Universität Bielefeld/IMW

Working Papers
Institute of Mathematical Economics

Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

Nr. 190

Fee Games: (N)TU–Games
With Incomplete Information

von
Joachim Rosenmüller

Oktober 1990

H. G. Bergenthal

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Universitätsstraße
4800 Bielefeld 1
Bundesrepublik Deutschland
Federal Republic of Germany

ABSTRACT

A fee–game is a special version of an NTU–game with incomplete information exhibiting certain side payment properties. We analyze the structure of the set of incentive compatible and individually rational mechanisms and, based on this analysis, we propose a further axiom for a value for (N)TU–games with incomplete information.

PREFACE

The first to discuss cooperative games with incomplete information of the players concerning the other players "types" were HARSANYI and SELTEN [2]. However, the idea of the "Bayesian incentive compatible mechanism" does not appear in their work . This idea was introduced by MYERSON [6] [7] [8] into the framework of NTU–games with incomplete information. Both authors considered a version of a value for such games to be defined by a suitable set of axioms. While MYERSON argues against the HARSANYI–SELTEN value (on the grounds of his "probability invariance axiom") he provides his own version of a axiomatized solution concept. However, this value in turn may have discontinuity properties (cf. [2]).

MYERSON did not attempt to formulate axioms for the HARSANYI–SELTEN value in the presence of mechanisms. This was done by WEIDNER [10]. It turned out that there is a canonical axiomatization of HARSANYI's and SELTEN's value – however, it is limited to the case that the types of the players are distributed in a stochastically independent way.

These facts, it would seem, indicate that there is still room for discussion of a "value for NTU–games with incomplete information".

Our model – already introduced in [9], differs by an essential feature from the previous ones: we do assume that players can agree upon one of *continuously many* decisions. Hence we try to model the generalized version of a characteristic function of an NTU– game with incomplete information.

In particular, we focus on what is called a fee–game. Such a game exhibits some side–payment properties; these properties appear in the *ex ante* and *ex post* situations. However, the introduction of incentive compatible mechanisms somehow blurres the side– payment nature of the game. For fee–games, we give a detailed description of the structure of incentive compatible and *in mediis* individually rational mechanisms. In particular, for the case of incomplete information on one side, the structure of those mechanisms which are not constant is precisely described. Based on this analysis, within the final section, we introduce the "*expected contract axiom*", which may be considered as a starting point for a further discussion of an axiomatic version of values for NTU– games with incomplete information.

SECTION 1 Unanimous CII–Games

Within this section we introduce the general model as well as some basic notation.

Definition 1.1: A (unanimous) cooperative game with incomplete information (*a CII–game*) is a 6–tupel

(1) $$\Gamma = (I,T,p;\underline{X},\underline{x};U)$$

Here, $I = \{1,...,n\}$ is called the set of individuals, agents, or *players.* $T = \prod_{i \in T} T_i$ is the cartesian product of finite sets T_i, $i \in I$, and T_i is the set of possible *types* of player i.

Let $(\Omega,\underline{F}.P)$ be an abstract probability space and let $\tau : \Omega \rightarrow T$ be a random variable selecting n–tupels of types; now p is a probability on T and our interpretation is that $p = P \circ \tau^{-1}$ is the distribution of τ. Of course, all relevant data will depend on p only (and not on the choice of $(\Omega,\underline{F},P)$), hence p is called the *distribution of types.*

Let $e = (1,...,1) \in \mathbb{R}^n_+$ then $ex = \sum_{i \in I} x_i$ and

(2) $$\underline{X} = \{x \in \mathbb{R}^n_+ \mid ex \leq 1\}$$

reflects the feasible *alternatives*, parameters, or "*contracts*" the players can agree upon. Within this paper we shall assume that contracts may only be reached unanimously, however, coalitions and their power can easily be introduced into the model. This, if players fail to agree upon some $x \in \underline{X}$, then they will have to bear the consequences of the status quo alternative $\underline{x}$. We assume $\underline{x} \in \underline{X}$.

Finally, the mapping

(3) $$U : I \times T \times \underline{X} \rightarrow \mathbb{R}$$

reflects the utilities; if chance chooses $t \in T$ and the players agree upon $x \in \underline{X}$, then player i's utility is $U_i^t(x)$.

As yet, we have not specified the rules of the game, this will be possible once we have the notion of a "mechanism".

Definition 1.2 Let Γ be a CII–game. A *mechanism is* a mapping

(3) $$\mu : T \to \underline{X}.$$

A mechanism μ is called *incentive compatible* (IC) if, for all $i \in I$ and all $t_i, s_i \in T_i$, we have

(4) $$E(U_i^\tau \circ \mu \circ \tau | \tau_i = t_i) \geq E(U_i^\tau \circ \mu(\tau_1,\ldots,s_i,\ldots\tau_n) | \tau_i = t_i).$$

(Later on we use $\mu^\tau = \mu \circ \tau$).

The concept of a ("Bayesian") IC mechanism goes back to HURWICZ [4], our version is close to the one adopted by MYERSON [7] in a different framework, see also [1]. The first authors to study CII–games were HARSANYI and SELTEN [2] – but in their context mechanisms are not present.

As particular version of a mechanism is the one we shall call *constant*, i.e., if $\mu^t = x$ for all $t \in T$ and some fixed $x \in \underline{X}$. E.g., the *status quo mechanisms* μ is defined by

(5) $$\underline{\mu}^t = \underline{x} \qquad (t \in T).$$

Now we are in the position to specify the rules according to which the game is to be played.

First of all, there is a *bargaining period.* Within this period players may agree upon a mechanism μ in which case they will register this mechanism with an agency – the referee or *court.* This agency is supposed to be very powerful in the sense that it is capable of eventually enforcing the mechanism agreed upon or – if no agreement has been reached – the status quo mechanism $\underline{\mu}$.

In the next step, *chance* chooses $t \in T$. Each player $i \in I$ observes his type $t_i \in T_i$.

Next, each player i *announces* his (alleged) type, say s_i *to the court*. This constitutes some $s \in T$, the court will choose $\mu(s) \in \underline{X}$ and player i receives utility $U_i^t(\mu(s))$.

According to the various stages of the development we consider basically three situations or "states of knowledge" with respect to the players and the court: "*ex ante*" – before the chance move, "*in mediis*" ("mediis in rebus" indicates the not so purely temporal aspect) – after the chance move and prior to the announcements, and "*ex post*" – after the announcements.

Now, whenever players agree upon – and register – some mechanism μ, then a game (in extensive/strategic form) Γ^μ arises in which players behave strategically by announcing types in view of the types they observed. Thus, in this game, strategies are mappings $\sigma_i : T_i \rightarrow T_i$. If "telling the truth" (i.e., the identity mapping) is a Nash equilibrium in Γ^μ the μ is IC and it is not hard to see that this definition is equivalent to the formal one presented in Definition 1.2 (i.e., to (4)). For the details, see [5] or [9].

Existence of IC–mechanisms does not constitute a problem (e.g. constant mechanisms are IC).

As a part of our story, we assume that the court will register only IC mechanisms. That is, registering mechanisms which in a sense induce players to lie about their true situation are considered to be against societies basic rules ("*contra bonos mores*"). From this it follows that the court *ex post* will be fully informed about the players true type. And, since "law enforcement", i.e. assigning $\mu(t)$ and the resulting utilities, takes place *ex post*, this provides implicitly a further argument for the restriction to IC mechanisms: the court wants to make sure that, when it enforces people to accept the results induced by μ, this is based on knowledge of the types and on the knowledge of truthfull announcement of these types.

There are further restrictions which the court wants to observe. If a player *in mediis*, i.e., after observing his own type, realizes that his conditional expectation of utility is worse than if no agreement had been reached at all, he will cry foul and try to get out of the agreement. Indeed, the court may not be able to enforce such mechanisms and hence it will restrict itself to the registration of *in mediis individually rational* (IR) IC mechanisms.

The formal definition is, of course, as follows.

Definition 1.3 Let Γ be a CII–game, the IC mechanism μ is *individually rational in mediis*, (for short: IR) if, for all $i \in I$ and all $t_i \in T_i$

(6) $$E(U_i^{\tau} \circ \mu^{\tau} \mid \tau_i = t_i) \geq E(U_i^{\tau}(\underline{x}) \mid \tau_i = t_i) = E(U_i^{\tau} \circ \underline{\mu}^{\tau} \mid \tau_i = t_i)$$

Let $\mathfrak{M}$ denote the set of all mechanisms obeying this definition, i.e.

(7) $$\mathfrak{M} = \mathfrak{M}(\Gamma) = \{\mu : T \rightarrow \underline{X} \mid \mu \text{ is IC and IR}\}.$$

Note that we have implicitly adopted the viewpoint that recontracting in mediis is not possible. Nevertheless the players, when bargaining ex ante about mechanisms, will have the situation *in mediis* in mind. This may cause them in advance to reach mechanisms that are elements of $\mathfrak{M}$. So, by assuming that mechanisms outside of $\mathfrak{M}$ are "immoral" (*contra bonos mores*) we restrict ourself to situations in which the players in principle would not decide differently if bargaining took place after the chance move. Some authors seem to indicate that it makes no difference as to whether bargaining takes place ex ante or in mediis. In our present setup we avoid discussing this question; nevertheless it might be argued that the model does not change when bargaining takes place *in mediis*.

A similar discussion could center around the question as to whether a mechanism should be ex post individually rational. (Formally this would mean that for all $i \in I$ and for all $t \in T$ $U_i^t(\mu^t) \geq U_i^t(\underline{x})$). Indeed, if IC mechanisms are employed then not only the court might be aware of the true types of the players as selected by chance. It could very well happen that the court enforces a payoff upon a player which *ex post* is worse than if he had not joined an agreement at all and, as the power of the court is much higher than compared to an ordinary bargaining situation with complete information, enforcing such kind of outcome again could be considered immoral. Much can be said in favor of this point. On the other hand it is clearly seen that a lot of situations can occur in which players value their *in mediis* expected payoff much higher than the expected payoff of the status quo alternative and nevertheless end up regretting the agreement with a positive probability. In such cases agreements would not be reached *ex ante* or *in mediis* even so players would like to contract. We have adopted the viewpoint that, presently, we do not want to insist on *ex post* individually rational mechanisms. Maybe, this càn be supported by the assumption that the court keeps the complete information of the true types of the players as a secret. Of course, this is a rather weak argument: after all, some agreements are not in accordance with good customs just because some agency keeps the violation of certain principles underlying the idea of "*bonis mores*" as a secret. On the other hand, disclosing the players' type might also violate certain basic principles of society.

SECTION 2 FEE–GAMES

A fee–game is a special type of a CII–game. Some of its properties reflect the fact that "side payments" be permitted: *ex post* it is seen that a game with transferable utility is at hand once the types are revealed. *Ex ante*, the expected utilities of parameters reveal the T.U.–character. However, once IC mechanisms are agreed upon, the side payment character is blurred.

Recall that $e = (1,\ldots,1) \in \mathbb{R}^n_+$ and $\underline{X} = \{x \in \mathbb{R}^n_+ \mid ex \leq 1\}$. We write

(1) $$\partial\underline{X} = \{x \in \underline{X} \mid ex = 1\}.$$

Definition 2.1

1. $b^0 \in \mathbb{R}^n_+$ is called a *fee* if $eb^0 < 1$ holds true.

2. If b^0 is a fee, then the *Standard T.U.–game* (*STU–game*) or bargaining situation (à la NASH) *generated by* b^0 is the pair (O, V_{b^0}) with

(2) $$V_{b^0} := \{x - (ex)\, b^0 \mid x \in \underline{X}\}.$$

The interpretation is obvious: a unit of money can be distributed among the players by agreement. However, registration bears some cost which is proportional to the total amount of money. And each player i is allotted a specific share of this cost, more precisely, for a unit of money which he obtains by agreement he has to pay a "fee" b^0_i.

Note that (O, V_{b^0}) formally is a (unanimous) NTU–game in the ordinary sense (full information) but players register a contract $x \in \underline{X}$ (and not a utility n–tupel).

In this setup, fees depend on characteristics of the players. It may be that player i pays an expert or lawyer to work out the details of the contract and that this payment depends on the size of the contract. It could also be a feasible interpretation to assume that the total payment $ex = \sum_{i \in I} x_i$ goes to the court (as to cover the expenses of registration) and that the court charges players according to circumstances which are based on certain properties/characteristics of players. In German courts at least it is possible to obtain a special tariff at court by pointing out (and proving!) that personal circumstances are justifying such remedy ("Armenrecht").

Now to some trivial technical proceedings.

Remark 2.2 Given a fee b^0, the mapping

(3)
$$U^0 : \underline{X} \to \mathbb{R}^n$$
$$U^0(x) = x - (ex)\, b^0$$

yields $V^{b^0} = \{U^0(x) \mid x \in \underline{X}\}$; thus $U_i^0(x)$ reflects player i's utility if $x \in \underline{X}$ is agreed upon. In particular, $U^0(\underline{x}) = U^0(0) = 0$ is the utility vector of the *status quo point.* The individually rational part of the Pareto frontier, i.e., the set $\{U^0(x) \mid x \in \partial\, \underline{X}, U^0(x) \geq 0\}$ is a simplex with extreme points

(4)
$$(1 - eb^0)\, e^i \qquad (i \in I),$$

where $e^i = (0,...,1,...,0)$ denotes the i'th unit vector. The inverse images of these extremals (under $(U^0)^{-1}$, that is) are the vectors

(5)
$$a^{0i} := (1 - eb^0)\, e^i + b^0,$$

the convex hull of which, say

(6)
$$C^0 := cvH(\{a^{01} \mid i \in I\})$$

is as well a simplex. All faces of C^0 are parallel to the corresponding faces of $\partial\underline{X}$. The vectors a^{0i} convey all necessary information, thus they may equally be used to describe the STU–game generated by b^0. Indeed, we have

Lemma 2.3 Let $(\overset{\circ}{a}^i)_{i \in I}$ be n linearly independent vectors, $\overset{\circ}{a}^i \in \partial \underline{X}$ (i∈I), such that the faces of the simplex generated are parallel to the corresponding faces of $\partial\underline{X}$. Then there is a unique fee $b^0 \in \mathbb{R}^n_+$ such that

$$\overset{\circ}{a}^i = (1 - eb^0)\, e^i + b^0 \qquad (i \in I)$$

holds true.

Proof: Trivial; b^0 is obtained via the unique affine mapping that throws the a^i into multiples of unit vectors, i.e., we have

(7) $$b^0 = \frac{1}{n} \sum_{i \in I} \overset{\circ}{a}^i - \lambda \frac{e}{n}$$

with

(8) $$\lambda = \overset{\circ}{a}^1_1 - \overset{\circ}{a}^2_1 = \dots = \overset{\circ}{a}^i_i - \overset{\circ}{a}^k_i \quad (i \in I,\ k \in I,\ k \neq i)$$

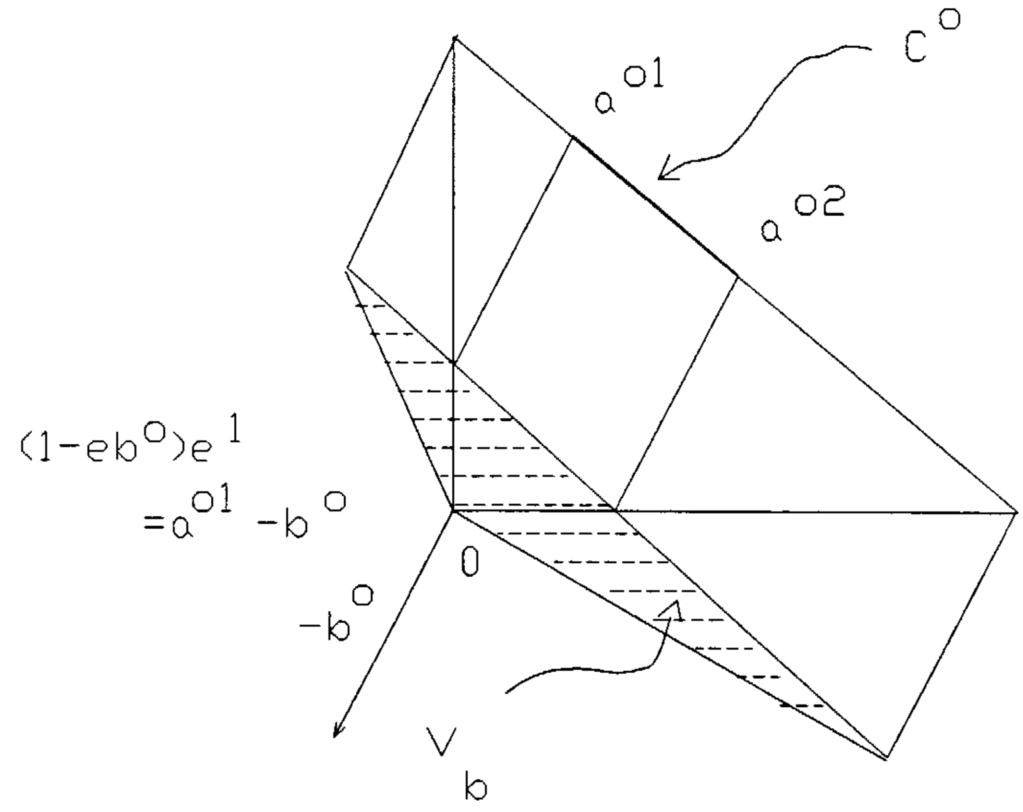

Fig.1
The STU–game generated by a fee

Without making this a formal definition we shall call a simplex $C^0 \subseteq \underline{X}$ which is spanned by its vertices $(\overset{0}{a}{}^i)_{i\in I}$, satisfying the conditions of 2.3, via (6) an *admissible subsimplex* of $\underline{X}$.

We are now in the position to formally describe "fee games" – CII–games that are generated by type depending fees.

Definition 2.4 1. A *fee schedule* is a mapping

$$b : I \times T \to \mathbb{R}_+$$

(i.e., a matrix) such that, for every $t \in T$, b^t is a fee.

2. A CII–game $\Gamma = (I,T,p;\underline{X},\underline{x};U)$ is a *fee–game* if

(8) 1. $\underline{x} = 0$

2. There is a fee schedule b such that for every $t \in T$

(9) $$U^t(x) = x - (ex)b^t.$$

First let us pause for some interpretation.

In this model chance chooses $t \in T$ and, eventually, player $i \in I$ will pay $(ex)b_i^t$ towards the "expenses of the contract" if the agreement $x \in \underline{X}$ is reached, thus his utility is $x_i - (ex)b_i^t$. Again we may assume that these are his costs for his expert or lawyer or else that the court is charging a fee for in order to register – and enforce – a contract (or mechanism).

Nevertheless it might be necessary to justify the selection of the fee vector by a chance move in a more detailed manner: e.g., the "expenses of the contract" might result from a more elaborate economical activity which makes the money to be distributed available, thus expenses or costs may as well result from economical activities and depend on random influences.

Note that each player observes only his own b_i^t; in particular the court does not know the actual fee that has been selected by chance. However, if the court registers a mechanism $\mu \in \mathfrak{M}$, then obviously the advantages are (at least) two–fold: *in mediis*, when players announce their observed type to the court, none of them will start arguing on the grounds that their basic rights are violated – since μ is individually rational. And *ex post*, when the court knows all the types announced, it is clear that these types are the true ones – thus the court can indeed act out the agreement specified by μ, μ^t that is, based on the knowledge of the true t. We may, therefore, attempt to further complete our story of how the game is played by assuming that, as a part of the registered agreement, the total amount of money $ex = \sum_{i \in I} x_i$ is first transferred into the courts custody, where after the amount $x_i - (ex)b_i^t$ is allotted to player i. (Of course the term "*in mediis*" does not apply from the viewpoint of the court as it does not observe any results of the chance move.)

If we feel that the court cannot be assumed to have enough power in order force player to *pay* towards the final distribution of money ($x_i - (ex)b_i^t$ might be negative), then it might choose to *collect* $\max_t b_i^t$ from player i together with the registration – and repay whatever necessary eventually.

Now, in order to add some geometrical interpretation, observe that in the *ex post* situation players have been involved in the STU–game

$$(0, V^t)$$

with $V^t = V_{b^t} = \{U^t(x) \mid x \in \underline{X}\}$ (see Definition 2.1). In view of Lemma 2.3 and the subsequent discussion it follows that a fee–game is equivalently described by a system $(C^t)_{t \in T}$ of admissible subsimplices of $\underline{X}$. E.g., the paradigm of a two person CII game is a system of intervals via $\underline{X}$, (see Figure 2) each of which gives rise to an STU–game (as indicated in Figure 1).

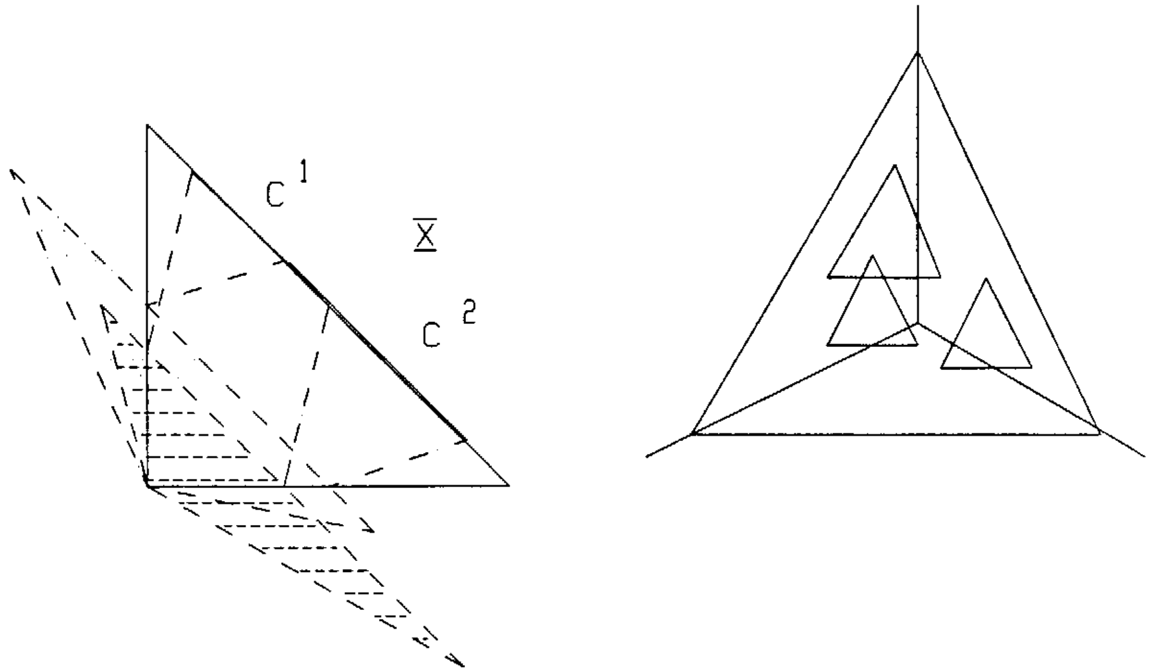

Fig. 2
Paradigm of the fee–game

Next, it is obvious that *ex ante* and *in mediis* players visualize an STU–game as well. For, if we define

$$V^{\underline{X}} := \{E\, U^{\tau}(x) \mid x \in \underline{X}\} \tag{10}$$

and

$$V^{\underline{X} | t_i} = \{E(U^{\tau}(x) \mid \tau_i = t_i) \mid x \in \underline{X}\}, \tag{11}$$

then clearly, we have

Lemma 2.5 $(0, V^{\underline{X}})$ and $(0, V^{\underline{X}|t_i})$ are STU–games generated by Eb^τ and $E(b^\tau \mid \tau_i = t_i)$ respectively. Thus, $V^{\underline{X}} = V_{Eb^\tau}$ and $V^{\underline{X}|t_i} = V_{E(b^\tau | \tau_i = t_i)}$.

Proof: Trivial E.g., $E(b^\tau(\tau_i = t_i)$ is at once recognized to be a fee and we have, for $x \in \underline{X}$

$$E(U^\tau(x) \mid \tau_i = t_i) = E(x - (ex)b^{\tau'} \mid \tau_i = t_i) = x - (ex)\, E(b^\tau \mid \tau_i = t_i).$$

Finally, it turns out that in our present model the introduction of IC–mechanisms in general amounts to a restriction of the available utilities. To this end define

(12) $$V^{\mathfrak{M}} := \{E\, U^\tau \circ \mu^\tau \mid \mu \in \mathfrak{M}\}$$

and

(13) $$V^{\mathfrak{M}|t_i} := \{E(U^\tau \circ \mu^\tau \mid \tau_i = t_i) \mid \mu \in \mathfrak{M}\}.$$

Now, if $W_+ = W \cap \mathbb{R}^n_+$ denotes the nonnegative part of a subset W of $\mathbb{R}^n$, then we have

Lemma 2.6 $V^{\mathfrak{M}}$ and $V^{\mathfrak{M}|t_i}$ are nonempty, compact, convex polyhedra satisfying

$$V^{\mathfrak{M}} \subsetneq V^{\underline{X}}_+,\ V^{\mathfrak{M}|t_i}_+ \subsetneq V^{\underline{X}|t_i}_+ .$$

Proof: Both sets are nonempty containing the utility vector $0 \in \mathbb{R}^n$ which is obtained via the status quo mechanism $\underline{\mu}$.

Next, the set of mechanisms has the canonical linear structure of $\underline{X}^T \subsetneq (\mathbb{R}^n)^T$ and the inequalities defining IC–mechanisms ((4) in SEC.1, that is) can be rewritten as

$$(14)\qquad \sum U_i^{(t_i,t_{-i})}(\mu^{(t_i,t_{-i})})\,p(t|t_i) \geq \sum_{t_{-i}\in T_{-i}} (U_i^{(t_i,t_{-i})}(\mu^{(s_i,t_{-i})})\,p(t|t_i).$$

Here $t = (t_i,t_{-i}) \in T_i \times \prod_{j\neq i} T_j =: T_i \times T_{-i}$ and $(p(\cdot\,|t_i)$ is the conditional distribution of τ given $\tau_i = t_i$, i.e. $p(\cdot\,|t_i) = P(\tau = \cdot\,|\tau_i = t_i)$. Since U^t is affine, (14) constitutes a system of linear inequalities ("in μ"). Similarly the conditions defining mechanisms in $\mathfrak{M}$ (i.e. (4) in Definition 1.2 and (6) in Definition 1.3) constitute a system of linear inequalities. Hence $\mathfrak{M}$ is a compact convex polyhedron. The same is now true for $V^{\mathfrak{M}}$ and $V^{\mathfrak{M}|t_i}$ since e.g. the latter one is the image of $\mathfrak{M}$ under the composition of the affine mappings

$$\underline{X}^T \to (\mathbb{R}^n)^T;\ (\mu^t)_{t\in T} \to U_1^t(\mu^t),\dots,U_n^t(\mu^t))_{t\in T}$$

and

$$(\mathbb{R}^n)^T \to \mathbb{R}^n;\ (u_1^t,\dots,u_n^t) \to (\sum_{T^{-i}} u_1^t\,p(t|t_i),\dots,\sum_{T^{-i}} u_n^t\,p(t|t_i)).$$

Next, let us prove that $V^{\mathfrak{M}} \subseteq V^{\underline{X}}_{+}$. The elements of $V^{\mathfrak{M}}$ are positive, since by definition for $\mu \in \mathfrak{M}$

$$E(U_i^\tau \circ \mu^\tau) = \sum_{t_i\in T_i} E(U_i^\tau \circ \mu^T \mid \tau_i = t_i)\,P(\tau_i = t_i) \geq 0 \cdot \Sigma\,P(\tau_i = t_i)$$

(cf. Definition 1.3). Since $V^{\underline{X}}_{+}$ is a simplex spanned by 0 and $(1 - e\,E\,b^\tau)\,e^i$ $(i \in I)$ (cf. Lemma 2.5 and Fig.1), it is sufficient to show that, for $\mu \in \mathfrak{M}$, $E\,U^\tau \circ \mu^\tau$ is dominated by some element on the Pareto surface $\partial V^{\underline{X}}$ of $V^{\underline{X}}$. This, however, follows at once since

$$E\,U^\tau \circ \mu^\tau = E(\mu^\tau - (e\,\mu^\tau)\,b^\tau)$$

$$= \quad E(e\,\mu^\tau)\left[\frac{\mu^\tau}{e\mu^\tau} - b^\tau\right] \leq E\left[\frac{\mu^\tau}{e\mu^\tau} - b^\tau\right]$$

$$= \quad E\,\frac{\mu^\tau}{e\mu^\tau} - E\,b^\tau \in \partial V^{\underline{X}},$$

with an obvious modification for the case that $P(\mu^\tau = 0) > 0$. This proves the first relation of Lemma 2.6. The second can be done analogously. Note, however, that $V^{\mathfrak{M}|t_i}$ may contain elements with negative coordinates $j \neq i$.

Definition 2.7 1. μ is Pareto efficient (P.E.)

a) *ex ante*, if $E\,U^\tau \circ \mu^\tau$ is P.E. in $V^{\mathfrak{M}}$

b) *in mediis*, if there is no $\tilde{\mu} \in V^{\mathfrak{M}}$ such that, for all $i \in I$ and $t_i \in T_i$

$$E(U^\tau \circ \mu^\tau \mid \tau_i = t_i) \leq E(U^\tau \circ \tilde{\mu}^\tau \mid \tau_i = t_i)$$

holds true with *strict* inequality for at least some i and t_i

3) *ex post*, if there is no $\tilde{\mu} \in V^{\mathfrak{M}}$ such that, for all $i \in I$ and $t_i \in T_i$

$$U_i^t(\mu^t) \leq U_i^t(\tilde{\mu}^t)$$

holds true with *strict* inequality for at least one i and t_i.

2. μ^t is Pareto efficient in V^t if $U^t(\mu^t)$ is located on the Pareto surface of V^t, i.e., $\mu^t \in \partial V^t$ or, equivalently if $e\mu^t = 1$.

Note that (as in other models) P.E. *ex ante* implies P.E. *in mediis* implies P.E. *ex post.* For, e.g., if μ is not P.E. ex post, then $\tilde{\mu}$ as in 3. yields $\tilde{\mu}$ as in 2. by integrating with $p(\cdot \mid t_i)$.

Note also, that is much more to ask for "$\mu^t \in \partial V^t$ for all $t \in T$" as compared to 3. Indeed, this is a very strong requirement and, as it turns out, if it is satisfied than, generically, μ is a constant mechanism.

Theorem 2.8 Assume n = 2. Let I,T; $\underline{X}$,$\underline{x}$ and U be fixed. There is an open and dense set of probability distributions with the following properties: if $\mu \in \mathfrak{M}$ is such that μ^t is P.E. in V^t for every $t \in T$, then there is $\hat{x} \in \underline{X}$ with $\mu^t = \hat{x}$ $(t \in T)$.

Proof: First, fix $i \in I$, t_i, $s_i \in T_i$. By IC we know

(15) $$E(U_i^\tau \circ \mu^\tau \mid \tau_i = t_i) \geq E(U_i^\tau \circ \mu^{\tau_{-i}, s_i} \mid \tau_i = t_i)$$

and as $\mu^t \in V^t$ for all t we have $e\mu^t = 1$ $(t \in T)$, hence (15) reads

$$E(\mu_i^\tau - b_i^\tau \mid \tau_i = t_i) \geq (\mu^{\tau_{-i}, s_i} - b_i^\tau \mid \tau_i = t_i)$$

i.e.,

(16) $$E(\mu_i^\tau \mid \tau_i = t_i) \geq (\mu^{\tau_{-i}, s_i} \mid \tau_i = t_i)$$

which implies

$$E\, \mu_i^\tau \geq E(\mu_i^{\tau_{-i}, s_i})$$

for all $i \in I$ and all $s_i \in T_i$. Since $\mu^t \in V^t$ $(t \in T)$, summation over i yields

$$1 \geq E\, \mu_1^{s_1, \tau_2} + E_{\mu_2}^{\tau_1, s_2} = \sum_{t_2 \in T_2} \mu_1^{s_1, t_2}\, p^{t_2} + \sum_{t_1 \in T_1} \mu_2^{t_1, s_2} \qquad (s_1 \in T_1,\ s_2 \in T_2)$$

Multiplying each of these inequalities with $p^{s_1} = P(\tau_1 = s_1)$ and then summing over $s_1 \in T_1$ yields

$$1 \geq \sum_{s_1 \in T_1} \sum_{t_2 \in T_2} \mu_1^{s_1, t_2}\, p^{s_1}\, p^{t_2} + \sum_{t_1 \in T_1} \mu_2^{t_1, s_2}, \quad (s_2 \in T_2)$$

so that the same procedure performed with p^{t_2} finally results in

(17)
$$1 \geq \sum_{s_1 \in T_1} \sum_{t_2 \in T_2} \mu_1^{s_1,t_2} p^{s_1} p^{t_2} + \sum_{t_1 \in T_1} \sum_{s_2 \in T_2} \mu_2^{t_1,s_2}, p^{t_1} p^{s_2}$$
$$= \sum_{t \in T} (\mu_1^t + \mu_2^t)\, p^{t_1,t_2} = 1 .$$

Now, in view of (17) it follows that non of the inequalities employed can be a strict one, that is, in view of (16) we may conclude that

$$E(\mu_i^\tau \mid \tau_i = t_i) = E(\mu^{\tau_{-i},s_i} \mid \tau_i = t_i) \quad (i \in I, s_i \in T_i, t_i \in T_i)$$

hence

(18)
$$\sum_{t_2 \in T_2} \mu_1^{t_1,t_2} p^{t_1,t_2} = \sum_{t_2 \in T_2} \mu_1^{s_1,t_2} p^{t_1,t_2}$$

and

(18)
$$\sum_{t_1 \in T_1} \mu_2^{t_1,t_2} p^{t_1,t_2} = \sum_{t_1 \in T_1} \mu_1^{t_1,s_2} p^{t_1,t_2} .$$

Let us focus on (18). For any s_1 and t_1 the $|T_2|$–vectors $\mu_1^{s_1\cdot}$ and $p^{t_1\cdot}$ satisfy

(20)
$$(\mu_1^{s_1\cdot} - \mu_1^{s_1'\cdot})\, p^{t_1\cdot} = 0 .$$

These are for fixed s_1 and s_1' exactly $|T_1|$ equations. If $|T_1| \geq |T_2|$ then 0 is the only solution of (20) (i.e. of the linear system $x\, p^{t_1\cdot} = 0$ $(t_1 \in T_1)$) if the matrix $p^{\cdot\cdot}$ has full rank $|T_2|$ – which is the case for an open and dense set of matrices (distributions). If $|T_2| \geq |T_1|$, we draw the same conclusion from (19). Thus, (20) yields $\mu_1^{s_1\cdot} = \mu_1^{s_1'\cdot}$ and, as $\mu_1^{\cdot\cdot} + \mu_2^{\cdot\cdot} = 1$, we see that all $\mu_i^{t_1 t_2}$ are equal, q.e.d.

Remark 2.9 Suppose that in addition to $n = 2$ we require player 2 to have just one type, i.e.

(21) $$T = T_1 \times \{*\} .$$

Then it can be verified that the statement of Theorem 2.8 holds true for *all* probability distributions (instead of almost all).

SECTION 3 Incomplete information on one side

In this section we consider the special case where only two players are involved in the game, such that player 2's type is deterministic. Thus, player 1 *in mediis* has full information since he observes his own type and knows his opponents type.

Of course this is a severe restriction. In particular, various results of other authors apply to this restricted model – e.g. WEIDNER [10] (who is dealing with the case that types are distributed independently) should be taken into account.

Nevertheless, we feel that open questions can first be attached by restricting ourselves to this simple (?) case.

To be more precise, we assume throughout this section that T is of the special shape

$$T = T_1 \times \{*\} .$$

Then, it makes sense to write μ^{t_1} instead of $\mu^{(t_1,*)}$ $(t_1 \in T_1)$. Similarly, since player 1 is fully informed *in mediis* it is clear that $V^{\underline{X}|t_1} = V^{(t_1,*)}$ $(t_1 \in T_1)$, again we write V^{t_1} for this quantity. Similarly we use p^{t_1} instead of $p^{(t_1,*)}$ $(= p(t_1|*) = P(\tau_1 = t_1) = \cdots)$ etc.

Next, note that our special case allows for player 1 to induce an "ordering" of two types: If $\alpha,\beta \in T_1$ and $b_1^\alpha < b_1^\beta$ (i.e. he pays a smaller fee in case his type is α) then, for all $x \in \underline{X}$, $x \neq 0$

$$U_1^\alpha(x) = x_1 - (ex)\, b_1^\alpha < x_1 - (ex)\, b_1^\beta = U_1^\beta(x) .$$

Lemma 3.1 Let $T = \{\alpha,\beta\} \times \{*\}$ and assume that $b_1^\alpha < b_1^\beta$. Suppose that $\mu \in \mathfrak{M}$ is *ex ante* P.E. and satisfies $\mu^\alpha \neq \mu^\beta$. Then the following statements hold true

1. μ^α is V^α–Pareto efficient

2. $U_1^\alpha(\mu^\beta) = U_1^\alpha(\mu^\alpha)$

3. $U_1^\beta(\mu^\alpha) < U_1^\beta(\mu^\beta)$

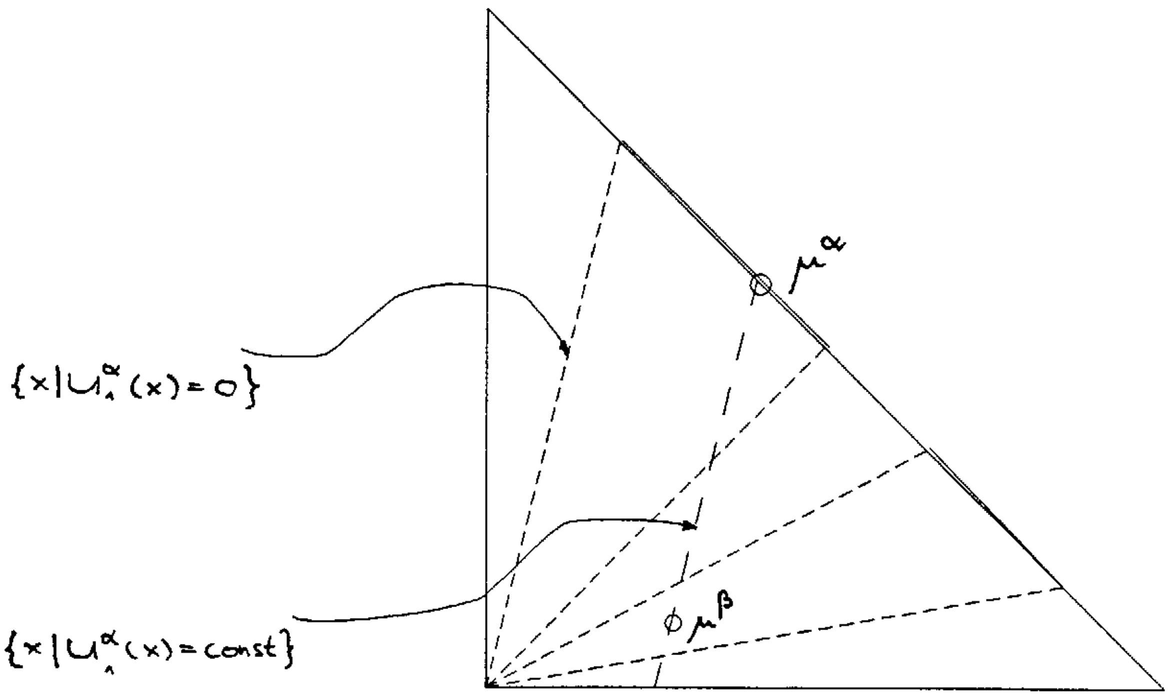

Fig.3
General position of non–constant $\mu \in \mathfrak{M}$

Proof: Observe that U_α^1 is linear with gradient

$$h_1^\alpha = (1 - b_1^\alpha, - b_1^\alpha),$$

thus the vector

$$\epsilon_1^\alpha := (b_1^\alpha, 1 - b_1^\alpha) > 0$$

satisfies

$$U_1^{\alpha}(\epsilon_1^{\alpha}) = h_1^{\alpha}\,\epsilon_1^{\alpha} = 0 \quad \text{(cf. Fig.4)}.$$

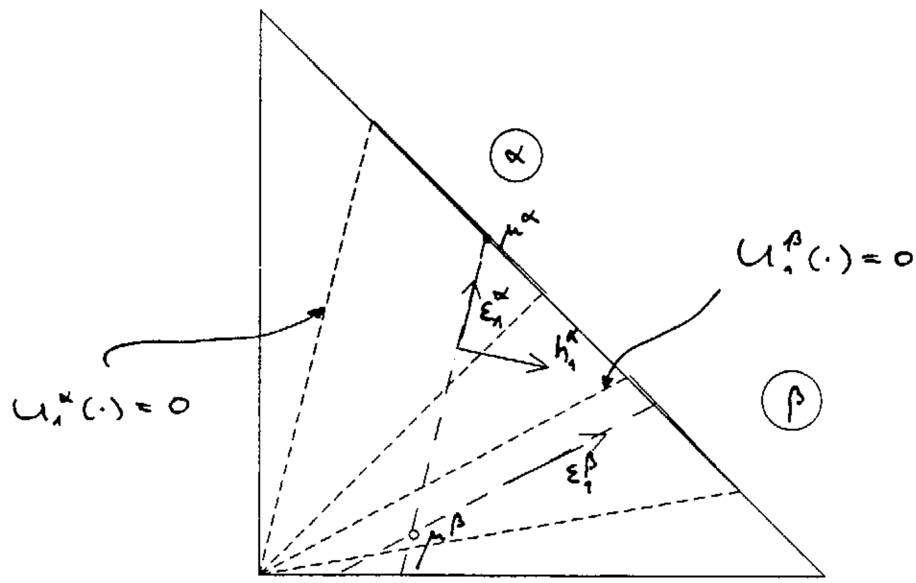

Fig.4
Gradients of utilities

Moreover,

$$U_1^{\beta}(\epsilon_1^{\alpha}) = h_1^{\beta}\,\epsilon_1^{\alpha}$$

(1) $$= (1-b_1^{\beta}, -b_1^{\beta})\,(b_1^{\alpha}, 1-b_1^{\alpha})$$

$$= b_1^{\alpha} - b_1^{\beta} < 0\,.$$

In order to prove the first statement, assume μ^{α} is not P.E. in V^{α}, then, for $\epsilon > 0$ sufficiently small let

$$\mu^{\alpha,\epsilon} = \mu^{\alpha} + \epsilon\,\epsilon_1^{\alpha} \qquad \epsilon\,\underline{X}$$

As μ is IC we observe that

(2) $$U_1^\alpha(\mu^{\alpha,\epsilon}) = U_1^\alpha(\mu^\alpha) \geq U_1^\alpha(\mu^\beta)$$

and in view of (1)

(3) $$U_1^\beta(\mu^{\alpha,\epsilon}) = U_1^\beta(\mu^\alpha) + \epsilon\, U_1^\beta(\epsilon_1^\alpha)$$

$$< U_1^\beta(\mu^\alpha) \leq U_1^\beta(\mu^\beta)$$

(again using that μ is IC.) Now, (2) and (3) show that $(\mu^{\alpha,\epsilon}, \mu^\beta) =: \mu^\epsilon$ is IC. Moreover, as

(4) $$U_2^\alpha(\epsilon_1^\alpha) = (-\,b_2^\alpha, 1\!-\!b_2^\alpha)\,(b_1^\alpha, 1\!-\!b_1^\alpha) = 1 - b_1^\alpha - b_2^\alpha = 1 - eb^\alpha > 0$$

it is seen that $\mu^\epsilon \in \mathfrak{M}$ and (again viewing (2) and (4)) that (no matter what P) μ^ϵ is an *ex ante* Pareto improvement compared to μ.

In order to prove the second statement, assume

$$U_1^\alpha(\mu^\beta) < U_1^\alpha(\mu^\alpha)\,.$$

Now, for small $\epsilon > 0$, consider

$$\mu^{\beta,\epsilon} := \mu^\beta + \epsilon\, \epsilon_1^\beta$$

Then

(5) $$U^\alpha{}_1(\mu^{\beta\epsilon}) < U_1^\alpha(\mu^\alpha)$$

is still true for small $\epsilon > 0$. Moreover

(6) $$U_1^\beta(\mu^\alpha) \leq U_1^\beta(\mu^\beta) = U_1^\beta(\mu^\beta + \epsilon\epsilon_1^\beta)$$

(as $U_1^\beta(\epsilon_1^\beta) = 0$). Again, from (5) and (6) it is derived at once that $\mu^\epsilon := (\mu^\alpha, \mu^{\beta,\epsilon}) \in \mathfrak{M}$.

Finally, in order to verify the third statement, assume that

$$U_1^\beta(\mu^\alpha) = U_1^\beta(\mu^\beta)$$

Then in view of (1), we have

$$h_1^\beta(\mu^\alpha - \mu^\beta) = 0$$
$$h_1^\alpha(\mu^\alpha - \mu^\beta) = 0.$$

But h_1^β and h_1^α are linear independent in view of

$$\begin{vmatrix} h_1^\beta \\ h_1^\alpha \end{vmatrix} = \begin{vmatrix} 1 - b_1^\beta, - b_1^\beta \\ 1 - b_1^\alpha, - b_1^\alpha \end{vmatrix}$$

$$= b_1^\beta - b_1^\alpha < 0.$$

Lemma 3.2 Let $T = T_1 \times \{*\}$ and let $\mu \in \mathfrak{M}$ be *ex ante* P.E. If, for $\alpha,\beta \in T_1$,

$$b_1^\alpha = b_1^\beta$$

then $\mu^\alpha = \mu^\beta$.

Proof: In this case, $U_1^\alpha(\cdot) = U_1^\beta(\cdot)$. Hence, using IC, yields

$$\begin{aligned} & U_1^\alpha(\mu^\alpha) \geq U_1^\alpha(\mu^\beta) = U_1^\beta(\mu^\beta) \\ (7) \qquad & \geq U_1^\alpha(\mu^\beta) = U_1^\beta(\mu^\beta) \geq U_1^\beta(\mu^\alpha) \\ & = U_1^\alpha(\mu^\alpha) \end{aligned}$$

where all inequalities have to be equations.

It follows that μ^α and μ^β both are multiples of $\epsilon_1^\alpha = \epsilon_1^\beta$, assume (w.l.g.) that $\mu^\alpha = \lambda_1\, \epsilon_1^\alpha > \lambda_2\, \epsilon_2^\alpha = \mu^\beta$.

Then

(8) $$U_2^\alpha(\epsilon_1^\alpha) = (-\,b_2^\alpha, 1-b_2^\alpha)\,(b_1^\alpha, 1-b_1^\alpha) = 1 - b_2^\alpha - b_1^\alpha > 0$$

and

(9) $$U_2^\beta(\epsilon_1^\alpha) = U_2^\beta(\epsilon_1^\beta) > 0,$$

thus

(10) $$U_2^{t_1}(\mu^\alpha) > U_2^{t_1}(\mu^\beta)$$

for $t_1 = \alpha,\beta$. It follows that $\bar{\mu} = (\mu^\alpha,\mu^\alpha)$ is IC (by (7)) and an *ex ante* improvement for player 2 (by (10)), thus $\bar{\mu} \in \mathfrak{M}$ follows at once as well as the fact that $\bar{\mu}$ Pareto dominates μ *ex ante*. The assumption $L_1 > \lambda_2$ must, therefore be wrong, q.e.d.

Theorem 3.3 Let $T = T_1 \times \{*\}$ and consider $\mu \in \mathfrak{M}$, μ *ex ante* P.E.. Let

$$T_1^l \quad (l = 1,\dots,L)$$

denote the disjoint subsets of T_1 given by the requirements

$$b_1^\alpha = b_1^\beta \qquad (\alpha,\beta \in T_1^l) \qquad\qquad (l = 1,\dots,L)$$

$$b_1^\alpha < b_1^\beta \qquad (\alpha \in T_1^l,\ \beta \in T_1^{l+1}) \qquad\qquad (l = 1,\dots,L-1)$$

such that

$$T_1 = \sum_{l=1}^{L} T_1^l$$

(using Σ instead of $\cup$ for disjoint unions). Then the following holds true:

1. μ^{α} is V^{α} Pareto efficient $(\alpha \in T_1^1)$

2. $U_1^{\alpha}(\mu^{\beta}) = U_1^{\beta}(\mu^{\beta}) \qquad (\alpha \in T_1^l,\ \beta \in T_1^{l+1},\ l = 1,\ldots,L-1)$

3. $U_1^{\beta}(\mu^{\alpha}) < U_1^{\beta}(\mu^{\beta}) \qquad (\alpha \in T_1^l,\ \beta \in T_1^{l+1},\ l = 1,\ldots,L-1)$

Proof: Use the techniques supplied by 3.1 and 3.2 (cf. Fig.5).

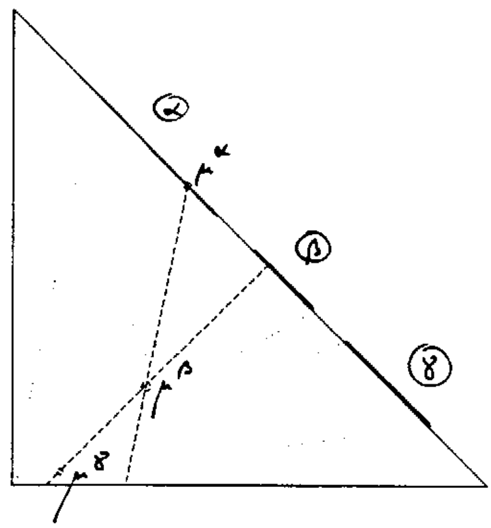

Fig.5
Three types in Figure 3

Theorem 3.4 Let $T = \{\alpha,\beta\} \times \{*\}$ and $b_1^{\alpha} < b_1^{\beta}$. Suppose $\mu \in \mathfrak{M}$ is *ex ante* P.E. and satisfies $\mu^{\alpha} \neq \mu^{\beta}$. Then the following holds true.
If $E\ U_2^{\tau} \circ \mu^{\tau} > 0$, then $U_1^{\beta}(\mu^{\beta}) = 0$.

Proof: For small $\eta > 0$ define

(11) $$\mu^{\alpha,\eta} := \mu^{\alpha} + p^{\beta}\,(\eta, -\eta)$$

(12) $$\mu^{\beta,\eta} := \mu^{\beta} + p^{\alpha} \frac{\eta}{-U_1^{\beta}(\epsilon_1^{\alpha})}\,\epsilon_1^{\alpha} + p^{\beta} \frac{\eta}{-U_1^{\alpha}(\epsilon_1^{\beta})}\,\epsilon_1^{\beta}$$

Assuming $U_2^{\alpha}(\mu^{\alpha}) > 0$, $U_1^{\beta}(\mu^{\beta}) > 0$ it is clear that $\mu^{\cdot,\eta} = (\mu^{\alpha,\eta}, \mu^{\beta,\eta})$ is individually rational *in mediis* for player 1 ($U_1^{\alpha}(\mu^{\alpha,\eta}) > U_1^{\alpha}(\mu^{\alpha})$!) (for small $\eta > 0$).

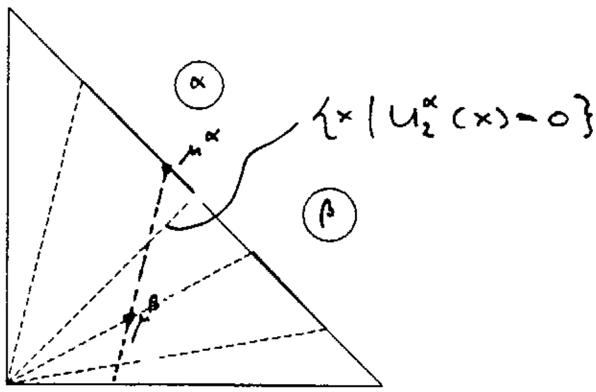

Fig.6
General position of non–constant $\nu \in \mathfrak{M}$
– the final version –

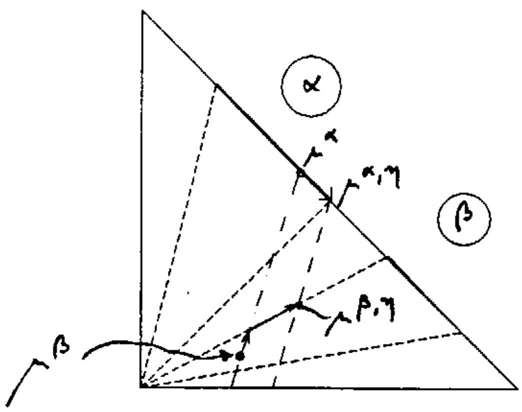

Fig.7
The proof for Fig.6

Moreover, $\mu^{\cdot,\eta}$ is IC. For, in view of 3.1, we know that

$$U_1^\beta(\mu^\alpha) < U_1^\beta(\mu^\beta)$$

and this relation is preserved with respect to $\mu^{\cdot,\eta}$ for small η. In addition (again using 3.1)

$$U_1^\alpha(\mu^{\beta,\eta}) = U_1^\alpha(\mu^\beta) + p^\beta\eta = \ldots$$

(since $U_1^\alpha(\epsilon_1^\alpha) = 0$)

$$\ldots = U_1^\alpha(\mu^\alpha) + p^\beta\,\eta = \ldots$$

(by 3.1)

$$\ldots = U_1^\alpha(\mu^{\alpha,\eta}) .$$

Thus, $\mu^{\cdot,\eta}$ is IC and hence $\mu^{\cdot,\eta} \in \mathfrak{M}$. We are now going to show that $\mu^{\cdot,\eta}$ is an *ex ante* improvement (strictly for player 2) for both players – this proves also $\mu^{\cdot,\eta} \in \mathfrak{M}$ (since for player 2 *ex ante* = *in mediis*) and hence establishes the Theorem.

Now

(13) $$E\,U_1^\tau \circ \mu^{\tau,\eta} = E\,U_1^\tau \circ \mu + p^\alpha p^\beta\,\eta + p^\beta p^\alpha \frac{\eta}{-U_\beta^1(\epsilon_1^\alpha)} \cdot U_\beta^1(\epsilon_1^\alpha)$$
$$= E\,U_1^\tau \circ \mu$$

and

$$E\,U_2^\tau \circ \mu^{\tau,\eta} = E\,U_2^\tau \circ \mu - p^\alpha p^\beta\,\eta + p^\beta p^\alpha \frac{\eta U_2^\beta(\epsilon_1^\alpha)}{-U_1^\beta(\epsilon_1^\alpha)}$$
$$+ (p^\beta)^2 \frac{\eta U_2^\beta(\epsilon_1^\beta)}{-U_1^\alpha(\epsilon_1^\beta)}$$

Next, compute

(15) $$U_1^\alpha(\epsilon_1^\beta) = b_1^\beta - b_1^\alpha > 0$$

and

(16) $$U_2^\beta(\epsilon_1^\beta) = 1 - b_2^\beta - b_1^\beta = 1 - eb^\beta > 0 ,$$

hence, it suffices to show

(17) $$\frac{U_2^\beta(\epsilon_1^\alpha)}{-U_1^\beta(\epsilon_1^\alpha)} \geq 1 .$$

However,

$$\begin{aligned} U_2^\beta(\epsilon_1^\alpha) &= (-b_2^\beta, 1-b_2^\beta)\,(b_1^\alpha, 1-b_1^\alpha) \\ &= 1 - b_2^\beta - b_1^\alpha \\ &> b_1^\beta - b_1^\alpha = -U_1^\beta(\epsilon_1^\alpha) \end{aligned}$$ q.e.d.

There are a few conclusions we can easily draw in view of this result.

Corollary 3.5 Let $T = \{\alpha,\beta\} \times \{*\}$ and $b_1^\alpha < b_1^\beta$. Suppose $\mu^\alpha \in \underline{X}$ is V^α efficient (i.e. $e\mu^\alpha = 1$) and $U_1^\alpha(\mu^\alpha) \geq 0$. If $\mu^\beta \in \underline{X}$ exists such that $(\mu^\alpha, \mu^\beta) \in \mathfrak{M}$ is *ex ante* P.E., then it is uniquely defined.

Proof: As the geometrical situation is cleared by Theorem 3.4 (and represented by Fig.6), we shall restrain ourselves to some verbal–geometrical argument. Clearly, if $\mu^\beta \neq \mu^\alpha$, then μ^β is uniquely defined as the intersections of the two lines $\{x \mid U_1^\alpha(x) = U_1^\alpha(\mu^\alpha)\}$ and $\{x \mid U_1^\beta(x) = 0\}$ as depicted in Fig.6. If these lines do not intersect in $\underline{X}$ then $\mu^\beta = \mu^\alpha$ constitutes $\mu \in \mathfrak{M}$ if and only if $E\ U_2^\tau(\mu^\alpha) \geq 0$. Thus, it is not hard to give precise conditions so that μ^β exists and to actually provide a formula for its computation.

Corollary 3.6 ("Unique implementation")

Let the conditions of 3.4 and 3.5 be satisfied. Let $u \in V^{\mathfrak{M}}$ be $V^{\mathfrak{M}}$–Pareto efficient (i.e., located on the Pareto surface of $V^{\mathfrak{M}}$). Then there exists *uniquely* $\mu \in \mathfrak{M}$ such $E\ U^\tau \circ \mu^\tau = u$.

Proof: Again we shall argue verbally and claim that the details are obtained by inspecting 3.4 and Fig.6. We distinguish three cases.

Case 1: Assume that u is in $V^{\underline{X}}$ Pareto efficient. Then, clearly, μ^t must be V^t efficient for $t = \alpha,\beta$. By Theorem 2.8 – or rather Remark 2.9 – μ is constant, i.e., $\mu^t = \hat{x}$ for some $\hat{x} \in \underline{X}$. Since

(18) $$u = E\, U^{\tau}(\hat{x}) = \hat{x} - (e\hat{x})\, Eb^{\tau} = \hat{x} - Eb^{\tau},$$

we obtain $\hat{x} = u - Eb^{\tau}$

Case 2: Now, if u is not efficient in $V^{\underline{X}}$, then $\mu^{\alpha} \neq \mu^{\beta}$. We shall distinguish Case 2a and Case 2b according to Theorem 3.4.

Case 2a: $U_2^{\alpha}(\mu^{\alpha}) > 0$. Then $U_1^{\beta}(\mu^{\beta}) = 0$.

It is not hard to see that the four linear equations

(19) $$\begin{aligned} u &= U^{\alpha}(\mu^{\alpha}) + U^{\beta}(\mu^{\beta}) \\ &= \mu^{\alpha} - (e\mu^{\alpha})\, b^{\alpha} + \mu^{\beta} - (e\mu^{\beta})\, b^{\beta} \end{aligned}$$

(20) $$e\,\mu^{\alpha} = 1 \qquad \text{(by 3.3)}$$

(21) $$U_1^{\alpha}(\mu^{\beta}) = U_1^{\alpha}(\mu^{\alpha}) \qquad \text{(by 3.3)}$$

(22) $$U_1^{\beta}(\mu^{\beta}) = 0 \qquad \text{(by 3.4)}$$

determine μ^{α} and μ^{β} uniquely.

Case 2b: If $U_2^{\alpha}(\mu^{\alpha}) = 0$, then we argue analogously to 2a. q.e.d.

Corollary 3.7 Let the conditions of 3.4 and 3.5 be satisfied. Let $\hat{x} \in \underline{X}$ be such that $\hat{\mu}$, as defined by $\hat{\mu}^t = \hat{x}$ ($t \in T$) satisfies $\hat{\mu} \in \mathfrak{M}$. Then

$$U_1^{\alpha}(\mu^{\alpha}) < U_1^{\alpha}(\hat{x}),\ U_1^{\beta}(\mu^{\beta}) = 0 \text{ and } E\ U_1^{\tau} \circ \mu^{\tau} < E_1^{\tau}\ U_1^{\tau} \circ \hat{\mu}^{\tau}.$$

Thus, in the present situation, the expectation for player 1 with respect to a non-constant mechanism in $\mathfrak{M}$ is worse than compared to any constant mechanism in $\mathfrak{M}$.

Again, the proof rests on a geometrical argument: inspection of Fig.6 shows that, for player 1, $U_1^{\alpha}(\mu^{\alpha}) < U_1^{\alpha}(\hat{x})$ (for $\hat{x}$ must be "in mediis IR" in situation β) while $U_1^{\beta}(\mu^{\beta}) = 0 < U_1^{\beta}(\hat{x})$. q.e.d.

It is now easy to visualize the shape of $V^{\mathfrak{M}} \subseteq V^{\underline{X}}_{+}$ (cf. Fig.8).

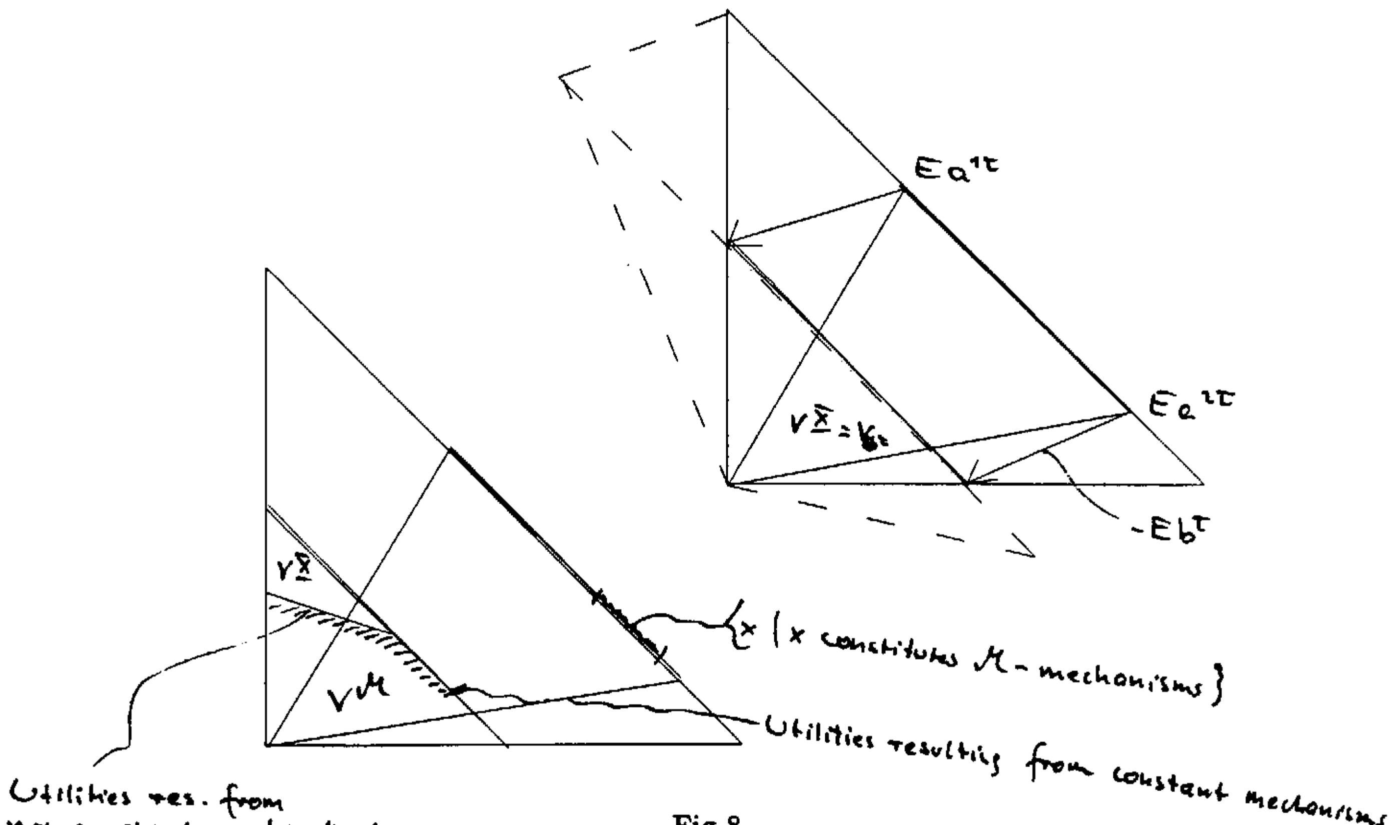

Fig.8
The shape of $V^{\mathfrak{M}}$

Theorem 3.8 Let $T = T_1 \times \{*\}$ and let T_1^l (l=1,...,L) denote the number of the decomposition of T_1 indicated in 3.3, i.e. $b_1^\alpha < b_1^\beta$ ($\alpha \in T_1^l$, $\beta \in T_1^{l+1}$ l=1,...,L–1). Then the following holds true:

If $E\, U_2^\tau \circ \mu^\tau > 0$ for $\alpha \in T_1^1$ then $U_1^\beta(\mu^\beta) = 0$ for $\beta \in T_1^L$.

Proof: For notational convenience, we shall restrict ourselves to the case that $T_1 = \{\alpha,\beta,\gamma\}$ and $b_1^\alpha < b_1^\beta < b_1^\gamma$.

For sufficiently small $\eta > 0$ take $\mu^{\cdot,\eta}$ to be defined by the following

$$\mu^{\alpha,\eta} = \mu^\alpha + p^\beta\, p^\gamma(\eta, -\eta)$$

$$\mu^{\beta,\eta} = \mu^\beta + p^\beta\, p^\gamma \eta \frac{\epsilon_1^\beta}{U_1^\alpha(\epsilon_1^\beta)}$$

$$\mu^{\gamma,\eta} = \mu^\gamma + p^\alpha\, p^\beta \eta \frac{\epsilon_1^\beta}{U_1^\gamma(\epsilon_1^\beta)} .$$

It is not hard to see that this constitutes an *in mediis* i.r. mechanism.

In order to show that the IC–property is preserved during the transition from μ to $\mu^{\cdot,\eta}$, we have to distinguish 3 cases.

a) ("α,β") Player 1 does not pretend to be β when α is his type; because

$$U_1^\alpha(\mu^{\alpha,\eta}) = U_1^\alpha(\mu^\alpha) + p^\beta\, p^\gamma\, \eta$$

$$U_1^\alpha(\mu^{\beta,\eta}) = U_1^\alpha(\mu^\beta) + p^\beta\, p^\gamma\, \eta$$

b) ("α,γ") Player 1 does not pretend to be γ when α is his type. This follows for small $\eta > 0$ since $U_1^\alpha(\mu^\alpha) > U_1^\alpha(\mu^\gamma)$ holds true –

which is a consequence of Theorem 3.3.

c) ("β,γ") Player 1 does not pretend to be γ when β is his type, because

$$U_1^\beta(\mu^{\beta,\eta}) = U_1^\beta(\mu^\beta)$$

$$U_1^\beta(\mu^{\gamma,\eta}) = U_1^\beta(\mu^\gamma)\ .$$

Thus, we have established $\mu^{\cdot,\eta} \in \mathfrak{M}$. It remains to show that this mechanism constitutes a Pareto improvement versus μ.

Now, writing $\pi := p^\alpha\, p^\beta\, p^\gamma$ we observe that player 1 has expectation

$$\begin{aligned} E\ U_1^\tau \circ \mu^{\tau,\eta} \quad &= E\ U_1^\tau \circ \mu^\tau \\ &+ \pi\,\eta + 0 + \pi\,\eta \cdot (-1) \\ &= E\ U_1^\tau \circ \mu^\tau\ , \end{aligned}$$

while player 2 has expectation

$$\begin{aligned} E\ U_2^\tau \circ \mu^{\tau,\eta} \quad &= E\ U_2^\tau \circ \mu^\tau \\ &- \pi\,\eta \\ &+ (p^\beta)^2\, p^\gamma\, \eta \frac{U_2^\beta(\epsilon_1^\beta)}{U_1^\alpha(\epsilon_1^\beta)} \\ &+ \pi\,\eta \frac{U_2^\gamma(\epsilon_1^\beta)}{-U_1^\gamma(\epsilon_1^\beta)}\ . \end{aligned}$$

However, the quotient $\dfrac{U_2^\beta(\epsilon_1^\beta)}{U_1^\alpha(\epsilon_1^\beta)}$ has already been established to be positive (see (15) and (16), this term appears in the proof of 3.4 as well). Thus it remains to show that

(18) $$\frac{U_2^{\gamma}(\epsilon_1^{\beta})}{-U_1^{\gamma}(\epsilon_1^{\beta})} \geq 1$$

– which is the exact analogue to (17), q.e.d.

Note that, structurally, the case $|T_1| = 3$ is slightly different compared to $|T_1| = 2$. While the proof seems a bit easier and more straight forward, the result is of course slightly weaker: If player 2 in his "worst type" has positive utility, than player 1 in *his* worst type has zero utility – nothing is said about the intermediate types.

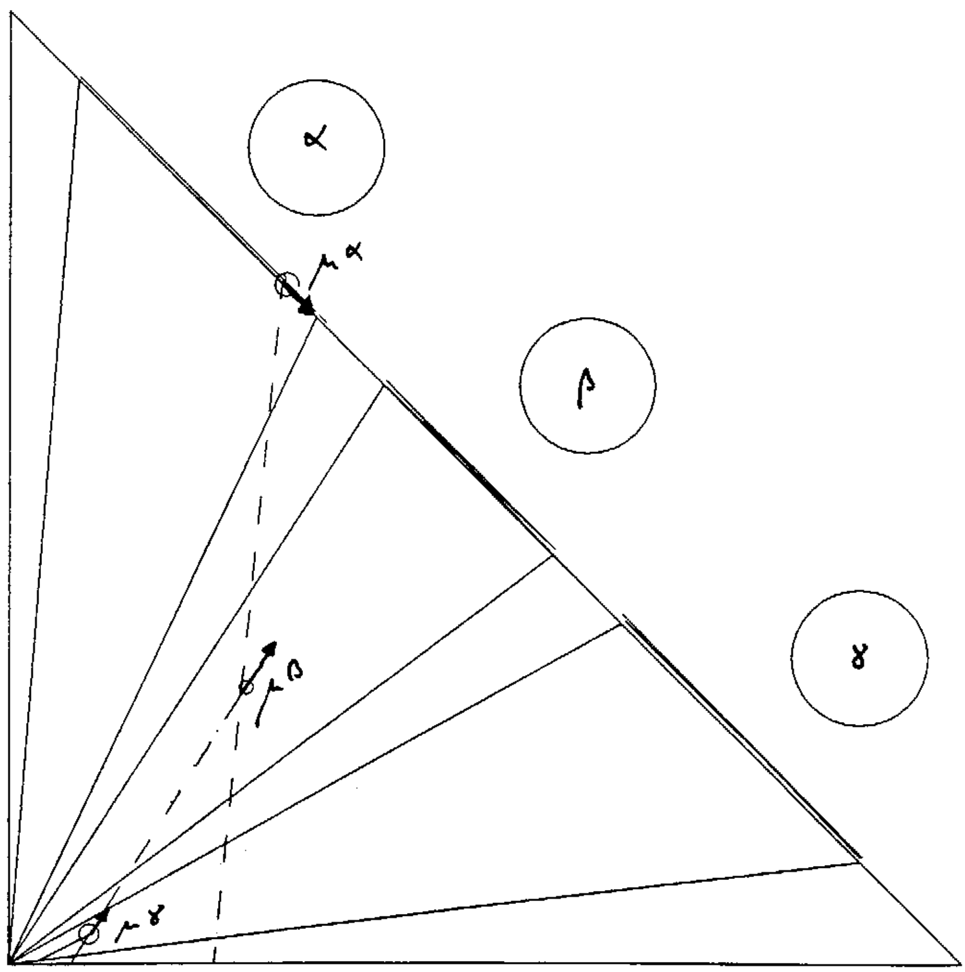

Fig.9
Three types in Figure 6

SECTION 4 The expected contract

Consider the Standard T.U–game $(0,V^t)$, i.e., the *ex post* version when $t \in T$ has been realized by the chance move. A "value" (e.g., the NASH–bargaining solution) in the "classical" context of complete information is a mapping, say

$$\chi : \{u,W) \mid \mathbb{R}^n \supseteq W \text{ compact, convex; } u \in W\} \ .$$

Given some elementary axioms (symmetry etc.) such a value (e.g. the NASH–solution) will satisfy

$$\chi(0,V^t) = \frac{e}{n}(1-eb^t) =: \bar{u}^t \tag{1}$$

(where $V^t = V_{b^t}$ (see 2.1) and $e = (1,...,1)$).

$\bar{u}^t$ can be considered a (rather undisputable) solution for the *ex post* situations . And the *ex ante* expectation of players if in each *ex post* situation $\bar{u}^t$ could be chosen (i.e. in the case of full information) is given by

$$E\,\chi(0,V^\tau) = E\,\bar{u}^t = \frac{e}{n}(1-e\,Eb^\tau) + \chi(0,V_{Eb^t}) = : \bar{u} \ . \tag{2}$$

This formula (which is of course resting heavily on linearity) shows that the above mentioned *ex ante* expectation is as well the value of the "expected" STU game $(0,V_{Eb^\tau})$. (Of course $V_{Eb^\tau} = V^{\underline{X}}$ (cf. 2.5)).

Clearly, given complete information, players would register the mechanism implied by

$$\begin{aligned} \bar{x}^t \quad &:= \frac{e}{n}(1-b^t) + b^t \\ &= \bar{u}^t + b^t = (U^t)^{-1}(\bar{u}^t) \\ &= \frac{1}{n}\sum_{k=1}^{n} a^{tk} \end{aligned} \tag{3}$$

(cf. 2.3 for the meaning of the a^{tk}). This motivates the following.

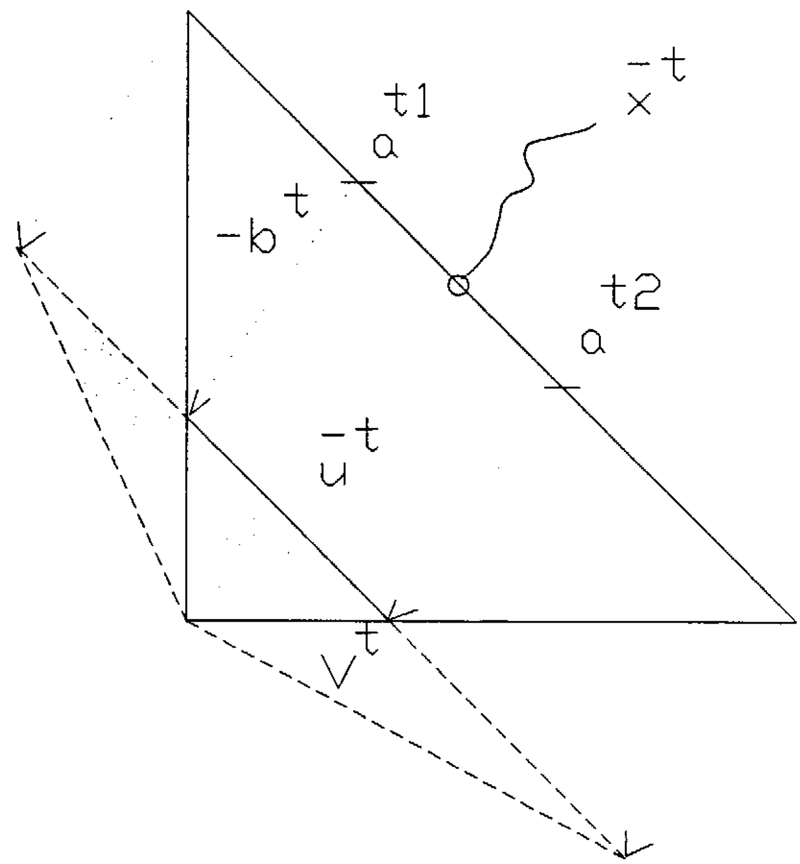

Fig.10
The expected contract

Definition 4.1 The *expected contract* is given by

$$\bar{x} = E\,x^{\tau} = \frac{e}{n}\,(1\!-\!e\,Eb^{\tau}) + Eb^{\tau}\,.$$

Note that the expected utility of $\bar{x}$ equals $\bar{u}$, i.e.

(4)
$$\begin{aligned} E\,U^{\tau}(\bar{x}) &= E(\bar{x} - (e\,\bar{x})b^{\tau}) \\ &= \bar{x} - Eb^{\tau} \\ &= \frac{e}{n}\,(1\!-\!e\,Eb^{\tau}) = \bar{u} \end{aligned}$$

that is, the expected contract implements the ex ante expectation of a situation with complete information.

Thus, if players had no information, they could register $\bar{x}$ and have the same expected utility (ex ante) as in a world of truth speaking individuals in which, after each choice of t, everybody would reveal his type and the fair $\chi(0,V^t)$ would result by an obvious agreement.

Now, suppose players, being aware that the world is not the ideal place they (we all) would like it to be, agree upon the implementation of BIC mechanisms.

When bargaining *ex ante*, they might want to have a mechanism which is *ex ante* Pareto efficient. In that case, there should be no reason for deviating from the result that – in *ex ante* expectation – was presented by the fairness considerations "with full information" – provided of course, such BIC mechanism exists at all.

Now we have

Lemma 4.2 Assume $n = 2$. Let I,T; $\underline{X}$, $\underline{x}$ and U be fixed. There is an open and dense set of probabilities (distribution of τ) such that the following holds true:

Let μ be BIC such that

$$E\, U^{\tau}(\mu^{\tau}) = \bar{u}$$

Then $\mu^t = \bar{x}$ $(t \in T)$.

That is, the only way to implement a fair *ex ante* utility is by means of the (constant) mechanism that chooses the expected contract.

Proof: It is easily seen that μ is V^t–Pareto efficient for all $t \in T$. Therefore, we may choose the dense and open set according to the result of Theorem 2.8. Then, as μ is BIC there is $\hat{x} \in \underline{X}$ such that

$$\mu^t = \hat{x} \quad (t \in T)$$

holds true. Now, as

$$\begin{aligned}\bar{x} - Eb^{\mathcal{T}} &= \bar{u} = E\, U^{\mathcal{T}} \circ \mu^{\mathcal{T}} \\ &= E\, U^{\mathcal{T}}(\hat{x}) = \hat{x} - (e\hat{x})\, Eb^{\mathcal{T}} \\ &= \hat{x} - Eb^{\mathcal{T}},\end{aligned}$$

we conclude that $\hat{x} = \bar{x}$, q.e.d.

The simple result of Lemma 4.2 bears some consequences in view of a discussion regarding the concept of a "bargaining solution" or "value for CII–games". For our present purpose it suffices to fix I,T,$\underline{X}$ and $\underline{x} = 0$, hence a value would be a mapping

$$\psi : \{(p, U) \mid \Gamma = (I, T; p; \underline{X}, \underline{x}; U) \text{ is CII}\} \to \underline{X}^{T}$$

satisfying certain axioms.

We do not want to discuss the axioms of symmetries. Also, in the present ("side payment") context, an axiom of covariance with respect to linear rescaling of utility cannot be formulated – note, however, that rescaling the elements of $\underline{X}$ and U is also covariant, thus there is no problem in discussing "Fee–games with varying currency".

However, we want a solution to yield, at any instant, an ex ante Pareto efficient mechanism $\mu \in \mathfrak{M}$. More precisely (if we write $\psi(\Gamma)$ instead of ψ (p,U) by obvious reasons and use $\mathfrak{M}(\Gamma)$ as an obvious extension of notation):

Axiom $\mathfrak{M}$:

"For any Γ, $\psi(\Gamma) \in \mathfrak{M}(\Gamma)$ and $\psi(\Gamma)$ is *ex ante* Pareto efficient"

Furthermore, we believe that the expected contract $\bar{x}$ should be chosen if it induces a mechanism in $\mathfrak{M}$. That is

Axiom $\bar{x}$: (The expected contract axiom)

"Let $\bar{\mu} = \bar{\mu}(\Gamma)$ be the (constant) mechanism induced by $\bar{x}$ (= $\bar{x}(\Gamma)$), i.e.,

$$\bar{\mu}^t = \bar{x} \quad (t \in T) .$$

If $\bar{\mu} \in \mathfrak{M}$, then $\psi(\Gamma) = \bar{\mu}$.

There are two motivations we may offer for this axiom, and we shall discuss them in the light of the case $T = \{\alpha, \beta\} \times \{*\}$.

In this case, player 2 (since *ex ante* and *in mediis* is the same for him) initially viewed $V^{\underline{X}}$. However, player 1 convinced him, that he would not accept anything but IR *in mediis* and this led them to discuss mechanisms – IC and IR mechanisms, that is. So now, player 2 views $V^{\mathfrak{M}}$. Suppose it turns out that $\bar{u} \in V^{\mathfrak{M}}$. Player 2, a firm supporter of the IIA–Axiom, argues that $V^{\mathfrak{M}} \subseteq V^{\underline{X}}_{+}$ (Lemma 2.6) and it turns out that $\bar{u}$ is uniquely "implemented" by $\bar{\mu}$ (Lemma 4.2).

Player 1, if he wants to argue against this, may propose different *constant* contracts, to start with. However, any constant μ is generated by some $x \in \underline{X}$. But then the discussion may as well be conducted in the ex ante situation. The only difference (with respect to the "no information" case) is that player 1 knows he will not regret his decision to contract with player 2 *in mediis* – but the IC property is not needed at all. Thus, all arguments directed against $\bar{x}$ can be discussed from the *ex ante* viewpoint (this is an IIA argument) – and it is hard to see any such arguments as $\bar{x}$ satisfies all symmetry, efficiency (and covariance) requirements.

Next, player 1 could argue for nonconstant mechanisms.

Lemma 4.3 Let $T = \{\alpha,\beta\} \times \{*\}$ and $b_1^{\alpha} < b_1^{\beta}$. Assume $\mu \in \mathfrak{M}$ to be *ex ante* Pareto efficient and $\mu^{\alpha} \neq \mu^{\beta}$. Also assume that $\bar{\mu} \in \mathfrak{M}$. Then

$$U_1^{\alpha}(\mu^{\alpha}) < U_1^{\alpha}(\overline{x}),\ U_1^{\beta}(\mu^{\beta}) = 0,$$

and

$$E\ U_1^{\tau} \circ \mu^{\tau} < E\ U_1^{\tau} \circ \overline{\mu}^{\tau} = \overline{u}_1\ .$$

Proof: This is a consequence of Corollary 3.7.

Clearly, this settles all attempts of player 1 to argue towards nonconstant mechanisms: he will be worse of in any situation. no matter whether he considers *ex ante or in mediis* situations.

Are there values satisfying **Axiom $\overline{\mathbf{x}}$**? As is easily seen, any mechanism "implementing" the NASH–value of $V^{\mathfrak{M}}$ will do this (and in our "standard case", $T = \{\alpha,\beta\} \times \{*\}$ this mechanism is uniquely defined). However, it needs some more axiomatic field work to characterize a value for CII games. We will deal with this question in a further discussion.

LITERATURE

[1] D'Aspremont, C. and Gérard–Varet, L.–A.:
Incentives and incomplete information.
Journal of Public Economics 11 (1979), pp. 25 – 45

[2] Harsanyi, J.C. and Selten, R.:
A generalized Nash solution for two–person bargaining games with incomplete information.
Management Science 18 (1972), pp. 80 – 106

[3] Holmström, B. and Myerson, R.B.:
Efficient and durable decision rules withincomplete information.
Econometrica 51 (1983), pp. 1799 – 1819

[4] Hurwicz, L.:
On informationally decentralized systems.
Decision and Organisation. (R. Radner and B. McGuire eds.)
North Holland (1972), pp. 297 – 336

[5] Myerson R.B.:
Incentive compatibility and the bargaining problem.
Econometrica 47 (1979), pp. 61 – 73

[6] Myerson, R.B.:
Two–person bargaining problems with incomplete information.
Econometrica 52 (1984), pp. 461 – 487

[7] Myerson, R.B.:
Cooperative games with incomplete information.
International Journal of Game Theory 13 (1984), pp. 69 – 96

[8] Myerson, R.B. and Satterthwaite, M.A.:
Efficient mechanisms for bilateral trading.
Journal of Economic Theory 29 (1983), pp. 265 – 281

[9] Rosenmüller, J.:
Remarks on cooperative games with incomplete information.
Working paper 166, Inst. of Math. Econ. (IMW),
University of Bielefeld (1988), 44 p.

[10] Weidner, F.:
On the Harsanyi–Selten Value.
Working paper 180, Inst. of Math. Econ. (IMW),
University of Bielefeld (1989), 44 p.