STABLE SETS
IN NON-COOPERATIVE GAME THEORY

Dries Vermeulen
Department of Economics
Maastricht University
E-mail: d.vermeulen@eca.unimaas.nl

1. Introduction

Generally speaking, game theory deals with the analysis of situations in which a number of people, each of them having his own interest in the outcome of the situation, must decide what to do. In the area of non-cooperative game theory these people (the players) are supposed to be unable or unwilling to commit themselves to taking specific actions in the given situation (the game). This non-cooperative character of the players forces us to search for solutions of the game that are self-enforcing.

In their paper of 1986 Kohlberg and Mertens argued that a convincing self-enforcing solution must necessarily satisfy a number of conditions. A solution that is designed to meet all conditions is called a stability concept. They also showed that such a concept is inevitably set-valued. Since then a number of stability concepts have been proposed. Although there is nowadays a wide variety of stability concepts, the two that are most commonly known are the ones defined by Mertens (1980) and Hillas (1990).

In this paper the general framework is introduced first. Then some of the conditions proposed by Kohlberg and Mertens are reviewed. Finally some of the results concerning the stability concepts of Mertens and (especially) Hillas are given.

2. The name of the game

Each of the players has a finite number of choices in the particular type of non-cooperative game considered in this paper. Further, the players are supposed to make their choices simultaneously. (To put it a little differently, no player has any information concerning the choices of the other players when he has to make his own choice.) After these choices are made, each player receives a payoff. The obvious goal of any player of the game is to obtain a payoff that is as high as possible. Such a game is usually called a one-shot game, or normal form game.

Example 1. If there are two players involved in the game, we can represent a normal form game by what is called a bimatrix. For example,

$$
\begin{bmatrix}
1,2 & (0, 0) \\
(0, 0) & (0, 0)
\end{bmatrix}
$$

is a 2 x 2-bimatrix game. Player I (the row player) has to choose between the first and the second row, player II (the column player) between the first and the second column. The
choices of the players determine an entry of the bimatrix. If for instance player 1 chooses his first row and player II chooses the first column, we end up in the entry (1, 2). The first number is the payoff to player I, so he gets one unit. Similarly, player II gets two units.

In order to give a general definition, let \( n \) be a positive natural number. The player set \( (1, \ldots, n) \) of an \( n \)-person normal form game \( \Gamma \) (from now on simply called a game) is denoted by \( N \). Each player \( j \in N \) is assumed to have a finite set \( A_j \) of (pure) strategies and a payoff function \( u_j : A \rightarrow \mathbb{R} \). Thus, using the shortened notation \( u = (u_i)_{i \in N} \), the game can be written as \( \Gamma = (A, u) \).

Given a game \( \Gamma = (A, u) \), a player, say \( j \in N \), can decide to randomize between the choices in \( A_j \) that are available to him. The game that results when randomization is explicitly allowed is called the mixed extension of \( \Gamma \). In this paper we will simply identify the game \( \Gamma \) with its mixed extension.

In the mixed extension each player \( i \in N \) is assumed to choose a mixed strategy \( x_i = (x_{i1}, \ldots, x_{in}) \), from the set \( \Delta(A_i) \) of probability distributions on \( A_i \). So, the coordinates of \( x_i \) are not negative and their sum equals one. Given the strategy profile \( x = (x_i)_{i \in N} \) of choices of the players, player \( j \) calculates his expected payoff as follows. A pure strategy profile \( x = (x_i)_{i \in N} \) is played with probability \( \prod_i u_{ij} \). So, player \( j \) expects to get a payoff equal to \( u_j(x) = \sum_{i \in N} \prod_i u_{ij} x_{ij} \). When \( x \) is played. Clearly, the payoff function of this mixed extension is a function from the set \( \Delta_N = \prod_i \Delta(A_i) \) of \( (\Delta) \)-strategy profiles to \( \mathbb{R} \).

**Example 2.** A mixed strategy of player I in the bimatrix game of example 1 can be written as \( (p, 1 - p) \) for \( 0 \leq p \leq 1 \). Playing such a strategy \( (p, 1 - p) \) means that the first row is chosen with probability \( p \) and the second one with probability \( 1 - p \). Playing \( (1, 0) \) obviously means that the first row is chosen with certainty.

The central issue in this paper is the search for a 'reasonable' solution for each game. This means that we need to answer two questions. First of all, what do we mean by a solution? Secondly, what do we consider to be a 'reasonable' solution? The first question answered in the following definition. We specifically allow sets of strategy profiles in the solution, because the conditions for 'reasonable' solutions in the next section necessarily yield set-valued solutions as we will see in example 6.

**Definition 1.** A solution \( \sigma \) is a rule that assigns to each game \( \Gamma \) a collection \( \sigma(\Gamma) \) of (usually closed and non-empty) subsets of the space of strategy profiles of the game \( \Gamma \). The elements of \( \sigma(\Gamma) \) are called the solution sets of the game \( \Gamma \).

**Example 3.** The map \( \tau \) that assigns the collection \( \{x \mid x \text{ is a strategy profile of } \Gamma \} \) to a game \( \Gamma \) is a solution by definition. However, this is not very 'reasonable' one. If we look for instance to the game of example 1, it is clear that the strategy pair \( (0, 1), (1, 0) \) is not a good deal for player I. If he agrees to play this profile his payoff will be zero, while he will get one if he decides to play \( (1, 0) \) instead (provided that player II will stick to the agreement).

Thus, we need to specify what 'reasonable' means for a solution. And, although this question does not seem to pose too much problems, it turns out to be a very difficult one indeed. In the next section we will give some examples of possible answers to this question.

3. Requirements for solutions

Usually in non-cooperative game theory 'reasonable' is interpreted as what is called 'self-enforcing'. A set of strategy profiles of a given game is called self-enforcing if no player of the game wishes to deviate to a strategy profile outside the set once it has been agreed not to play any of the profiles outside this set. Such a condition is obviously necessary for a solution set of a non-cooperative game, since the players are not obliged to stick to any agreement as we supposed in the model. So, the motivation to do so anyway must lie within the agreement itself. The first natural formalization of self-enforcingness is the equilibrium concept introduced by Nash in 1952.

**Definition 2.** For a game \( \Gamma = (A, u) \), let \( (x, y_i) \) denote the strategy profile where player \( i \) uses the strategy \( y_i \in A_i \) and his opponents use the strategies in \( x \in \Delta_A \). For player \( i \) and a strategy profile \( x \)

\[
\beta(x) := \{ y_i \in \Delta(A_i) \mid u_i(y_i, x) \geq u_i(x_i, x) \ \text{for all} \ x_i \in \Delta_i \}
\]

is the set of best replies of player \( i \) against \( x \). A strategy profile in the set

\[
\beta(x) := \prod_i \beta_i(x)
\]

is also called a best reply against \( x \). A (Nash) equilibrium is a strategy profile \( x \) that is a best reply against itself. In other words, \( x \) must be an element of \( \beta(x) \). The associated solution assigns to each game \( \Gamma \) the collection \( \{ x \mid x \text{ is an equilibrium of } \Gamma \} \).

At first sight, equilibria seem to catch the spirit of self-enforcingness quite nicely. After all, each player plays 'as good as he can' (i.e., a best reply) given the strategy choices of his opponents in a Nash equilibrium. However, some Nash equilibria are not as convincing as we would expect a 'reasonable' solution to be. This will be explained in the next.

**Example 4.** First note that the strategy pair \( (0, 1), (0, 1) \) is a Nash equilibrium of the bimatrix game in example 1. This is evident, once we realize that any strategy of player I (player II) is a best reply against the strategy \( (0, 1) \) of player II (player I).

However, suppose that player II plays a given strategy \( (q, 1 - q) \). Then the expected payoff of player I is \( (1 - q) \) when he plays his first row, while he will get zero when he plays his second row. So, playing \( (1, 0) \) will always give him at least the payoff he gets when he plays \( (0, 1) \), the above-mentioned equilibrium strategy. For this reason, \( (0, 1) \) is called a weakly dominated strategy.

For this reason attempts have been made in the past to find methods to eliminate 'bad' equilibria like the one in the above example. The idea was to refine the collection of equilibria of a game to a smaller set of acceptable equilibria by imposing extra conditions. Probably the two most notorious examples of such refinements of the Nash equilibrium solution are perfect Nash equilibrium defined by Selten (1975) and proper Nash equilibrium defined by Myerson (1978).

Traditionally, a definition of a refinement was given, after which one tried to prove that it did not suffer from flaws like the one described above. Unfortunately all refinements introduced so far have their own specific shortcomings.

In their seminal paper of 1986, Kohlberg and Mertens broke with this tradition. They first composed a list of (initially seven) properties they thought to be essential for any reasonable solution. After that they tried to find a solution that satisfies all of their
requirements. Although it lies outside the scope of this contribution to describe all seven requirements in detail, we will describe some of them.

**Existence.** This basic requirement states that every game should have at least one solution set.

**Admissibility.** The admissibility of a strategy profile can be seen as a natural strengthening of the notion of undominatedness. The requirement states that, given a game, any strategy profile in any solution set of the game should be admissible, and in particular not use (weakly) dominated strategies.

Example 4 shows that Nash equilibria are not always admissible since, as is already said, weakly dominated strategies are not used in admissible profiles. On the other hand, it is a well-known fact that perfect equilibria are admissible (see the Appendix for a precise definition of perfect equilibria and admissibility). We will nevertheless show that perfect equilibria do not satisfy the next requirement.

**Deletion of a Bad Strategy.** The philosophy behind this requirement is as follows. Suppose that a player of a game has a "bad" pure strategy, for instance a dominated one. Then it is first of all reasonable to suppose that this strategy is not used in any solution set of that game, like we did in the previous requirement. One can however take the argument even further. Since every opponent of this player knows that the pure strategy under consideration is a bad one, the opponents also know that the player will never use his bad strategy. Thus one might argue that any solution set of the game should also be a solution set of the game that results when the bad strategy is eliminated from the set of pure strategies of the player in question.

**Example 5.** We will show that perfect equilibria do not survive the deletion of a weakly dominated strategy. Consider the 2 × 3-bimatrix game

\[ \Gamma = \begin{pmatrix} (1,2) & (1,1) & (0,1) \\ (1,1) & (0,0) & (0,0) \end{pmatrix} \]

The strategy pair \((0,1),(1,0,0)\) is perfect since the sequence \(((1,1,1,1,1,1))\) is a perfect and converging to \((0,1,1,0,0))\. Note that the third column is dominated by the first column for player II. The game that results from the deletion of the third column is

\[ \Gamma' = \begin{pmatrix} (1,2) & (1,1) \\ (1,1) & (0,0) \end{pmatrix} \]

Furthermore, \((0,1),(1,0)\) is the strategy pair in this new game that corresponds to the original strategy pair \((0,1),(1,0,0)\). However, playing the second row has now become a weakly dominated strategy for player I. Hence, the strategy pair cannot be perfect, since a perfect equilibrium is admissible, and does therefore not use a weakly dominated strategy.

**Invariance.** In general, invariance means that "similar" games should have the "same" solution sets. In this case, the similarity of two games refers to deletion of pure strategies of one (larger) game that are duplicates of other (possibly mixed) strategies. The resulting of this strategy profile yields a smaller (second) game. The connection between these two games is given in

**Definition 3.** Let \( \Gamma = (A, u) \) and \( \Gamma' = (B, v) \) be two games. A map \( f = (f_i)_{i \in M} \) from \( A \) to \( B \) is called a reduction map from \( \Gamma \) to \( \Gamma' \) if for every player \( i \),

\[ f_i(\Delta_i(B)) = \Delta_i(A) \]

is linear and onto and

\[ f_i(u_i) = f_i(u) \circ f \]

The two games \( \Gamma \) and \( \Gamma' \) are similar in the sense that for every player \( j \) and every strategy \( x_j \in \Delta(B) \) of this player in the game \( \Gamma' \), the strategy \( f_j(x_j) \) gives him the same payoff in \( \Gamma' \) as \( x_j \) does in \( \Gamma' \), no matter what the other players do (using the functions \( f_i \) to transform their strategies from one game to the other). The two games do however have different strategy spaces. So, now the problem is, what do we mean by the "same" solution sets? This question can be answered as follows.

**Definition 4.** A solution \( \sigma \) is called invariant if for any pair of games \( \Gamma \) and \( \Gamma' \), and any reduction map \( f \) from \( \Gamma' \) to \( \Gamma \) we have

\[ \sigma(\Gamma) = \{ f(S) \mid S \in \sigma(\Gamma') \} \]

and

\[ f^{-1}(T) = \{ S \in \sigma(\Gamma') \mid f(S) = T \} \]

for all \( T \in \sigma(\Gamma) \).

These equalities state that every solution set of the larger game \( \Gamma' \) projects (via \( f \)) onto a solution set of the smaller game \( \Gamma \) and, moreover, that every strategy profile of the larger game that projects into a solution set \( T \) of the smaller game is an element of a solution set \( S \) of the larger game that projects onto \( T \).

4. Stable sets

The main problem of Kohlberg and Mertens was that in 1986 no solution was known that satisfied all of their requirements. Thus, the central question in their paper was:

**Does there exist a solution that satisfies all requirements?**

A solution that is designed to satisfy all requirements is usually called a **stability concept**. The solution sets of such a concept are called **stable sets**.

Although Kohlberg and Mertens did not manage to find a stability concept that satisfies all requirements, at least two observations can be made concerning the way they searched for one. First of all they showed that their requirements inevitably led to a set-valued solution.

**Example 6.** Consider the game \( \Gamma \) introduced in Example 5. Let \( S \) be a solution set of this game according to some solution \( \sigma \) satisfying all requirements. We will show that \( S \) contains at least two strategy profiles. To this end, note that, by existence and admissibility,

\[ \{((1,0),(1,0))\} \]

is the unique solution set of \( \Gamma' \), since playing the first row (column) is the only undominated strategy of player I (player II). Thus, since \( \sigma \) satisfies the deletion of a bad strategy by assumption, the set \( S \) must contain the strategy pair \((1,0),(1,0,0))\). However, the same line of reasoning applied to the deletion of the (dominated) second column shows that \( S \) must also contain \(((0,1),(1,0,0))\). For this particular reason solutions are allowed to be set-valued in this paper.

Secondly, they introduced a specific method to generate stability concepts. Roughly speaking, they constructed the game \( \Gamma \) neighborhood of 'perturbed games' and then said that a set of strategy profiles of \( \Gamma \) is stable if for any possible perturbation there is at least one strategy profile in the set that survives the perturbation.
Formally, let $\Gamma = (A, u)$ be a game. Let $P$ be a set of 'perturbed games', each equipped with some metric $d$. It is assumed that each perturbed game $\Gamma' \in P$ has a non-empty set $E(\Gamma')$ of equilibria and that $\Gamma$ is an element of $P$.

**Definition 5.** A closed set $S \subseteq A$ is called a $P$-set of the game $\Gamma$ if for any neighborhood $V$ of $S$ there exists a number $\delta > 0$ such that $E(\Gamma') \cap V$ is not empty if $d(\Gamma', \Gamma) < \delta$.

**Remark.** The sets satisfying this condition are not themselves called stable since, for example, $A$ also satisfies the condition. The sets that actually will be called stable also need to be sufficiently small. The selection of sufficiently small $P$-sets can be done in various ways.

Obviously, the resulting stability concept depends on the choice of the specific set $P$ and the metric $d$. Using a variation on this theme Mertens (1989) was able to construct a stability concept that did indeed satisfy all requirements.

**Example 7.** In 1990 Hills (see also Mertens) introduced a stability concept. He first identified a game $\Gamma = (A, u)$ with its best reply correspondence $\beta_\Gamma : A \rightarrow \Delta A$, defined by $\beta(x) := \prod_{a \in A} \beta_\Gamma(x(a))$.

The set of perturbed games was chosen to be the collection $C$ of all compact and convex-valued upper hemimetric (uhc) correspondences $\phi : A \rightarrow \Delta A$.

It is not difficult to show that $\beta$ is such a correspondence. So, the game $\Gamma$ is indeed an element of $C$ given the above identification. Further, for a correspondence $\phi \in C$, the set of fixed points $Fix(\phi) = \{ x \in A \mid x(\phi(x)) = x \}$ serves as the set of equilibria of the 'perturbed game' $\Gamma$. This is in agreement with the identification of $\Gamma$ with $\beta$, since $Fix(\beta)$.

The metric on $C$ is based on the Hausdorff distance $d_H$ with, for $X, Y \subseteq A$,

$$d_H(X, Y) := \inf \{ \varepsilon > 0 \mid X \subseteq B_\varepsilon(Y) \text{ and } Y \subseteq B_\varepsilon(X) \}.$$ 

Then the pointwise Hausdorff metric $d$ on $C$ is defined by

$$d(\phi, \psi) := \sup \{ d_H(\phi(x), \psi(x)) \mid x \in A \} \quad (\phi, \psi \in C).$$

Now a closed set $S \subseteq A$ is called a C-set of the game $\Gamma$ if for any neighborhood $V$ of $S$ there exists a number $\delta > 0$ such that $Fix(\phi) \cap V$ is not empty if $d(H, \phi(x)) < \delta$. Finally, $S$ is in the sense of Hills.

**Results.** Hills (1990) showed that the solution that assigns to each game $\Gamma$ its collection $r(\Gamma)$ of stable sets in the sense of Hills satisfies EXISTENCE, ADMISSIBILITY, and, moreover proved that it also satisfies a strong variant of DELETION of a BAD STRATEGY.

In my thesis a certain type of stable sets is introduced whose definition stays closer to known game theoretic notions while it satisfies the same requirements. Looking back, this is not so surprising, since Hills and I were able to prove that both definitions yield the same stability concept. Using this equivalence we also showed where a stable set of the sense of Mertens contains a stable set in the sense of Hills.

I also managed to construct a counterexample for the INVARIENCE of this stability concept. The example moreover shows that this is not due to Hills' particular choice $C$ of perturbed games. It is merely caused by the choice of minimal C-sets as final solution sets. The following selection method however does yield an invariant solution.

Let $\Gamma$ be a game and let $S$ be a C-set of this game. Let us call $S$ extendable if, for all games $\Gamma' \in P$ and all reduction maps $f$ from $\Gamma'$ to $\Gamma$, the set $f^{-1}(S)$ is a C-set of $\Gamma'$.

Now a non-empty and closed set $S$ of strategy profiles of $\Gamma'$ is called stable if it is extendable, connected and consists entirely of perfect equilibria. It can be shown that this solution satisfies all requirements of Kohlberg and Mertens.

5. Appendix

**Definition of admissibility.** Let $\Gamma = (A, u)$ be a game. A strategy profile $x$ in $A$ is called completely mixed if all coordinates $x_i$ are positive. For player $i$, a strategy $y_i$ is an admissible best reply against a strategy profile $x$ if there is a sequence $(x^n_k)_{k \in N}$ of completely mixed strategy profiles converging to $x$ such that $y_i(x^n_k)$ is a best reply against $x^n_k$ for all $k$.

We say that a solution $\sigma$ satisfies admissibility if, for every game $\Gamma$, every solution set $S$ is in $\sigma(\Gamma)$ and every strategy profile $x$ in $S$, the strategy $x_i$ is admissible for every player $i$.

**Definition of perfect equilibria.** Let $\eta > 0$ and let $x \in A$ be a completely mixed strategy profile. Then $x$ is called $\eta$-perfect if for all players $i \in N$ and for all $a \in A_i$ we have that $x_i(a) \leq \eta$ whenever $a$ is not a best reply against $x$.

A strategy profile $x \in A$ is called perfect if there exist a sequence $(x^n_k)_{k \in N}$ of positive real numbers converging to zero and a sequence $(x^n_k)_{k \in N}$ of strategy profiles in $A$ converging to $x$ such that $x^n_k$ is $\eta_k$-perfect for all $k$.

**References**


