In favour of CSR as an extended corporate governance model: social contract, conformism and evolution

Lorenzo Sacconi

University of Trento and EconomEtica

CS(tk)R as an extended model of corporate governance

- **Definition**: who runs a firm (entrepreneurs, directors, managers) have responsibilities that range
  - from the fulfilment of fiduciary duties towards the owners
  - to the fulfilment of analogous fiduciary duties towards all the firm's stakeholders
- **The scope of CSR**

  CSR extends the concept of fiduciary duty **from a mono-stakeholder** perspective (where the sole relevant stakeholder is the owner of the firm) **to a multi-stakeholder one** in which the firm owes fiduciary duties to *all* its stakeholders (the owners included

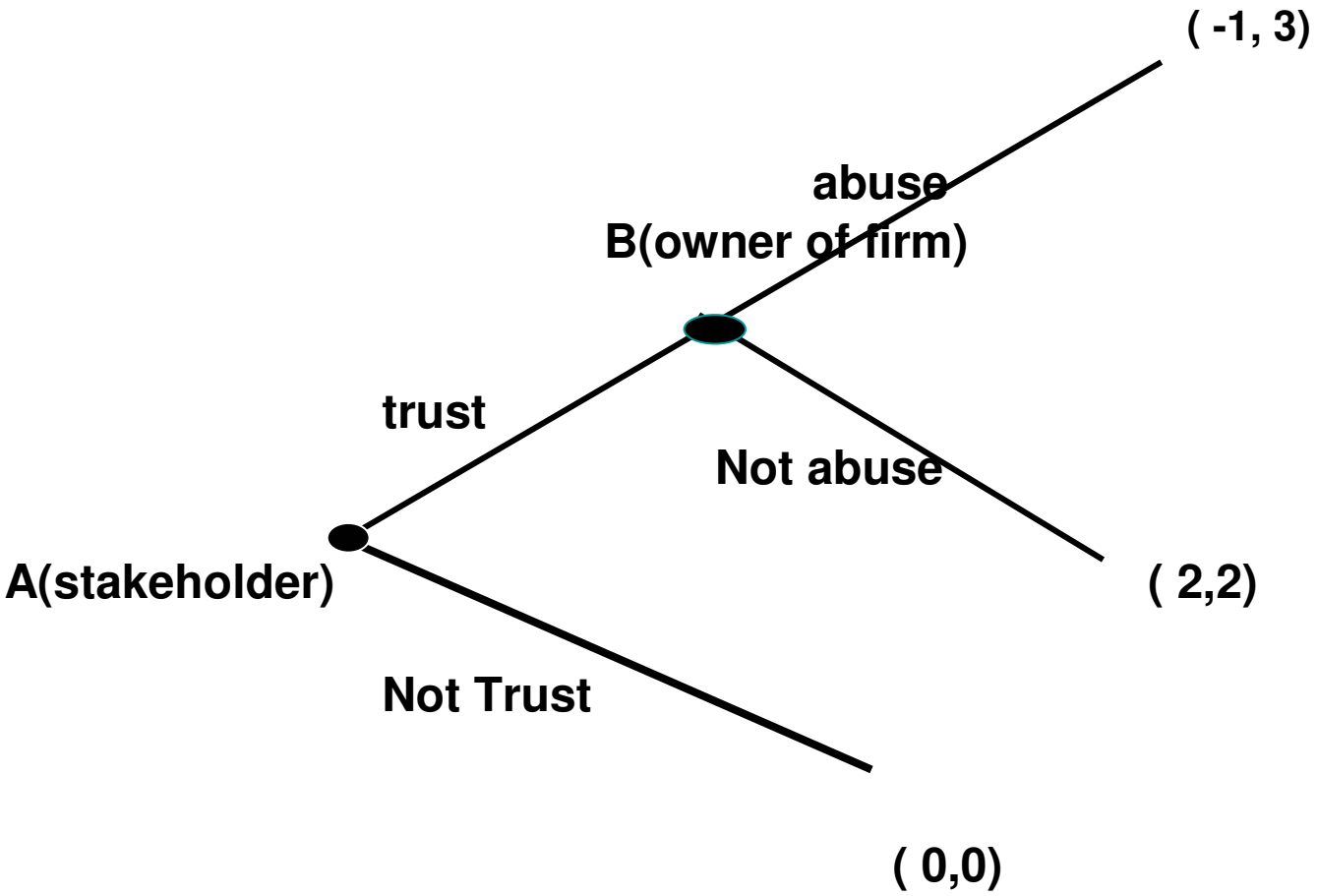
The economist's skeptics view:
Please, do not care explicitly for CSR, ethical norms or management systems; simply enlightened self-interest will indirectly do the job in the long run

- Caring directly for multiple objectives (and multiple fiduciary duties) would enlarge too much management discretion and reduce its accountability as a agents
- A single-argument objective-function (total shareholder value in the long run) is more manageable by limitedly rational managers and makes then more accountable
- Managing according TSV **in the long run** eventually implies to fulfill stakeholders' claims (instrumentally, not as right-holders)
- Enlightened self-interest **in the long run** provides the relevant incentive (motivation) for implicit (unintentional) CSR conduct

Why prof. Jensen is wrong

- In any significant game representing the long run interaction between the firm and stakeholders there are too many equilibria,
- Some of these allow opportunist firms to gain higher payoffs than the perfectly fair equilibrium
- hence an enlightened self-interested firms in the long run would prefer those equilibria in which it abuses substantially of its stakeholder trust
- In order to select fair equilibria, an explicit normative principle (an objective function) must be assumed
- An ex ante normative choice does not imply that there is also the ex post incentive to comply with the ex ante chosen equilibrium
  - Belief formation
  - Complex motivations

The trust game



Only one Nash Equ :(0 0,) Trust is impossible in one-shot relationship

Why does the trust game represent the interaction between the firm and its stakeholder?

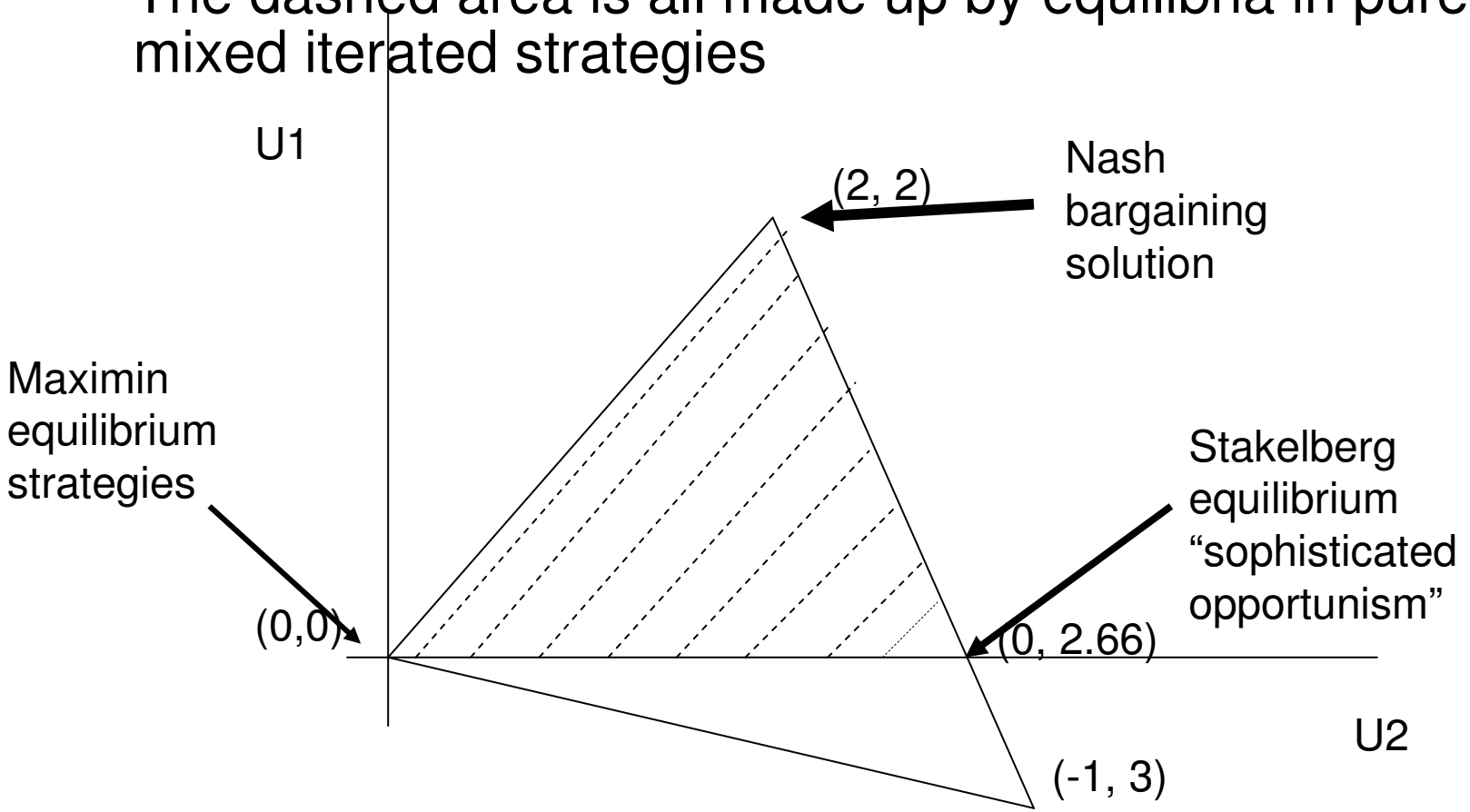
- The game represent a given control structure where the owner of the firm may exercise discretion
- By entering the stakeholder (call she Eve) make a specific investment
- the contract is incomplete , hence Eve may trust the owner of the firm (Adam)
- Adam may abuse or not of his authority
- By abusing Adam appropriates all the surplus produced by their joint specific investments

The game of reputation

- **Reputation effects are modeled by repeating the game of trust infinite times amongst a long-run player (the firm) a possibly infinite short run players who enter each time the game (stakeholders)**
  - Short run players update their beliefs over the possible "types" of the long run player given an "a priori" probability
  - One (just one) of the long run player's equilibrium strategies is simulating the completely "honest" type in order to support its reputation
  - From some point on the stakeholders will trust the firm and will start to enter, for this expected utility exceeds not entering
  - Then the best response of the firm is to continue supporting his reputation by not abusing

BUT: the equilibrium set of the repeated trust game contains too many equilibria

- as in the repeated PD, if all the repeated strategies are permitted, many equilibria are possible
- The dashed area is all made up by equilibria in pure and mixed iterated strategies



A repeated equilibrium of refined abuse

- Consider the player B-type that make a commitment on the mixed strategy *(2/3-ε **a**, 1/3+ε **no-a**)* (with ε as small as possible, practically nil)
  - B may develop a reputation for being this type by playing the two pure strategies with the attached probability throughout all the repetitions of the game
- Hence player Ai necessarily enters (average positive payoff 3ε = nearly zero)
  - this gives B an average expected payoff 2.66-ε.
  - Then player B's best response is to stick to this type
- Hence the preferred (by B) mixed strategy equilibrium is that in which player B (the firm) abuses two third of times, appropriating the largest part of the surplus

The normative problem

- Which justification can we given for agreeing ex ante on the fair outcome (equilibrium)?
- This would work as a Justification for the current control structure
- In order to choose (1,1) the firm must have a "stakeholder objective function" (mutual advantage not one-sided maximisation)

The cognitive problem

- An ex ante agreement does not constraint behaviour
- Neither generate **common knowledge** about the ex post behaviour (this would require know actual ex post behaviour)
- But can give the basis for developing the relevant belief about the other player behaviour (Default reasoning)

The motivation problem

- Reputations can be of many kinds (types).
- A company may develop a reputation for abusing trust of its employees, customers, suppliers, and capital-lenders only to the extent that they are **indifferent** between maintaining their relations with the firm and withdrawing from them.
- Stakeholders activism refuses to give in to this conduct, and actively countervail hypocritical corporate conduct.
- May the recent behavioural turn in behavioural microeconomics help explaining these stakeholders' conducts?
- If the owner and the stakeholder agree on a principle do they develop motivations that explain that the stakeholder refuses to give in ?

Deduction of the "stakeholder balancing principle" from a theory of the constitutional contract

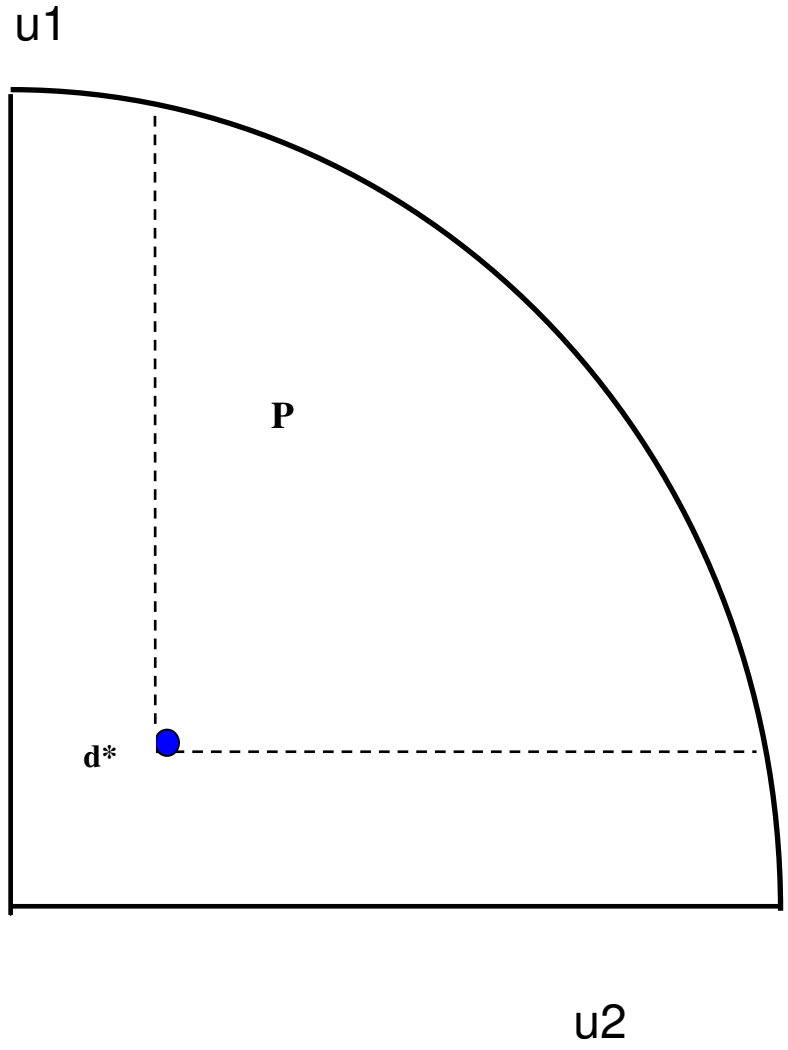
- **The model of constitutional contract of the firm rests on an analogy between**
  - **social contract theories** used to justify 'by agreement' both the 'legal constitution' (Buchanan 1979) and the mutually advantageous rules of morals (Gauthier 1986)
  - the **economic theory of efficient choice of the control structure of firms**, based on the idea of contractual incompleteness, (Williamson 1975, Grossman and Hart 1986, Hart and Moore 1990)

The model

- **There is a two-step collective decision-making among potential members of the coalition S.**
  - At time t = 0 the allocation of rights is decided (not only ownership and control but also redress), and this determines the control structure exerted over the productive coalition S
  - At time t = 1 the right-holding individuals undertake investment decisions with a view to subsequent transactions
  - At time t = 2 events occur which are unforeseen by the initial contract.
  - At time t = 3 a new bargaining game begins, defined for each allocation of rights and for every set of investment decisions.
- **This problem is modelled as a compounded bargaining game Gc on the constitutional and post-constitutional decision**,
  - First: a constitutional bargaining game is carried out at time t = 0, where chosen is a set of strategies (rights) by means of which …..
  - second: a subsequent game can be played at time t = 3 within the limits of the given constitution

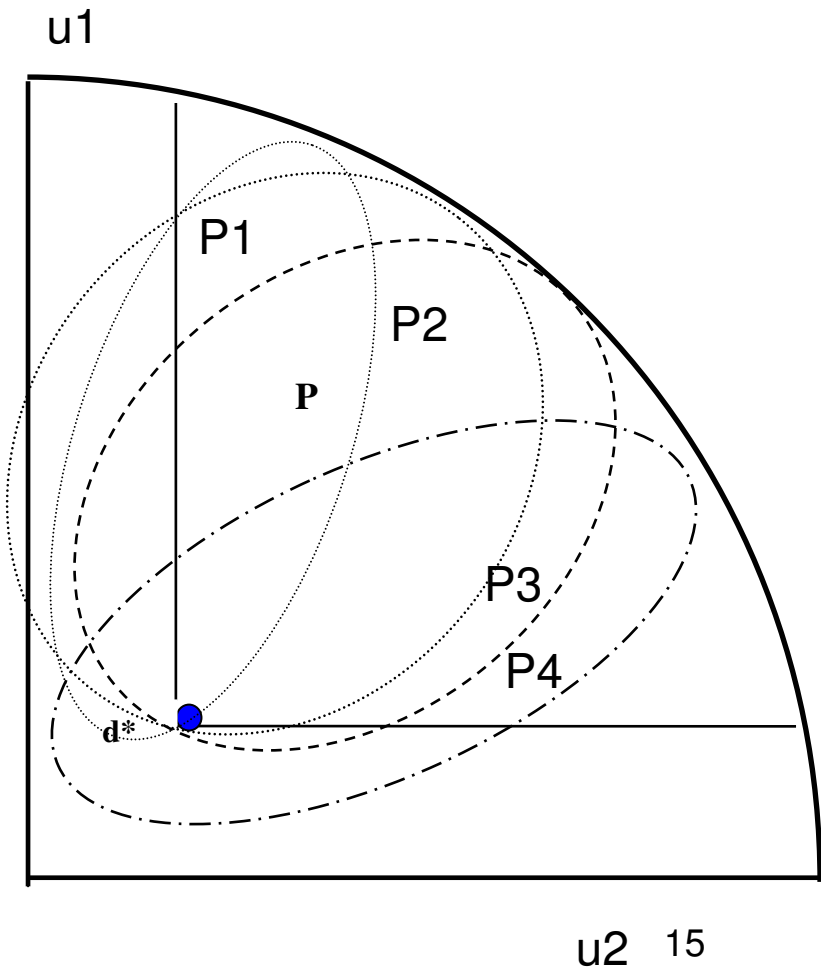
The Gc game

- Gc outcome space P consists of the 'state of nature' equilibrium d* + all the other 'state of nature', possible outcomes + their (linear) combinations
- Agreements over a constitutions can generate whatever outcome that were previously only virtually possible
- The state of nature is a *non-cooperative* game, whilst the Gc is a *cooperative* bargaining game
- It is a thought experiment the players may enter if they want to solve the sub-optimality of GN by agreement



A distinctive feature of constitutional choice

- Players simply choose a **subset I** of the set of joint strategies admissible in **Gc.**
- Each subset of the Gc strategies space is a limitation on the players' freedom
- Thus the choice of any subset coincides with the choice of a 'constitution'
- Each subset (constitution) in turn defines a cooperative sub-game **Gi** whose outcome space **Pi** is a subset of the outcome space P of Gc
- These are a coalition games in which the players negotiate on how much they obtain from cooperation according their "constitutional rights"



The constitutional contract is worked out by backward induction

- **As a whole the individuals take part in a sequential game**
- **First**: they start by solving the post-constitutional games Gi defined for each constitution
- Given hypothetically each sub-game, the players calculate the payoff assigned to them by the *Shapley value*
- $Vi = \sum_{S} [(s-1)!(n-s)!/ n!] [v(S) – v(S-\{i\})]$
- For each Gi there is a well defined solution of the coalition problem such that $\sigma i \geq d^*$

Moving backwards to the initial phase of the constitutional choice….

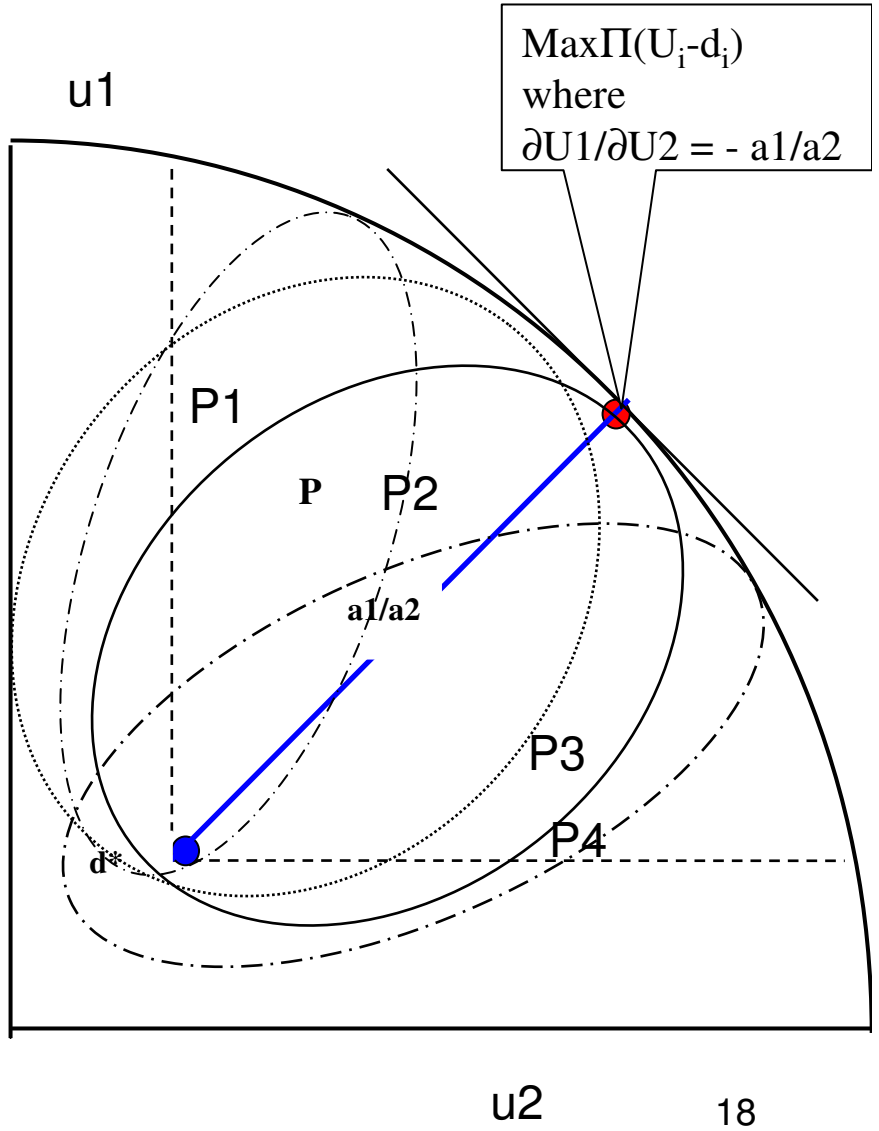
- **Second** : The Gc choice must be made unanimously by all the members of S.
- If this agreement is not reached players are doomed to play the unprofitable 'state of nature' game with solution d*
- Gc is the typical cooperative bargaining game
- The most accredited solution is Nash bargaining solution (N.B.S).

$$Max \Pi_i (Ui - d^*_i)$$

- It follows from different sets of very general postulates (Nash 1950, Harsanyi -Zeuthen 1977)

Nash solution for the choice of the firm constitution

- In Gc the solution has to be reached within a symmetrical outcome space generated by all the *logically possible* subsets of the set of strategies of Gc itself
- All the points in this space are understood as solutions for possible post-constitutional games.
- The N.B.S selects a constitution such that the post-constitutional game will distribute equal parts of the cooperative surplus calculated with respect to the Gc outcome space

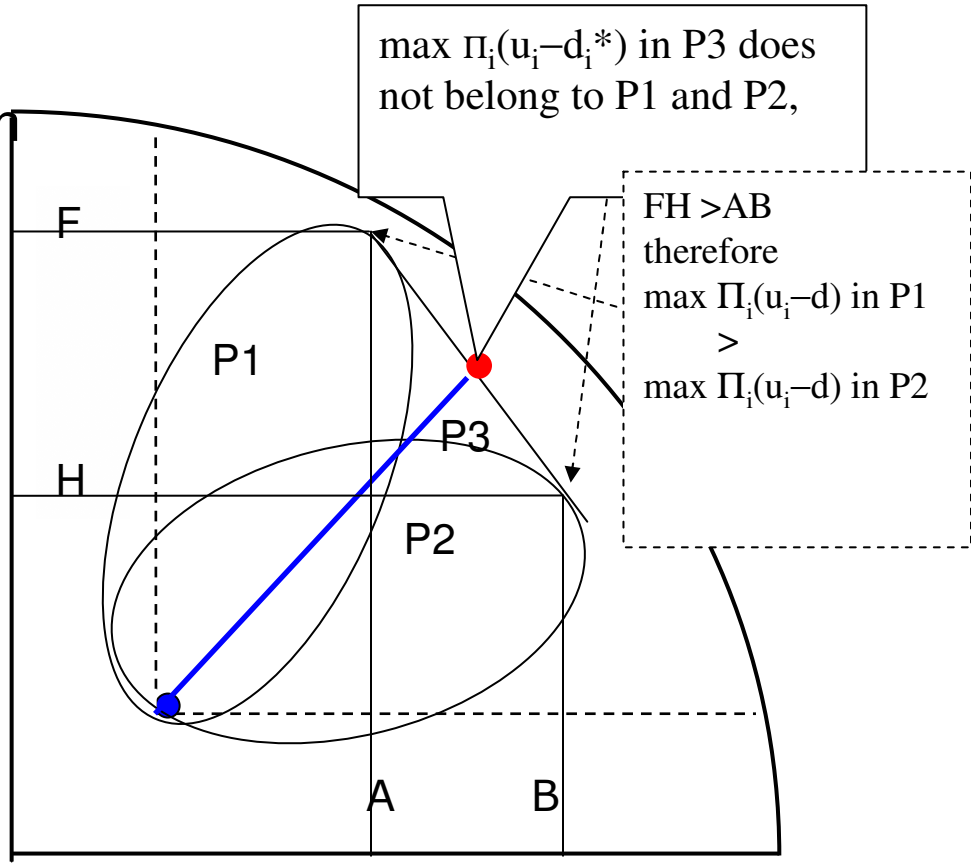


Dealing with exclusive property rights

- Thus far *every logically* possible constitution has been considered
- More realistic is the hypothesis that only a certain of restrictions on the set of all the strategies of Gc are institutionally feasible.
- Only exclusive allocations of property rights on all the physical assets of the firm are institutionally feasible.
  - **Control structures allow assigning all authority to some party, but not intermediate or equal degrees of authority,**
- the **N.B.S. relative to the all-inclusive payoff space of GC may not coincide with the solution of any of the institutionally feasible sub-games, (the choice must fall within the set of *institutionally feasible* solutions**)

A two players case

- There is one feasible constitution G1 (which assigns ownership to 1) with payoff space P1, whose solution is more efficient than that of the alternative sub-game G2 with payoff space P2.
- Ownership must be given to 1
- However, 1 must still take account of 2's claims and compensate him.
- Hence the solution must be calculated within the payoff space P3 generated as the convex hull of the combinations of outcomes associated with the feasible constitutions P1 and P2.
- This requires utility side-payments by which 1 compensates 2 until the cooperative surplus is distributed according to the criterion of NBS.

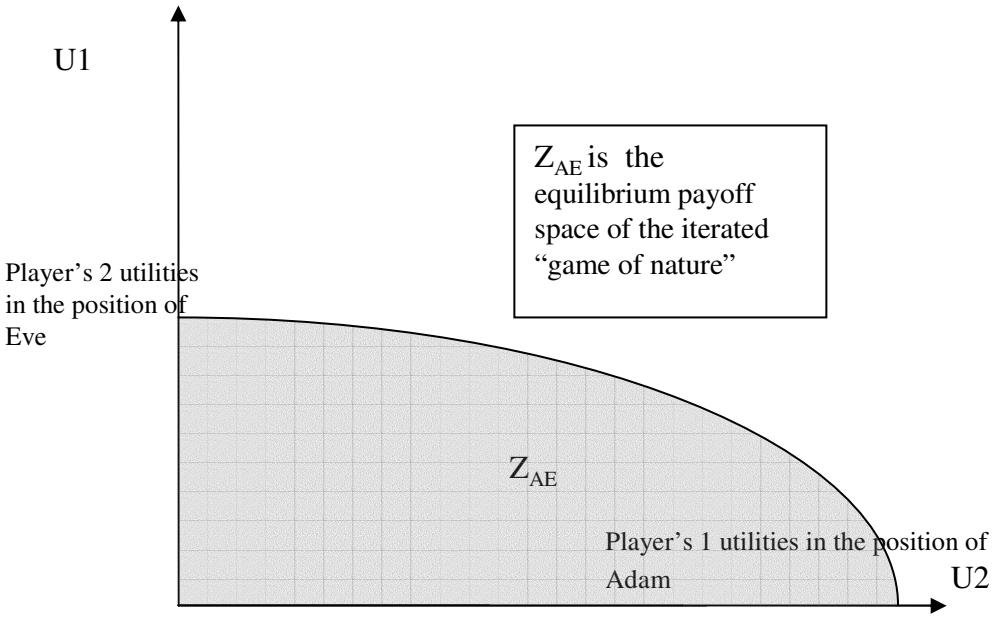


Some difficulties in the constitutional contract of the firm

- Instability of the equitable solution based on utility side payments when only asymmetric outcome spaces are feasible (property rights)
- The convex combination of points in P1 and P2 may not correspond to any feasible outcome
- The utility side payment is an outcome corresponding to a point in the convex combination of P1 and P2 ,
- No implementation mechanisms may exist for it

Binmore's theory of social contract

- The Game of life is a repeated game with multiple equilibria
- "Original position" is thought experiment for stable (independent on random exchange of position) selection mechanism within the equilibrium set



Binmore's theory :Original position and symmetry

- each player consider the entire set of possible outcomes as if he/her were in the condition to occupy both the role of Adam and Eve
- translation of the payoff space XAE: For each "physical" outcome of the original game (XAE) there is a symmetric translation that generates a symmetric outcome (a point in XEA) with the players' position reversed,
- Empathetic preferences allow to use the same utility units under the translation

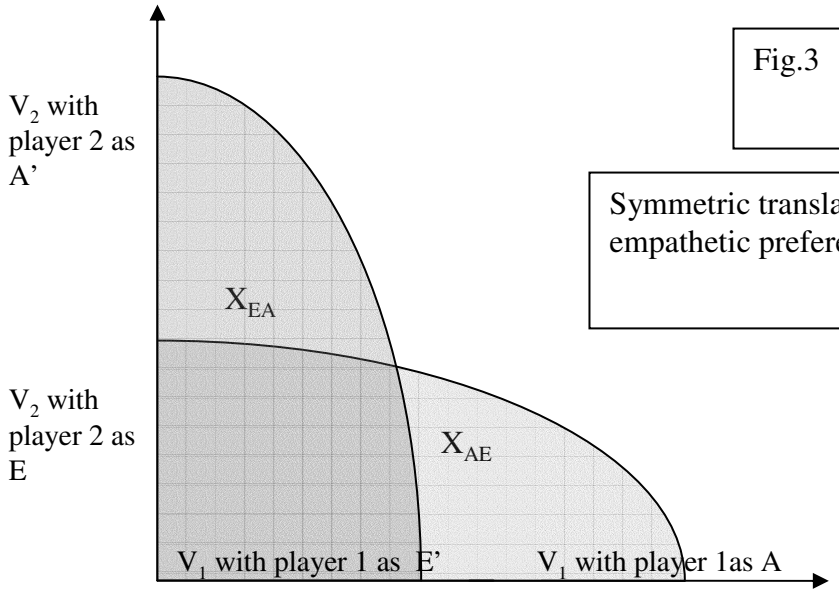


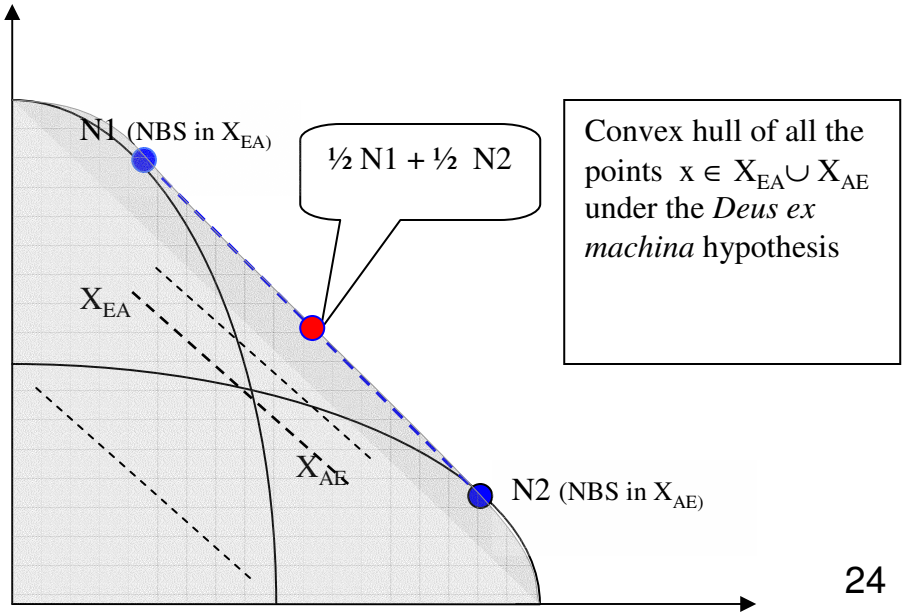
Fig.3

Symmetric translation under empathetic preferences

Binmore's theory: God provides for a convexity

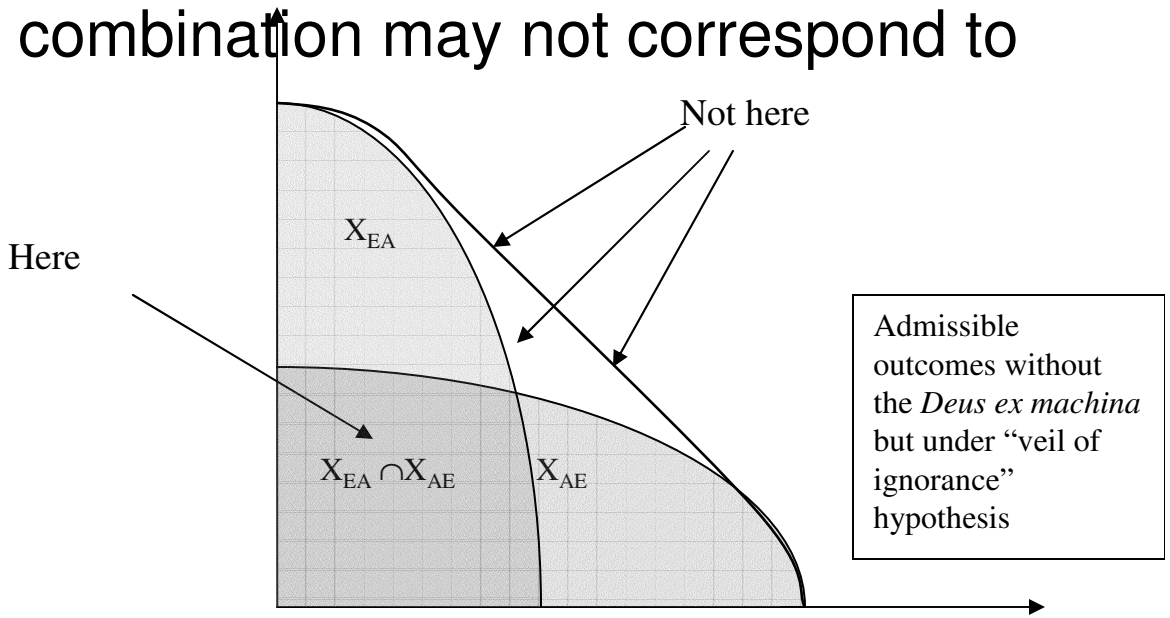
- "Veil of ignorance": only equal probability combinations of any outcome with its symmetric translation can be considered
- **Deus ex machina hypothesis** : some external all encompassing mechanism guarantees that whichever agreement on a convex combination will be implemented
- **Results**: equal probability combinations of utilitarian solutions or NBS

Fig.4



Binmore's theory: No Deus ex machina

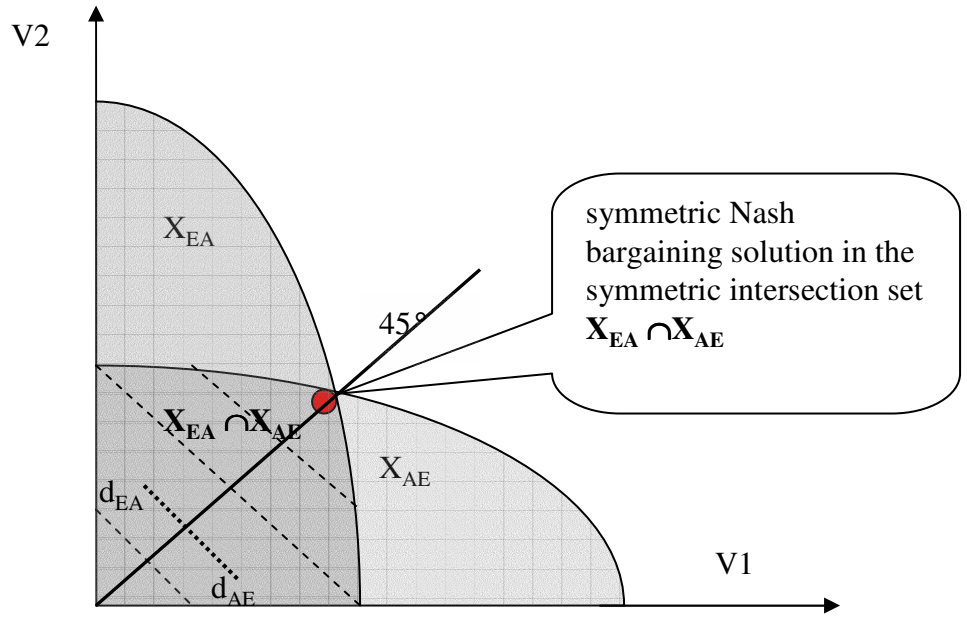
- Keep the veil of ignorance but skip the "Deus ex machina" : God not ready to enforce every ex ante agreement
- state of nature does not allows for an all encompassing enforcement mechanism
- need to consider of ex post stability (self-enforceability)
- Only ex post self-enforceable outcomes are feasible ex ante agreement
- What lie in the convex combination may not correspond to any feasible outcome



Egalitarian solution

- Restriction to the symmetric intersection sets: only here convex combination corresponds to equilibria no matter the result of the lottery
- The solution must lies on the bisector
- In asymmetric space NBS predicts the egalitarian solution

Fig.6



Rawls vindicated (also for non kantians)

- egalitarian solution corresponds to the Rawlsian maximin.
- Eve's payoffs are those attached to the disadvantaged player both as E or E', and they are maximised under the positions permutation,.
- egalitarianism basically depends on the requirement of *ex post stability* plus the ex ante requirement asking to make judgments under the veil of ignorance
- It is just because we cannot hypothesise an external enforcer, given empathetic preferences, that we are constrained to make and agreement within the basically symmetrical subset intersection XAE $\cap$XEA

Back to the constitutional choice amongst feasible governance structure of the firm

- If
  - the “veil of ignorance” hypothesis is introduced ,
  - but the *Dues ex machina* hypothesis is rejected,
  - given any set of feasible outcome spaces,

  Constitutional Choice selects the firm governance structure corresponding to the bargaining sub-game with the feasible outcome space P* such that the ***egalitarian solution*** in P* *dominates* egalitarian solutions of the alternative feasible spaces.

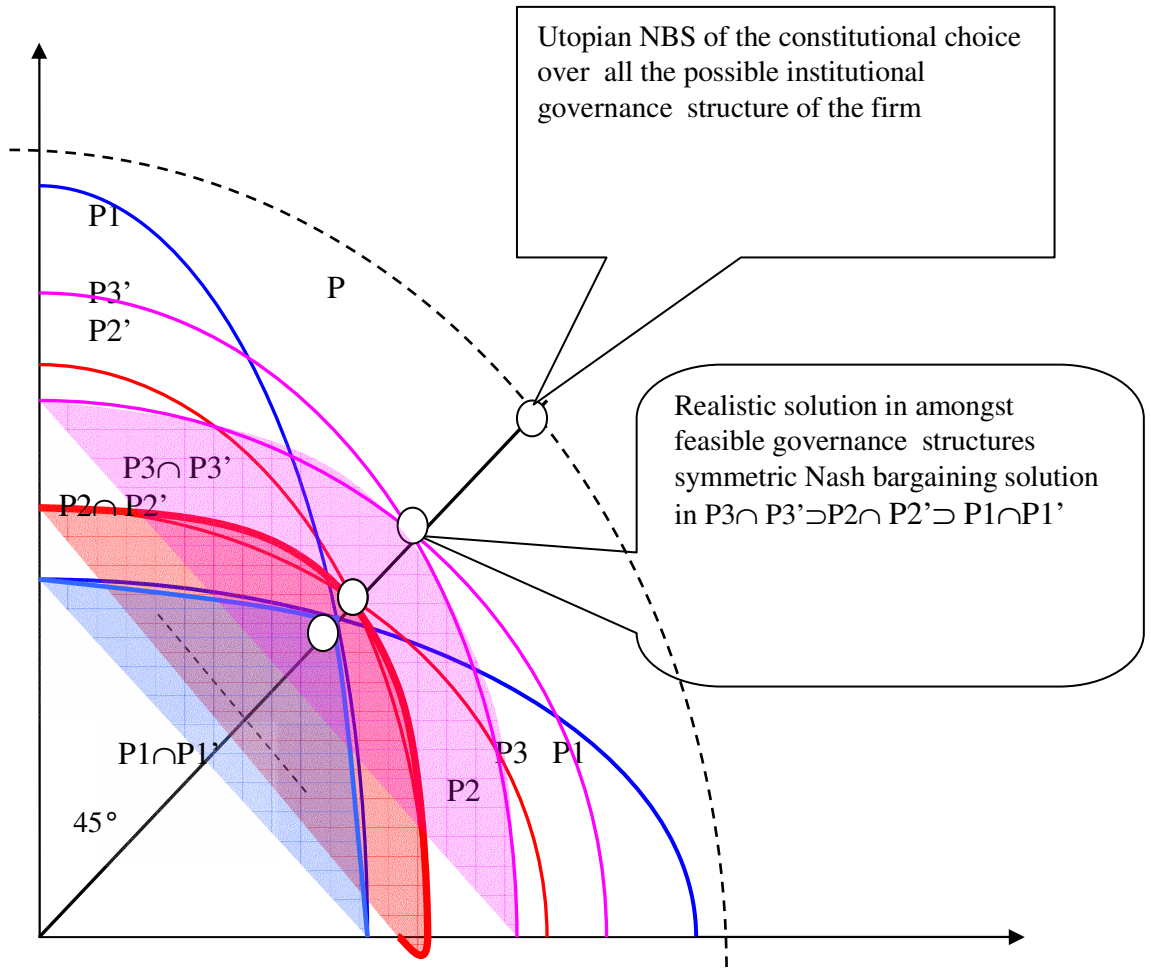
Moreover

- given any two feasible outcome spaces P1 and P2 and their symmetric translations P1’ and P2’,
- no matter how any other characteristics of the spaces is specified,
- if $P1 \cap P1' \supset P2 \cap P2'$

then $\sigma 1 > \sigma 2$, (where σ is the egalitarian solution within a given outcome space), and hence inclusiveness of the symmetric intersection is the only relevant characteristic

Egalitarian NBS within symmetric intersection spaces are monotonic

Fig.3



Corollaries: efficiency

- Consider the two feasible outcome spaces P1 ad P2
- (i) P1 includes both the maximal utilitarian solution and the highest solution in terms of Kaldor- Hicks efficiency
- but nevertheless (ii) P2, with its symmetric translation, generates an intersection set that includes the intersection of P1 and its own symmetric translation.
- Then any rational social contract must prefer the constitution of the firm corresponding to the outcome space P2 - no matter the efficiency properties of P1.
- In fact under the “veil of ignorance” the Utilitarian and Kaldor-Hicks solutions are not feasible
- The feasible “intersection” of P1 and P1 is less efficient than P2.
- Hence, equality constraints efficiency

Corollaries: freedom of Choice and spontaneous order

- Much new-institutional theorising about governance forms is based the implicit postulate that institution design cannot go further than prescribing outcomes interpretable as spontaneous orders.
  - normative presumption that freedom of choice must be respected
  - But also because only spontaneous orders are self-enforcing norms, such that they do not require the intervention of an external *Deux ex machina* who would constraint individual freedom
- But a mild libertarian would not reject that individual agents may enter the original position in order to make an assessment of possible spontaneous order outcomes

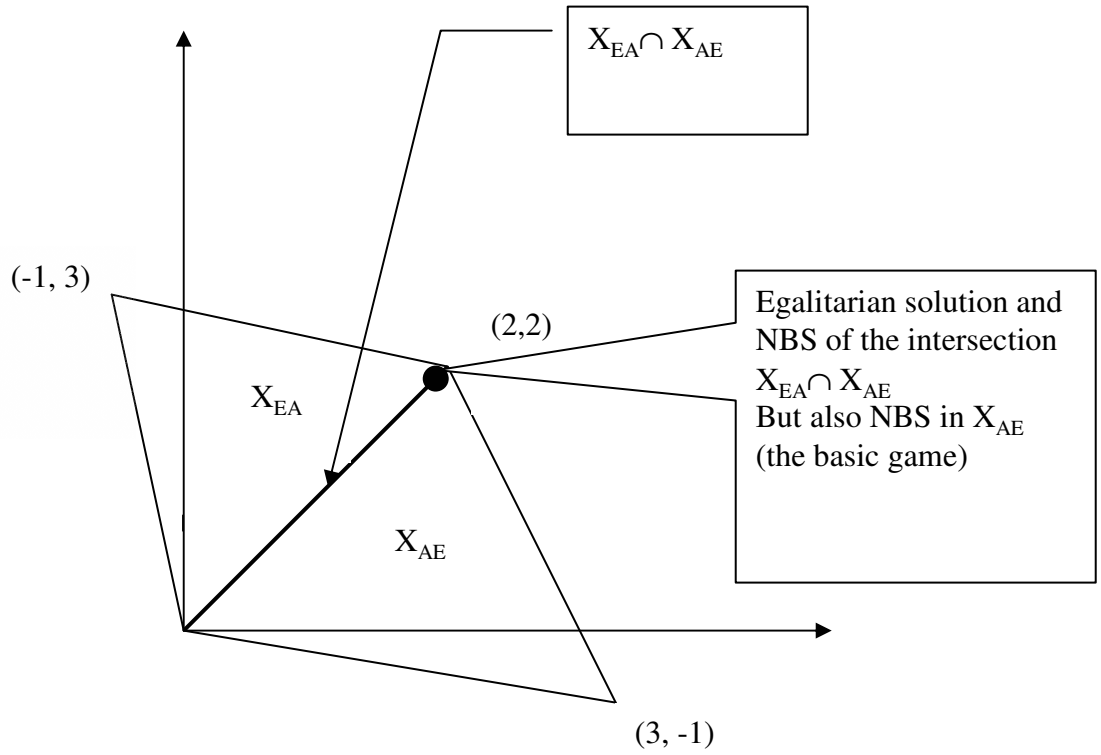
Corollaries freedom of Choice and spontaneous order

- However constraining the moral point of view with care for freedom of choice and stability (no *Deus ex machina*) has dramatic consequence to the libertarian point of view:

- only governance structures allowing for egalitarian payoffs allocations are acceptable.

- Far from ostracizing the “mirage of social justice” in the small scale society constituted by stakeholders of a firm, a moderate libertarian **cannot be but egalitarian** in the selection of the firm governance structure.

Back to the trust game: which equilibrium should be justified according to the contractarian theory?



Intuition: reputation would not work be without the reinforcement conformist preference and motivation

- Many stakeholders have preferences not purely self-interested or valuing only material advantages (consequences).
  - –These stakeholders also place importance on the firm's reciprocity in complying company's duties, especially if agreed upon in a public announced code.
  - – any deviation from the CSR standard (commitment), is punished more than would be the case if simple material interest were concerned.

C) The theory of conformist preference and its formal model

1) General form of the overall utility function

$$V_i(\sigma) = U_i(\sigma) + \lambda_i F[T(\sigma)]$$

- U_i is material utility for states σ (described as consequences)
- λ_i is an **exogenous psychological** parameter (a **disposition**) that expresses how important the ideal component is within the motivational system of player i
- T is a fairness (to be specified) principles defined over states σ
- F is a function (to be specified) of the fairness principle expressing both the agent's conditioned conformity an other individuals' reciprocal conformity to T

2) The exact form of the fairness-function *T*,

represents formally the ideal (the solution of the game under "original position" or ideal game)

=> Contractarian characterization of the principle *T*: Nash bargaining solution, i.e. *Nash social welfare function N*

$$T(\sigma) = N(U_1, \ldots U_N) = \prod_{i=1}^{N} (U_i - c_i)$$

- A contractarian principle reflects **non consequentialist** reasons to act
- the principle of fairness is impartially agreed on the basis of the different agent-relative reasons to act of each players

3) Definition of the two personal indexes of conformity (to specify *F*)

a) Player *i* personal index of conditional conformity (varying from 0 to -1) :

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

b_i^1 = belief of player *i* over player *j*'s action

$T^{MAX}\left(b_i^1\right)$ = maximum attainable by the function T given i's belief over j's strategy,

$T^{MIN}\left(b_i^1\right)$ = minimum attainable by the function T given i's belief over j's strategy,

b) Estimation function of player *j* index of reciprocal conformity to the ideal (varying from 0 to -1)

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

b_i^2 = player *i*'s *second order* belief over the belief of player *j* over the choice of player *i*

4) The Ideal component of the utility function

• These indexes are compounded in the following **ideal component** of the utility function:

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

• Hence The level of j's adhesion to a moral principle, as estimated by *i*, represents the **marginal incentive** for *i* to act according to the conformist motivation

- if i completely conforms and expect that j conforms too, then the value of its ideal utility is

λ×1×1 = λ

- If i not completely conform and expect not complete conformity on the part of j, then ideal utility is

(1-x) (1-y) λ<λ

- If conformity is nil at least for one player then ideal utility is

(1-1)(1-y) λ =0

How conformist preferences prevent refined abuse

- To be parsimonious, assume that only stakeholders adhere to the ideal of the socially responsible firm
- This presupposes that the firm has at least signalled a commitment to such an ideal, but *not necessarily* that it has developed conformist preferences for reciprocal compliance with it.
- Hence set to 0 the λ_B parameter in the manager's or entrepreneur's utility functions (weight of conformity within player B utility function)
- Stakeholder A on his part has a positive weight λ_A

Beliefs

Player A's relevant first-order and second-order beliefs are:

b_A^1 = *(2/3**a**, 1/3* ***no-a)*****,** in short *(2/3,1/3)*

b_A^2 = ***e,*** b_A^2 = ***no-e***

We define player A's overall utility function for two situations

A) when she believes that player B will abuse with probability 2/3 and not abuse with probability 1/3, while she has the second-order belief that player B predicts that **she (player A) will enter**.

B) *When she* believes that player B will play the mixed strategy *(2/3,1/3),* but she will **not** play the entry strategy, so that her second-order belief is that she herself ***does not enter*** and the firm predicts that she will not enter (formally b_A^2 = ***no-e***).

Calculating player A conformity indexes

Case A* : player A strategy ***e***, given beliefs *(2/3,1/3)*

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

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

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

- these 0-levels of the conformity are better understood as degrees of deviation from complete compliance
- In both the A* and A** cases, player B's mixed strategy (*2/3,1/3*) nullifies any effort that player A might make to enhance the level of ideal attainment. Whatever player A does, in fact, the level of *T* is always 0.
- Thus A has **no responsibility** for any deviation from the maximum feasible level of *T*, given B's choice.

Calculating Player B expected conformity index

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

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

- Here B's strategy implies a marked deviation from maximal conformity conditional on A's behaviour of entering.
- The deviation can be *imputed entirely to player* B's decision to play his mixed strategy instead of his ***no-a*** strategy.
- In this case player B does not conform with the ideal at a significant level, and this results in the negative value assumed by his expected conformity index.

Player B's (expected) conformity index (2)

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

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

– Given his belief ***no-e***, player B cannot significantly deviate from the ideal, hence he is not accountable for a deviation from the maximal ideal's value given **no entry** by player A.

– Comparing B* and B** shows that the *intention* to exploit player A's acquiescence *implies that B has a significant responsibility* for a deviation from (non-conformity with) the ideal only conditional on the expectation that in effect player A will give in

Player A overall utility function

- Player A's overall utility values for the two alternative strategies ***e*** and ***no-e*** respectively, given that he predicts player B will use strategy (*2/3,1/3*) :
- Paying "**enter**": the material payoff' is (practically) 0, whereas her conformist utility is based on indexes A* and B*.
- Thus player A's overall utility for strategy ***e*** is

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

- Playing "**no enter**" : her material payoff is again 0, whereas conformist utility is given by indexes A** and B**,
- Thus player A's overall utility for strategy ***e*** is

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

A straightforward conclusion

- Player A with conformist preferences refuses to give in to the mixed equilibrium strategy of the repeated trust game.
- Granted that λ_A is positive, this result typically follows from the opportunistic nature of player B's mixed strategy type
- the logic of strategic choice under conformist preferences reverses the result of standard strategic calculation

A more complete picture

- a complete explanation of the endogenous observance of voluntary CSR norms would be based on the analysis of the equilibria emerging from an evolutionary trust game
- Hp1: the firms population, from which players are selected at random, is a mix of types:
  - **enlightened self-interested idiosyncratically committed to a code of ethics with $\lambda = 0$**
  - **sophisticated opportunists (playing the mixed equilibrium strategy), $\lambda = 0$**
  - **conformist ideology-driven firms (with $\lambda \sim 1$)**
- Hp 2: stakeholders are selected from a population also composed by a mix of types
  - **just endowed with materialistic preferences or**
  - **endowed with ideal - conformist preferences.**
- The result may be quite counterintuitive to the economists' wisdom that cooperation emerge from enlightened self-interest.

An insight about evolution

- When there are just enlightened-egoist firms and consequentialist stakeholders:
  - **A cooperative evolutionary equilibrium emerges based on reputation**
- By mutation, sophisticated opportunist firms now enter:
  - **the equilibrium of refined abuse emerges (enlightened egoists are displaced by sophisticated opportunists)**
- By mutation, conformist stakeholders now enter:
  - **An equilibrium emerges whereby sophisticated opportunist simulate enlightened egoists**
- Last, by a new mutation, conformist firms enter:
  - They are more efficient in accumulating reputation
    - They give a more reliable initial signal about the the honest type
    - Their opportunity cost for cooperation are lower for, given internal expectations of reciprocal conformity, they gain psychological utility which countervails a lower discount rate for future payoffs (shadow of the future)
  - **A psychological equilibrium of reciprocity emerges**

Economists may be wrong in predicting ethical behavior because it is profitable in the long run

- There are two main basins of attraction separating types
- In the region where materialist STKs prevail, sophisticated firm fare better than enlightened (equally well than the conformist)
- in the region where there is a sufficient number of conformist STKs, conformist firms displace both enlightened and sophisticated opportunists
- **Result**: those firm who gain intrinsic utility from reciprocal conformity are better utility maximiser, an then they occupy the ecological niche before retained by the enlightened
- Enlightened egoists would disappear in the long run.