

A Rawlsian view of CSR as a Multistakeholder model of corporate governance

By

Lorenzo Sacconi

Definition of CSR

- 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

Fiduciary duties

- **A subject has a legitimate interest but is unable to make the relevant decisions**
 - He (the *trustor*) delegates decisions to a *trustee* entitled with the authority to choose actions and goals.
 - For a fiduciary/authority relationship to arise, the trustor must accept it because it is functional to some interests he already holds
- **Hence these interest are legitimate claims over the trustee**
- **These claims impose fiduciary duties on the trustee who must be accountable for that**

Stakeholders

- individuals or groups who have essential interests “at stake” in the running of the firm both because
 - they make **specific investments** in the firm
 - *undergo* the ‘**external effects**’, positive or negative, of the transactions performed by the firm
- These are both “**primitive**” stakeholders considered in terms of the **urgency** of their interests (relevant for social welfare analysis) and **not** according to their power of threat

■ *The scope of CSR*

the model 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 transaction-cost rationale for extending fiduciary duties

□ Why do companies exist?

- Contracts are incomplete,
 - **unforeseen** contingencies
- Investments may be **specific**
- Behaviors are **opportunistic**: try to renegotiate incomplete contracts

□ Renegotiation induces the expectation that investments will be expropriated

- **it destroys incentives to make efficient investments**

.....extending fiduciary duties (continues)

- **the firm Corporate Governance structures** allocate residual rights of control to the owners , i.e. authority over the ex ante not contractible aspects of transactions
 - Renegotiation will **not threaten** them
 - their **investments** are **safeguarded** from the other stakeholders' opportunism

BUT.....

- The firm is team production: **many stakeholders cooperate by means of their specific investments** (human capital, social capital, trust etc.)

There is always “abuse of authority”

- Those who holds residual control **appropriate** the full surplus by **expropriating** other stakeholders' investments
- If fiduciary duties are only attached to ownership
 - Non-controlling stakeholders will **not** be **protected**,
 - while their contracts are nevertheless **incomplete** (contracts are not the solution)
- That's why “control structures are **always second best**”: some have the incentive to **over-invest**, others have the incentive to **sub-invest**
- **Equity and efficiency problems cannot be separated**

A first Rawlsian intuition: the maximin principle as the proper balancing criterion among different stakeholders claims (1)

- For mere **incentive reasons** those who are in the position to carry out the **most important investment** must be granted **residual control**,
 - which in general will **induce inequalities** and gives him the opportunity to **abuse** non-controlling stakeholders
- But, since the firm is a **joint venture for mutual advantage**, disadvantaged non-controlling stakeholders must also **benefit** from cooperation.

Rawlsian imaximin principle as the proper balancing criterion among different stakeholders claims (2)

- Non controlling (worst off due to abuse of authority) are granted the right **to veto** any control structure
 - **unless** it is not the **better one for the worst-off** stakeholder themselves (with respect to all the available alternatives)
- Thus, to legitimate a unilateral control structure, (wherein ownership is held by the stakeholder undertaking the most important investment)
 - the implementation of a **redress principle** is necessarily required.

A comprehensive structure of corporate governance (1)

- The firm's control structure legitimacy is granted if the residual control right is accompanied by **further fiduciary duties** owed the subjects **not controlling** the firm
- 'extended governance' should comprise:
 - the **residual control right** (ownership) allocated to the stakeholder with the **largest investments** at risk and with relatively **low governance** costs
 - **fiduciary duties** of those who effectively run the firm (directors and managers) owed **to owners**,
 - **fiduciary duties** of those in a position of authority in the firm (the owner and/or delegated directors and managers) owed to **non-controlling** stakeholders

A comprehensive structure of corporate governance (3)

- **Fiduciary duties owed to non controlling stakeholder include**
 - the obligation to run the firm so that
 - the company distributes to each *strict-sense-stakeholder* a **‘fair share’ of the surplus**
 - while the broad-sense stakeholders are **immunized** against negative externalities;
 - effective **accountability** to the non-controlling stakeholders in terms of relevant information related to their legitimate interests and rights
 - the right to be **represented** in corporate bodies where they can exercise effective **supervision** over the owner’s and directors’ compliance with their extended fiduciary duties

A comprehensive structure of corporate governance (4)

- According to this revision of the corporate governance structure, boards of directors or managers appointed by owners
 - owe a ***special fiduciary duty*** to the ‘residual claimants’ who have directly delegated authority to them (*via* a narrow fiduciary proviso).
 - This duty applies, however, only under the constraint of a ***more general fiduciary proviso*** relative to *all* the stakeholders
 - which is defined ***via duties owed to non-controlling stakeholders.***

Objective function of the SR firm

- Run any corporate activity in the way that it
 - **minimizes negative externalities** affecting stakeholders in the broad sense
 - **maximization of the joint surplus and its simultaneous fair distribution**, as established by the impartial cooperative agreement among the stakeholders in the strict sense
 - When more than one option is available in the above-defined feasible set, choose the one that **maximizes the *residual* allocated to owners** (the shareholders).

CSR as a matter of explicit **self-regulation**, **soft law** and **social norms**

- **Not** mere managerial **discretionary** decisions, but
- Rules of behaviour established through **explicit** social norm and standard
 - general **principles** of fair treatment for each company stakeholder,
 - principles of **inter-stakeholder** fair balancing,
 - **precautionary rules of behaviour**, so that fiduciary duties are put in practice by rules of conduct that preempt opportunistic behaviour in typical critical situations;

CSR as a matter of explicit self-regulation, soft law and social norms

- **Agreed upon by both firms and stakeholders through (voluntary) forms of multi-stakeholder social dialog (deliberative democracy)**
- **Self-imposed by firms on themselves without external legal enforcement, but instead by means of the internal adoption of statutes and codes of ethics , self-organization, training, auditing, and reshaping of incentive**
- **Monitored and verified by third-party independent civil society bodies;**

Complementarities: regulation and self-regulation

- Effective CSR self-regulation is a viable option only within an legal environment that does not obstruct it
- Such obstruction would occur in the case of **too narrow definitions** of the firm's objective-function such as
 - **shareholder value maximization as the sole corporate goal**
- The 2006 UK company law reform is an example of how the board duties may be enlarged to legitimate balancing decisions
- it opens the door to self-regulatory CSR standard that more precisely specifies CSR principles and guidelines
- On being asked to account for their decisions, boards would appeal to such codes in order to justify their behavior to stakeholders.

The original position perspective

- In order properly to assess the **implementation** of CSR, take the perspective of a hypothetical '**state of nature**'
 - It logically **precedes** historical legal constructs that may legally obstruct the emergence of such a normative model.
- Hence
 - Admitted that company laws do not obstruct proper self-regulation,
 - **endogenous beliefs, motivations and preferences** of economic agents (companies and stakeholders) **are the essential forces** driving the implementation of the CSR model of multi-stakeholder governance.

Aoki's definition of institution is appropriate

- An institution is “**a self-sustaining system of shared beliefs about a salient way in which the game is repeatedly played**”
 - It is a rule not in the sense of “rules exogenously given by the polity, culture or a meta-game”,
 - but in the alternative sense of “**rules as being endogenously created through the strategic interaction of agents, held in the minds of agents and thus self-sustaining**”
- In order for beliefs to be shared by agents in a self-sustaining manner (...) the content of the shared beliefs” must be “**a *summary representation (compressed information) of an equilibrium of the game*** (out of the many that are theoretically possible).
- The salient feature of an equilibrium may have corresponding **symbolic representation inside the minds** of agents and coordinate their beliefs” (Aoki 2001)

Limitation of Aoki's def. of institution

- Institutions in the above game-theoretical definition **only *ex post*** tell the best action to each player
 - **once the participants shared knowledge that they have already reached an equilibrium state**
- They tell players only **how to maintain the existing pattern of behavior** supported by the existing beliefs system.
- Institutions thus are **devoid** of any significant **normative** meaning and force.
- On the contrary, norms like constitutions or laws, ethical codes, shared social values, organizational codes of conduct and procedures **have primarily a prescriptive meaning**

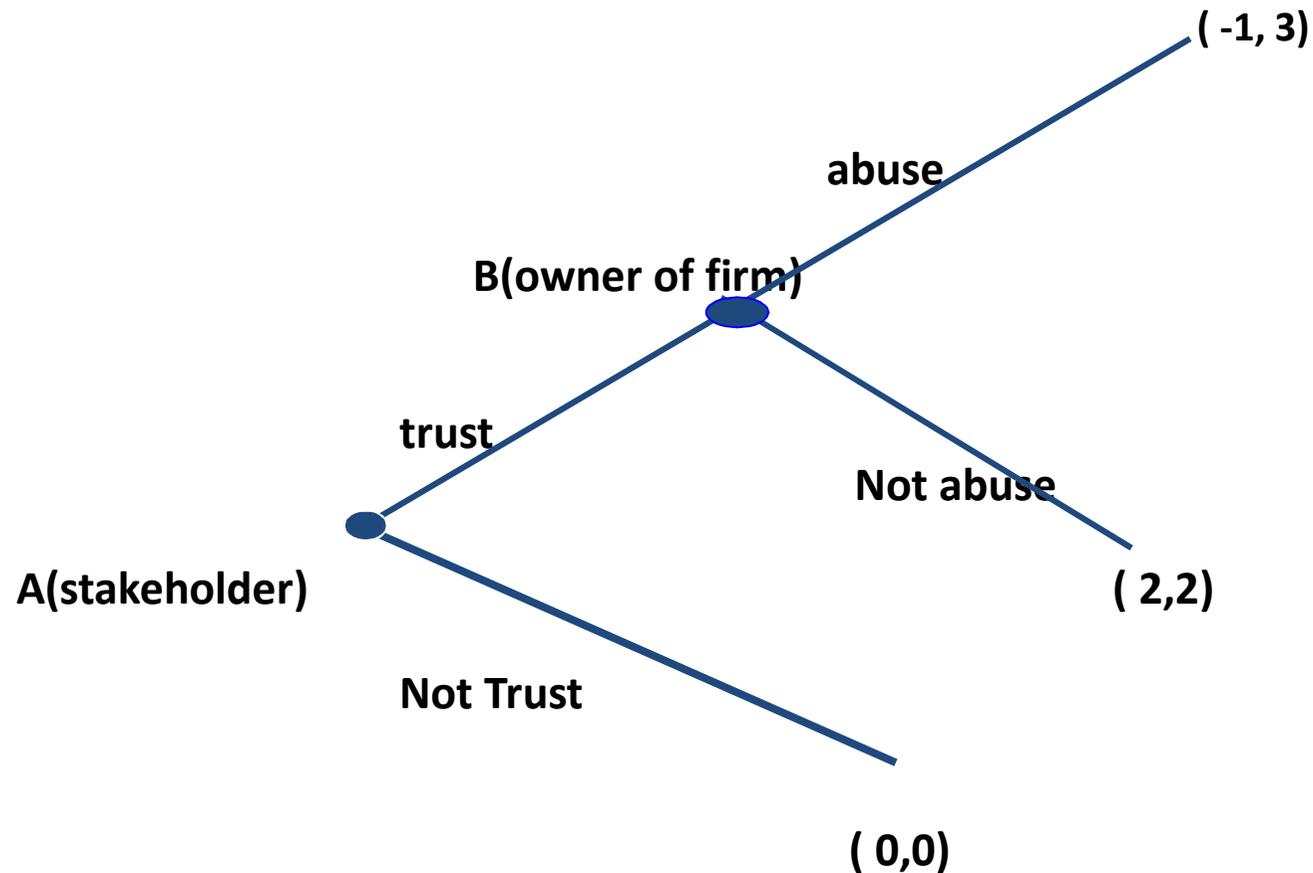
Integration of the definition of institution

- A **second** component of a proper definition of institution is the **mental representation of a norm**,
 - expressed by **utterances** in the players' language concerning **values** and **obligations**,
 - With **prescriptive** and **universalizable meaning** able to **justify** shared acceptance by all participants in a given interaction domain.
- It enters their **shared mental model** of how the game **should** be played and hence becomes the basis for their **coordination on a specific equilibrium**.
- The key point is explaining **how** a normative **beliefs system** (preceding the evolution of the corresponding equilibrium), **becomes accepted by all agents** in the relevant domain.

The social contract as integration of the definition of institution

- The best justificatory account for norms, entailing ex ante **shared acceptance**, is the ***social contract model***
- Contractarian norms result from a **voluntary agreement** in an hypothetical original choice situation which logically **comes before any exogenous** institution is over-imposed on a given action domain
- To define the agreement, the social contract model sets also aside threats, fraud and manipulation **because** these , resources **contingent on arbitrary historical institutions**
- Thus a norm arises and can be maintained only because of the **voluntary agreement** and adhesion of agents.

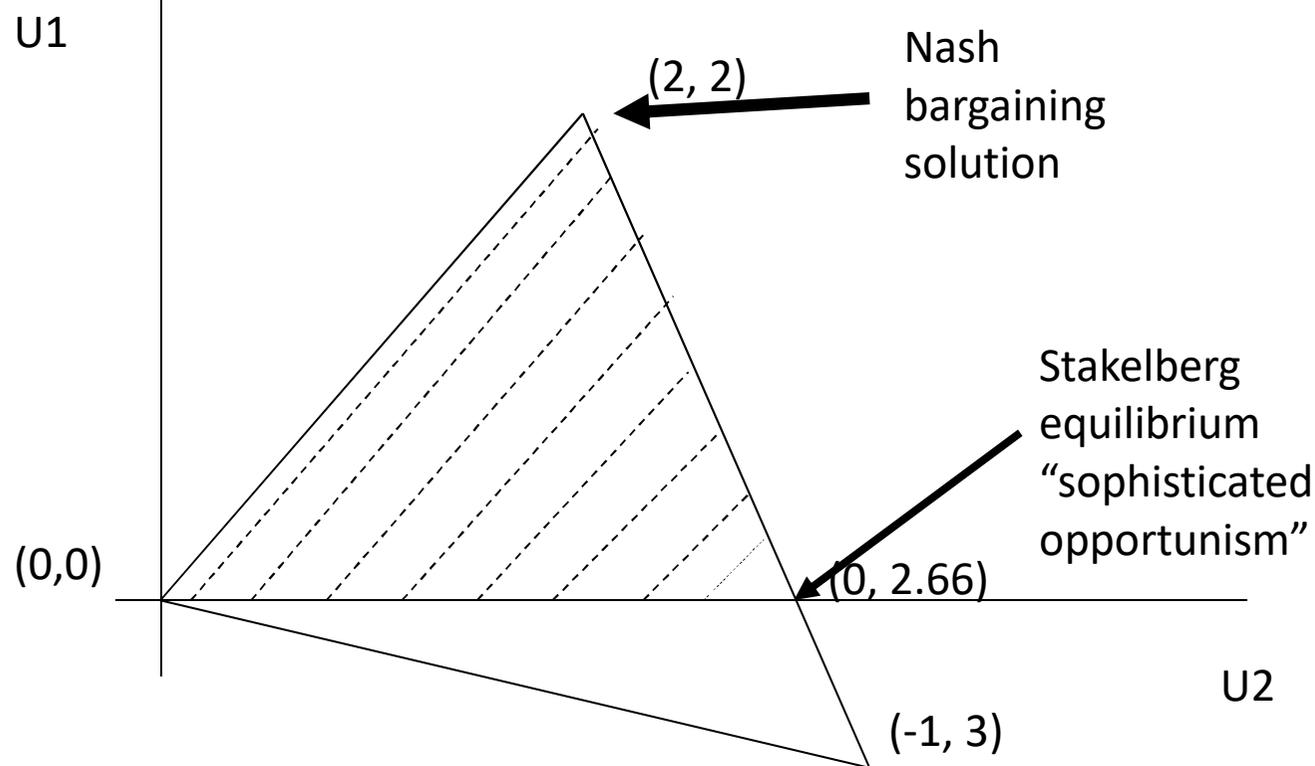
The structure of the stakeholders / firm interaction: a trust game



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

Multiplicity of equilibria in the repeated trust game

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



Given the interaction stakeholder/firm a Rawlsian SC plays **two roles**

- The *normative role*:

“what pattern of behavior the firm and its stakeholders **must select from the set of possible equilibrium patterns** when they put themselves under the *ex ante* standpoint of an **impartial agreement**?”

- The *motivational role*:

“**what and how many equilibrium patterns** of behaviors would *ex post* **retain their stability and motivational force** if firm and stakeholder were able to agree on a CSR standard in an *ex ante* perspective?”

1) The need for a normative balancing principle

- “Stakeholder” is *descriptive*:
 - there are **many** classes of individuals holding a stake in the firmbut
 - Stakeholder claims may also be **conflicting**
- Stakeholder theory needs a **normative** principle in order to say
 - **how to balance conflicting claims,**
 - **how to identify those claims that generate fiduciary duties**
- Thus a **normative *criterion*** is needed to find out a balance that
 - **Can be impartially accepted in advance by all the stakeholders**

The normative role of SC in the selection of an impartial equilibrium

- The function of a **fairness principle** is to give impartial reasons for **singling out a unique equilibrium** solution amongst the many possible.
- This will be a **particular equilibrium point** coinciding with an outcome rationally acceptable
- Note that the normative principle is used to single out an equilibrium point **within the equilibrium set** of the game
- The perspective is that of an *ex ante* impartial choice, **but it now concerns equilibria**, i.e. game solutions that are self-enforceable.

The constitutional contract theory on the control and responsibility structure of the firm

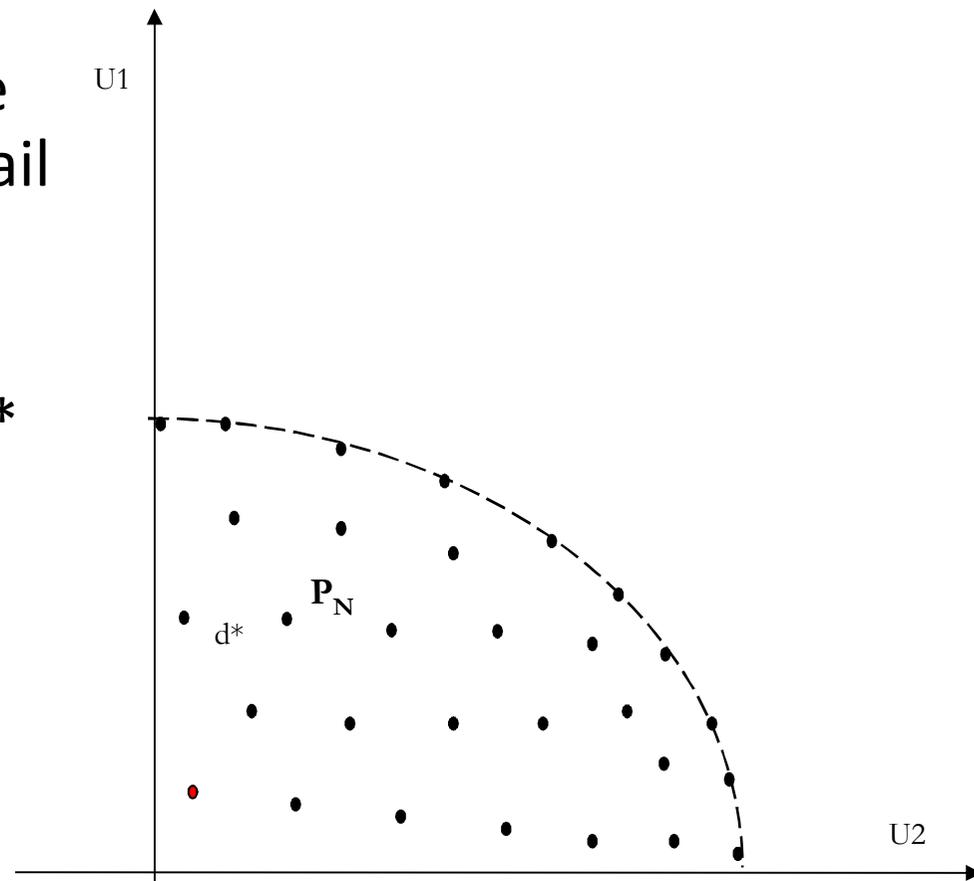
- A two-step collective decision-making among potential members of a coalition S .
 - At time $t = 0$ the **allocation of rights** is decided (not only ownership and control but also **redress**), and this determines the control structure exerted over the productive coalition S
 - At time $t = 1$ the 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, under the given each allocation of rights

The model

- This problem is modelled as a compounded bargaining game G_c of the constitutional and post-constitutional decisions
 - 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

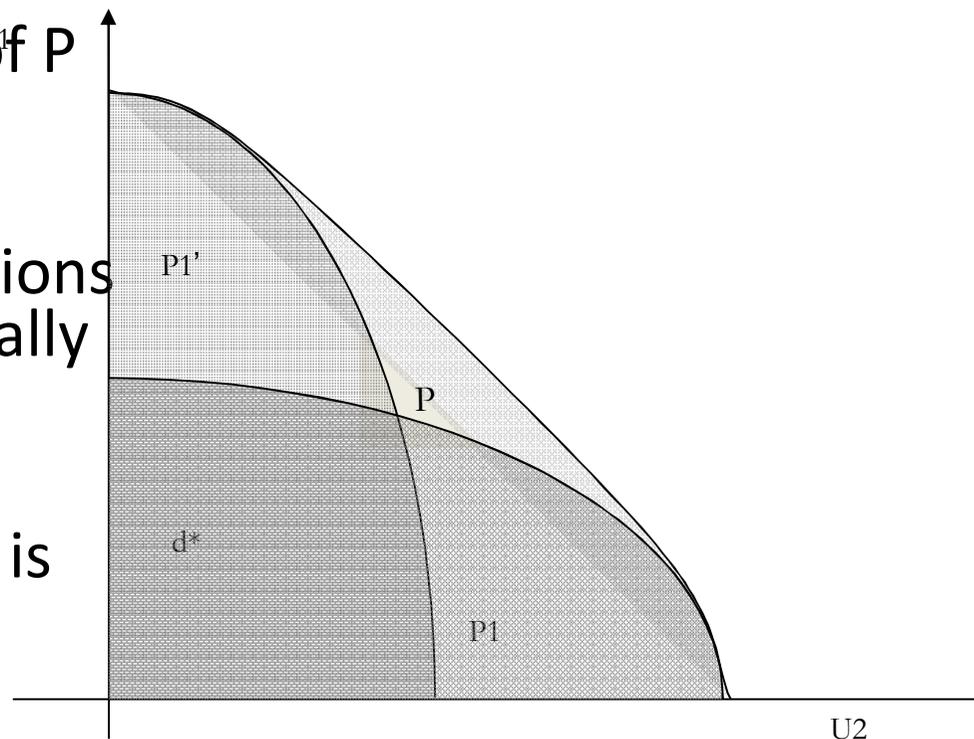
The GN game

- In the background there is a **non cooperative ‘state of nature game’** to which the players will resort if they fail to agree on a constitution.
- The ‘**state of nature game**’ admits a **single solution d^*** which is mutually disadvantageous to all parties,.
- the ‘state of nature’ arises when contracts are **left incomplete**, so that the parties undergo reciprocal opportunistic behaviour



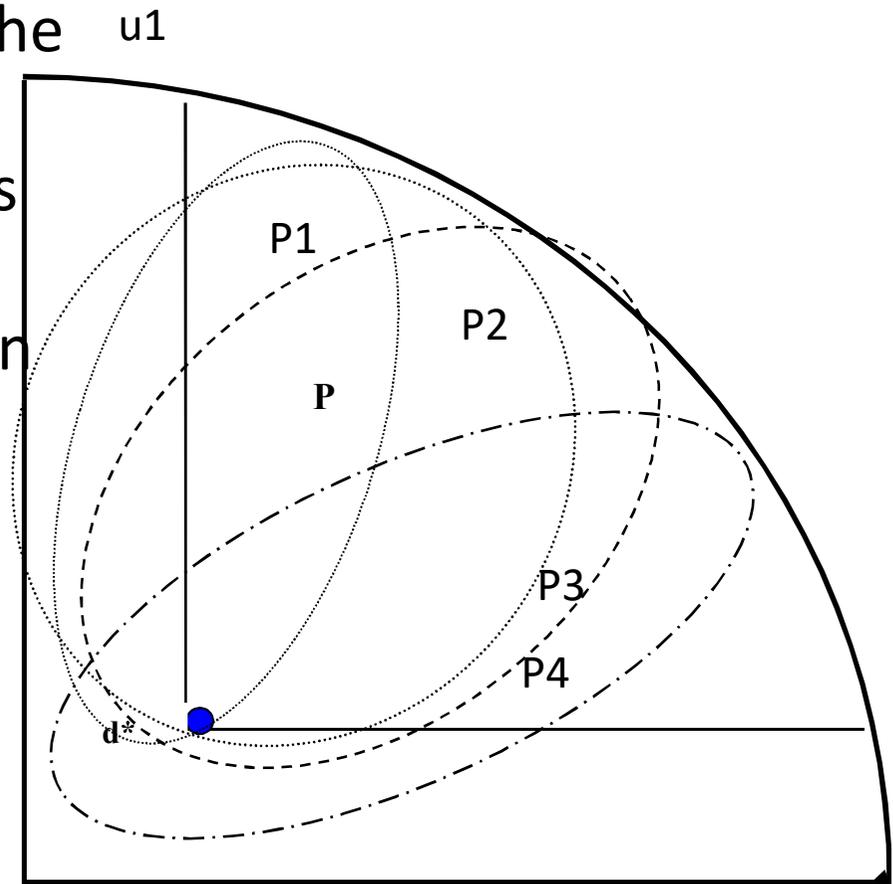
The Gc game

- **Gc** outcome space **P** consists of the 'state of nature' equilibrium d^* + the convex combinations all other possible outcomes
- Also symmetric translation of **P** are admitted (exchanges of position)
- Agreements over a constitutions can generate whatever logically possible outcome
- The former was a *non-cooperative* game, whilst Gc is *cooperative*
- players enter a **thought experiment** for they **assume** that **whatever possible agreement is implementable**



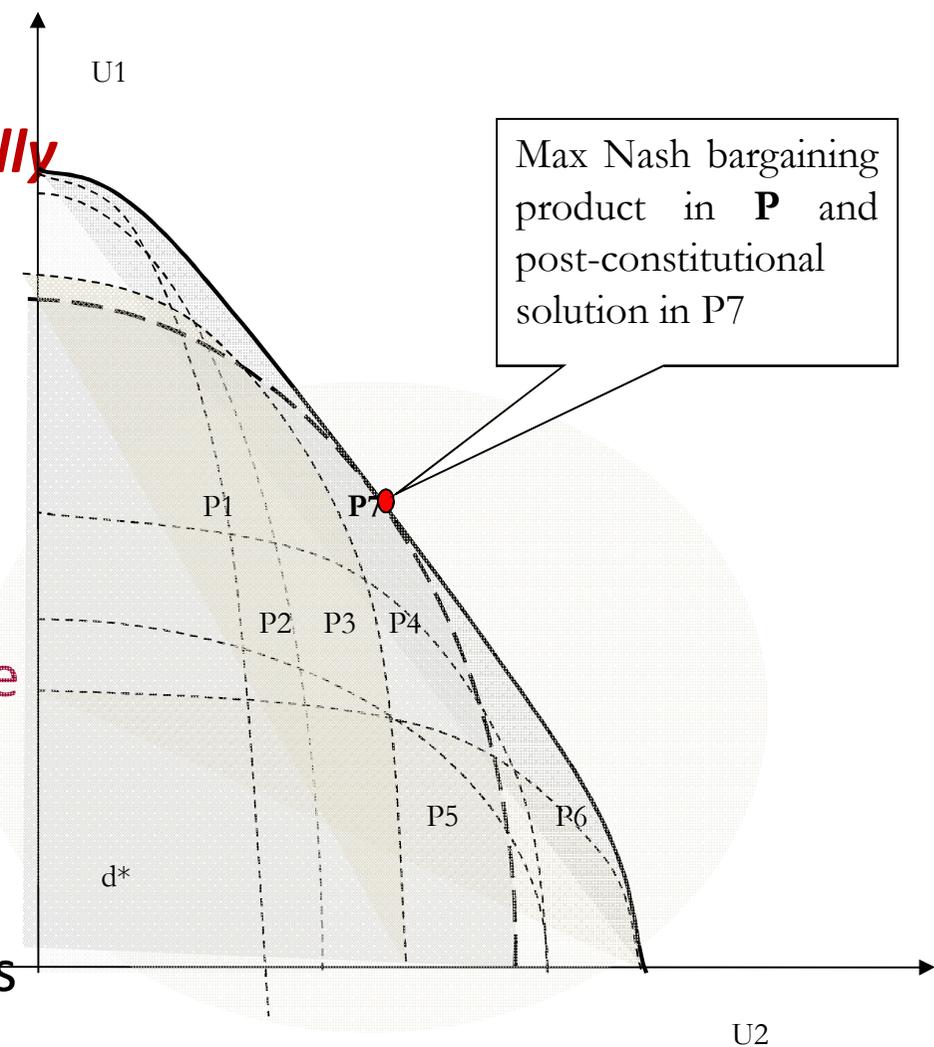
A distinctive feature of constitutional choice

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



Nash B.S. for the choice of the constitution

- In Gc the solution is to be found within the **symmetrical** outcome space generated **by all the logically possible** subsets of Gc
- All the points in P are seen as **solutions for possible** post-constitutional games.
- The selected constitution is such that the corresponding post-constitutional game will **distribute equally the cooperative surplus** calculated with respect to the symmetric Gc space
- This constitution distributes rights so that **no party has an ex post advantaged bargaining position**

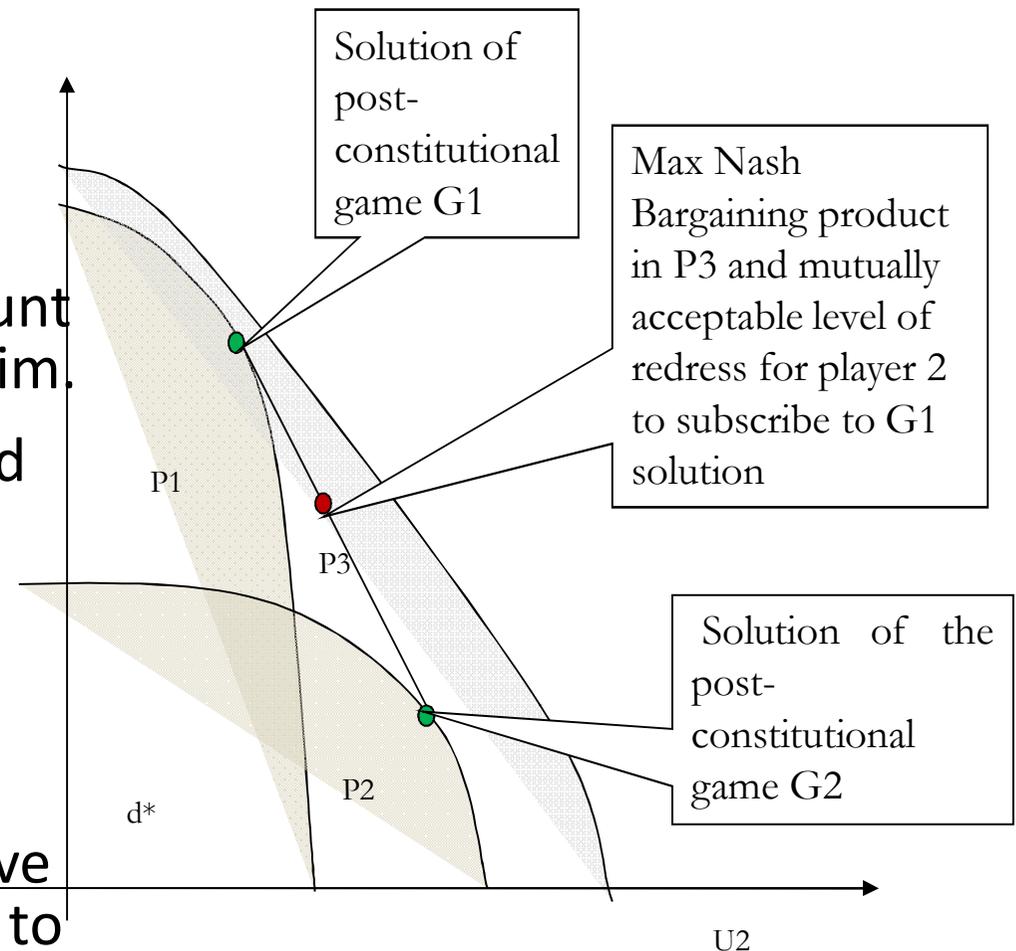


Institutional Feasibility

- So far **every logically possible** constitution has been considered (a world in which it would be possible to allocate decision rights in whatever proportion)
- More **realistic** is that only a **certain number** of restrictions on the set G_c are institutionally feasible.
- Assume that only **exclusive allocations** of property rights on the physical assets are institutionally **feasible**.
 - **Not** allowed **intermediate** degrees of authority,
 - feasible constitutions are such as **to bias** post-constitutional bargaining heavily in favour of one or other party.
- the N.B.S. **is not the same** than that relative to the all-inclusive payoff space of GC (the choice must fall within the set of **institutionally feasible** solutions)

Choosing constitutions under feasibility

- Two feasible constitution G1 and G2 are considered
- defined for G1 is **more efficient** than that of the alternative G^{U1}
- Ownership **must be given to 1**
- However, 1 must still take account of 2's claims and **compensate** him.
- The solution must be calculated **within the payoff space P3** generated as the **convex hull** of the combinations of outcomes belonging to P1 and P2.
- By a **utility side-payments** 1 redresses 2 until the cooperative surplus is distributed according to the **criterion of relative need**.

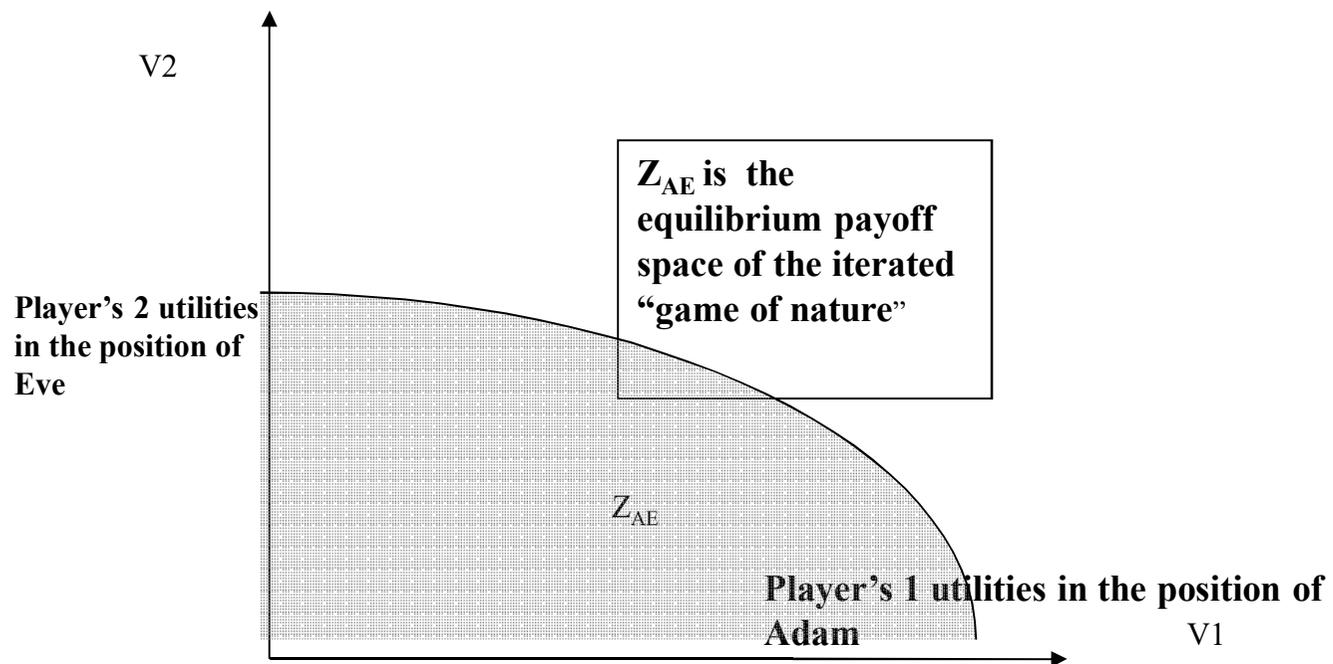


Difficulties in the constitutional choice among institutionally feasible governance structures

- **Instability** of the **equitable solution** based on utility side payments when the only underlying feasible outcomes are asymmetric outcome space (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** but **outside** both P1 and P2 ,
- **No implementation mechanisms may exist for it**

Binmore - Rawls theory of social contract

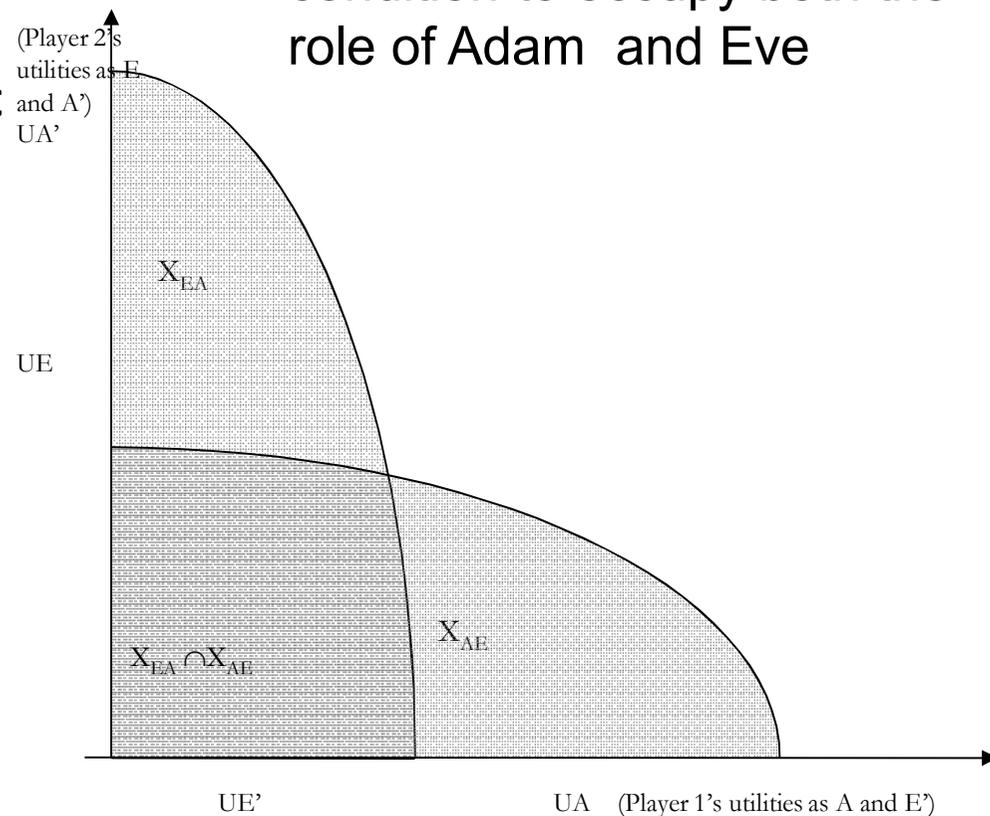
- The Game of life is a repeated game with multiple equilibria played by two player (1 and 2)that can take the social role and identity of Adam and Eve
- The “original position” is a thought experiment for the (stable under symmetric exchange of personal positions) selection of one equilibrium within the equilibrium set



Original position and symmetry

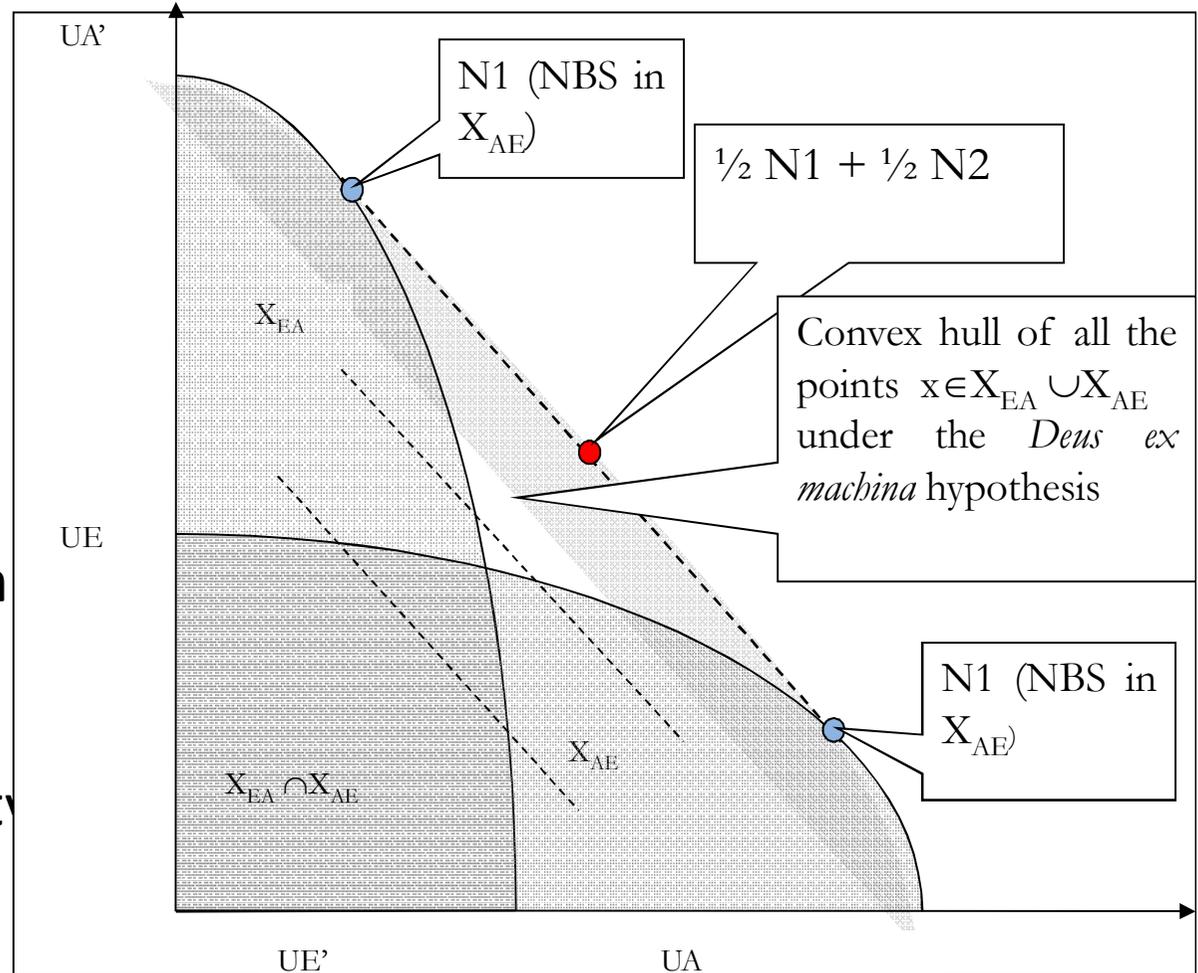
- **translation of the payoff space** X_{AE} : For each “physical” outcome of the original game (X_{AE}) there is a symmetric translation that generates a symmetric outcome (a point in X_{EA}) with the players’ position reversed,
- **Empathetic preferences** allow us to use the **same utility units** under the exchange of Adam and Eve positions between player 1 and player 2

□ each player (player 1 endowed with V_1 , and player 2 endowed with V_2) considers the entire set of possible outcomes as if he/her were in the condition to occupy both the role of Adam and Eve



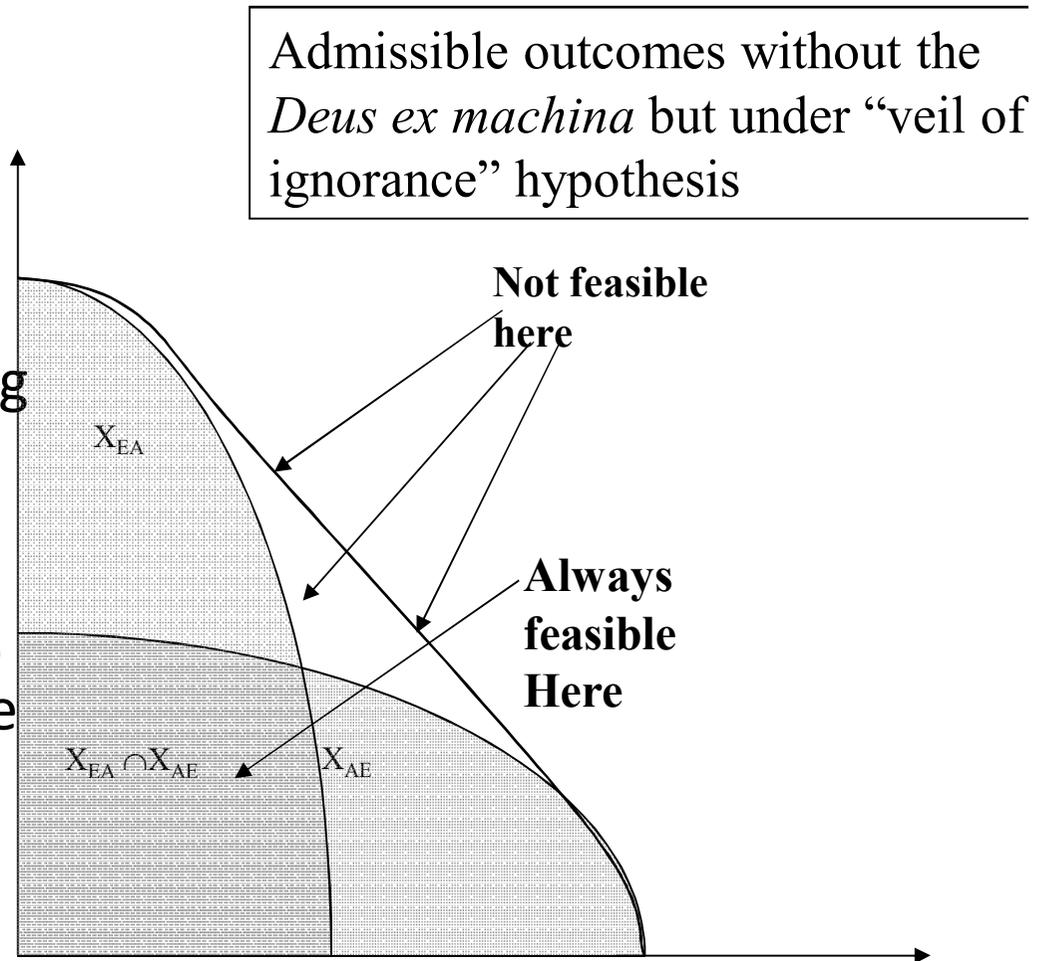
Equiprobability and convexity

- “Veil of ignorance”: only equal probability combinations of each outcome with its own symmetric translation must be considered
- **Deus ex machina hypothesis** : some external mechanism guarantees that **whatever agreement on a convex combination** will be implemented
- **Results**: equal probability combinations of utilitarian solutions or NBS



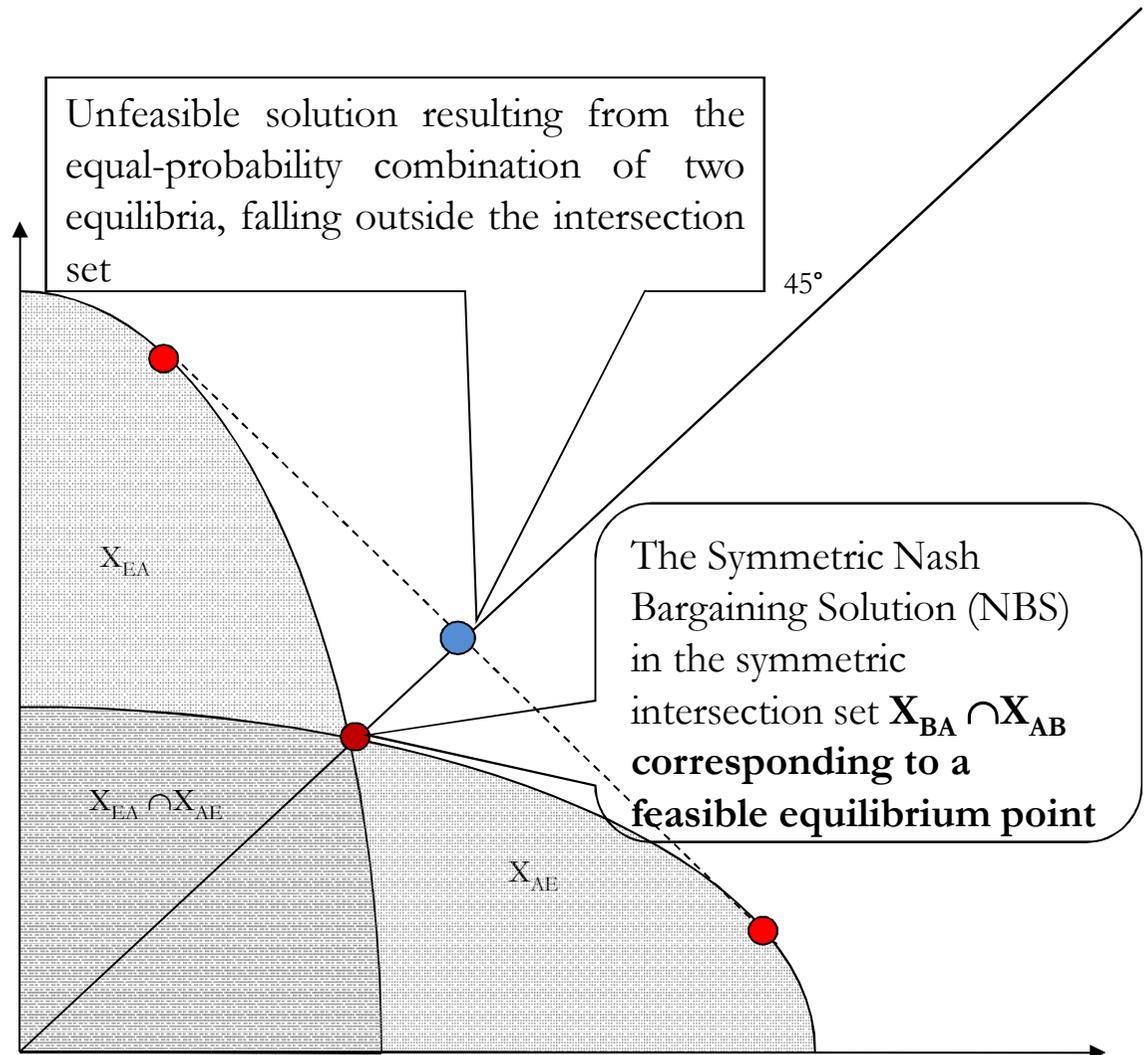
No Deus ex machina

- Unfortunately **God is not ready to enforce any ex ante agreement**: hence keep the veil of ignorance but skip the “Deus ex machina”:
- The state of nature does not allow for an all encompassing enforcement mechanism
- Need to consider ex post stability (self-enforceability): **only ex post self-enforceable outcomes** are feasible ex ante agreement
- What lies 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 lie on the **bisector**
- Even in asymmetric space NBS predicts the egalitarian solution



Rawls vindicated (also for the non kantians)

- The egalitarian solution corresponds to the Rawlsian maximin.
 - Eve's payoffs identify the disadvantaged player both as E or E', and they are maximised under solution
- Egalitarianism basically rests on the requirement of *ex post equilibrium* plus the *ex ante* requirement that asks making judgments acceptable under the "veil of ignorance"
 - Just because we cannot hypothesise an external enforcer, given empathetic preferences, we are constrained to make and agreement within the symmetrical subset intersection $XAE \cap XEA$

A Rawlsian theory of corporate governance

- Consider **two different institutionally *feasible* subsets** G1 and G2 derived from the all-inclusive set of the possible governance structures.
- By design define them as corresponding to two outcome spaces P1 and P2 consisting of Nash equilibria
- The “**veil of ignorance**” hypothesis is then introduced
 - i.e. players consider *each* feasible constitution from an impartial standpoint by allowing the **mutual replacement of the roles**
 - This means that a **symmetrical translation** with respect to the Cartesian axes is taken **for every** candidate outcome space,
- A solution **must be invariant** under the symmetric translation of the respective outcome spaces.

A Rawlsian theory of corporate governance(1)

- **PROPOSITION I:**
- Given any pair of feasible convex outcome sub-spaces P_1 and P_2 , relative to a pair of constitutions and their respective post constitutional sub-games G_1 and G_2 ,
 - if the “**veil of ignorance**” hypothesis is **introduced**,
 - **but** the “**Dues ex machina**” hypothesis is **rejected**,
 - **then** the Constitutional Choice selects a constitution corresponding to the bargaining sub-game G_i endowed with a feasible outcome sub-space P^*
 - **such that** the *egalitarian solution* in P^* dominates the **other egalitarian solution** belonging to the alternative feasible sub-space.

A Rawlsian theory of corporate governance(2)

- Given any two feasible convex outcome sub-spaces P_1 and P_2 and their symmetric translations P_1' and P_2' ,
- *no matter other characteristics of the relevant spaces,*

$$\sigma_2^* > \sigma_1^*$$

if and only if $P_1 \cap P_1' \subset P_2 \cap P_2'$

- where σ^* is the **egalitarian solution** within the respective outcome space P_i
 - and the **order relation** $>$ should be understood as ***strictly superior unanimous acceptance*** (strong Pareto dominance).
- **Inclusiveness of the symmetric intersection** is the **only property relevant** to the constitutional choice of sub games

A Rawlsian theory of corporate governance (two property regimes and their outcome spaces)

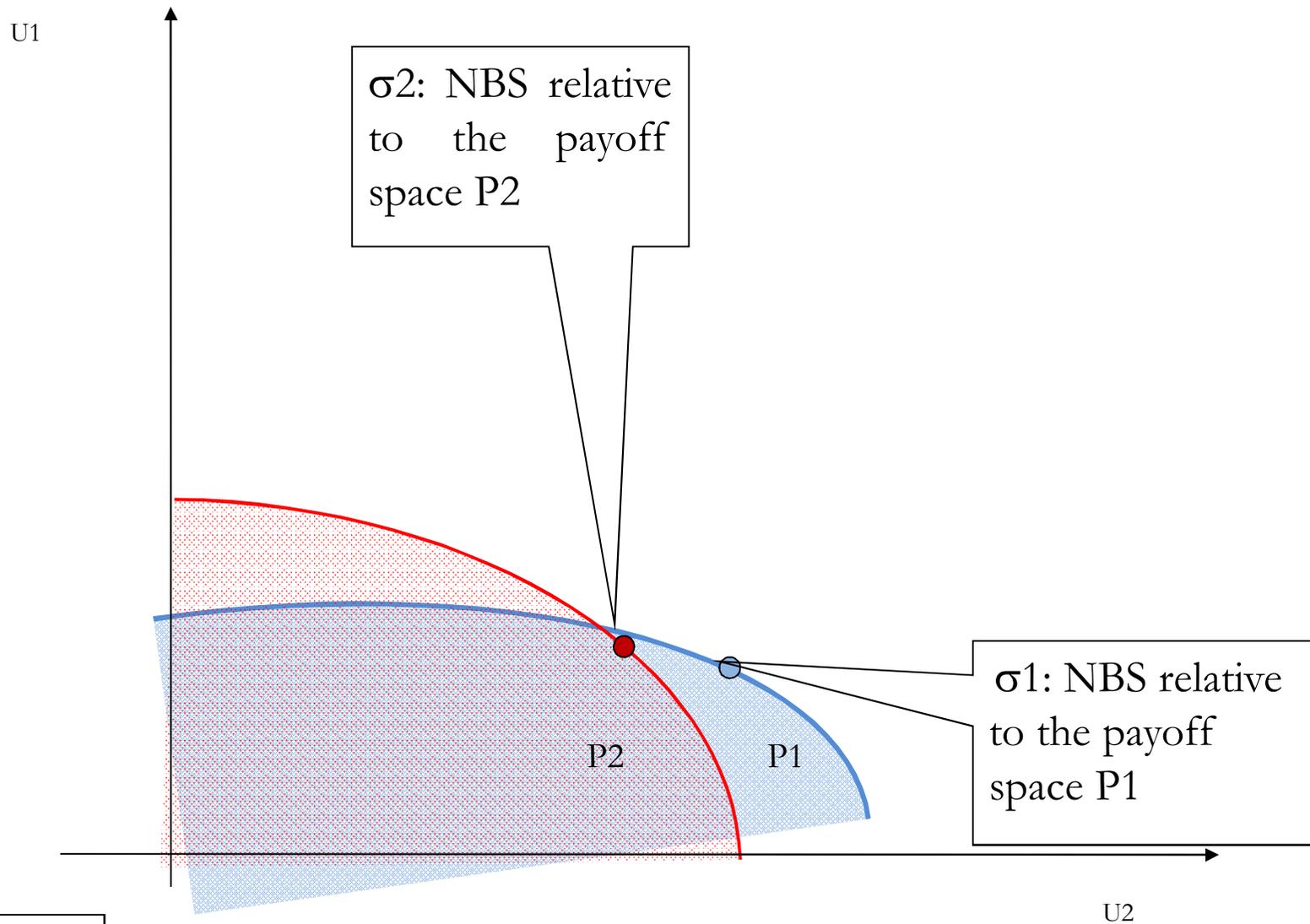
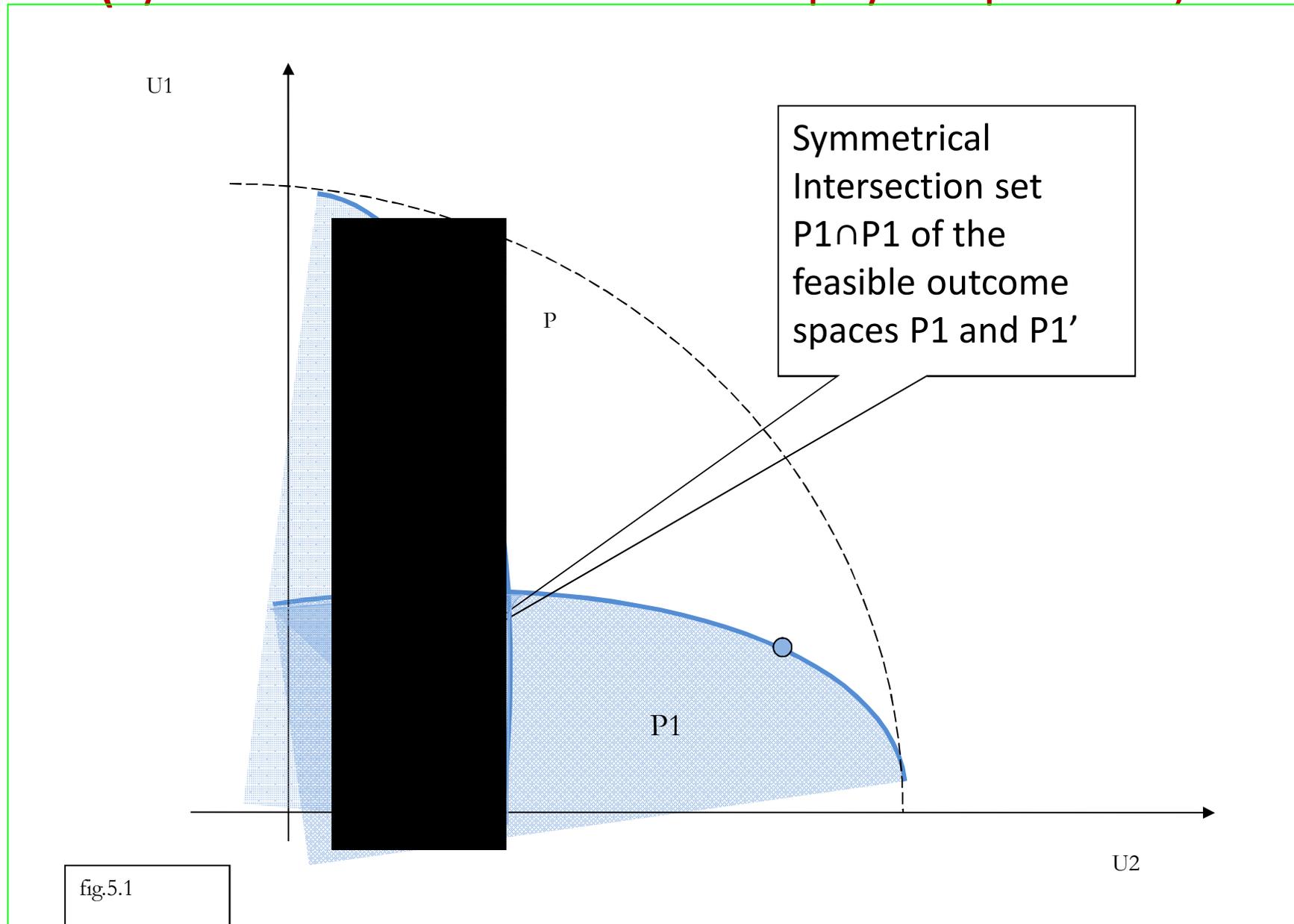


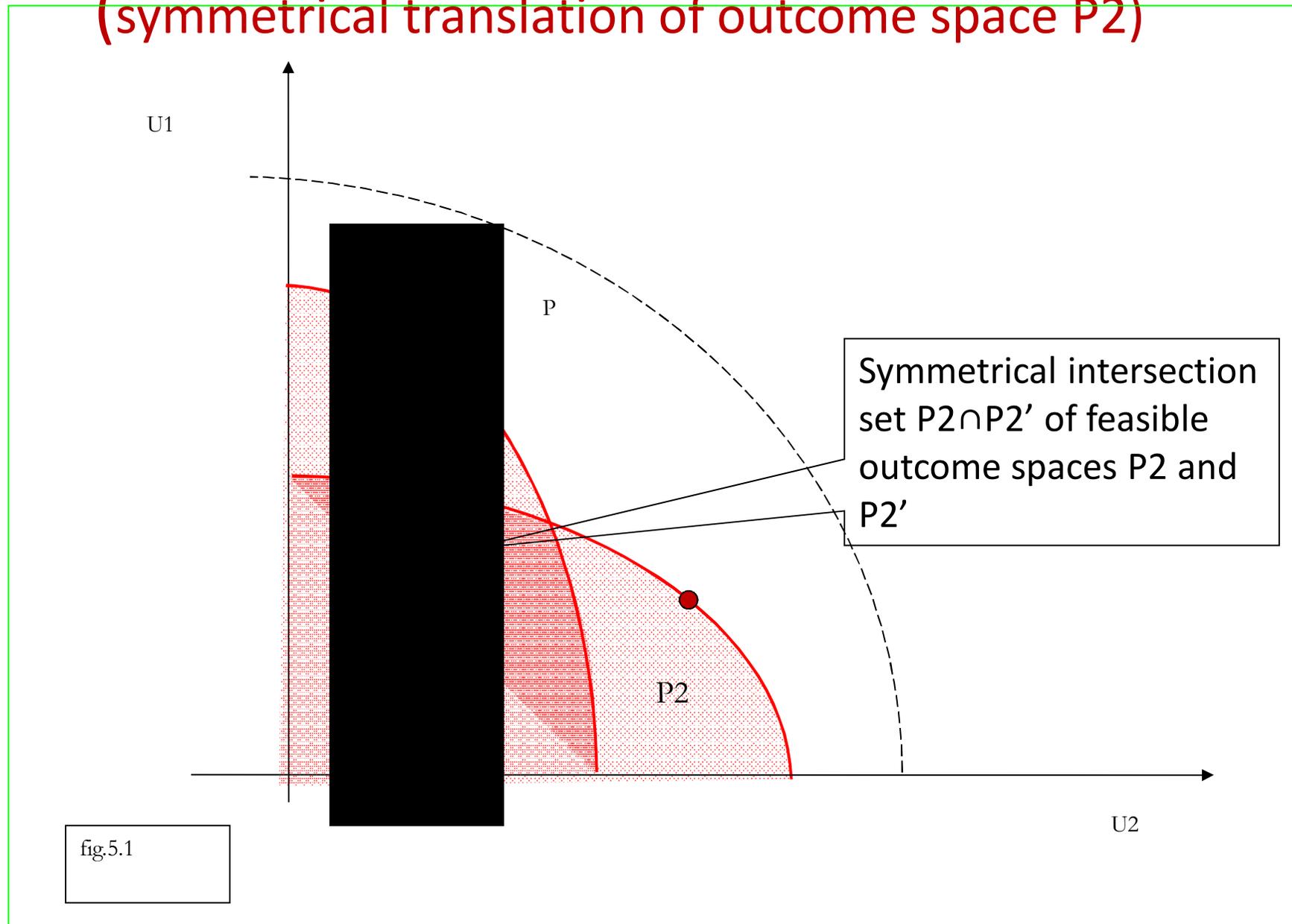
fig.5.1

A Rawlsian theory of corporate governance (symmetrical translation of the payoff space P_1)



A Rawlsian theory of corporate governance

(symmetrical translation of outcome space P2)



A Rawlsian theory of corporate governance (simultaneous symmetrical translation of both the outcome spaces P1 and P2)

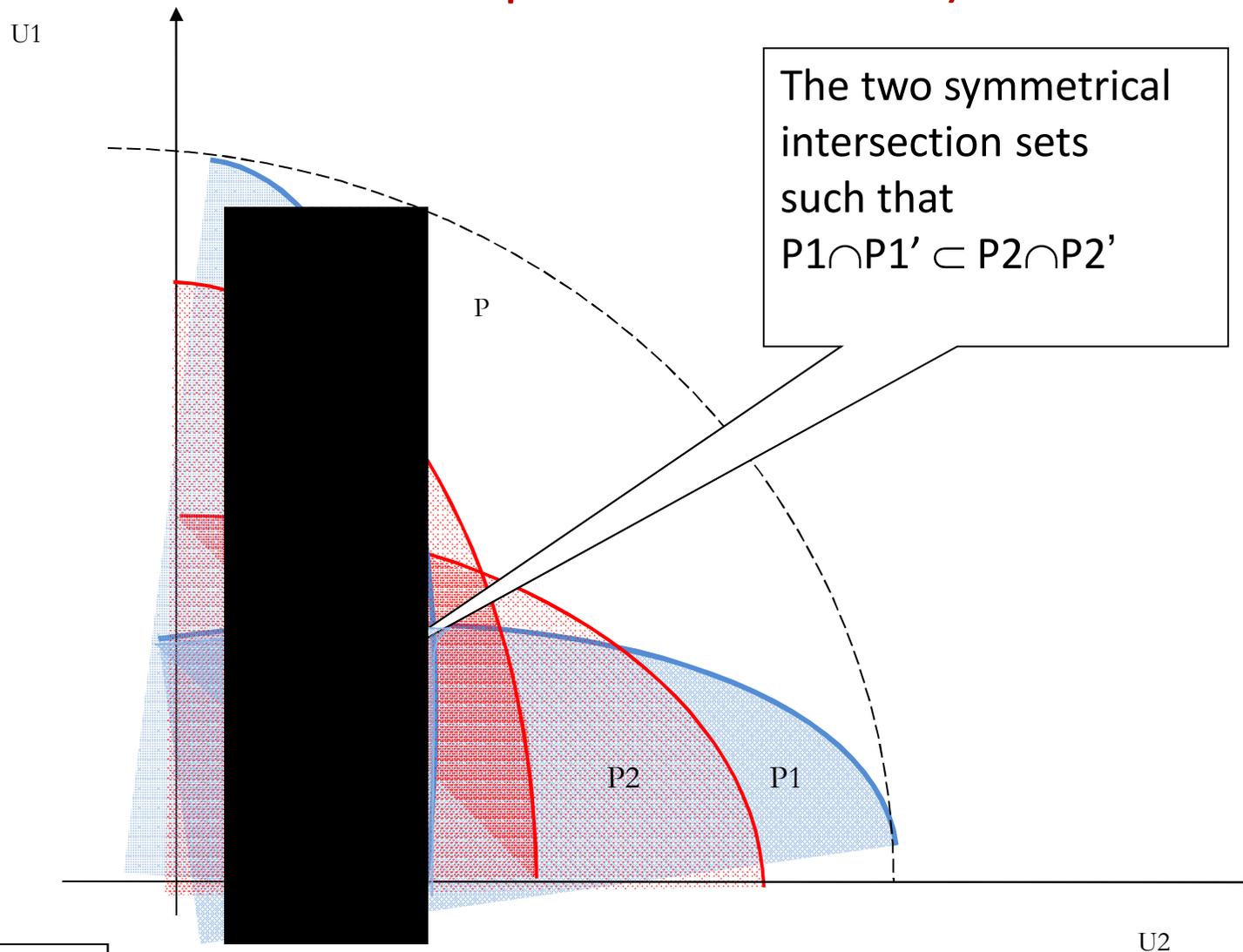
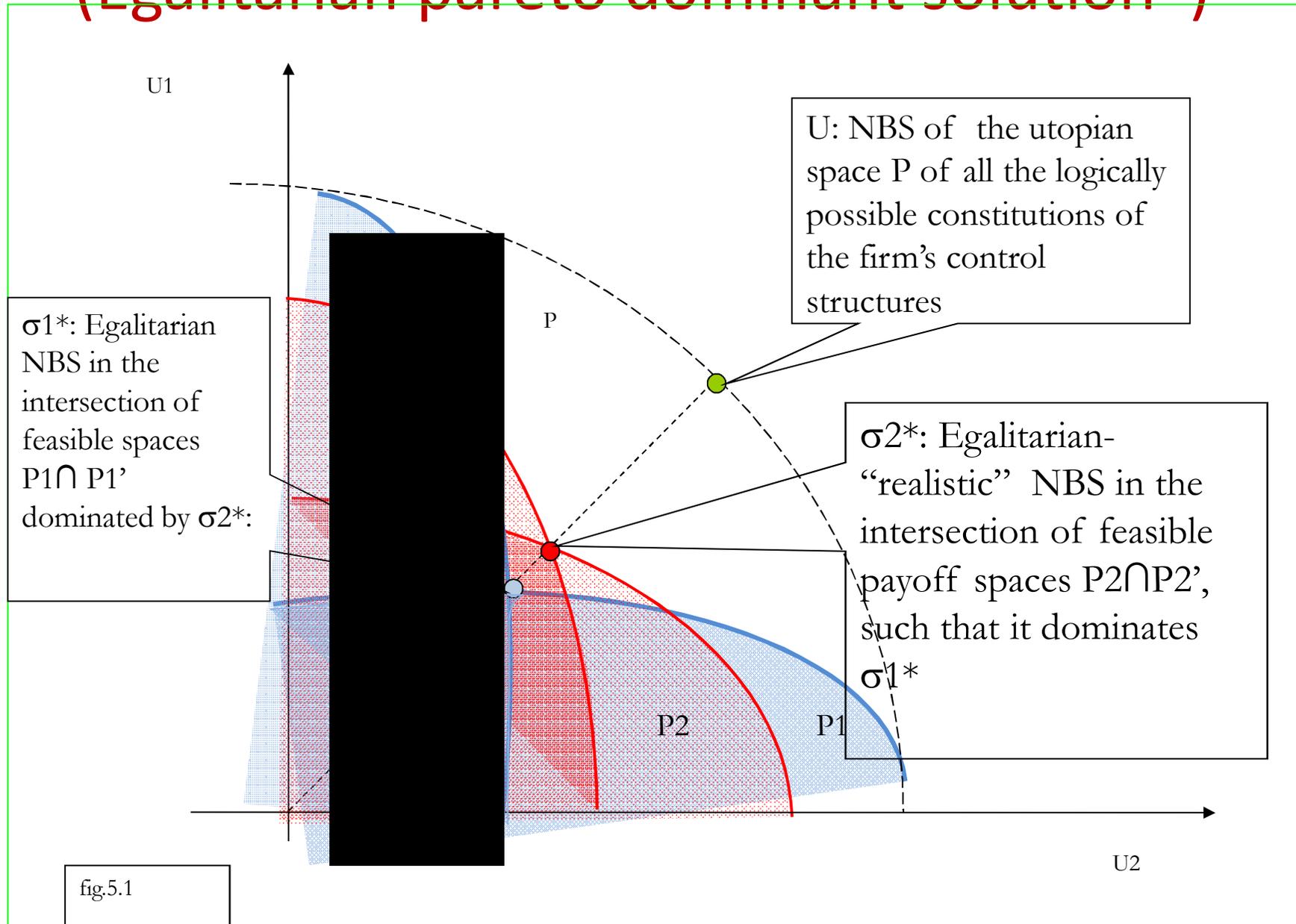


fig.5.1

A Rawlsian theory of corporate governance (Egalitarian pareto dominant solution)



A Rawlsian theory of corporate governance(3)

- Pareto-dominance **only between egalitarian solutions** is relevant for the unanimous choice of constitutions,
 - *no matter how other characteristics of the payoff spaces are settled.*
- The propositions states that
 - the level of unanimous acceptance of a constitution **dominates** the level of acceptance of another constitution
 - only if its **egalitarian** solution is Pareto-superior to the egalitarian solution of the alternative,

Challenging received wisdom: **equity comes before efficiency**

- Consider the two feasible outcome spaces P1 and P2
- P1 includes both the Utilitarian and Kaldor- Hicks solution
- but nevertheless P2, with its symmetric translation P2', generates an intersection set that **includes** the intersection of P1 and its own symmetric translation P1'.
- 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 both P2 and P2'

Challenging received wisdom: welfare VS fairness?

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

Challenging received wisdom: (Kaplow and Shavell)

- “Our argument for basing the evaluation of legal rules entirely on welfare economics, giving no weight to notions of fairness, derives from the fundamental characteristic of fairness-based assessment:
- such assessment does not depend exclusively on the effects of legal rules on individual’s well-being. As a consequence, satisfying notion of fairness can make individual worse-off, that is , reduce social welfare.
- This thesis is particularly compelling because also in important and simple situations, i.e. “symmetric contexts – those in which all individuals are identically situated – it is always the case that everyone will be worse off when a notion of fairness leads to the choice of different legal rule from that chosen under welfare economics” (p.52).

Consider the following case

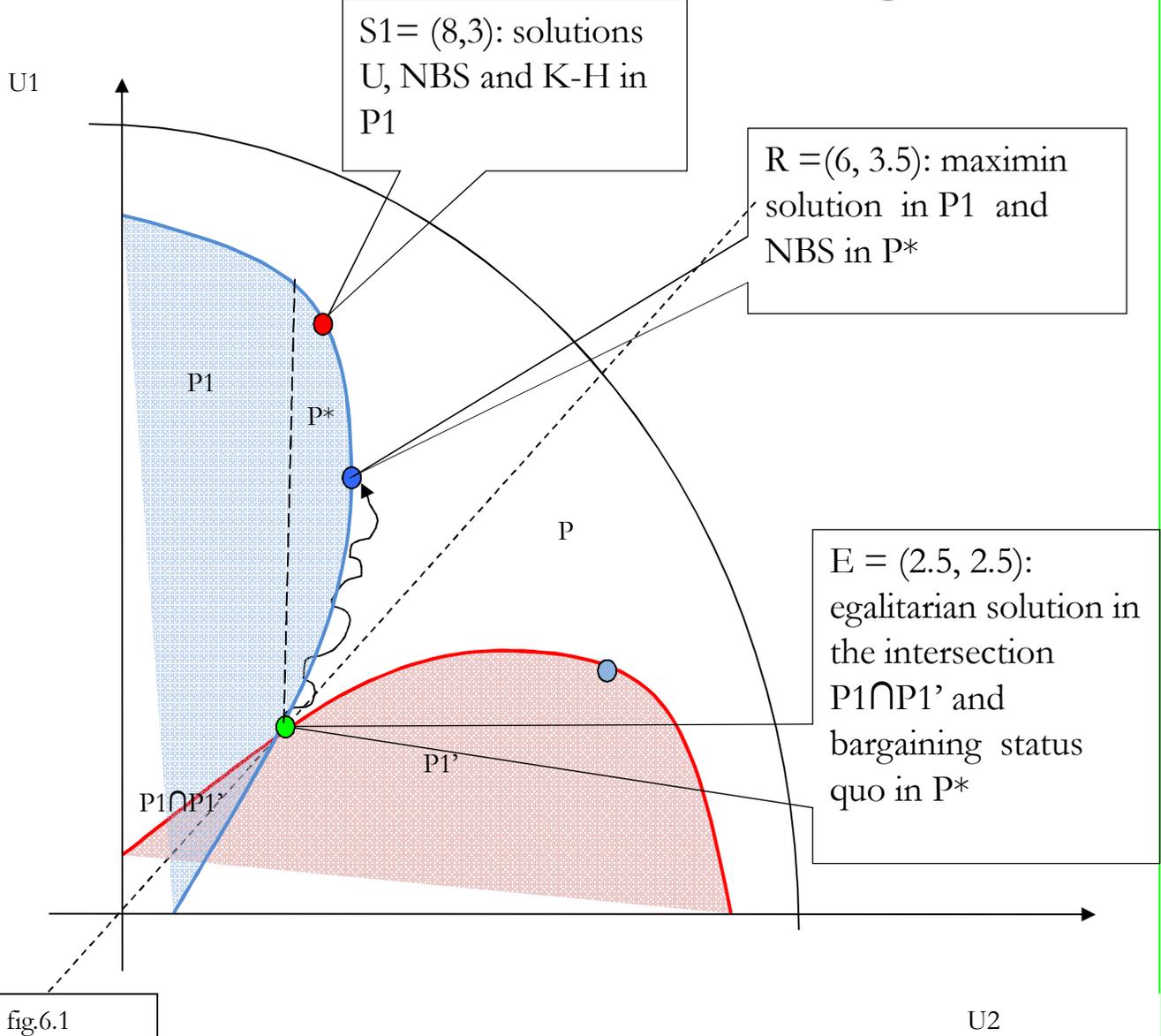


fig.6.1

U_2

Challenging received wisdom: (Kaplow and Shavell)

- The **feasible** payoff space P_1 is so asymmetric that by considering its translation P_1' , the intersection set is a very narrow region of the plan
- and the egalitarian solution in $P_1 \cap P_1'$ proves to be Pareto-dominated by S_1 , where both the maximal utilitarian solution and the maximum Nash bargaining product reside
- This seems to be a case where keeping to fairness makes every players worse off
- In fact, player 1 could try to convince player 2 to relinquish egalitarianism with the reasonable argument that there is a mutual advantage in switching to S_1 .

But this is not the case

- Giving egalitarianism priority over welfare maximization is perfectly reasonable because it allows selecting a specific PO improvement.
- Egalitarian solutions **constrain** Pareto efficiency in so far as egalitarianism is taken to be the proper starting point from which acceptable Pareto improvements are calculated through a “second thought” in bargaining
- This solution is the maximin point R on the north-east frontier of the space P1, where player 2’s payoffs (the disadvantaged player) are improved as much as possible, no matter what the marginal payoff improvement of player 1
- Pareto improvements with respect to E are achieved by moving along the frontier of P1, and they end as soon as no better improvement in player 2 payoff is possible.
- This solution dominates E, but it **makes sense only because** E is taken to be the appropriate **status quo**

Received wisdom: a **mild** libertarian cannot be but egalitarian (VS.Hayek)

- Much new-institutional theorising about governance is based the implicit postulate that institutions design cannot go further than 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 *Deus ex machina*
- But a mild libertarian would not reject that individual agents may enter the “original position”
- At least in order to make an assessment of possible spontaneous order outcomes under random permutation of social roles (A and E) taken by any player

A mild libertarian cannot be but egalitarian (VS.Hayek)

- However constraining the moral point of view with a care for freedom of choice and stability (no *Deus ex machina*) has dramatic consequences to the libertarian point of view:
 - **Only governance structures providing for egalitarian payoffs allocations are acceptable.**
- Far from ostracizing the “mirage of social justice” in the small scale society constituted by the stakeholders of a firm, a moderate libertarian **cannot be but egalitarian** in the selection of the firm governance structure.

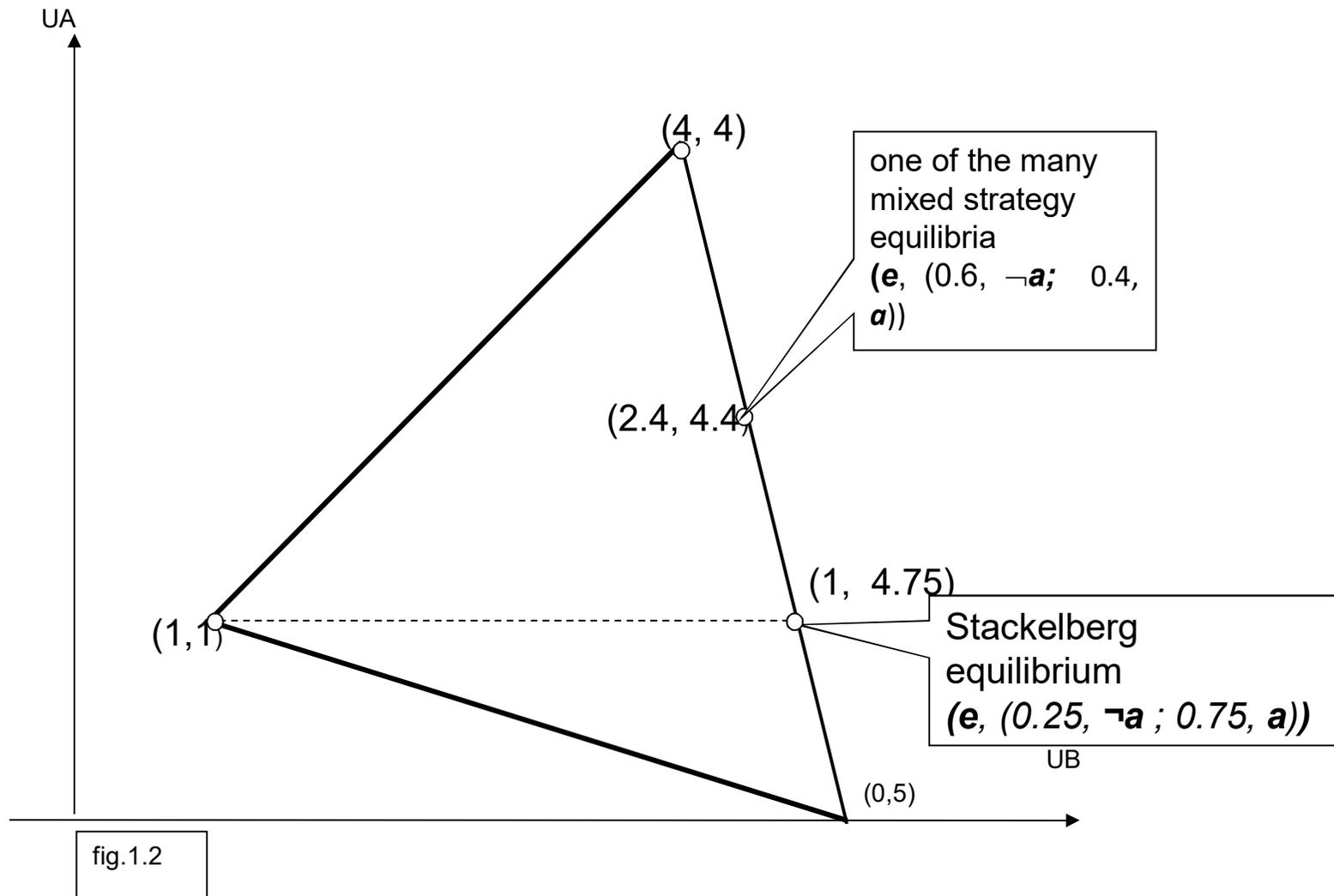
Unique selection of an equilibrium point in the trust game under role permutation

- Intersection coincides with the permutation axis from space X_{AE} to space X_{EA}
- i.e. the North-West boundary of the original payoff space
- The egalitarian solution in the symmetrical intersection $X_{AE} \cap X_{EA}$ coincides with the NBS of the original game (2,2)
- Applying the “veil of ignorance reasoning without “deus ex machina” provides a reason for selecting the intuitively fair outcome (2,2)

But ex ante justification does not answer the ex post stability problem

- Agreement behind the veil of ignorance solves the **justification** problem,
 - **unique solution**
- But what about the situation **beyond** the veil of ignorance?
- Why should the player **be sure** that other players will carry out the corresponding strategy when they are back to game of life?
- Equilibrium selection must guarantee **shared knowledge** that the equilibrium selected is in fact the equilibrium played **beyond the veil**
- But why should players believe that what they recognize to be **justified behind** the veil of ignorance is also going to be carried out **beyond** the veil of ignorance?

Again the problem of multiplicity in the repeated TG



Another equilibrium may be ex post focal to the firm

- the player B-type make a commitment on the mixed strategy $(\mathbf{e}, (0.25, \neg \mathbf{a} ; 0.75, \mathbf{a}))$
 - **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 A necessarily enters (average positive payoff is 1, equal to stay out)
 - this gives B an average expected payoff is 4.75
 - 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

2) Motivational role of Rawlsian SC. Are all the equilibria equally capable to provide incentives?

- May a norm agreed from an ex ante (pre-play) perspective, **affect the motivational force** exerted by different equilibria in a game?
- A positive answer would amount to a restriction on the set of equilibrium points that have **motivational force ex post** over the players' behavior.

Motivational role (continues)

- The conjecture is that a preference for equilibrium strategies **may depend** not just on their outcomes but also on the **level of conformity** that any equilibrium exhibits in regard to the agreed norm,
 - **if** the Social contract generates a modification in the players' payoffs in favor of those situations wherein no significant deviation from reciprocal conformity occurs,
 - **then** it may be the case that the overall motivational strength reinforcing an equilibrium behavior may be integrated by an additional motivational factor
 - **that** in the end confines overall motivational strength only to a **subset** of the possible Nash equilibria (**refinement**)

A truly Rawlsian theory of norm compliance: the sense of justice

□ Endogenous solution of the stability problem:

- when institutions are consistent with principles agreed under a veil of ignorance, we develop a sense of justice that carries with itself the **desire to stabilize** the institutions

□ Definition of the sense of justice:

“Given that a person’s capacity for **fellow feeling** has been realized by forming attachments (to lower level institutions) and given that a society’s institutions **are just** and are **publicly known to be just**, then this person **acquires** the correspondent **sense of justice** as he recognizes that he and those for whom he cares are the beneficiaries of these arrangements” (p.491.)

Conformist contractarian preferences

- **The “Conformist-contractarian preferences theory” tries to capture this philosophical view in a psychological game theoretical model**
- **References and previous works**
 - Genakoplos et al. (1989) Games and Econ Behav.
 - Rabin (1993) Amer. Econ. Review
 - Grimalda and Sacconi (2005) CPE
 - Sacconi (2006/7) JoBE
 - Sacconi and Faillo (2009) CPE

The “Conformist-contractarian preferences theory”

- Players are involved in a non cooperative game G
 - only suboptimal Nash equilibria are feasible
- In a pre-play communication stage players agree on a principles (a CSR norm) for solving the ensuing non-cooperative game
 - By an “ex ante” ideal bargaining game $B(G)$ under a veil of ignorance, players agree on a principle of justice T (viz. Nash bargaining product)
 - These agreements are not binding, so that they can be understood as “cheap talk”

Conformist contractarian preferences theory (continue)

- But players attach “motivational force” (a desire) to conformity to the CSR principle ,
- The effectiveness of the disposition to conform is conditional on the other players expected behavior
 - **conditional conformity** : how much player A contributes to full conformity , given A’s belief over player B’s action
 - **reciprocal expected conformity**: how much player A believes that player B contributes to full conformity, given A’s prediction of B’s belief over A’ action

Conformist preferences: elements of the fromal model

- **First**, a principle T (=Nash Bargaining Solution), which is a distributive criterion of material utilities.
 - Players **adopt T** (the norm) by **agreement** in a pre-play phase, under veil of ignorance, and employ it in the settlement of a **consistency ordering** over the set of possible states σ
 - The highest value of T is reached in situations σ where material utilities are **distributed** according to **maximal consistency** with the principle T (max NBS)

Conformist preferences: elements...(2)

- **Second**, an index of conditional **conformity**: the extent to which - given the other agents' expected actions - the first player is directly responsible for deviation from the maximum value of T .
- **Third**, an index of **reciprocal conformity**: the extent to which the *other* player is expected to be personally responsible for a deviation from the maximisation of T , given what he (is expected to) expects from the first player's behaviour.

Definition of the two personal indexes of conformity

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

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

b_i^1 = belief of player i over player j 's action
 $T^{MAX}(b_i^1)$ = maximum attainable by the function T given i 's belief over j 's strategy,
 $T^{MIN}(b_i^1)$ = 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 T (varying from 0 to -1)

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

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

Conformist preferences: elements... (3)

- **Fourth**, an exogenous parameter λ representing the **motivational force** of the agent's psychological **disposition** to act on the motive of reciprocal conformity with an agreed norm.
- **Five**, steps **two** and **three** coalesce in defining an overall **index F** of conditional and expected reciprocal conformity for each player in each state of the game.
 - This index operates as a weight (between 0 and 1) on the exogenous parameter λ deciding whether λ will actually affect or not (and, if so, to what extent) the player's payoffs.

The overall utility function in explicit form

The overall utility function V_i is the linear combination of the two components (material and ideal)

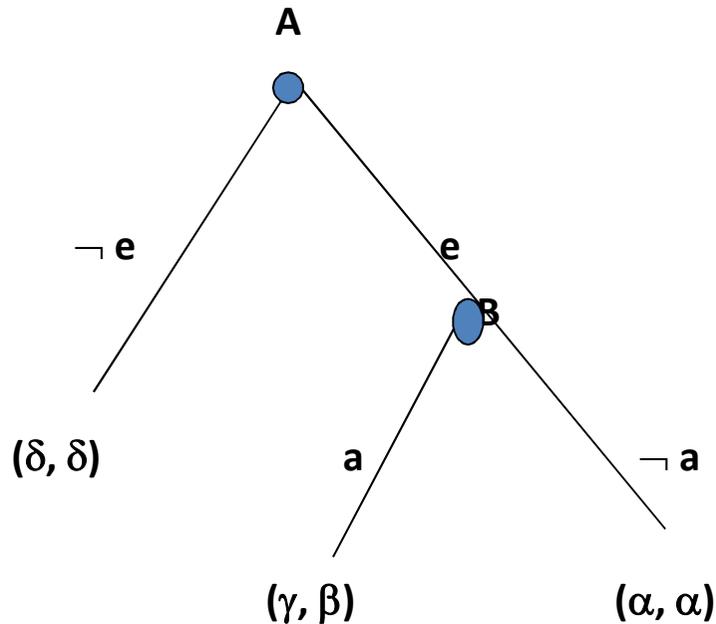
$$V_i(\sigma_i, b_i^1, b_i^2) = U_i(\sigma_i, b_i^1) + \lambda_i \left[1 + f_i(\sigma_i, b_i^1) \right] \left[1 + \tilde{f}_j(b_i^2, b_i^1) \right]$$

Material component
Psychological component

Weight of the psychological component
 $\in [-1,0]$.
Player i 's index of conformity. Given i 's beliefs (b_i^1) about j 's strategy
 $\in [-1,0]$.
Player j 's index of conformity (from the point of view of i). Given i 's beliefs about j 's beliefs (b_i^2) about i 's strategy.

- **NOTICE:** The appropriate notion of equilibrium is Psychological Nash Equilibrium (Geanakoplos et al. 1989): beliefs on how the game is played enter the player's utility payoff

What does it happen in the one shot TG with conformist preference ?



- Where $\beta > \alpha > \delta > \gamma$, note that $\gamma - \delta < 0$
- In order to calculate T: δ is taken as the status quo
 - T is $\Pi(u_i - \delta_i)$ for any state $(\delta, \delta), (\gamma, \beta), (\alpha, \alpha)$
 - $T(e, \neg a) = (\alpha - \delta)^2$, $T(e, a) = (\gamma - \delta)(\beta - \delta)$,
 - $T(\neg e, a) = T(\text{no-e, no-a}) = T(\text{no-e, } \sigma_i) = (\delta - \delta)^2 = 0$

The one shot trust game under conformist preferences

	$\neg a$	a
e	4,4	0,5
$\neg e$	1,1	1,1

Matrix (a): TG normal form

	$\neg a$	a
e	$(4-1)(4-1) = 9$	$(0-1)(5-1) = -4$
$\neg e$	$(1-1)(1-1) = 0$	$(1-1)(1-1) = 0$

Matrix (b): T values at each state

	$\neg a$	a
e	$(4+\lambda) = 6, (4+\lambda) = 6$	0, 5
$\neg e$	1,1	$(1+\lambda) = 3, (1+\lambda) = 3$

Matrix (c) : psychological TG with conformist utilities included with $\lambda = 2$, two psychol equ.

Conformity of mixed strategies

- Players assess how mixed strategies generate states, described **according** the distributive principle T
- when a player B mixed strategy is considered, conditional on entrance by A, **what players take care of is the probabilistic combination of two states described according to the value of NBF (=T)**
- “Randomizing” means that **two realisations of the distributive principle T** - $(\gamma-\delta)(\beta-\delta)$ and $(\alpha-\delta)^2$ - may occur
- Thus that the appropriate value of T related to a mixed strategies, given A’s entrance, is the **expected value** of the NBF given the probability mixture of the two states

$$T(e, \sigma_i) = (1-p) [(\gamma-\delta)(\beta-\delta)] + p(\alpha-\delta)^2$$

Mixed strategies and the repeated trust game under conformist preferences

- Payoff of mixed strategies reflects exactly the **frequency** in which strategies **produces states of affairs with different level of conformity** to the principle T
- For example, the pair of conformist payoffs (3.62, 5.62) corresponds to the following pair of repeated strategies:
 - Player B mixes his types $\neg a$ and a with frequency 0.6 and 0.4
 - Player A plays repeatedly his strategy e as long as he does not see player B employing **abuse** with a frequency higher than 0.4, but if this frequency is exceeded he switch to “ $\neg e$ forever”.
 - This produces a mixture of conformity and non conformity with the related conformist payoffs

Pure strategy equilibria

- **First**, the status quo point (1,1) is translated toward North-East along the bisector to a point with overall utilities (3,3), **which is also a psychological equilibrium**
- **Second**, thanks to the conformist weights $\lambda = 2$, the outcome (4,4) where the NBS is maximized translates to the point (6,6), **which is also a psychological equilibrium.**

About equilibria in player B's mixed strategies

- ❑ the entry strategy e of player A cannot be rewarded with any psychological conformist utility **until** the T expected value induced by a mixed strategy $\{(p, \neg a); (1-p, a)\}$ is not positive
- As long as this threshold is not exceeded, psychological payoffs **do not add** to the material payoffs of both players A and B under any of such pairs (mixed strategy, entrance)
- Then.....

1. Stackelberg equilibrium is ruled out

□ Note the importance of the

mixed strategy $\sigma_B^{0.25}$.

- This was player B's Stackelberg mixed strategy that would correspond to the preferred (by the firm) equilibrium strategy of the material repeated TG.
- It is **noticeable** that the pair $(\mathbf{e}, \sigma_B^{0.25})$ **is not any more an equilibrium** in the conformist (psychological) repeated TG
- **No** psychological utility is added to payoffs for this outcome

2. In many cases A's "giving in" is not a best reply to a mixed strategy

- The **threshold** that allows mixed strategies to gain support from psychological conformist utility is reached at the mixed strategy

$$\sigma_B^{0.307} = \{(0.307, \neg a); (0.693, a)\}.$$

- Here the expected value of T is **zero** for any A's choice, so A is **fully conformist** by choosing either **e** or **¬e**.
- Playing the mixed strategy is **partially conformist** also to player B, because the T value given A's entrance would be minimized by playing **a**.

2. In many cases “giving in” is not a best reply to a mixed strategy (follows)

□ **But** adding just a bit of psychological utility does not mean that B’s mixed strategy induces “enter” as A’s psychological best response .

– The player A’s overall payoff from $\neg e$ is **still higher** than the overall payoff from **giving in** to player B’s mixed strategy , i.e.

$$U_A \{[(0.307, \neg a), (0.693, a)], \neg e\} = 3 >$$

$$U_A \{[(0.307, \neg a), (0.693, a)], e\} = 1.84.$$

3. In general the B's best reply to "giving in" is **not** to abuse

- **Assume** player B has chosen a repeated mixed strategy
 - whereby he has been able to accumulate a reputation
 - That *for the first time* **induces** player A to enter
- **then** he (B) immediately would recognize the incentive to switch to a strategy that employs the strategy $\neg a$ with the highest frequency
- **It follows that**
 - player B's best reply to player's A entry is to **deviate** from any mixed strategy σ_B^n to $\neg a$.

MAIN PROPOSITION

- Given a repeated TG with pure and mixed strategies,
 - whereby a psychological game with conformist preferences is defined,
 - so that the motivational exogenous parameter λ is great enough to guarantee the existence of a psychological equilibrium in correspondence to $(\mathbf{e}, \neg \mathbf{a})$,
 - **Then:** the game's psychological equilibria are **only** the **two** in pure strategy $(\mathbf{e}, \neg \mathbf{a})$ and $(\neg \mathbf{e}, \mathbf{a})$,
 - and **no** equilibrium points in mixed strategies exist.

Refinement of the equilibrium set

- ❑ Even though generating a psychological game from a basic one shot Trust Game enables as usual the determination of new equilibrium points
- ❑ when we step from the one-shot TG to the **repeated** TG - where usually many equilibria are admitted
- ❑ **then** transforming the payoff space by means of conformist preferences has a powerful effect in **reducing** the psychological equilibria to a subset of the Nash equilibria (**Refinement**)