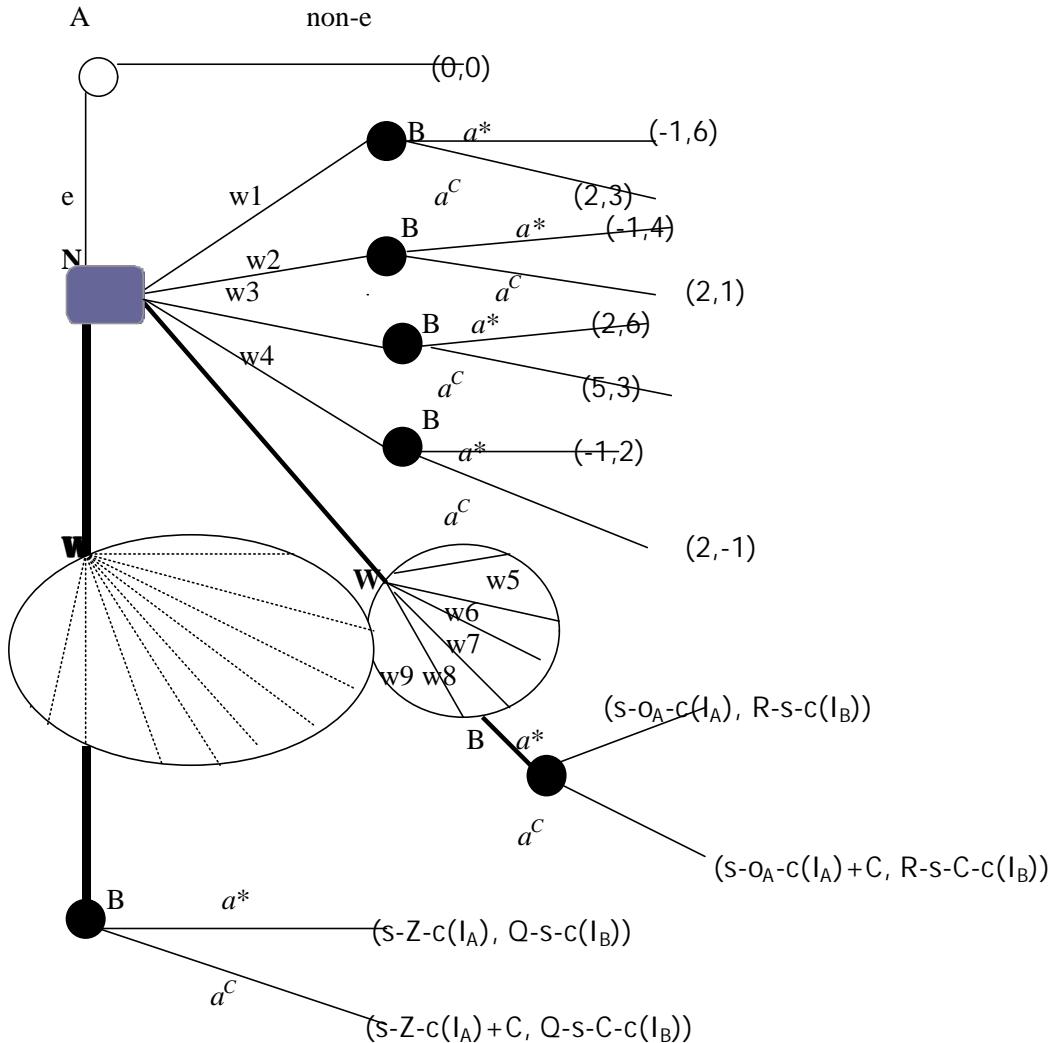
discretionarily order him to carry out later on, against some pre-established payments. If A chooses ‘enter’, thus, a relationship of authority is established. If ‘not’ is chosen the relationship is ended. In the course of the relationship between A and B unforeseen events can arise. They render the explicit contract, on the basis of which A decides to enter into a dependent relationship with B, ‘mute’. The contract is ‘mute’ because it cannot contain explicit and concrete provisos contingent on the occurrence of the unforeseen events. This is exactly why the contract inherently builds up an authority relationship: B will be enabled to decide ex post discretionarily what to command A to perform, without being any specific decision automatically sorted out by the ex-ante established provisos of the contract. But now A becomes afraid that B will abuse of his authority. Moreover he realizes that an abuse, if it were to come, would not be easily checked on the basis of any existing contractual rules or provisos. The problem is: *why A should ‘trust’ B and ‘enter’?*

## 2.2. The Game in Extended Form

The hierarchical transaction is depicted in fig. 1 as a game in extensive form. Player A (in short ‘A’) moves first. A has to choose out of a decision-set containing ‘entering’ ( $e$ ) or ‘not entering’ ( $non-e$ ) into an authority relationship with player B (in short ‘B’). By choosing  $e$  A surrenders to B the right to decide at a later moment in the game one action within the set  $\mathcal{A} = \{a^*, a^f\}$ . By choosing  $non-e$  the game is ended. If A chooses to enter, he incurs a specific investment  $I_A$  with a fixed sunk cost  $c(I_A)$ . If A enters, as a consequence B also incurs a specific investment  $I_B$ , with fixed sunk cost  $c(I_B)$ . Notice that in order to keep things simple we do not take these investments as separate strategic decisions within the game. They are assumed as the simple mechanical consequence of the only strategically relevant A’s choice to enter. However they shape the game in terms of the payoff-consequences of the strategic decisions ( $e$  or  $non-e$ ,  $a^*$  or  $a^f$ ) the players undertake. In fact, specificity of investments is meant to signify that

$$V(e, a, I_i) - c(I_i) > r, \text{ for } i = A, B$$

where  $r$  is the net benefit of any action  $a$ , given the investment  $I_i$  when it is taken outside the relationship between A and B and  $V(e, a, I_i)$  is the benefit of the action  $a$  ordered by B and carried out by A, given  $I_i$ . That is, with respect to any other transaction outside the A and B’s relationship, there is a surplus attached to completing the transaction through the couple of actions ( $e, a$ ) within the relationship between A and B after the investment  $I_i$  as been made. The global amount of wealth created by transacting between them, after the specific investment has been made, is higher than what could come out of transactions with any agent outside the relation.



(Fig .1. A game of hierarchical transaction with unforeseen contingencies)

Taking the utility function of the players linear in the money, this would also be true if measured in terms of each player's payoff functions. Moreover we assume that the relationship between A and B is symmetrically cooperative, in the sense that the investments of both of them contribute additional value to the surplus:

$$S = V(e, a, I_A, I_B) - c(I_A, I_B) > V(e, a, I) - c(I),$$

where  $I_i$  is the investment of A or B but not both. If A enters, he will receive a fixed salary

$$s - c(I_A) = p \geq 0$$

where  $p$  is a constant. We assume that A's fixed salary covers the cost of A's specific investment but possibly *not* the cost of the variable levels of effort that are requested by the two actions that B can ask A to perform contingent over what may occur during the game.

At move two in the game tree Nature chooses a state of the world out of two sets  $W$  and  $\Omega$ . States in  $W$  are defined as follows: each  $w_i \in W$  is one of the alternative, jointly exhaustive, possible descriptions of the world, describable by a given set of linguistic resources  $\mathbf{L}$ .  $\mathbf{L}$  is a formal language made up of

- N predicative letters  $P_1, \dots, P_n$
- M variables  $x_1, \dots, x_m$

- N individual constants  $q_1, \dots, q_n$ ,
- The usual logical connectives ( $\&$ ,  $\vee$ ,  $\neg$ ,  $\Rightarrow$ )
- All the formulae that can be generated by operating with the connectives and the rules of inference.

Formally said, each possible world  $w_i$  is one of the maximal sets of (well formed) formulae of **L**, that is one set of formulae that cannot be increased by any formula without introducing a contradiction in the conjunction of the formulae belonging to it. To exemplify, given that we have M free variables and N predicative letters, one state  $w_i$  will *first* consist, for each free variable, of the conjoint affirmation of as many of the N predicates as it doesn't imply a contradiction, and *second* of the conjunction of all of them over the set of M variables. Take U to be the domain of interpretation of the language **L**. Each state of the world  $w_i$  made up of one possible maximal set of terms in **L**, can be seen as one possible way of describing, as completely as possible by means of **L**'s linguistic resources, any object belonging to U. It is an alternative description to any other state  $w_j$  because for at least one variable, interpretable in the domain U,  $w_i$  affirms a predicate which is negated by other states  $w_j$ . To say differently, any state  $w_i$  is an alternative description of U within **L**. Within this framework we can define an event E that a given object has a certain property j as the subset of possible states such that each of these states, as expressed in **L**, affirms the predicate  $P_j$  about the individual variable  $x_i$  corresponding to the given object in U.

Let me now introduce the less conventional assumption that before playing the game players are *aware* that the set of possible states included in W may be 'incomplete'. By this I mean that they are *aware* that the language **L** is limited and some properties could eventually be exemplified by events that may occur, which cannot be exactly described by predicates included in **L**. Thus, the domain U will be 'completely' described by the states in W *only relatively* to what may be expressed through the resources of **L**. In general, however, the domain U will contain some properties that the language **L** will not completely account for, so that any  $w_i$  will not affirm nor negate these properties. However, let assume that the players are aware of that, but that they are not even able *ex ante* to say *which* properties these are nor *even to imagine them*. This means that the players are aware that, beyond the set W, there also may exist a set  $\Omega$  of ex ante unspecified and undermined states, whose elements contain some properties that are incompletely describable in terms of the existing language **L**. For any  $\omega_i \in \Omega$  players will be ex post able to find out some free variable, interpretable in U, such that they cannot exactly say whether some predicate in **L** is true or false for it. The reason is that predicates in **L** are inadequate for describing the unexpected proprieties of the state  $\omega_i$ , and this authorize us to call it an 'unforeseen state of the world'. Eventually they will enrich their language by new terms. But, as far as the current linguistic resources are concerned, players cannot find in **L** predicates that describe exactly what happens under  $\omega_i$ .

While in the context of W an event E is defined as the subset of all the states  $w_i \in W$  in which a certain property is definitely true, on the contrary in the case of the elements of  $\Omega$  an event E may be defined as the subset of those states  $\omega_i \in \Omega$  in which the given property can be affirmed only up to some degree. Being able to hold expectations only about events that they are able to formulate in **L**, there are not precise expectations on unforeseen events. Unforeseen events are things we cannot speak about directly within **L**. We may infer they existence from the fact that some ex ante known events appear with some vagueness under states in  $\Omega$ .

We are now able to define the state-set  $\Omega$ . States  $\omega_i \in \Omega$  are possible descriptions of our domain  $U$  other than  $w_i$ , containing properties that *ex ante the players are not in a position to anticipate, and that ex post they are not able to describe with fully precision by means of the linguistic resources included in  $\mathbf{L}$* .

At the 3<sup>rd</sup> move in the game tree player B chooses. B's decision consists of selecting one action to be ordered out of the set  $a = \{a^*, a^c\}$ , assuming that Player A, once entered, will perform whichever order B will have selected. (Remember that under this model, if B chooses a given action, then A will necessarily perform it later on. This permits us to focus on the unique strategic variable in the hand of A - deciding to enter or not the authority relationship.) Let define the two alternatives:

- action  $a^*$  maximises the overall cooperative benefit  $V(e, a, I_A, I_B) = R$  and takes entirely the return  $R$  net of the costs already sustained by B (that is  $c(I_B)$  and  $s$ );
- action  $a^c$  is equally efficient than  $a^*$ , but it also provides for an additional fixed compensation  $c$  to A.

Actions  $a^*$  and  $a^c$  represent variable costs to A as a function of the effort  $o_A$  spent by A on performing the required task. Without any loss of generality, I call  $o_A$  the cost of effort. In turn the cost  $o_A$  is conditional on the state of the world selected by Nature, so that the same action under different states may involve high, medium or low effort cost. Anyway B's choice may disregard these costs. This implies that, under some state of the world selected by Nature, B may choose an action that represents to A a cost hardly covered by the fixed salary  $p$ .

### 2. 3. Payoffs Under Foreseeable States

Consider now the players' payoffs under the case of foreseen states of the world. Remember that we assume that players have linear utility function in the monetary outcomes, so that in the game under  $W$  players A and B have the following payoffs functions respectively:

$$u_A(e, a^*, w_i) = s - c(I_A) - o_A; \quad u_A(e, a^c, w_i) = s + c - c(I_A) - o_A \\ u_B(e, a^*, w_i) = R - s - c(I_B); \quad u_B(e, a^c, w_i) = R - s - c(I_B) - c.$$

To find out the effective payoffs, it is necessary to go back to the states of the world on which they depend. Assume that for every  $w_i$  it is known by the players that:

- specific investments made by A and B are jointly indispensable to the production of the maximum surplus, and the resulting cooperative return is superadditive, that is

$$V(e, a, I_A, I_B) \geq V(e, a, I_i), \quad \text{for } i = A, B$$

$$V(e, a, I_A, I_B) \geq V(e, a, I_A) + V(e, a, I_B)$$

- parties make their specific investments  $I_i$  without free riding one another.

Next assume that there are three possible descriptions of "maximum return"  $R$  (high, medium and low) conditional on some characteristics of the states

$$R = \{R_+, R_{\gg}, R_-\}.$$

To say it differently, the state space  $W$  is partitioned in three events as far as maximum return is concerned. Similarly, assume that each action  $a$  may produce three possible levels of effort's cost (high, medium and low) contingent on some characteristics of the states.

$$\theta_A = \{\theta_A+, \theta_A\gg, \theta_A-\}$$

As above, the state space  $W$  is partitioned in three elements (events) as far as effort's cost is concerned. In sum, the language of our theory is able of describing nine states of the world by multiplication of the two set of events:  $w_1 = (R_+, o_{A+})$ ,  $w_2 = (R_{\gg}, o_{A+})$ ,  $w_3 = (R_+, o_{A\gg})$ ,  $w_4 = (R_-, o_{A+})$ ,  $w_5 = (R_{\gg}, o_{A+})$ ,  $w_6 = (R_{\gg}, o_{A-})$ ,  $w_7 = (R_+, o_{A-})$ ,  $w_8 = (R_-, o_{A\gg})$ ,  $w_9 = (R_-, o_{A-})$ . Because we do not want to concentrate here on the intrinsic vagueness of qualitative term as 'high', 'medium' and 'low', whereas we will be concerned in the next section with modelling vagueness introduced by unforeseen states of the world, we take some cardinal values as representative of the three classes. Take for example

$$R_+ = 12, R_{\gg} = 10, R_- = 8,$$

$$\theta_{A+} = 5, \theta_{A\gg} = 2, \theta_{A-} = 0$$

The additional compensation  $C$  is a two-valued variable conditional on player's A action. I take it to be  $C = 3$  if  $a^C$ ,  $C = 0$  if  $a^*$ .

Last, let the constants be

$$s = 5, c(I_A) = c(I_B) = 1$$

Payoffs can be then calculated according to the form of the utility functions given above, considering that nine states may occur under the choice of both actions  $a^C$  and  $a^*$  (in fig.1 only the payoffs associated to states  $w_1 = (R_+, o_{A+})$ ,  $w_2 = (R_{\gg}, o_{A+})$ ,  $w_3 = (R_+, o_{A\gg})$ ,  $w_4 = (R_-, o_{A+})$  are depicted.)

I will confine my discussion of the solution to this particular example for more concreteness, but result may be easily generalised. If A 'enters', B's dominant action is always to choose  $a^*$  without more compensation. In all the situations in which the burden  $\theta_A$  to A of the action  $a^*$  chosen by B is high, this implies that player A, if she 'enters', obtains the payoff (-1). Consequently in these cases ( $w_1, w_2, w_4$ ) A's best reply is *non-e*.

Contingent upon the states now considered, the game is the traditional *sequential 'Prisoner's Dilemma'* (PD game): the only feasible outcome is the sub-optimal Nash equilibrium (*non-e*,  $a^*$ ). In the remaining cases A's best reply is always *e*, with the Nash equilibrium outcome (*e*,  $a^*$ ).

### 3. The Game Under Unforeseen Contingencies

The aim of this section is to consider the interaction between the two players when Nature selects a state from the state-space  $\Omega$ . We show that, if the game is meant as one-shot, the solution is not only ex ante, but also ex post undetermined and ambiguous. In order to see that, first ask "what do the players know about the game under  $\Omega$ ?" *Ex ante* (before Nature makes its choice) independently of the state  $\omega_i$  chosen, they know that a generic return  $Q$ , described in terms of monetary gains, will follow (without loss of generality we can also assume that players know that  $Q$  may take its values within a specified monetary interval, for example [0,20]). Moreover a generic action, in terms of a number  $Z$  of hours to be worked, will be chosen by player B (as before, we may assume that players know that  $Z$  lies within a definite time interval, for example [0, 8]).

At the same time they are aware that, if nature selects any state  $\omega_i$  out of  $\Omega$ , this will imply the occurrence of events that they are not able to specify ex ante. In term of what they are able to express within the existing language, in each state  $\omega_i$  there will be ambiguity about:

- (i) *The joint nature of the return:* Is it still true that  $Q$  is the output of the joint cooperation of both players carried out through their simultaneous specific investments and actions (as it was true under states belonging to  $W$ )?
- (ii) *The size of the cooperative benefit:* How much, if any, of the observable  $Q$  could be attributed to such investments and how much could be due to exogenous change of the environment? (The employee could result not being necessary to the realization of surplus or, on the other hand, B may become superfluous.)
- (iii) *The existence of externality:* Are specific investments  $I_A$  and  $I_B$  in effect the fruits of independent decisions, or some of the players free rides the other's investment?
- (iv) *The size of the burden of B's orders to A:* What does any value of  $Z$  mean in terms of effort's cost of the action asked to A by B? (For example some unforeseen technological change would transform the burden of effort to a source of enjoyment.)

We must keep in mind that in any state  $\omega_i$  there will be ambiguity about (i)–(iv) not because there is some uncertainty concerning which state occurs – so that payers cannot separate states where (i)–(iv) are true from those where they are false. On the contrary, in *each*  $\omega_i$  it will be vague whether the propositions (i)–(iv) hold. This is so for players are incompletely able to describe *each*  $\omega_i$ , i.e. the characteristics required to check whether (i)–(iv) hold or not are not clearly specified in each  $\omega_i$ .

Ex post they learn the exact nominal value of the occurring  $Q$  and  $Z$  and the precise state  $\omega_i$  selected out of  $\Omega$  (and its possible alternatives). But what were ex ante true will still be true ex post. The player are ex post aware that in the presence of unforeseen states certain pieces of information remain vague, that is proposition (i)–(iv) are not definitely true or false in the particular state  $\omega_i$  selected out of  $\Omega$ .

As a matter of consequence the solution of the game under unforeseen contingency is indeterminate also in the ex post perspective. In spite of  $Q$  and  $Z$  being known ex post, the payoff functions have not determinate values conditional on states like  $\omega_i$ . Players are unable to understand their joint and separate contributions to the production of the surplus, whether each of them has effectively made a specific investment, whether the investment made by one free rides the other, and moreover how costly any action asked by B to A is in terms of A's effort costs. Understandably this will influence the players' incentives (Tirole 1999). This translates in the ambiguity of the two basic variables  $R$  and  $oA$  needed for specifying the players' payoff functions.

#### 4. Solution Theory: Social Contract and the Ethical Code

Now, take the perspective of a repeated game of reputation amongst a long-run player B and an infinite series of short-run players taking in turn the role of player A. At each repetition the stage-game just defined in the sections above will be played again and again (Kreps 1990, Fudenberg and Levine 1989, 1991). The repeated game gives an elementary model of what we may call a *firm*, that is a series of hierarchical transactions amongst an institutional hierarchical superior and a potentially infinite number of employees that

last potentially *ad infinitum*, i.e. even beyond the life-span of a single individual owner – what seems at least necessary in order to understand the model as speaking about an institution like the firm. Can this change of perspective influence the solution? It is apparent that the bulk of the argument rests on the capability of player B to undertake commitments over his possible two strategies, which become “types” in the eyes of the players successively taking the A’s role. In such a way these may function as the basis for the reputation effects mechanism. More precisely, what is needed is that player B may undertake commitments such that

- *Under the case W:* even though the prior probabilities were largely concentrated on states  $w_i$  with  $a_{A+}$  (high effort-costs), the sub-optimal outcome (*non-e,a\**) is avoided nevertheless.
- *Under the case W:* reputation also supports rational “entrance” by player B under the occurrence of the unforeseen states  $\omega_i$ .

The second requirement is of course the most demanding. The point is that a commitment is normally understood as a conditional strategy, which can be announced in advance to any play of the game, stating which of the B’s actions will be selected contingently on the occurrence of any possible world. But there is no way for B to announce a commitment conditional on unforeseen states of the world, as he cannot describe the concrete conditions under which a particular action will be undertaken. This problem could be by-passed by adopting the same action in all the unforeseen states whatsoever – but this would amount to an inefficient *unconditional* commitment.

The *solution* here suggested is resorting to the *Constitutional Contract* of the firm as expressed by two types of norms:

- *With respect to W:* an *explicit incomplete contract* which establishes what must be done contingently upon a restricted set possible worlds belonging to W, and leaves B full discretion in relation to the remaining states in W.
- *With respect to  $\Omega$ :* a code of ethics containing
  - (i) a set of explicitly stated constitutional, general abstract principles,
  - (ii) a set of interpretative rules that establish under what conditions a situation falls into the domain of a principle;
  - (iii) the conditions for admissibility of  $a^*$  or  $a^C$ .

Both types of norm are based on a unique abstract principle for solving games (i.e. a solution concept) that I identify with the *Nash Bargaining Solution (NBS)* for cooperative bargaining games (Nash 1950, Harsanyi 1977). For our game is non-cooperative, *NBS* is not the obvious solution concept to be used here. It must be understood as a normative principle which does not simply follow from the individual rationality assumptions for non-cooperative games, and is meant as a constitutional rule purposively *ex ante* deliberated in order to give shape to Player’s B (the one in the position of authority) commitments. Suppose the players are able to posit themselves under an *ex ante* hypothetical stand point – the Archimedean point (Gauthier 1986, Binmore 1991) - from which they agree on abstract principles that will regulate their hierarchical transactions in general. This is the hypothetical bargaining position from which they agree on the social contract of the firm. I assume that in such a hypothetical position they would recognize the *NBS* as the general abstract principle for solving whatever strategic interaction among them may involve the joint production of a surplus (Sacconi 2000). Moreover player B would accept (at least *ex ante*) to commit himself to follow this principle in any of the ensuing hierarchical transactions, where the joint production of a surplus

is involved. Thus, *NBS* is the solution to which the parties *would converge*, if they *were* able to reason according to the ‘*as if*’ format. Adopting *NBS* amounts to hypothesising that the player B will behave in the game – *in fact* a *non-cooperative* game – ‘*as if*’ it were a *cooperative* one. Last, in the context of the social contract, this solution must be calculated on the basis of a *fair status quo*, which establishes the maximum level of concession of the parties without recourse to the use of force or fraud, threat or parasitism by one party on another (Gauthier 1986).

Take first the game under  $W$ , where the solution concept may be translated into a concrete incomplete contract. Computation of *NBS* is straightforward in this case. *Pareto Optimality* is guaranteed by both  $a^*$  and  $a^c$  by assumption. Parties must obtain payoffs at least equal to what they gain from the fair status quo  $d$ . Given utilities linear in the monetary payments, the status quo payoffs are

$$u_A(d) = c(I_A) + o_A$$

$$u_B(d) = c(I_B)$$

The variable  $o_A$  figures in A’s outside option since otherwise A would leave the bargaining table in a worse condition than the one in which she joined it. *NBS* requires

$$\text{Max } \mathbf{P}_h (u_h(V(e,a,I_A,I_B) - u_h(d))), \quad \text{for } h = A, B$$

The players’ payoffs are given by participation in the cooperative return, respectively

$$u_A(V(e,a,I_A,I_B)) = s + c$$

$$u_B(V(e,a,I_A,I_B)) = R - s - c$$

Nash’s Bargaining solution therefore requires that one action  $a \hat{\rightarrow} a$  is chosen so that

$$\text{Max } [(s + c) - (c(I_A) + o_A)] \wedge [(R - s - c) - c(I_B)]$$

The cooperative return is  $R = V(e,a,I_A,I_B)$  and the surplus net of the *status quo* is

$$R - (c(I_A) + c(I_B) + o_A)$$

Were  $c$  a continuous variable, the solution would be choosing  $c$  such that

$$s + c = (c(I_A) + o_A) + \frac{1}{2} [R - (c(I_A) + c(I_B) + o_A)]$$

$$R - s - c = c(I_B) + \frac{1}{2} [R - (c(I_A) + c(I_B) + o_A)]$$

Given that the possible values of  $c$  contingent on the action  $a$  are only two, it follows that for all the states from  $w_5$  to  $w_6$  the *NBS* is satisfied by the action  $a^*$ . In states  $w_1$ ,  $w_2$ ,  $w_3$ ,  $w_4$  *NBS* requires that  $a^c$  be chosen. When Nash product is in tie under  $a^*$  and  $a^c$ , as it happens in  $w_4$ , I assume that action  $a^c$  must be used (this amount to say that the constitutional contract is biased in favour of player A, who must be convinced to enter the relationship.)

An incomplete contract, with discretionary power being given to B, therefore consists of a simple function that permits action  $a^*$ , provided that states of the world with respect to which it is explicitly prohibited do not occur:

$$f(a^* | w_i) = \begin{cases} 0 & \text{if } w_i = w_1, w_2, w_3, w_4 \\ 1 & \text{otherwise (that is for every other } w_i \in W) \end{cases}$$

Leaving B free to act according to his best reply in cases other than  $w_1$ ,  $w_2$ ,  $w_3$ ,  $w_4$ , the contract of delegation ensures that the solution is always in line with *NBS*. This contractual rule may be employed to

specify player's B conditional commitments. Compliance with them is observable in any  $w_i$  state of the world. As a consequence, player A may rest on B's reputation effects related to his conformity to these commitments in order to trust B and enter.

## 5. The Role of 'Vague' Constitutional Principles

In the case of  $\Omega$  it is not possible to calculate ex ante the *NBS* for each possible state of the world. Why should a code of ethics succeed where the explicit contract fails? *NBS* is only introduced here as a general abstract concept of solution, like a constitutional principle without any concrete reference to the particular states under which the game is played. Rather than ex ante calculating the Nash product for each possible state, mere abstract characteristics are stated as descriptive pre-conditions for a game can be submitted to the *NBS*. These are displaying *in the appropriate way* (i) the existence of a cooperative surplus and (ii) a fair *status quo* meant as the costs the parties independently undergo for making their investments and actions. Ex post, when unforeseen contingencies will have been brought about, we will be able to see whether the occurred state satisfies the requirements asking for a *NBS*-constraint over player B's choice.

In fact, general principles, *formulated in universal abstract terms* of the current language **L**, are terms of reference both for *ex ante foreseen* and for *ex ante unforeseen* or simply 'new' states of the world. The form of reference is a *membership function*. Due to generality, universality and abstractness of the constitutional principles, any state of the world, even though unforeseen, belongs to the domain of a principle at least *to some degree* (between full membership and nil). Moreover the requirements on the appropriate membership functions can *always* be specified, both in the ex ante and ex post perspective, as they are not contingent upon the description of ex ante unforeseen contingencies. Thus, by means of its general principles, the code of ethics *is never mute* in relation to events, including events that are *ex ante* unforeseen.

There is of course an important price in terms of vagueness to be paid to this all encompassing property of general and abstract principles. Unforeseen events in particular introduce vagueness in the membership relationship. I take for granted, even if this can be not always the case, that we mostly are able to clearly adjudicate a state of the world to the domain of application of a general principle as far as this state can be completely described within the same language by which also the principle is formulated. The hypothesis is that most vagueness is associated to unforeseen facts, because of the imprecision of the language **L** in ex post describing events that were *ex ante* unforeseen.

To state the point analytically, encode by P an *abstract constitutional principle* and by E the event 'the constitutional principle P is fulfilled'. The event E is equivalent to summing up all the *descriptive* features required to fulfil the principle. The statement is that every unforeseen state can belong to the event E. Simply, this belonging must be defined in terms of *vague* membership. This allows for a fuzzy set modelling of unforeseen contingencies.

Take an ordinary set of objects  $\Omega$ . Intuitively mean  $\underline{E}$  as a fuzzy subset of  $\Omega$  such that the elements of  $\Omega$  are member to  $\underline{E}$  *up to a certain degree*. Degrees are given by membership functions  $\mu_{\underline{E}}(\omega_i)$  with domain the set  $\Omega$  and co-domain the real line [0,1]. Thus a fuzzy set  $\underline{E}$  is the set of ordered pairs such that any element of the given reference set  $\Omega$  is associated to the membership function assigning the degree up to which it shares the property referred to by the fuzzy set.

$$\underline{E} \equiv \{(\omega_i, \mu_{\underline{E}}(\omega_i)) \mid \omega_i \in \Omega\}$$

Consequently it may be written

$$\underline{E} = \{(\omega_1, \mu_{\underline{E}}(\omega_1)), \dots, (\omega_n, \mu_{\underline{E}}(\omega_n))\}$$

Let the set  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  be the state-space, i.e. the set of all the possible alternative descriptions of the state of the world and let  $P(\Omega)$  be the power set of the subsets of  $\Omega$ , which is usually meant to be the collection of all the possible events, meant as subsets of  $\Omega$ . The bulk of our approach to unforeseen contingencies is the following proposition:

*Let  $\Omega$  be the set of unforeseen states of the world. Then its power set  $P(\Omega)$ , defining the set of all the possible events, is the collection of the fuzzy subsets of  $\Omega$ , which are derived by associating to any  $\omega_i \in \Omega$  its membership function to any fuzzy set representing any given event. Events defining the domain of application of universal, abstract a general principles formulated in a given normative language  $\mathbf{L}$  induce membership functions with positive (not zero) degrees of membership for any state  $\omega_i \in \Omega$ .*

Under this interpretation the elements of any fuzzy set  $\underline{E}$  are meant as alternative descriptions of the state of the world, which are unforeseen *ex ante* and only partially describable *ex post*. An event  $\underline{E}$  denotes the existence of an unforeseen contingency as it corresponds within  $\Omega$  to properties that only vaguely can be traced back to the event  $\underline{E}$ . As a result, when the state space consists of unforeseen states of the world, its power set (the set of events) consists of fuzzy sets, i.e. vague events.

True, I do not characterise unforeseen states as such directly in fuzzy terms. What is properly modelled as fuzzy set are the events we are able to express in our exiting language, whose members are unforeseen states of the world. By understanding unforeseen states through their imperfect membership function into fuzzy sets, we capture the impact of their being *ex ante* unforeseen on the precision and completeness of their *ex post* description by means of the existing language. To say it differently, being *ex ante* unforeseen implies vagueness in the *ex post* description, due to the inadequacy of the *ex ante* language  $\mathbf{L}$ .

We are now able to draw important conclusions about the availability of commitments in the repeated game played under unforeseen contingencies. Although the player is not *ex ante* in a position to describe the set  $\Omega$ , he is nevertheless *aware* of its eventual existence. He cannot undertake commitments contingent upon concrete descriptions of elements in  $\Omega$ , for these are unforeseen. *But* he can undertake commitments regarding the occurrence *ex post* of a *certain membership relationship* between any generic element of  $\Omega$  and the set  $\underline{E}$ . i.e about the degree up to which any generic  $\omega_i$  is a member of  $\underline{E}$ . Notice that *ex ante* the player does not need any description of the particular  $\omega_i$ . He only needs to be aware that a generic  $\omega_i$  may be selected out of  $\Omega$ , to know the set of conditions compounded in  $\underline{E}$ , and to establish the requirement that *if* any generic  $\omega_i$  would belong to  $\underline{E}$  at least up to some given degree, then the principle P (the *NBS* in our case) is to be carried out *ex post*. Thus B undertakes commitments on the basis of principles and criteria that can be understood in the *same terms* both *ex ante* and *ex post*. The occurrence of unforeseen contingencies no more creates any ambiguity and non-observability problem concerning compliance with them.

## 6. First Step in the Ethical Procedure of Deliberation: Vagueness of the Constitutional Principle

In this section I begin defining operationally the code of ethics as a procedure of deliberation able to decide for every state, included the unforeseen  $\omega_i \in \Omega$ , whether *ex post* a given action (for example  $a^*$ ) is permissible in the light of a constitutional principle established *ex ante*. It fits the intuition of “procedural rationality” as the decision is the output of a reasoning process addressed to respond to the decision maker’s limited knowledge of the alternatives (Simon 1972). The procedure of deliberation contains many steps in order to decide whether:

- in each state  $\omega_i$  occurs facts such that it is a situation (call it ‘event  $E1$ ’) to which *NBS* does apply;
- in the same state  $\omega_i$  occurs facts such that , assuming that B chooses  $a^*$ , then the outcome is *efficient/fair* (call it ‘event  $E2$ ’).
- given any  $\omega_i$ , the permissible actions is  $a^*$  or  $a^C$ .

First a better understanding of the event  $E1$  is needed. Let  $\mathbf{G}$  be the game that I have just described in sec.2 and 3, and let  $G_c$  be the codification for the conditions required for a game can be treated as one element of the class of cooperative bargaining games with fair status quo.  $E1$  stands for the set of states where it is (at least vaguely) true that

$$\mathbf{G} \text{ is a } G_c.$$

More precisely,  $E1$  is understood as the set theoretic counterpart of a sentence resulting from the conjunction of the two following properties expressed in the language of our game  $\mathbf{G}$ :

- (i) *The return Q has the nature (regardless of the extent) of a cooperative benefit;*
- (ii) *Each player's specific investment cost (plus the effort cost in the case of player A) coincides with his fair status quo.*

To begin with (i), define a return being the cooperative benefit produced by the two players through their specific investments as follows

$$R_\Gamma(\omega_i) = V(e, a, I_A, I_B), = R^*$$

where  $\forall C \neq A, B$  and for each  $a \neq a$ ,  $V(e, a, I_A, I_B) \geq V(e, a, I_C) = V(e, a, I_B) \geq V(e, a, I_C)$ . Take  $R_\Gamma(\cdot)$  to be the return function defined for the game  $\Gamma$  in any state  $\omega_i$ . The set theoretic counterpart of  $R^*$  is

$$R^* = \{\omega_i \in \Omega \text{ s.t. } R_\Gamma(\omega_i) = R^*\}$$

which is the definition of the event that ‘return in game  $\Gamma$  is  $R^*$ ’. Consider however that *ex post* each unforeseen state does not admit a clear-cut description of the economic *nature* of the return. It only allows for an account in terms of a numerical value  $Q \in [0, 20]$  which is the clearly describable aspect of return in any  $\omega_i$ . Then the question is whether the observable returns  $Q$  are also cooperative benefits  $R^*$ . In order to be expressed formally this question needs a set theoretic definition of the event that the return in unforeseen contingencies may at least be described by a number  $Q$

$$Q^* = \{\omega_i \in \Omega \text{ s.t. } q_\Gamma(\omega_i) = Q \in [0, 20]\}$$

where  $q_\Gamma(\cdot)$  is the return-numerical-description-function that for each unforeseen state *ex post* selects a number for the return in game  $\Gamma$ . Thus the first property is whether the intersection of the two above defined sets is not empty

$$R^{**} = Q^* \cap R^{*\perp} \emptyset$$

where

$$R^{**} = \{\omega_i \in \Omega \text{ s.t. } q_\Gamma(\omega_i) = Q \in [0, 20] \text{ & } R_\Gamma(\omega_i) = R^*\}$$

Notice that it is assumed that each state  $\omega_i \in \Omega$  has the superficial describability property  $Q \in [0, 20]$ , so that the set  $Q^*$  coincides with  $\Omega$ . Thus the mentioned condition reduces to

$$\Omega \cap R^{*\perp} \emptyset$$

Moreover, because all the  $\omega_i$  are describable by  $q(\cdot)$  but they not necessarily contain the affirmation  $R_\Gamma(\omega_i) = R^*$ , the first condition we are interested in becomes whether the event that ‘the observable return is of cooperative nature’ is one of the possible event included in the power set of possible events defined on  $\Omega$

$$R^* \in P(\Omega),$$

which reduces to the condition that there are states  $\omega_i \in \Omega$  that have a positive membership in the set  $R^*$ .

Now, consider more formally property (ii). To be precise, define a cooperative two-player bargaining game  $G_c$  by the ordered pair  $(R^*, D^*)$ , where  $R^*$  is the payoff space (i.e. we assume that it coincides with the cooperative return of the two players' joint investments and efforts), and  $D^*$  is the two players' fair *status quo*. If the game  $\mathbf{G}$  may be interpreted as a cooperative bargaining game, we want to know whether the fair status quo for each player is

$$D^* = [c(I_A) + o_A, c(I_B)]$$

In order to find out the set theoretic counterpart of this property, define  $d^*(\cdot)$  as the function that for any  $\omega_i \in \Omega$  selects the fair *status quo*  $D^*$  of the cooperative game that may be construed out of  $\omega_i$ , so that  $D^* = d^*_\Gamma(\omega_i)$  if the cooperative game played in the state  $\omega_i$  has the parameters of our game  $\Gamma$ . Then

$$D^* = \{\omega_i \in \Omega \text{ s.t. } d^*_\Gamma(\omega_i) = [c(I_A) + o_A, c(I_B)]\}$$

is the event (understood as a set of states) that the pair  $[c(I_A) + o_A, c(I_B)]$  coincides to  $D^*$  in the cooperative bargaining game construable out of the game  $\mathbf{G}$ .

Remember however that, as for return, the nature of player's A effort in each unforeseen state is not clearly ex post described. On the contrary we only have a superficial description of it in terms of the number of hours  $Z \in [0, 8]$  spent at work by A. The event that any unforeseen state, as far as A's effort is concerned, may be ex post described by some  $Z \in [0, 8]$  can be expressed

$$Z = \{\omega_i \in \Omega \text{ s.t. } z_\Gamma(\omega_i) = Z \in [0, 8]\}$$

where  $z_\Gamma(\cdot)$  is the numerical-hours-at-work-description function that for each  $\omega_i \in \Omega$  says how long player A is asked to work if he enters the game  $\mathbf{G}$ . Thus the second property we are interested in corresponds to the question whether the situations superficially describable by the  $Z$  values are also situations where the players' specific investments and effort defines the fair *status quo*  $D^*$  of the cooperative bargaining game construable out of  $\mathbf{G}$ . In set theoretic terms

$$D^{**} = D^* \cap Z \neq \emptyset$$

where

$$D^{**} = \{\omega_i \in \Omega \text{ s.t. } z_\Gamma(\omega_i) = Z \in [0, 8]\} \text{ & } d^*_\Gamma(\omega_i) = [c(I_A) + o_A, c(I_B)]$$

Because we assume that  $Z \in [0,8]$  gives a precise even though superficial description of the relevant aspect of each  $\omega_i \in \Omega$ ,  $Z$  coincides to  $\Omega$ . Thus the condition above reduces to

$$\Omega \cap D^* \neq \emptyset$$

and, as above, this implies that the relevant condition is that the event that ' $[c(I_A) + o_A, c(I_B)]$ ' is the fair status quo in  $\mathbf{G}$  is one of possible events defined on  $\Omega$

$$D^* \in P(\Omega),$$

or alternatively that at least some  $\omega_i \in \Omega$  have positive membership in the set  $D^*$ .

Therefore the event  $E1$ , corresponding to the joint two properties (i) and (ii), is the set of states resulting from the intersection of the two sets defined above

$$E1 = R^* \cap D^*$$

Remember that there is no vagueness about the two joint properties in any of the states  $w_i \in W$ , that is

$$\forall w_i \in W, \mu_{E1}(w_i) = 1.$$

But consider now the membership of any state  $\omega_i \in \Omega$  in the event  $E1$ .  $\Omega$  is an ordinary set of elements that, although unknown *ex ante*, *ex post* turns out to be univocally describable as regards of the variables  $Q$  and  $Z$ . But, because of the unexpected features of states  $\omega_i$ , these characteristics cannot be clearly traced back to the more basic properties defining  $E1$ . Therefore *ex post* we understand the membership of any unforeseen state  $\omega_i$  in  $R^*$  and  $D^*$  in terms of fuzzy membership functions, as suggested in sec.5.

*Take first the set of ordered pairs*

$$\underline{R}^* = \{(\omega_i, \mu_{\underline{R}}(\omega_i)) \mid \omega_i \in \Omega\}$$

where the membership function  $\mu_{\underline{R}}(\omega_i) = x \in [0,1]$  associates to each  $\omega_i$  a degree of membership in the fuzzy set  $\underline{R}^*$ . This expresses for any  $\omega_i$  the degree of vagueness that the return  $Q$  is a cooperative benefit  $R^*$ . Then take the second set of ordered pairs

$$\underline{D}^* = \{(\omega_i, \mu_{\underline{D}}(\omega_i)) \mid \omega_i \in \Omega\}$$

defined by the membership function  $\mu_{\underline{D}}(\omega_i) = r \in [0,1]$  that associates to each  $\omega_i$  a degree of membership in the fuzzy set  $\underline{D}^*$ . It expresses for any  $\omega_i$  the vagueness that the players' specific investments costs, together with the  $Z$  value describing superficially A's effort, constitutes the ' $[c(I_A) + o_A, c(I_B)]$ ' fair *status quo* of the cooperative bargaining game construable out of  $\mathbf{G}$ . Finally consider the set of ordered pairs

$$\underline{R}^* \cap \underline{D}^* = \{(\omega_i, \mu_{\underline{R}^* \cap \underline{D}^*}(\omega_i)) \mid \omega_i \in \Omega\}$$

It defines the fuzzy set resulting from intersection of the two fuzzy sets  $\underline{R}^*$  and  $\underline{D}^*$  in terms of membership function  $\mu_{\underline{R}^* \cap \underline{D}^*}(\omega)$ . It results that the event  $E1$  must be more properly understood as a fuzzy set  $\underline{E1}$ . Moreover,

$$\underline{E1} = \underline{R}^* \cap \underline{D}^*$$

stands for the event that initial conditions asked for the application of *NBS* are satisfied, which - because its constituting elements are unforeseen states - is understood as a vague event. The membership of any state of

the world  $\omega_i$  in the fuzzy intersection event  $E1$  then is calculated by the MIN operator (corresponding to fuzzy intersection)

$$\mu_{E1}(\omega_i) = \text{MIN}(\mu_{R^*}(\omega_i), \mu_{D^*}(\omega_i)).$$

In order to decide whether to treat any state  $\omega_i$  ‘as if’ it were a situations to which Nash’s solution must be applied, now let introduce an  $a$ -cut set.  $a$  must be understood as a vagueness threshold discriminating the states in which the game  $G$  sufficiently satisfies the conditions for being a Gc, from the remaining states.

For  $a = 0.5$  define

$$\mu_{E1a}(\omega_i) = \begin{cases} 1 & \text{if } \mu_{E1}(\omega_i) \geq 0.5 \\ 0 & \text{if } \mu_{E1}(\omega_i) < 0.5 \end{cases}$$

This condition can be stated *ex ante*, so as to commit player B to treat the game as the appropriate domain of application of NBS whenever an unforeseen state were to belong to  $E1$  at least up to degree 0.5. Moreover its fulfilment can be checked *ex post*, when the players are also able to establish which of the *ex ante* unforeseen states exceed the ethical threshold of admissibility, simply by calculating the *crisp* (non fuzzy) set

$$E1_a = \{\omega_i \in \Omega \mid \mu_{E1}(\omega_i) \geq 0.5\}$$

For example, *ex post* the players may learn about  $\Omega$  and, given the imprecise knowledge on each  $\omega_i$  due to its unexpected features, they generate the following fuzzy set

$$\mu_{E1}(\omega_i) = \begin{array}{ccccc} \omega_1 & \omega_2 & \omega_3 & \omega_4 & \omega_5 \\ \hline 0.1 & 0.6 & 0.8 & 0.4 & 0.9 \end{array}$$

Then the  $a$ -cut set is

$$\mu_{E1a}(\omega_i) = \begin{array}{ccccc} \omega_1 & \omega_2 & \omega_3 & \omega_4 & \omega_5 \\ \hline 0 & 1 & 1 & 0 & 1 \end{array}$$

The crisp set  $E1_{0.5} = \{\omega_2, \omega_3, \omega_5\}$  is the ‘admissible’ set of states, and the players are allowed to conclude by default that NBS is applicable to it. The important implication is that commitments can be undertaken and verified on this basis:

- *Ex ante* B undertakes the commitment to treat the game as the domain of application of NBS in case any state  $\omega_i$  should come out, which *ex post* falls into the ‘admissible’  $a$ -cut set  $E1_{0.5}$ .
- *Ex post* players learn the degree of membership of each state  $\omega_i$  in the set  $E1$  and determine without any ambiguity whether NBS has effectively to be applied according to the above stated condition.

This fuzzy information can be used by each players in the role of A in order to check whether player B treat any stage-game under consideration appropriately. This is not enough however to define B’s commitments to use particular actions  $a \hat{I} a$ .

## 7. Second Step in the Ethical Procedure of Deliberation: Fuzzy Measures of Surplus and Effort

By  $E2$  we mean (in words) that ‘the level of cooperative benefit  $R_j^*$  associate to the observable return  $Q$  and the player’s A level of effort  $e_{Aj}$  are such that the outcome is *efficient/fair* in the sense of NBS, if the action  $a^*$  is

chosen'. To define the event  $E2$  in set theoretic terms, we must consider a subset of  $E1_\alpha$  - call it event  $E2$  - and actions contingent on states  $(a^*|\omega_i)$ . Thus, event  $E2$  is defined as the set of states where the conjunction of any two properties  $R_j^*$  and  $o_{Aj}$  necessarily implies that  $NBS$ , defined on the cooperative payoff space  $R^*$  and the fair status quo  $D^*$  (containing the given  $o_{Aj}$ ), is maximised if action  $a^*$  is chosen when any of such states occurs. That is, for  $h = A, B$ , and  $N$  levels of possible cooperative return  $R_j$  and possible effort cost  $o_{Aj}$ ,

$$E2 = \{(\omega_i) \in E1_\alpha \text{ s.t. } (R_j^* \& o_{Aj}) \Leftrightarrow (a^*|\omega_i) = \text{ArgMax}_{a \in I_A} \mathbf{P}_h [u_h(f_a(R_j^*)) - u_h(D^*)]\}$$

where  $R_j^*$  and  $o_{Aj}$  are the levels of cooperative benefit and of player's  $A$  effort cost occurring in state  $\omega_i$ . Moreover the  $NBS$  is defined by player's  $h$  utility function  $u_h$  ranging over the possible allocations of the cooperative benefit  $R_j^*$  calculated through the distribution function  $f_a(R_j^*)$  associated to action  $a^*$  such that, for utilities linear in the monetary payoffs,

$$\begin{aligned} u_B(f_a(R_j^*)) &= R_j^* - s \\ u_A(f_a(R_j^*)) &= R_j^* - (R_j^* - s) \end{aligned}$$

and by the player's  $h$  utility function  $u_h$  ranging over the values of fair status quo  $D^*$  of form  $[c(I_A) + o_{Aj}, c(I_B)]$  as the level  $o_{Aj}$  changes according to the different  $\omega_i$ .

The possibility that any  $(a^*|\omega_i)$  has an *efficient/fair* outcome depends on some proportion between the size of the cooperative return  $R_j^*$  (i.e. the *extent* of  $R^*$ ), and the size of effort  $o_{Aj}$  associate to the task  $a^*$  in state  $\omega_i$ . Players however, when any  $\omega_i \in \Omega$  occurs, do not see clearly  $R_j^*$  and  $o_{Aj}$  but their observable substitutes  $Q$  and  $Z$ , and they remain in a condition of vagueness about the meaning of these two pieces of information in terms of the relevant variable  $R^*$  and  $o_A$ . Thus they proceed by considering separately the two pieces of vague information before being able to compound them in order to evaluate whether and how any given  $\omega_i$  belongs to  $E2$ .

The first component of vagueness of event  $E2$  is indeterminacy of the causal relationship that links any size of return  $Q$  (i.e.  $Q_i \in [0, 20]$ ) to any size of the cooperative surplus  $R_j^*$ . In fact, whatever the level of cooperative benefit  $R_j^*$  considered, the set of the observable returns  $Q$  (i.e. states  $\omega_i \in \Omega$  exhibiting a level of  $Q_i$  each) compatible with the given level of cooperative benefit, is a *fuzzy set*.

- *Example 1.* As before, assume that  $Q$  can vary between 0 and 20, i.e. there are at least 20 states  $\omega_i$ . Moreover assume the  $R_j^*$  are not single values of cooperative benefit but the following five intervals

$$R^{*-} = 0 \leq R^* \leq 6; \quad R^{*-} = 7 \leq R^* \leq 9; \quad R^{*\gg} = 10 \leq R^* \leq 12; \quad R^{*+} = 13 \leq R^* \leq 16; \quad R^{*++} = 17 \leq R^* \leq 20.$$

Each interval of benefit defines therefore a fuzzy set of states. An example of the fuzzy sets defined by the various intervals of vague cooperative benefit is given in fig. 2. Fig. 2 represents a situation in which we know that substantial part of the return is a cooperative benefit, but its *extent* is vague. Vagueness is delimited by logical incompatibility between certain levels of observable return and certain sizes of cooperative benefit. On the contrary vagueness increases as the size of observable returns exceeds the given level of cooperative benefit, so that other sizes become more compatible with the observed ones. In fact, when the observable return is between 0 and 6 there is no vagueness that cooperative benefit must lies in the interval  $R^{*-}$ . But when the observable return becomes larger the case becomes more and more vague as

other sizes of  $R^*$  seem more possible. At the same time a very large size like  $R^*_{++}$  is incompatible to nearly all the levels of observable return, but its vagueness is at minimum when the return is very similar to the required value of  $R^*$ .

$\omega_1$ $Q=0,$	$\omega_6$ $Q=6,$	$\omega_8$ $Q=8,$	$\omega_{10}$ $Q=10,$	$\omega_{12}$ $Q=12,$	$\omega_{15}$ $Q=15,$	$\omega_{16}$ $Q=16,$	$\omega_{18}$ $Q=18,$	$\omega_{20}$ $Q=20$
$\mu_{R^*--}(\omega_i)$	1	1	0.8	0.7	0.2	0.15	0.1	0
$\mu_{R^*-}(\omega_i)$	0	0	0	0.7	0.5	0.2	0.15	0.05
$\mu_{R^*\approx}(\omega_i)$	0	0	0	0.2	0.6	0.3	0.2	0.15
$\mu_{R^*+}(\omega_i)$	0	0	0	0	0	0.6	0.7	0.5
$\mu_{R^*++}(\omega_i)$	0	0	0	0	0	0	0.6	0.8

(Fig.2. Each state, characterised by a specified observable ex post return  $Q_i$ , is associated to every class of fuzzy cooperative benefit)

The second component of vagueness of  $E2$  is the meaning of the variable  $Z$  in terms of  $o_A$ .  $Z$  is a description of  $a^*$  in physical terms (hours required to carry out the task). But in order to understand whether and how a state belongs to the event  $E2$  it is necessary to translate the value of  $Z$  into a measure of A's effort cost  $o_A$  (the tiredness or enjoyment experienced by A out of any action  $a$ ). This evaluation is inevitably vague, because of the novelty of events occurring in  $\omega_i$ , some of which may translate a high value of  $Z$  into a low effort-cost or, on the contrary, a low value of  $Z$  into a high effort-cost.

- *Example 2.* Let us consider as in section 2 only three possible values of effort-cost for  $a^*$

$$o_{A+} = 5, o_{A\gg} = 2, o_{A-} = 0$$

These values are immaterial per se, and can be taken as numerical exemplification for 'high', 'medium' and 'low' effort-cost of a given task. However taking specified, even if arbitrary, numbers for the effort-cost entering the utility functions of the players will make later possible the computation of NBS. Moreover take as given a value of  $Z$  (say  $Z = 8$  hours), in order to exemplify how the same value of  $Z$  may mean different levels of effort-costs. In fig.3 fuzzy membership functions are understood as expressing the possibility that, given the value of  $Z$ , in each of the states of the world (consider three of these) it happens that A must face each of the possible effort levels.

The example presents the case of three states that are inversely related to the three levels of player's A effort. Take  $\omega_1$  first. This is a state where some unexpected features make the task very hard, so that it is quite clear that the effort spent is high, whereas it is much more vague that it is low. On the contrary, in state  $\omega_3$  occurs some unexpected change that makes very easy the task, so that it is quite clear that effort is low, whereas it is much more vague that the effort requested is high.

	$w_1$	$w_2$	$w_3$
$\mu_{\underline{\omega}A_z}(\omega_i) =$	0.2	0.4	0.9
$\mu_{\underline{\omega}A_{\sim}}(\omega_i) =$	0.6	0.9	0.3
$\mu_{\underline{\omega}A_+}(\omega_i) =$	0.9	0.2	0.1

(Fig. 3. Each state, characterized by the same value of Z, is associated to every level of fuzzy effort-cost)

Examples 1 and 2 illustrate how the pieces of information  $Q$  and  $Z$ , which are foreseeable ex ante and are observable ex post, may give only vague knowledge about the relevant variables for evaluating whether in a given state the action  $a^*$  satisfies the event  $E2$ . This vague knowledge concerning any level  $j^{\text{th}}$  of cooperative return  $R_j^*$  and player's A effort  $o_{Aj}$  can be summarised by the corresponding fuzzy sets

$$\underline{R}_j^* = \{(\omega_i, \mu_{R_j^*}(\omega_i)) \mid \forall \omega_i \in EI_a\}$$

$$\underline{O}_{Aj} = \{(\omega_i, \mu_{O_{Aj}}(\omega_i)) \mid \forall \omega_i \in EI_a\}$$

## 8. Vagueness of Efficient/Fair Outcomes: Third Step in the Ethical Procedure of Deliberation

How the two components of vagueness mentioned above must be combined in order to express an overall – albeit vague – judgement? Consider the conjoint properties  $(R_j^* \& o_{Aj})$  where as above  $R_j^*$  and  $o_{Aj}$  represent the  $j^{\text{th}}$  levels of cooperative benefit and effort. The event corresponding to  $(R_j^* \& o_{Aj})$  is a fuzzy set defined as

$$\underline{R}_j^* \cap \underline{O}_{Aj} = \{(\omega_i, \mu_{R_j^* \cap O_{Aj}}(\omega_i)) \mid \omega_i \in EI_a\}$$

whose membership functions are calculated as follows

$$\mu_{R_j^* \cap O_{Aj}}(\omega_i) = \text{MIN} [\mu_{R_j^*}(\omega_i), \mu_{O_{Aj}}(\omega_i)]$$

This establishes how vague it is that in any given unforeseen state the conjunction of the two properties  $R_j^*$  and  $o_{Aj}$  occurs.

Notice that vagueness refers to statements like '*in  $w_1$  the cooperative return lies in the interval  $R_+$  and the effort associated with the action  $a^*$  is at level  $o_{A+}$* '. But, on the syntactic level, we are perfectly entitled to endorse inference like

'if  $(R_{i+}^* \& o_{A+})$  holds, then from  $a^*$  follows p'

where p may be some formal property of the payoffs. Vagueness lies at the semantic level, that is at the level where we ask whether, in any  $\omega_i$ ,  $R_{i+}^*$  and  $o_{A+}$  are *true*, whereas it does not jeopardize the syntactic inference that, if  $R_{i+}^*$  and  $o_{A+}$  are assumed, then from  $a^*$  it follows that p. Let p be the following property of maximum Nash product

$$a^* = \text{ArgMax}_a \bar{\mathbf{I}}_a \mathbf{P}_h (u_h(f_a(R_j^*)) - u_h(D^*)) \quad (\text{for } h = A, B)$$

such that the solution payoffs are  $[s - (c(I_A) + o_{A+}); (R_j^* - s) - c(I_B)]$ .

For every interval of cooperative benefit and every level of effort we can calculate whether  $p$  or  $\neg p$ . Calculations of the combinations  $(R_j^* \& o_{Aj})$  such that it is true that ‘if  $a^*$  is chosen then  $p$ ’, are summarised in the following table:

0£ $R_j^*$ £ 6, 7£ $R_j^*$ £ 9, 10£ $R_j^*$ £ 11, 12£ $R_j^*$ £ 16, 17£ $R_j^*$ £ 20					
$o_{A^-} = 0$	$p$	$p$	$p$	$\neg p$	$\neg p$
$o_{A^-} = 2$	$p$	$p$	$\neg p$	$\neg p$	$\neg p$
$o_{A^-} = 5$	$p$	$\neg p$	$\neg p$	$\neg p$	$\neg p$

(Fig.4 For any couple of parameters the figure shows when it is true that the Nash Product is maximised if player B chooses  $a^*$ )

Notice that  $R_j^*$  and  $o_{Aj}$  are intervals and numbers and not fuzzy sets as such. In order to understand whether the conjunctions of the two parameters’ values, which imply  $p$ , hold, we must study the corresponding intersections of fuzzy sets. Thus, take the union of all the fuzzy sets intersections defining events corresponding to joint properties  $(R_j^* \& o_{Aj})$  such that that  $p$  follows from  $a^*$ .

$$\mathbf{U}[(R_j^* \cap O_{Aj})^*] = \\ \{(R_{-}^* \cap O_{A-}) \cup (R_{-}^* \cap O_{A-}) \cup (R_{\approx}^* \cap O_{A-}) \cup (R_{-}^* \cap O_{A\approx}) \cup (R_{-}^* \cap O_{A\approx}) \cup (R_{-}^* \cap O_{A+})\}$$

We call this fuzzy set event  $E2$ . Intuitively  $E2$  is the union of all the situations such that choosing  $a^*$  under any of them satisfies the *NBS*, i.e. it is the event that *NBS* is satisfied (up to some degree) by choosing  $a^*$ . Thus we see that the proper notation for the above-mentioned event  $E2$  is the fuzzy set  $E2$ . Formally  $E2$  states for each state  $\omega_i$  its membership to the union of all the events like  $(R_j^* \cap O_{Aj})^*$  for which the *efficient/fair NBS* outcome is implied when  $a^*$  is chosen. By definition, degrees of membership of any state  $\omega_i$  into the set  $\mathbf{U}[(R_j^* \cap O_{Aj})^*]$  are assigned by the MAX operation,

$$\mu_{E2}(\omega_i) = \mu_{\mathbf{U}[(R_j^* \cap O_{Aj})^*]}(\omega_i) = \\ \text{MAX}[\mu_{(R_{-}^* \cap O_{A-})}(\omega_i), \mu_{(R_{-}^* \cap O_{A-})}(\omega_i), \mu_{(R_{\approx}^* \cap O_{A-})}(\omega_i), \mu_{(R_{-}^* \cap O_{A\approx})}(\omega_i), \\ \mu_{(R_{-}^* \cap O_{A\approx})}(\omega_i), \mu_{(R_{-}^* \cap O_{A+})}(\omega_i)]$$

Having this definition at hand, we are now able to work out formal conditions for committing player B to carry out action  $a^*$  whenever doing that in an unforeseen state satisfies *NBS* at the *appropriate* level. Within  $E1_\alpha$ , the set  $E2_\beta$  is defined as the  $\beta$ -cut set of those states (unknown *ex ante*) whose grade of membership to the event  $E2$  is not lower than threshold  $\beta = 0.5$

$$E2_\beta = \{\mu_{E2}(\omega) \geq 0.5 \mid \forall \omega \in E1_\alpha\}$$

Membership in this set is defined by the non-fuzzy membership function

$$\mu_{E20.5}(\omega_i) = \begin{cases} 1 & \text{if } \mu_{\underline{E}2}(\omega_i) \geq 0.5 \\ 0 & \text{if } \mu_{\underline{E}2}(\omega_i) < 0.5 \end{cases}$$

We therefore have a criterion for saying in which states the action  $a^*$  must be considered admissible, or be treated ‘as if’ it were *efficient/fair*. This criterion may be announced ex ante and meant as the basis for committing B in face of the expected ambiguity of the relevant information under unforeseen states. Moreover it can be verified ex post as the membership functions take their values when the unforeseen states take place in practice.

## 9. Default Inference of Admissible Actions

Let  $\pi$  be an evaluation function of actions, with domain the set {admissible, inadmissible} and co-domain B’s conditional choice set, i.e. actions conditional upon states of the world. Take the rule of inference

$$\forall \omega_i \in \Omega, \text{ If } \mu_{\underline{E}2\beta}(\omega_i) \neq 0 \text{ then } \pi(a^*|\omega_i) = \{\text{admissible}\}$$

This rule is basically needed in order to define commitments conditional upon the occurrence of unforeseen states of the world, which may be undertaken ex ante and can be verified ex post, as far as vagueness concerning the satisfaction of a given solution concept does not exceeds a given threshold.

The above scheme of inference is worked out by analogy with what in AI literature is normally called *default reasoning* (Reiter 1980, Ginsberg 1987). A statement is allowed among the default conclusions of a deductive theory when it follows from the base of knowledge plus statements about the “normal course” of the relevant matter (default rules of inference). Even though there is no proof that these statements are true, nevertheless they are added to the base of knowledge only because they are “consistent” in the simple sense that no refutations of them at moment are known. Because defaults make possible deriving ‘syllogistically’ conclusions from the knowledge base of the theory (plus the rule of inference), in this way a theory may be extended. An extension is the overall set of sentences that can be added to the knowledge base by undergoing it to all the admissible default rules of inference. Statements derived by default inference are valid ‘until proved otherwise’ (i.e. literally by default). The resulting set of sentences extending the base theory is clearly *non monotonic* - default extensions of a theory can be eliminated by the emergence of new evidence (McDermott and Doyle 1980). The typical example of default inference is

B(x): M(F(x))

---

∴ F(x)

where M stands for ‘it is consistent that...’. B(x) and F(x) may be read respectively as “Tweety is a Bird”, “Tweety Flies”, and the premise M(F(x)) is introduced for normally birds fly and there is at that moment no evidence that contradicts that they fly. The result is a default inference rule that allows concluding that Tweety flies even if Tweety is a Penguin, which clearly reveals that the system is non-monotonic (the knower will eventually come to learn that Penguins do not fly).

There is a basic analogy between default reasoning and the rule given above for evaluating actions under unforeseen contingencies. The fuzzy-based inference establishes the admissibility of an act in a state if we *do not have proof of the incompatibility* (this would amount to say that  $\mu_{E\beta}(\omega_i) = 0$ ) between an unforeseen state of the world and a proposition of our base theory. The statement about the admissible action then follows, even though it remains vague under the current state whether the action fits the solution theory (we only ask that vagueness does not exceed a given threshold.)

Let illustrate the default inference rule applicable to our case:

- (1)  $\forall \Gamma, (\Gamma \text{ is } Gc) \rightarrow (\text{NBS solves } \mathbf{G})$
  - (2)  $M ((\omega_i \in E1_a) \rightarrow E1)$
  - (3)  $M ((\omega_i \in E2_b) \rightarrow E2)$
  - (4) If  $(a^* | \omega_i) = \text{ArgMax}_{a \in A} \Pi_i(u_i - d_i)$  then  $(a^* | \omega_i) = \{\text{admissible}\}$
- 
- ∴  $(a^* | \omega_i) = \{\text{admissible}\}$

Premises 1 and 4 are sentences of our knowledge base, that is statements of our solution theory such that “all games treatable as a cooperative bargaining game are solved according to *NBS*” and “if an action in a state satisfy the *NBS*, then that action is admissible”. Premises 2 and 3 are typical default reasoning assumptions (default rules of inference), based on consistency. The first says that it is consistent with our knowledge to apply *NBS* in the state  $\omega_i$  (or “normally” states like  $\omega_i$  are such that the game is *Gc*). The second says that it is consistent (it is “normally” true that) that action  $a^*$  in state like  $\omega_i$  satisfies *NBS*. All that is assumed in so far as it does not contradict the information at disposal of the players. As far as any state  $\omega_i$  belongs to the events  $E1_\alpha$  and  $E2_\beta$ , it follows that there is not information that the two properties are false in that states.

## 10. Back to Reputation Effects

We now have an appropriate basis for developing the reputation game under unforeseen contingencies. The game defined in section 2 now simply becomes the stage-game (called a *play*) of a repeated game. The stage-game is played repeatedly by the long-run player B against an infinite series of short-run players playing each for a single period (from now on I shall refer to them as player  $A_i$  on the basis of their index of entry  $i$ ). Thus the strategies of the long-run player B for the repeated game will become rules which establish his choice between  $a^*$  and  $a^c$  for each play, contingently on the previous history of the game up to the present play. Each  $A_i$  will decide from the standpoint of the stage game in which he is taking part, whether to enter or not (*e, non-e*) contingently on the previous history.

The basic request is that ex post any  $A_i$  must be able to check whether B conform to his commitment. In order to see how this is so, let define player's B types. Quite unconventionally, each type is

not associated to performance of an idiosyncratic action, but to fulfilment or non-f fulfilment of the conditions laid down by the code. There are three *types* of player B

*Type θ1* takes on the commitment to follow the ethical code in each play.

$$\theta_1 = \begin{cases} (a^c | \omega_i), \forall \omega_i \text{ such that } (a^* / \omega_i) \text{ is inadmissible} \\ (a^* | \omega_i) \text{ otherwise} \end{cases}$$

*Type θ2* always adopts  $a^*$  - his dominant action - in each state of the world

$$\theta_2 = \forall \omega_i, (a^* / \omega_i)$$

*Type θ3* tries to make his behaviour ‘unpredictable’

$$\theta_3 = \begin{cases} (a^c / \omega_i) \text{ with prob 0.25, } \forall \omega_i \text{ such that } (a^* / \omega_i) \text{ is inadmissible} \\ (a^* / \omega_i) \text{ otherwise} \end{cases}$$

The code signals the *possibility* of a type, *but* compliance with the code depends on reputation effects. In order the mechanism of reputation to be put at work, all that is required is that the prior probability assigned to the type  $\theta_1$  - who respects the code - should not be zero (plus the hypothesis that player B is not short-sighted). Assume *prior* probabilities of the types are as follows:

$$p(\theta_1) = p, \quad p(\theta_2) = q, \quad p(\theta_3) = r$$

where  $p$  can be very small and  $1-p = q+r$ . small The updating rule is Bayesian. Read  $(a^c \cap \omega_i | \theta_1)$  as ‘the occurrence of the action  $a^c$  in the state  $w_i$  given that player B is the type  $\theta_1$ ’. Likelihood functions of actions, for each state of the world given each type, are the following.

In the case of the *type θ1*:

$$p(a^c \cap \omega_i / \theta_1) = \begin{cases} 1 \text{ if } (a^* / \omega_i) = \text{inadmissible} \\ 0 \text{ if } (a^* / \omega_i) = \text{admissible} \end{cases}$$

$$p(a^* \cap \omega_i / \theta_1) = \begin{cases} 0 \text{ if } (a^* / \omega_i) = \text{inadmissible} \\ 1 \text{ if } (a^* / \omega_i) = \text{admissible} \end{cases}$$

In the case of the *type*  $\theta_2$ :

$$p(a^c \cap \omega_i / \theta_2) = \begin{cases} 0 & \text{if } (a^* / \omega_i) = \text{inadmissible} \\ 0 & \text{if } (a^* / \omega_i) = \text{admissible} \end{cases}$$

$$p(a^* \cap \omega_i / \theta_2) = \begin{cases} 1 & \text{if } (a^* / \omega_i) = \text{inadmissible} \\ 1 & \text{if } (a^* / \omega_i) = \text{admissible} \end{cases}$$

In the case of the *type*  $\theta_3$ :

$$p(a^c \cap \omega_i / \theta_3) = \begin{cases} 0.25 & \text{if } (a^* / \omega_i) = \text{inadmissible} \\ 0 & \text{if } (a^* / \omega_i) = \text{admissible} \end{cases}$$

$$p(a^* \cap \omega_i / \theta_3) = \begin{cases} 0.75 & \text{if } (a^* / \omega_i) = \text{inadmissible} \\ 1 & \text{if } (a^* / \omega_i) = \text{admissible} \end{cases}$$

Note that learning is allowed only when the code establishes  $(a^* / \omega_i) = \text{inadmissible}$ .

The conditional probability of *types*, given the act undertaken in the preceding stage game, with reference to each state of the world  $\omega_i$  is calculated by *Bayes rule*. For example in the case of *type*  $\theta_1$

$$p(\theta_1 | \omega_i \cap a^c) = \frac{p(a^c \cap \omega_i | \theta_1) p(\theta_1)}{\sum_{j=1}^3 p(a^c \cap \omega_i | \theta_j) p(\theta_j)}$$

If player B chooses  $a^*$  in a state in which the code establishes the inadmissibility of  $a^*$ , then the conditional probability of *type*  $\theta_1$  is nil. After enough (say  $n$ ) observations of  $a^c$ , in states in which the code requires it, the conditional probability of *type*  $\theta_1$  increases to the extent that decision to enter ( $\theta$ ) becomes appealing for each  $A_{n+1}$ .

Let  $N$  be the number of periods that players  $A_i$  employ in order to update the probabilities of the *types*, until a probability distribution over the actions of B is generated that induces player  $A_{n+1}$  to enter.  $M < N$  is the number of times in which the players  $A_i$  effectively learn, given that sometimes (precisely  $N - M$  times) states may occur for which the code requires  $a^*$ . Let us suppose that until now player B acted as if he were the *type*  $\theta_1$ , that is, he used  $a^c$  in the  $M$  periods in which the code requires it. Given that players  $A_i$  are

rational, they will have refrained from playing  $e$  until the period  $N$ , but from the period  $N + 1$  they will begin to enter.

To choose his best reply player B will compare the expected payoffs from at least two strategies:

- *strategy s1:* after having used  $a^c$  in all the first  $M$  periods in which the code established the inadmissibility of  $a^*$ , in the  $N + 1^{th}$  period (or any of the succeeding periods) in which the code again requires the use of  $a^c$ , use  $a^*$  instead.
- *strategy s2:* after having used  $a^c$  in all the first  $M$  periods in which the code required it, from the  $N + 1^{th}$  period and in all succeeding periods, when the code again requires it, continue to use  $a^c$ .

Fudenberg and Levine's results (1988, 1991) guarantee that, for a discount rate  $\delta \sim 1$ , the upper bound of the equilibrium set is given by an equilibrium that, except for an initial period of  $N$  plays, gives player B the stage game *Stackelberg's payoff* for the whole duration of the repeated game. To keep calculations simple, take the ex post perspective, when  $\Omega$  has been revealed, and consider the subset of  $\Omega$  of those states  $\omega \in E1_\alpha$  such that  $a^*$  does not maximise *NBS*, but nevertheless the player's A payoff remains positive. I assume that the probability mass over such subset of states measures zero.<sup>1</sup> Thus, in any state  $\omega_i$  we have to be concerned with, if  $a^*$  is incompatible to *NBS*, the *Stackelberg payoff* of player B is certainly given by the pairs of strategies  $(e, a^c)$ . In fact if B's binding commitment fell on  $a^*$ ,  $A_i$ 's best reply would be *non-e*, with the payoff 0 for B. Thus the optimal choice of binding commitments for a leader à la Stackelberg would always coincide with behaviour that conforms to the ethical code, or with the type  $\Theta_1$ .

This permits stating the following proposition without the need of a formal proof – which would be essentially an adaptation from (Fudenberg and Levine 1989), see however (Sacconi 2000).

**PROPOSITION:** *Take the game of hierarchical transaction as the stage-game of an infinitely repeated game between a hierarchical long-run superordinate and an infinite series of short-run subordinates, in the presence of unforeseen events. Then an ethical code, to which is associated a 'compliant type' of the hierarchical superordinate with positive initial probability, allows the superordinate to induce a Nash equilibrium in the repeated game, such that (i) the total payoff of the hierarchical superordinate is identical to that he would obtain if he were able to take on binding commitments à la Stackelberg except for the  $N$  initial periods spent in accumulating reputation; (ii) in no period does the payoff of the hierarchical subordinates differ from that which conforms to the ethical code.*

## 11. Limitations and Final Remarks

Some apparently limiting assumptions are implicit in our result. First, it is assumed that two agents  $A_i$  and  $A_{i+1}$  who participate consecutively in the repeated game, must share the same judgement regarding the observed action  $a^*$ . When a state that was not specified *ex ante* occurs, their judgement regarding the compatibility of the state with the two conditions of admissibility must be the same, that is to say the vagueness of the relationship between the state and the two events  $E1$  and  $E2$  must be measured identically by the two players. This assumption is assured by interpreting vagueness as objective indeterminacy or as vague knowledge, but not as subjective belief. It should be noted that both vague knowledge and subjective belief (subjective probability) intervene, but at very distinct points of the argument: when the state  $\omega_i$  occurs, its compatibility with the events  $E1$  and  $E2$  is the subject of vague evaluation but inter-subjectively invariable.

Since the code contains a default rule of inference, the judgement about the admissibility of the action  $a^*$  observed *ex post* is neither vague nor uncertain, but univocal although not monotonic. If the action in such circumstances is inadmissible B's behaviour appears to incompatible with the hypothesis that he is a *type* who respects the code. Only at this point does probability come into play: each  $A_i$  and  $A_{i+1}$  will have an initial belief expressed via a subjective prior probability distribution regarding the *types* of B. The *type* is established without any vagueness, since it coincides with the behaviour prescribed by the default inference or its negation. Each *pure type* (i.e the *ethical* and the *non-ethical*) chooses  $a^*$  when the conditions regarding the membership of states in the sets  $E1_\alpha$  and  $E2_\beta$  are respectively (1,1) or (1,0). Then, the probability of all the *type* varies according to the evidence collected *ex post* via likelihood functions.

*Second*, in order that B may be able to foresee and anticipate the evolution in the beliefs of the other players, it is necessary that B comes to the same vague judgement as they do regarding the compatibility of the state known *ex post* with the events  $E1$  and  $E2$ . What is required is not necessarily that B has at the beginning the same piece of information and therefore expresses the same judgement about the compatibility between the states and the sets  $E1$  and  $E2$ . What is required is simply that B should have access to the common judgements expressed by  $A_i$  and  $A_{i+1}$ , since it is from these judgements that  $A_i$  and  $A_{i+1}$ 's short-run behaviour derives. Therefore the only relevant thing is the vague judgement in the state of information shared by  $A_i$  and  $A_{i+1}$ .

Notice that strong assumptions such as the existence of a common distribution of fuzziness, or – worst – that it is common knowledge, are not necessary. What is really needed is much less than this: only any couple of players  $A_i$  and  $A_{i+1}$  must share the same fuzziness distribution on the unforeseen states, which have already transpired when they participate in the game. This same distribution must be also known by B at stage  $i$  and  $i+1$ , without asking however that B agrees on it being an accurate representation of the vagueness at hand. On the other hand, player B has strong reasons for employing it when he is implementing the rules of the code. In fact, only if his behaviour conforms to what the code asks him, in the light of how things are seen by any couple of players  $A_i$ ,  $A_{i+1}$ , he will benefit from the reputation effects that induce those players, participating in the stage games in the role of A, to “enter”. To put it in other words, it is in B's best interest to take the fuzziness distribution assigned by any couple  $A_i$  and  $A_{i+1}$  for granted, in order to maintain the reputation effects mechanism at work.

## REFERENCES

- AL-NAJJAR N.I., ANDERLINI L. AND L.FELLI: *Incomplete Contracts in a Complex World*, mimeo, 2000.
- ANDERLINI L. AND L.FELLI: *Bounded Rationality and Incomplete Contract*, Georgetown University, mimeo, 2000.
- ARROW K.: "Business Codes and Economic Efficiency", in Beuchamp T. and N Bowie (eds), *Ethical Theory and Business*, 3rd ed., Englewood Cliffs, N.J. (Prentice Hall), 1988.
- BENSON G.C.S.: "Codes of Ethics", *Journal of Business Ethics*, 8, (1989).
- BILLOT A.: *Economic Theory of Fuzzy Equilibria*, Berlin (Springer), 1991.
- BINMORE K.: "Game Theory and The Social Contract", in R.Selten (ed.), *Game Equilibrium Models in Economics, Ethics and Social Sciences*, Berlin (Springer) 1991.
- BINMORE K.: *Just Playing*, Cambridge Mass. (The MIT Press), 1998.
- BROCK H.W.: "A Game Theoretical Account of Social Justice", *Theory and Decision*, 11 (1979), pp. 239-265.
- CENTER FOR BUSINESS ETHICS: "Are Corporation Institutionalizing Business Ethics?", *Journal of Business Ethics*, 5 (1986), pp.77-89.
- COLEMAN J.S. (1990) *The Foundation of Social Theory*, Harvard, (Belknap Press).
- COLEMAN J.: *Risks and Wrongs*, Cambridge (Cambridge University Press), 1992.
- DUBOIS D. and H.PRADE: "Non-Standard Theories of Uncertainty in Plausible Reasoning", G.Brewka (ed.) *Principle of Knowledge Representation*, (CSLI Publications), 1996.
- DUBOIS D. and H.PRADE: "Possibilistic Logic and Plausible Inference" in G.Coletti, D.Dubois and R.Scozzafava (eds.), *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence*, New York (Plenum Press), 1995, pp.209-226.
- FUDENBERG D. and D.LEVINE: "Reputation and Equilibrium Selection in Games with a Patient Player", *Econometrica*, 57 (1989), pp.759-778.
- FUDENBERG D. and D.LEVINE: "Maintaining Reputation when Strategies are Imperfectly Observed", *Review of Economic Studies*, 59 (1991), pp.561-579.
- GAUTHIER D.: *Morals by Agreement*, Oxford (Clarendon Press), 1986
- GINSBERG M.L.: *Reading in Nonmonotonic Reasoning*, Mogan Kaufmann Publisher Inc. Los Altos California, 1987.
- GEFFNER H.: *Default Reasoning, Causal and Conditional Theories*, Cambridge Mass. (The MIT Press), 1992.
- GROSSMAN S. and O.Hart: "The Costs and Benefit of Ownership: A Theory of Vertical and Lateral Integration", *Journal of Political Economy*, 94 (1986), pp. 691-719.
- HARSANYI J.C.: *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge (Cambridge University Press), 1977.
- HART O. and J.MOORE: "Property Rights and the Nature of the Firm", *Journal of Political Economy*, 98 (1988), pp.1119-1158.
- HART O. and J.MOORE: "Foundations of Incomplete Contracts", *Review of economic Studies*, 66, (1999), pp.115-138.

- KREPS D.: "Corporate Culture and Economic Theory" J.Alt and K.Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge (Cambridge University Press), 1990.
- MANSUR Y.M.: *Fuzzy Sets and Economics*, Aldershot Hampshire (Edward Elgar), 1995
- MCDERMOTT D. and J.DOYLE: "Nonmonotonic Logic I", *Artificial Intelligence*, 13 (1980), pp.41-72.
- MOLANDER E.: "A Paradigm for Design, Promulgation and Enforcement of Ethical Codes", *Journal of Business Ethics*, 6 (1987), pp.619-663.
- NASH J.: "The Bargaining Problem", *Econometrica*, 18 (1950), pp.155-162.
- REITER R.: "A Logic for Default Reasoning", *Artificial Intelligence*, 13 (1980), pp.81-132.
- SACCONI L.: "Codes of Ethics as Contractarian Constraints on the Abuse of Authority within Hierarchies: A Perspective from the Theory of the Firm", *Journal of Business Ethics*, 21 (1999), pp.189-202.
- SACCONI L.: *The Social Contract of the Firm. Economics, Ethics and Organisation*, Berlin (Springer), 2000.
- SEN A.: *Moral Codes and Economic Success*, London (LSE, ST-ICERD discussion papers n.49), 1993.
- SIMON H.: "From Substantial to Procedural Ratioanlity" C.C McGuire and R.R.Radner (eds.) *Decision in Organisation*, Amsterdam North Holland
- SIMS R.R.: "The Institutionalisation of Organizational Ethics", *Journal of Business Ethics*, 10 (1991), pp.493-506.
- TIROLE J., Incomplete Contracts: Where do We stand?, *Econometrica*, 67, 4, (1999) pp.741-781.
- VAMBERG V.J.: "Organizations as Constitutional Order", in *Constitutional Political Economy*, 3 (1992), pp.223-255.
- WILLIAMSON O.: *The Economic Institutions of Capitalism*, New York (The Free Press), 1986.
- WILLIAMSON T.: *Vagueness*, London (Routledge), 1994.
- ZADEH L.A.: "Fuzzy Sets As a Basis for the Theory of Possibility", *Fuzzy Sets and Systems* 1 (1978), pp.3-28.
- ZADEH L.A.: "Fuzzy Sets", *Information and Control*, 8 (1965), pp.338-353.
- ZIMMERMAN H.J.: *Fuzzy Set Theory and Its Applications*, 2nd revised ed., Dordrecht-Boston (Kluwer Academic Press), 1991.

## Notes

I am grateful to one anonymous referee for a number of detailed suggestions improving the preliminary version of the paper. I also acknowledge the support of the MURST received under two national research projects in Law & Economics.

<sup>1</sup> (*Department of Economics, Università di Trento and CELE, Centre for Ethics, Law & Economics, Università Cattaneo, Castellanza*)

End Note:

Introducing probabilities of unforeseen states of the world would ask for a clarification about the meaning of prior probability of states that ex ante were unforeseen, i.e. were not included within the states space. I leave this question to a further discussion. However see (Sacconi 2000, pp. 207-8).