On the Inconsistency of Equilibrium Refinement*

Werner Güth†

Abstract

Consistency and optimality together with converse consistency provide an illuminating and novel characterization of the equilibrium concept (Peleg and Tijs, 1996). But (together with non-emptiness) they preclude refinements of the equilibrium notion and selection of a unique equilibrium (Norde et al., 1996). We suggest two escape routes: By generalizing the concept of strict equilibrium we question the practical relevance of the existence requirement for refinements. To allow for equilibrium selection we suggest more complex reduced games which capture the inclinations of "players who already left".

*The author is thankful for comments by Martin Dufwenberg, Jörg Oechssler and Klaus Ritzberger and the constructive advice of two anonymous referees and the editor in charge.

†Max Planck Institute for Research into Economic Systems, Strategic Interaction Group, Kahlaische Str. 10, 07745 Jena, Germany

1. Introduction

Consistency is a compelling property of many solution concepts for cooperative games (see, for instance, Harsanyi, 1959; Lensberg, 1988; Maschler, 1990; Thomson, 1990). For strategic games consistency of the equilibrium concept is well-known (e.g. Aumann, 1987) and provides a novel and illuminating characterization of the equilibrium concept for strategic games (see Peleg and Tijs, 1996).[1]

A solution $s = (s_1, ..., s_n)$ of a game with player set $N = \{1, ..., n\}$ is **consistent** if for every non-empty subset M of N the strategy constellation $s_M = (s_i)_{i \in M}$ is a solution of the reduced game whose active player set is M whereas all other players $j \notin M$ are constrained to their strategies s_j. Thus in view of the reduced game players $j \notin M$ have already left after choosing their strategies s_j. Intuitively a consistent solution allows the players in M to reconsider their choice when the players not in M have already decided. Consistent solutions thus have a **decentralization property** in the sense of being immune against reconsiderations by smaller subgroups.[2] Clearly, equilibria s are consistent: If one does not want to deviate from s in the original game, a unilateral deviation is also unprofitable in the reduced game. The consistency axiom, optimality and converse consistency (all consistent strategy constellations have to be included in the solution set) jointly characterize the equilibrium concept (Peleg and Tijs, 1996).

Recommending to play one of the many possible equilibria is often not very helpful.[3] More specific advice[4], however, requires to select one equilibrium as the

[1] A more trivial characterization is by optimality and true expectations (see, for instance, Aumann and Brandenburger, 1995, see also Nash, 1951).

[2] It is interesting that sequential rationality also applies to subgroups (of agents) but not to all subgroups (of agents) as required by consistency.

[3] To illustrate how little is gained by just predicting equilibrium behavior consider the **ultimatum game**. Here a positive sum c can be distributed among player 1 and 2. First player 1 demands d with $0 \leq d \leq c$. Then player 2 can accept, what yields d for 1 and $c - d$ for 2, or not what implies 0-payoffs for both. Here (in addition to $d = c$ and general rejection) any allocation of c among the two players is an equilibrium outcome (the responder, player 2, must reject any offer below the equilibrium offer which he accepts). The example illustrates the need to select a unique equilibrium if game theory is supposed to resolve strategic uncertainty.

[4] Refusing to refine the equilibrium concept (see, for instance, Aumann, 1996) would have to justify why certain rationality requirements, namely those characterizing the equilibrium concept, are acceptable and more demanding ones not. The axioms characterizing risk dominance for 2x2-games with two strict equilibria are, for instance, rather convincing (see Harsanyi and Selten, 1988) and may be supplemented to cover larger classes of games.

solution of the game or at least to refine the equilibrium notion. A consistent selection of a unique equilibrium solution is, however, impossible: If a solution concept satisfies non-emptiness, optimality, and consistency, it must be the equilibrium concept (Norde, Potters, Reijnierse, Vermeulen, 1996). Thus refining the concept of equilibria questions either existence for every (finite) game, optimality or consistency.[5] If one wants to maintain that equlibirium refinements and even selection are reasonable one has to challenge non-emptiness or consistency (since optimality is hardly debatable).

One escape allows players to "leave" the game in a set-valued way: In the reduced game the active players i in M do not necessarily know the strategies s_j of the players $j \notin M$, but only the sets containing them (Dufwenberg, Norde, Reijnierse, Tijs, 2001). In our view, this sacrifices the decentralization property of consistency, namely that the behavior of players $j \notin M$ is already determined in the s, M-reduced game.

Our approach is two-fold: On the one hand we take the compelling refinement of strict equilibria which violates non-emptiness but defines a closed subclass of games for which it is consistent. Solutions of minimal formations coincide with strict equilibria, whenever the latter ones exist, and avoid non-emptiness. It will be argued that this questions the practical relevance of "non-emptiness".

Equilibrium selection poses an even more serious challenge: The strategies of the "gone" players do not provide enough information to reveal their inclinations how to select among solution candidates. Usually one selects one equilibrium rather than another by taking into account the incentives of **all** players. Players, who have already "left", thus must leave some information how "inclined" they are to rely on a certain equilibrium. Our second attempt is to weaken the consistency requirement by suggesting a more demanding (more informative) definition of reduced games.

After introducing some definitions in section 2 the refinement of strict equilibria is discussed and generalized in section 3, 4 and 5. Equilibrium selection is the topic of sections 6 and 7. Section 8 concludes.

[5] A different "attack" against certain equilibrium refinements introduces small noise in sequential games but neglects mixed strategy-equilibria (see Bagwell, 1995, van Damme and Hurkens, 1997, and Güth, Kirchsteiger, Ritzberger, 1998).

2. Definitions

Let $G = (S_1, ..., S_n; u_1(\cdot), ..., u_n(\cdot))$ be a finite n-person **game** in normal form, e.g. the agent normal form in case of an extensive form game.[6] For $i = 1, ..., n$ the strategy set of player i is S_i and $u_i(s)$ for all strategy vectors $s = (s_1, ..., s_n) \in \bigtimes_{j=1}^{n} S_j$ the payoff of player i. Let Γ denote the class of games G with $n \geq 1$. Since we are only interested in games with a non-empty set of equilibria, a game G in Γ is always understood as its mixed extension for which Nash (1951) has proved non-emptiness.

A **solution** $\varphi(\cdot)$ for Γ assigns a set $\varphi(G) \subset S = \bigtimes_{j=1}^{n} S_j$ to all games G in Γ. For $s \in \varphi(G)$ and $M \subset N = \{1, ..., n\}$ with $M \neq \emptyset$ the s, M-**reduced game** of G is the normal form game $G^{s,M} = \left((S_i)_{i \in M}; (\widetilde{u}_i(\cdot))_{i \in M}\right)$ with $\widetilde{u}_i(\widetilde{s}_M) = u_i(\widetilde{s}_M, s_{-M})$ for all $\widetilde{s}_M \in S_M = \bigtimes_{j \in M} S_j$ where $s_{-M} = (s_k)_{k \notin M}$ is the behavior of the non-active players in $G^{s,M}$ according to s. If for any $G \in \Gamma$ and any $s \in \varphi(G)$ and set M with $\emptyset \neq M \subset N$ also $G^{s,M}$ is contained in Γ, the class Γ of games G is φ-**closed**.

The solution $\varphi(\cdot)$ for Γ is **consistent** if Γ is φ-closed and if $s_M \in \varphi\left(G^{s,M}\right)$ for all $s \in \varphi(G)$ and M with $\emptyset \neq M \subset N$.

Converse consistency of the solution $\varphi(\cdot)$ for the φ-closed class Γ of games requires that any s with $s_M \in \varphi\left(G^{s,M}\right)$ for all sets M with $\emptyset \neq M \subsetneqq \{1, ..., n\}$ satisfies $s \in \varphi(G)$.

Thus any candidate, which is immune against all ways of reducing the game by some players becoming non-active, must be an element of the solution set. Since the Norde et al. (1966) impossibility result does not rely on converse consistency, it will be largely neglected in the following.

Optimality in one person-games with $n = 1$ means that the solution $\varphi(G)$ is the set of all $s_1 \in S_1$ for which $u_1(s_1)$ is maximal, i.e. $\varphi(G) = \arg\max \{u_1(s_1) : s_1 \in S_1\}$.

[6] Introducing strategy trembles, which we partly will also use, is a way to guarantee sequential rationality in the underlying extensive form game (see Selten, 1975).

An **equilibrium** $s \in G$ satisfies $u_i(s) \geq u_i\left(\widetilde{s}_i, (s_j)_{j \neq i}\right)$ for all $\widetilde{s}_i \in S_i$ and $i = 1, ..., n$. Let $E(\cdot)$ be the solution correspondence on Γ given by the set $E(G)$ of equilibria for all $G \in \Gamma$. The theorem of Peleg and Tijs (1996) states that $E(\cdot)$ is the only solution function on Γ satisfying the three axioms above, i.e. consistency, converse consistency, and optimality.

Non-emptiness of a solution $\varphi(\cdot)$ on Γ is fulfilled if $\varphi(G) \neq \emptyset$ for all games $G \in \Gamma$. Norde et al. (1996) prove that non-emptiness, optimality[7], and consistency of a solution correspondence $\varphi(\cdot)$ on Γ imply $\varphi(\cdot) = E(\cdot)$. This seems to suggest that every **refinement** $\varphi(\cdot)$ of $E(\cdot)$, i.e. every correspondence φ with $\emptyset \neq \varphi(G) \subset E(G)$ for all $G \in \Gamma$ with $\varphi(G) \neq E(G)$ at least for some games $G \in \Gamma$, must violate consistency.

If one accepts that not all equilibria are reasonable, requiring consistency seems like asking for too much. How important is "non-emptiness"? And is there a weaker consistency axiom which does not preclude a refinement or selection concept $\varphi(\cdot)$ for Γ?

3. Strict equilibria

For a game $G \in \Gamma$ an equilibrium $s \in S$ is **strict** if for all players $i = 1, ..., n$ and all strategies $\widetilde{s}_i \in S_i$ with $\widetilde{s}_i \neq s_i$ one has

$$u_i(s) > u_i(\widetilde{s}_i, s_{-i}).$$

s_2 / s_1	s_2^1	s_2^2
s_1^1	1, -1	-1, 1
s_1^2	-1, 1	1, -1

Figure 1

[7] Actually Norde et al. use a weaker version meaning that a non-empty set of optimal strategies is selected in one person-games.

The matching penny-game in Figure 1 illustrates that a game $G \in \Gamma$ may not have a strict equilibrium. Let us nevertheless study the subclass $\Gamma^* \subset \Gamma$ such that for all games G in Γ^* the set $E^*(G)$ of strict equilibria is non-empty. Since $s \in E^*(G)$ is strict, every player $i \in M$ will lose by unilaterally deviating from s_M in the reduced game $G^{s,M}$ for all non-empty subsets M of N, a property which only holds for strict equilibria. This implies

Proposition 1: For all games $G \in \Gamma^*$, all strict equilibria $s \in E^*(G)$ and any M with $\emptyset \neq M \subset N$:

(i) $s_M = (s_i)_{i \in M} \in E^*\left(G^{s,M}\right)$ and

(ii) $G^{s,M} \in \Gamma^*$.

(iii) For the solution function $\varphi(\cdot) = E^*(\cdot)$ on Γ^* the subclass Γ^* of Γ is φ-closed and thus the solution function $\varphi(\cdot) = E^*(\cdot)$ on Γ^* is both, consistent and converse consistent.[8]

If for any subset M of N with $\emptyset \neq M \neq N$, any $s \in E^*(G)$ for $G \in \Gamma^*$ the cardinality of $E^*\left(G^{s,M}\right)$ is 1, this even implies that one can consistently select a unique equilibrium solution $s^* \in E^*(G)$, namely by simply selecting one strict equilibrium as the solution of G. This property holds for an important albeit rather special class of games.[9]

Unanimity games can be solved by the so-called Nash-bargaining solution (Nash,1953) which is the first sophisticated attempt of equilibrium selection.[10] We will show

[8] The consistency of $E^*(\cdot)$ follows from the one of $E(\cdot)$. Since $s_i \in E^*\left(G^{s,M}\right)$ for all $M = \{i\}$ with $i \in N$ implies $s \in E^*(G)$, the solution function $E^*(\cdot)$ is also converse consistent.

[9] The subclass of unanimity games is important since all systematic attempts of equilibrium selection up to now satisfy some kind of "Nash-property", i.e. of selecting the Nash-bargaining solution in unanimity games. Such concepts are Harsanyi and Selten (1988), Güth and Kalkofen (1989), Güth (1992), and Harsanyi (1995a and 1995b). A still to be generalized novel approach is Selten (1995) who starts by solving only bipolar games. Whereas Harsanyi and Selten (1988) restrict themselves to selecting the (Nash, 1953) bargaining solution for unanimity games with complete information, Güth and Kalkofen (1989) as well as Harsanyi (1995a and 1995b) try to satisfy the generalized Nash-property as suggested by Harsanyi and Selten (1972).

[10] Nash (1953) presents a simple non-cooperative bargaining model and a constructive selection procedure by which he determines one of its equilibria as the solution (see Güth and Kalkofen, 1989, for a more detailed discussion) in addition to his influential axiomatic characterization.

that this solution is consistent (see also Lensberg, 1988). An n-person **unanimity game** G is defined by $\left(K, \left(U^k\right)_{k\in K}\right)$ where

$$K = \{k_1, ..., k_L\} \text{ with } L \geq 2$$

is a finite index set and where for every $k \in K$ the vector

$$U^k = \left(U_1^k, ..., U_n^k\right) \text{ with } U_i^k > 0 \text{ for all } i \text{ and } k \in K$$

is the vector of individual payoffs resulting from the unanimous choice of $k \in K$ whereas all players receive 0-payoffs otherwise. Thus the normal form assumes $S_i = K$ and payoffs $u_i(s) = U_i^k$ for $s = (k, ..., k)$ for $i = 1, ..., n$ and 0-payoffs if at least two players rely on different strategies. Clearly, all unanimous strategy vectors $s = (s_1, ..., s_n)$ with $s_i = k$ for $i = 1, ..., n$ for some $k \in K$ are strict equilibria of the unanimity game G what proves $G \in \Gamma^*$.

The unanimity game G is called "**generic**" if there exists an index $k^* \in K$ such that

$$\prod_{i\in N} U_i^{k^*} > \prod_{i\in N} U_i^k \text{ for all } k \in K \text{ with } k \neq k^*.$$

The Nash-**bargaining solution** (Nash, 1953) of a generic unanimity game $\left(K, \left(U^k\right)_{k\in K}\right)$ is the strategy vector $s^* = (k^*, ..., k^*)$. Let $\Gamma^0 \subset \Gamma^*$ be the class of (generic) unanimity games and define the solution function on Γ^0 by $\varphi(G) = \{s^*\}$ for all games $G \in \Gamma^0$ by the Nash-bargaining solution s^* of game $G \in \Gamma^0$.

Proposition 2:

(i) The subclass $\Gamma^0 \subset \Gamma$ is φ-closed.

(ii) The Nash-bargaining solution $\varphi(\cdot)$ on Γ^0 is consistent, i.e. for every non-empty subset M of N the vector $k_M^* = (s_i^*)_{i\in M}$ with $s_i^* = k^*$ for all $i \in M$ is the Nash-bargaining solution of the s^*, M-reduced game.

Proof: For every $G \in \Gamma^0$ and its Nash-bargaining solution $s^* \in E^*(G)$ the games $G^{s^*,M}$ for $\emptyset \neq M \neq N$ are like trivial generic unanimity games with just one strict equilibrium, namely s^*_M. Thus the class Γ^0 is φ-closed and s^*_M is the Nash-bargaining solution of $G^{s^*,M}$. ■

The consistency of strict equilibria is implied by the consistency of equilibria. The essential message is the $\varphi(\cdot)$-closedness of Γ^0 and the consistency of the Nash-bargaining solution, i.e. of the special selection concept for unanimity games. Thus at least for special subclasses of games consistent equilibrium selection is possible. Since for $M \neq N$ the s^*, M-reduced games of $G \in \Gamma^0$ have just one strict equilibrium, the problem of capturing the inclinations of "gone" players, when solving the s^*, M-reduced games, does not arise.

This illustrates how by restricting the class of games one can avoid the impossibility theorem of Norde et al. (1996). Our first approach is to generalize the concept of strict equilibria such that non-emptiness is guaranteed for Γ instead for Γ^* only.

4. Solutions of minimal formation

What we question here is the practical relevance of the non-emptiness requirement of Norde et al. (1996). More specifically, the refinement concept of strict equilibria, whose consistency has just been discussed, will be generalized to demonstrate the argument. Take $G = (S_1, ..., S_n; u_1(\cdot), ..., u_n(\cdot)) \in \Gamma$ and consider the game $F = (F_1, ..., F_n; u_1(\cdot)|_F, ...u_n(\cdot)|_F) \in \Gamma$ with strategy sets F_i satisfying

$$\emptyset \neq F_i \subset S_i \text{ for } i = 1, ..., n$$

and payoff functions $u_i(\cdot)|_F$ which, for $i = 1, ..., n$, are the restriction of $u_i(\cdot)$ to the strategy combinations in

$$F = \mathop{\times}_{i=1}^{n} F_i.$$

If in the context of the larger game G the smaller game F is closed[11] with respect to the best reply correspondence of G we say that F is a formation (or curb set,

[11] For any mixed strategy vector $q = (q_1, ...q_n)$ in F, i.e. $q_j(s_j) = 0$ for all $s_j \notin F_j$, all best replies q^*_i in G of all players $i = 1, ..., n$ satisfy $q^*_i(s_i) = 0$ for all $s_i \notin F_i$.

see Basu and Weibull, 1991) of G (Harsanyi and Selten, 1988). A formation is called minimal if it contains no proper subformation. If F' and F'' are formations of G, also $F' \cap F''$ is a formation of G if $F' \cap F'' \neq \emptyset$. Thus for any game $G \in \Gamma$ there exists an unambiguous collection

$$\left\{F^1, ...F^L\right\}$$

of minimal formations with $L \geq 1$ (since G is a formation of G there exists at least one minimal formation). If $s \in E^*(G)$, i.e. if s is a strict equilibrium of G, we identify s with the minimal formation

$$F(s) = \left(\{s_1\}, ..., \{s_n\}; u_1(\cdot)\mid_{\{s\}}, ..., u_n(\cdot)\mid_{\{s\}}\right)$$

of G. By the theorem of Nash (1951) every minimal formation F^l with $l = 1, ..., L$ contains at least one equilibrium which, by the definition of formations, is also an equilibrium of G. We will later on (in section 5) discuss and try to justify the

Assumption: For all games $G \in \Gamma$ and each minimal formation F^l of G we can define a unique equilibrium solution s^l of F^l.

Clearly, for all games $G \in \Gamma$ every strict equilibrium s of G is the solution of the minimal formation $\{s\}$. For other minimal formations we so far simply assume that they can be solved uniquely. For $G \in \Gamma$ the game $G^r \in \Gamma$ results if we substitute[12] its minimal formations $F^1, ...F^L$ by their solutions $s^1, ..., s^L$, i.e.

$$G^r = (S_1^r, ..., S_n^r; u_1(\cdot)\mid_{S^r}, ..., u_n(\cdot)\mid_{S^r})$$

where for all players $i = 1, ..., n$

(i) $\left\{\begin{array}{l}\text{if } s_i \in S_i \text{ and } s_i \notin F^l \text{ for all } l = 1, ..., L, \text{ then } s_i \in S_i^r \\ \text{if } s_i \in S_i \text{ and } s_i \in F^l \text{ for some } l = 1, ..., L, \text{ then } s_i \text{ is substituted by } s_i^l \in S_i^r\end{array}\right.$

as well as

(ii) $S^r = \underset{i=1}{\overset{n}{\times}} S_i^r$ and $u_i(\cdot)\mid_{S^r}$ being the restriction of $u_i(\cdot)$ to S^r.

[12] For the game of matching pennies (Figure 1) with just one minimal formation, namely the game itself, and just one equilibrium this results in the trivial game where both players must use their equilibrium strategy since the game has no proper subformation.

Proposition 3: For all games $G \in \Gamma$ one has $G^r \in \Gamma^*$, i.e. substituting G by G^r avoids non-emptiness for $\varphi(\cdot) = E^*(\cdot)$, if the Assumption holds.

Proof: For every $G \in \Gamma$ there exists at least one minimal formation F^1 of G. Thus by the Assumption G^r of G contains a strategy vector s^1 which is the solution of F^1. If a player $i = 1, ..., n$ unilaterally deviates from s^1 in G^r, this implies a deviation to some strategy $s_i \notin F^1$ according to the definition of G^r. Since F^1 is closed with respect to best replies, player i must lose, i.e. s^1 is a strict equilibrium in G^r and $G^r \in \Gamma^*$.$\blacksquare$

5. Solving minimal formations

If the assumption in section 4 holds, the non-emptiness of the compelling refinement of strict equilibria seems to be rather minor since we can substitute minimal formations by their solutions and thereby guarantee existence. Furthermore, the assumption would be trivially granted if for all games $G \in \Gamma$ all minimal formations F^l have just one equilibrium s^l. Although up to now the problem never occurred in applications, there can be games $G \in \Gamma$ whose minimal formations contain multiple equilibria. The game $G \in \Gamma$ of Figure 2 has no proper subformation since all four pure strategies are best replies when the other player uses all four strategies with equal probability. Thus there exists only one minimal formation $F^1 = G$. Furthermore, any strategy vector such that both players use their 1st and 2nd as well as their 3rd and 4th strategy with equal probability is an equilibrium. This illustrates

Remark 4: Solving minimal formations can require equilibrium selection.

Selecting an equilibrium solution for a minimal formation with multiple equilibria poses quite a challenge since, by the definition of a minimal formation, these equilibria have to be non-strict. It is in these (rare) cases where we must pay a price to overcome the problem. What we suggest is to use a familiar idea when trying to select among non-strict equilibria, namely to induce strictness by arbitrarily small payoff perturbations (only when solving minimal formations with multiple equilibria). The non-existence of strict equilibria in minimal formations F^l of a game $G \in \Gamma$ is avoided if one does not solve the game F^l directly but as a limit of **payoff perturbed games** (see Harsanyi, 1973 and 1975, for an earlier use of payoff perturbances).

s_1 \ s_2	s_2^1	s_2^2	s_2^3	s_2^4
s_1^1	1, -1	-1, 1	0, 0	0, 0
s_1^2	-1, 1	1, -1	0, 0	0, 0
s_1^3	0, 0	0, 0	1, -1	-1, 1
s_1^4	0, 0	0, 0	-1, 1	1, -1

Figure 2

Let for small but positive ε game $G(\varepsilon) = (S_1, ..., S_n; u_1^\varepsilon(\cdot), ..., u_n^\varepsilon(\cdot))$ be the ε-payoff perturbed game of an arbitrary game $G = (S_1, ..., S_n;\ \ u_1(\cdot), ..., u_n(\cdot)) \in \Gamma$. Here for $i = 1, ..., n$ and for any mixed strategy vector $q = (q_1, ..., q_n)$ in $G(\varepsilon)$ the payoff $u_i^\varepsilon(q)$ is defined by

$$u_i^\varepsilon(q) = u_i(q) + \varepsilon \sum_{s_i \in S_i} \ln q_i(s_i).$$

Clearly, as long as ε is positive, any equilibrium $q^\varepsilon = (q_1^\varepsilon, ..., q_n^\varepsilon)$ of $G(\varepsilon)$ will be strict, i.e. ε-payoff perturbed games $G(\varepsilon)$ of G avoid the non-existence of strict equilibria. Rather than solving G directly one could determine the limit solution $q^* = \lim_{\varepsilon \to 0} q^\varepsilon$ by deriving the solutions q^ε of its ε-payoff perturbed games $G(\varepsilon)$ for which strict equilibria exist of which one can be selected as the solution of $G(\varepsilon)$. The limit solution q^* of G is always an equilibrium of game G but may not be strict (see section 7). We thus avoid non-emptiness by allowing for non-strict equilibria as solution candidates which can be justified as limits of strict equilibria. Possible emptiness in the sense of Norde et al. (1996) applies only for the limit game itself which is just seen as an extreme idealization and therefore of no practical relevance.

One may object that the ε-perturbed games $G(\varepsilon)$ of G are structurally very different from finite strategic games[13]. Payoff perturbances are rather ad hoc and mainly justified by their practical consequence of guaranteeing strictness. We therefore recommend to limit their application to rare problems where they or similar ideas seem unavoidable. Practically they matter very little. Altogether the non-emptiness of $\varphi(\cdot) = E^*(\cdot)$ seems to be a minor problem if one accepts the more general idea of solutions of minimal formations whose existence can be guaranteed.

6. Equilibrium selection

So far we have shown that the non-emptiness of strict equilibria is of minor importance. For games, where the problem is the multiplicity and not the existence of strict equilibria, a solution function $\varphi(\cdot)$ on Γ with $|\varphi(G)| = 1$ poses a more serious problem: To select uniquely and reasonably a solution also in the reduced games $G^{s,M}$ with $s \in \varphi(G)$ it is often not enough to know the strategies s_j of the "gone players" $j \notin M$. Any convincing theory of equilibrium selection will try to account for the inclinations of all players when selecting one rather than another solution candidate. The reduced games should therefore not only capture the strategies of the "gone players" but also their inclinations. We will illustrate this for the class of binary games with two strict equilibria.

Let $G = \Gamma^*$ be an n-person game with

$$S_i = \{X_i, Y_i\} \text{ for } i = 1, ..., n$$

[13]Klaus Ritzberger (personal communication) misses the multi-linearity of the payoffs in the mixed strategies. Nevertheless most theories of equilibrium selection rely on payoff perturbances but not necessarily of the same kind (see Nash, 1953, Harsanyi and Selten, 1988, Güth and Kalkofen, 1989, and Güth, Ritzberger and van Damme, forthcoming).

and two strict equilibria $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ relying on different strategies for all n players. To illustrate the problems when trying to select consistently a unique equilibrium solution we rely on the simple idea that the solution should be that strict equilibrium for which the product of unilateral deviation dividends is maximal (see Güth, 1992).[14]

Denote for $i = 1, ..., n$ these dividends by

$$x_i = u_i(X) - u_i(Y_i, X_{-i})$$

and

$$y_i = u_i(Y) - u_i(X_i, Y_{-i})$$

where $X_{-i} = (X_j)_{j \neq i}$ and $Y_{-i} = (Y_j)_{j \neq i}$. We say that X is the **unilaterally stable solution**[15] of G if

$$\prod_{i \in N} x_i > \prod_{i \in N} y_i$$

whereas this solution is Y if the inequality is reversed. A game G for which the (reversed) inequality holds is called a **"generic binary game"** with the two strict equilibria X and Y. Let $\Gamma^2 \subset \Gamma^*$ denote the subclass of all generic binary games G with two strict equilibria X and Y.

The usual definition of an s, M-reduced game $G^{s,M}$ of G implies

$$G^{s,M} = \left((S_i)_{i \in M}, \left(\widetilde{u}_i^{s,M}(\cdot) \right)_{i \in M} \right)$$

with $S_i = \{X_i, Y_i\}$ and

$$\widetilde{u}_i^{s,M}(\widetilde{s}_M) = u_i(\widetilde{s}_M, s_{-M})$$

for $i \in M$ and all strategy constellations $\widetilde{s}_M \in \underset{k \in M}{\times} S_k$. If $\varphi(\cdot)$ is specified as the set $\varphi(G) = \{X\}$ when X is unilaterally stable, and $\varphi(G) = \{Y\}$ when Y is unilaterally stable, the subclass Γ^2 of Γ^* is not φ-closed since the properly reduced

[14] For generic unanimity games this general selection idea selects the Nash-bargaining solution since for $s = (k, ..., k)$ the unilateral deviation dividend of player i is U_i^k for all $i = 1, ..., n$ and $k \in K$.

[15] Unilateral deviation stability is solely focusing on deviation losses and may fail, like risk dominance (Harsanyi and Selten, 1988), to select a unique payoff dominant strict equilibrium. This, however, is not a (major) weakness. Its drawback, in our view, is that it neglects the risks of deviations by more than one player. Its major advantage is its simplicity which is why it is applied here.

games $G^{s,M}$ for $s \in \varphi(G)$ do not have to contain two strict equilibria (as in case of unanimity games $G \in \Gamma^2$). Define therefore the subclass $\Gamma^{1/2} \subset \Gamma^*$ as the class of binary games with one or two strict equilibria. By assuming that $\varphi(\cdot)$ selects the unique strict equilibrium in binary games $G \in \Gamma^{1/2} \diagdown \Gamma^2$ we extend $\varphi(\cdot)$ to the subclass $\Gamma^{1/2} \subset \Gamma^*$ which is φ-closed.

The following example demonstrates that the usual definition of s, M-reduced games renders unilateral stability as inconsistent. Game G with $n = 3$ and payoffs

X_3	X_2	Y_2
X_1	4, 4, z	2, 2, 0
Y_1	2, 2, 0	6, 6, 0

Y_3	X_2	Y_2
X_1	0, 0, 0	4, 0, 0
Y_1	0, 4, 1	5, 5, 5

satisfying $z > 0$ has the two strict equilibria $X = (X_1, X_2, X_3)$ and $Y = (Y_1, Y_2, Y_3)$. The strict equilibrium X is the unilaterally stable solution of G if $z > 5/4$. For the $X, \{1, 2\}$-reduced game $G^{X,\{1,2\}}$ of G the unilaterally stable solution is, however, $s_M = (Y_1, Y_2)$ regardless of the value z. The inclination of player 3 for selecting X is completely neglected in the reduced game $G^{X,\{1,2\}}$. This illustrates that unilaterally stable strict equilibria are inconsistent.

Thus consistency of unilateral stability $\varphi(\cdot)$ for $\Gamma^{1/2}$ requires another definition of reduced games. For any $G \in \Gamma^{1/2}$, any strategy vector $s \in \varphi(G)$ and any non-empty, proper subset M of the player set $N = \{1, ..., n\}$ define the impact of unilateral deviation losses by "gone players" as

$$f(s, M) = \prod_{j \notin M} \max\{u_j(s) - u_j(\widetilde{s}_j, s_{-j}), 0\}$$

where $\widetilde{s}_j \neq s_j$ for all $j \notin M$. Extending $f(\cdot,\cdot)$ to $M = N$ by $f(\cdot, N) \equiv 1$ one always has $f(\cdot,\cdot) > 0$ due to $s \in \varphi(G)$ and $G \in \Gamma^{1/2}$, $f(X, M) = \prod_{j \notin M} x_j$ and $f(Y, M) = \prod_{j \notin M} y_j$. The $f(s, M)$-reduced game $G^{f(s,M)}$ of G is the binary $| M |$-person game

$$G^{f(s,M)} = \left((S_i)_{i \in M}, (\widetilde{u}_i(\cdot))_{i \in M}\right)$$

with $S_i = \{X_i, Y_i\}$ and

$$\widetilde{u}_i(\widetilde{s}_M) = f(s, M)^{\frac{1}{|M|}} \cdot u_i(\widetilde{s}_M, s_{-M}) \qquad (\star)$$

for all $i \in M$ and all strategy combinations $\widetilde{s}_M$ of active players in $G^{f(s,M)}$. The definition of $f(s, M)$ follows from the requirement that players $i \in M$ perceive the intentions of players $j \notin M$ as captured by the solution function $\varphi(\cdot)$ of unilateral stability. It illustrates how the solution function $\varphi(\cdot)$ suggests a definition of $f(s, M)$-reduced games $G^{f(s,M)}$. Let us call a solution f-**consistent** if it is consistent with respect to its $f(s, M)$-reduced games.[16]

Proposition 5: For all generic binary n-person games $G \in \Gamma^{1/2}$ the unilaterally stable strict equilibrium solution is f-consistent.

Proof: If the $f(s, M)$-restricted game $G^{f(s,M)}$ of $G \in \Gamma^{1/2}$ for some proper subset M of N has no other strict equilibrium than the M-restriction s_M of the unilaterally stable solution s of G, then s_M is also the unilaterally stable strict equilibrium solution of $G^{f(s,M)}$ since $f(s, M)$ is positive and since thus s_M is the only strict equilibrium of $G^{f(s,M)}$.

If $G^{f(s,M)}$ for $M \neq N$ would have another strict equilibrium, say $\widetilde{s}_M$, in addition to s_M, then due to $f(s, M) > 0$ the strategy constellation $(\widetilde{s}_M, s_{-M})$ must be a strict equilibrium of G. Since $(\widetilde{s}_M, s_{-M})$ differs from both, X and Y, this contradicts our definition of binary games G with two strict equilibria X and Y relying on different strategies for all players. ■

[16] An alternative definition, substituting ($\star$), completely overturns the usual definition of reduced games in the sense that "gone" players $j \notin M$ do not leave their strategies but only their inclinations. Assume

$$\widetilde{u}_i(\widetilde{s}_M) = f(\widetilde{s}, M)^{\frac{1}{|M|}} u_i(\widetilde{s}_M, \widetilde{s}_{-M}) \qquad (+)$$

for all strict equilibria $(\widetilde{s}_M, \widetilde{s}_{-M})$ of $G \in \Gamma^{1/2}$ and $\widetilde{u}_i(\widetilde{s}_M) = 0$ otherwise, i.e. if $\widetilde{s}_M$ cannot be completed to a strict equilibrium of G. Thus the "gone players" $j \notin M$ let the players i in M decide for them knowing that due to $f(\cdot,\cdot)$ these will take into account their incentives. Also for (+) Proposition 5 is true: For $M \neq N$ and $\widetilde{s}_M = X_M$, respectively Y_M one obtains

$$\prod_{i \in M}\left[\widetilde{u}_i(X_M) - \widetilde{u}_i\left(Y_i, (X_l)_{\substack{l \in M \\ l \neq i}}\right)\right] = \prod_{j \notin M} x_j \prod_{i \in M} x_i = \prod_{l \in N} x_l$$

and

$$\prod_{i \in M}\left[\widetilde{u}_i(Y_M) - \widetilde{u}_i\left(X_i, (Y_l)_{\substack{l \in M \\ l \neq i}}\right)\right] = \prod_{j \notin M} y_j \prod_{i \in M} y_i = \prod_{l \in N} y_l.$$

If $s = X$ is unilaterally stable in G, its M-restriction X_M is therefore unilaterally stable in $G^{f(X,M)}$.

Different solution functions $\varphi(\cdot)$ will require different functions $f(\cdot,\cdot)$ to render them as f-consistent. Since different solutions weigh strategic aspects differently, their functions $f(s,M)$ will usually react differently to strategic aspects. A more natural definition of $f(\cdot,\cdot)$ for a solution function $\varphi(\cdot)$ makes its f-consistency more desirable.[17] Here we have only illustrated how the consistency requirement can be weakened by introducing $f(s,M)$-reduced games and how this modifies the decentralization property of consistency.

7. General games

As in section 5 where we already relied on familiar ideas of equilibrium selection we will just mention the more or less standard techniques (mostly introduced by Harsanyi and Selten, 1988) when (due to the specific steps (v), (vi) and (vii) below) generalizing unilateral deviation stability. The procedure is as follows:

(0) Take an arbitrary game $G \in \Gamma$!

(i) Perturb G ε_k-uniformly[18], i.e. all minimal choice probabilities (Selten, 1975) are ε_k where $\varepsilon_k > 0$ is sufficiently small; for $i = 1, ..., n$ and any $s_i \in S_i$ represent the choice of s_i with maximal probability as a pure strategy (see Harsanyi and Selten, 1988, for details); denote by G_k the resulting (ε_k-uniformly perturbed) game!

(ii) Repeatedly and for all players simultaneously for G_k eliminate (strictly) inferior strategies (see the concept of rationalizable strategies as discussed by Pearce, 1984) and substitute duploid strategies (which yield the same payoff for all possible strategy combinations of the other players) by their centroid[19] (which assigns the same positive probability to all strategies in the same duploid class) till no further reduction of game G_k is possible!

[17] Like in cooperative game theory where usually different concepts rely on different definitions of reduced games different theories of equilibrium selection will focus on different strategic aspects and will therefore rely on divergent ways of capturing "inclinations of gone players".

[18] Whereas positive minimum choice probabilities were introduced by Selten (1975) to define a necessary condition for sequential rationality, namely perfectness, uniform trembles are already aiming at an unbiased selection among perfect equilibria.

[19] The idea here is that different duploid strategies differ only in strategically irrelevant aspects with no payoff implications whatsoever.

(iii) If all minimal formations $F_k^1, ..., F_k^L$ of G_k have just one equilibrium, namely $s_k^1, ..., s_k^L$, go to (iv); otherwise subject any F_k^l with multiple equilibria to a sequence $\varepsilon \searrow 0$ of small payoff perturbations[20] (see section 5 above) and apply this procedure (starting with (i)) to the games $F_k^l(\varepsilon)$; denote by $s_k^l = \lim_{\varepsilon \searrow 0} s_k^l(\varepsilon)$ the limit of the solutions $s_k^l(\varepsilon)$ which this procedure yields for the games $F_k^l(\varepsilon)$!

(iv) Solve G_k by solving its transformed game G_k^r (see section 4) resulting from substituting minimal formations by their solutions!

(v) For any two strict equilibria s_k^l and $s_k^{l'}$ of G_k^r with $l, l' = 1, ..., L$ and $l \neq l'$ define the comparison game $G_k^r(l, l') \in \Gamma^{1/2}$ with, for all active players i, strategy set

$$S_i^k(l, l') = \left\{ \left(s_k^l\right)_i, \left(s_k^{l'}\right)_i \right\},$$

i.e. player i can only use the strategies suggested by the two solution candidates, and payoff function

$$u_i(\cdot) = u_i(\cdot) \Big|_{\times_{j=1}^{n} S_j^k(l,l')} !$$

(vi) Compute the product of unilateral deviation dividends (see section 6) for the two strict equilibria s_k^l and $s_k^{l'}$ of $G_k^r(l, l')$ and denote by $R_k(l, l')$ their ratio!

(vii) For any solution candidate $s_k^{l'}$ determine the vector

$$R_k(l') = (R_k(l_1, l'), ..., R_k(l_{L-1}, l'))$$

with non-increasing components $R_k(l, l')$ with $l \neq l'$, choose the strict equilibrium $s_k^* = s_k^{l^*}$ of G_k^r as the solution whose vector $R_k(l^*)$ is lexicographically minimal (see Güth, 1992)!

(viii) Choose ε_{k+1} such that $\varepsilon_k > \varepsilon_{k+1} > 0$ and go to (i)!

(ix) Determine the solution s^* of G as the limit solution[21]

$$s^* = \lim_{k \nearrow \infty} s_k^* !$$

[20] Whereas strategy trembles are used to guarantee perfectness of equilibria, payoff trembles are supposed to induce their strictness (see Harsanyi, 1973 and 1975 as well as Harsanyi and Selten, 1988).

[21] The mathematical problem of existence and uniqueness of such a limit is discussed by Harsanyi and Selten (1988).

A rigorous proof that this procedure determines for all games $G \in \Gamma$ a unique solution $s^*(G)$ which by its derivation is a uniformly perfect[22] equilibrium is very demanding, especially proving that the limits, to be determined in steps (iii) and (ix), always exist. These difficulties or similar ones arise, however, whenever trying to develop a general theory of equilibrium selection. We can therefore point to more elaborate treaties of equilibrium selection which discuss the mathematical difficulties in more detail (e.g. Harsanyi and Selten, 1988).

Notice that the comparison games, defined on stage (v), are all contained in the $\varphi(\cdot)$-closed class $\Gamma^{1/2}$ for which f-consistency has been discussed in section 6. Thus the algorithm demonstrates how problems of equilibrium selection in general games (in Γ) can be boiled down to simpler ones, e.g. in special classes of games (like $\Gamma^{1/2}$), for which f-consistent equilibrium selection is possible.

8. Final remarks

Consistency of equilibria and equilibrium selection could become relevant for studies of endogenous timing which allow players to act early or later where this is partly assumed to be publicly observable and partly not (e.g. van Damme and Hurkens, 1999, Spencer and Brander, 1992, Sadanand and Sadanand, 1996, Güth, Ritzberger and van Damme, forthcoming, Hamilton and Slutsky, 1990). Consistency would imply robust behavioral plans in the sense that expecting others to decide early will not question the plans of players who wait. The outcome would not depend on timing dispositions which are either exogenously given (e.g. by the extensive or stage form) or endogenously determined.

In non-cooperative game theory it is quite commonly accepted that the solution should be an equilibrium of the game, but that not all equilibria are reasonable solution candidates (see, however, Aumann, 1996). To refine the equilibrium notion or even to select uniquely an equilibrium solution, however, contradicts

[22] The refinement concept of uniformly perfect equilibria is implicitly required by Harsanyi and Selten (1988). Its existence can be proved by employing the same technique as in Selten (1975).

consistency if the other axioms of the Norde et al. (1996)-impossibility result are considered as undebatable. Impossibility results tell us that our aspirations are too naive. Especially, a consistent way of equilibrium selection where the "gone" players $j \notin M$ just leave their strategy is not only impossible but also undesirable since such players may have good reasons to strive for one equilibrium rather than the other(s).

We first have questioned the practical relevance of the non-emptiness requirement by exploring the generalization of strict equilibria via solutions of minimal formations. When discussing equilibrium selection our weaker consistency requirement allows "gone" players to "leave" their incentives for selecting one equilibrium versus the other(s). Unilateral stability is, of course, only an example how to select among equilibia. For other concepts a similar justification seems possible by appropriately modifying the definition of reduced games. Thus there exists an alternative to the (set valued-) weakening of consistency by Dufwenberg et al. (2001) who also try to escape the impossibility result by Norde et al. (1996).

One may object against the consistency axiom that it pays attention to all non-empty subsets M of player set N in the same way. Like the active set of players in a proper subgame of an extensive form game also certain subsets M of N can have the property that the optimal behavior of every member i of M depends only the choices of the players in M (where according to the procedure in section 7 this property has to hold only in ε_k-uniformly perturbed games with $\varepsilon_k > 0$). Harsanyi and Selten (1988) refer to such sets M as cells.

Clearly when solving a reduced game for a player set M, which is a proper subcell of N, the inclinations of the "gone players" $j \notin M$ will not matter much since they do not influence the best reply structure of the reduced game. Restricting consistency by applying it to reduced games for cells M of N only would, however, mean to essentially give up consistency at all. The decentralization property when solving reduced games for cells M follows already from the cell property. In our view, one therefore should generally try to capture the inclinations of "gone players" $j \notin M$ and let them become less important by the cell property when the set M satisfies it.

References

[1] Aumann, R. J. (1987): Game theory, in: *The New Palgrave, A Dictionary of Economics*, Vol. 2, J. Eatwell, M. Milgate, and P. Newman (eds.), London: Macmillan, 460 - 482.

[2] Aumann, R. J. (1996): On the state of the art in game theory: An interview with Robert Aumann, in: *Understanding Strategic Interaction - Essays in Honor of Reinhard Selten*, W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, E. van Damme (eds.), Heidelberg/Berlin: Springer-Verlag, 8 - 34.

[3] Aumann, R. J. and A. Brandenburger (1995): Epistemic conditions for Nash equilibrium, *Econometrica*, 63 (5), 1161 - 1180.

[4] Bagwell, K. (1995): Commitment and observability in games, *Games and Economic Behavior*, 8, 271 - 280.

[5] Basu, K., J. W. Weibull (1991): Strategy subsets closed under rational behavior, *Economics Letters 36,* 141 - 146.

[6] Dufwenberg, M., H. Norde, H. Reijnierse, and S. Tijs (2001): The consistency principle for set-valued solutions and a new approach to normative game theory, *Mathematical Methods of Operations Research,* 54, 119 - 131.

[7] Güth, W. (1992): Equilibrium selection by unilateral deviation stability, *Rational Interaction, Essays in Honor of John C. Harsanyi*, R. Selten (ed.), Heidelberg: Springer-Verlag, 161 - 189.

[8] Güth, W. and B. Kalkofen (1989): Unique solutions for strategic games - Equilibrium selection based on resistance avoidance, *Lecture Notes in Economics and Mathematical Systems*, 328.

[9] Güth, W., G. Kirchsteiger, and K. Ritzberger (1998): Imperfectly observable commitments in n-player games, *Games and Economic Behavior*, 23, 54 - 74.

[10] Güth, W., K. Ritzberger and E. van Damme: On the Nash Bargaining Solution with Noise, *forthcoming European Economic Review.*

[11] Hamilton, J. H. and S. M. Slutsky (1990): Endogenous timing in duopoly games: Stackelberg or Cournot equilibria, *Games and Economic Behavior*, 2, 29 - 46.

[12] Harsanyi, J. C. (1959): A bargaining model for cooperative n-person games, *Contributions to the Theory of Games IV*, A. W. Tucker and R. D. Luce (eds.), Princeton: Princeton University Press, 324 - 358.

[13] Harsanyi, J. C. (1973): Games with randomly disturbed payoffs: A new rationale for mixed strategy equilibrium points, *International Journal of Game Theory*, 2, 1 - 23.

[14] Harsanyi, J. C. (1975): The tracing procedure, *International Journal of Game Theory*, 4, 61 - 94.

[15] Harsanyi, J. C. (1995a): A new theory of equilibrium selection for games with complete information, *Games and Economic Behavior*, 8 (1), 91 - 122.

[16] Harsanyi, J. C. (1995b): A new theory of equilibrium selection for games with incomplete information, *Games and Economic Behavior*, 10 (2), 318 - 332.

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

[18] Harsanyi, J. C. and R. Selten (1988): *A general theory of equilibrium selection in games*, Cambridge Mass: M.I.T. Press.

[19] Lensberg, T. (1988): Stability and the Nash-solution, *Journal of Economic Theory*, 45, 330 - 341.

[20] Maschler, M. (1990): Consistency, in: *Game Theory and Applications*, T. Ichiishi, A. Neyman, and Y. Tauman (eds.), San Diego, California: Academic Press, 183 - 186.

[21] Nash, J. F. (1951): Non-cooperative games, *Annals of Mathematics*, 54, 286 - 295.

[22] Nash, J. F. (1953): Two-person cooperative games, *Econometrica*, 21, 128 - 140.

[23] Norde, H., J. Potters, H. Reijnierse, and D. Vermeulen (1996): Equilibrium selection and consistency, *Games and Economic Behavior*, 12, 219 - 225.

[24] Pearce, D. G. (1984): Rationalizable strategic behavior and the problem of perfection, *Econometrica* 52, 1029 -1050.

[25] Peleg, B. and S. Tijs (1996): The consistency principal for games in strategic form, *International Journal of Game Theory*, 25, 13 - 34.

[26] Sadanand, A. and V. Sadanand (1996): Firm scale and the endogenous timing of entry: A choice between commitment and flexibility, *Journal of Economic Theory*, 42, 516 - 530.

[27] Selten, R. (1975): Reexamination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory* 4, 25 - 55.

[28] Selten, R. (1995): An axiomatic theory of a risk dominance measure for bipolar games with linear incentives, *Games and Economic Behavior*, 8 (1), 213 - 263.

[29] Spencer, B. J. and J. A. Brander (1992): Pre-commitment and flexibility - Applications to oligopoly theory, *European Economic Review*, Vol. 36, 1601 - 1626.

[30] Thomson, W. (1990): The consistency principle, in: *Game Theory and Applications*, T. Ichiishi, A. Neyman, and Y. Tauman (eds.), San Diego, California: Academic Press, 183 - 186.

[31] van Damme, E. (1983): *Refinements of the Nash equilibrium concept*, Berlin/Heidelberg/New York: Springer Verlag.

[32] van Damme, E. and S. Hurkens (1999): Endogenous Stackelberg Leadership, *Games and Economic Behavior*, 28, 105 - 129.

[33] van Damme, E. and S. Hurkens (1997): Games with imperfectly observable commitment, *Games and Economic Behavior*, 21 (1/2), 282 - 308.