Consistent Pairs in Extensive Games *

Tilman Börgers
Department of Economics
University College London
Gower Street
London WC1E 6BT
U.K.

* This paper is a revision and expansion of parts of my earlier paper "On the Definition of Rationalizability in Extensive Games". I am grateful to Dieter Balkenborg, Peter Funk, Martin Hellwig and Joel Sobel for their comments. I gratefully acknowledge partial financial support from the Swiss National Science Foundation.

Abstract

The notion of consistent pairs, developed in Samuelson [8] and Börgers and Samuelson [5], is extended to perturbed 2 player extensive games. It is argued that this solution concept provides a natural formalization of "forward induction". It is shown, however, that the concept can lead to problems of non-existence and of multiplicity. Unlike similar examples in [8] and [5], the examples given here are robust under small changes of payoffs. This paper considers both uncorrelated and correlated perturbations. This makes it possible to study interactions between "forward induction" and "backward induction".

1 Introduction

The plan of this paper is to extend the notion of "consistent pairs", which was introduced for normal form games in Samuelson [8] and Börgers and Samuelson [5], to extensive games. There are two motivations for this extension. One is that, in extensive games, "consistent pairs" seem to provide a natural formalisation of the notion of "forward induction", which has occupied game theorists for some time, but for which no agreed mathematical definition exists.[1] The second motivation is that a result of Samuelson [8] and Börgers and Samuelson [5] can be strengthened once extensive games are considered. The result is that for some games no consistent pairs exist. In [5] and [8] this result is interpreted as an impossibility result concerning the notion of "perfectly rational" behaviour. A weakness of the examples which [5] and [8] use to prove this result is that they are not robust, i.e. one can change payoffs slightly, and consistent pairs do exist. In this paper we shall show that, once one considers extensive games, one can give *robust* examples of non-existence of consistent pairs.

The two points made in this paper are related to each other. It will be seen that the non-existence of consistent pairs in extensive games is due to intuitive problems that are intrinsic to the logic of "forward induction".

For finite 2 player games in normal form [5] and [8] called a pair of nonempty sets of strategies, one set for each player, a "consistent pair" if each player's set is the set of all best responses to beliefs the support of which is equal to the other player's set. The strategy sets referred to in this definition are supposed to be interpreted as the sets containing the two players' "rational" strategies. The definition then requires that for each player a strategy is called "rational" if and only if it is a best response to a belief which attaches positive probability to rational strategies of the other player only, and which gives each rational strategy positive probability. The first of these two restrictions for beliefs reflects that rationality is assumed to be common knowledge, the second has the intuitive interpretation that a priori a player does not regard any of the rational strategies of the other player as impossible.[2]

To extend consistent pairs to finite 2 player extensive games one could simply determine the consistent pairs of the normal forms of these games. This would be in line with the view of Kohlberg and Mertens [7] that normative game theoretic solution concepts should depend on the normal form of an extensive game only. If this view were adopted the non-robustness of the examples in [5] and [8] would not necessarily be a problem because non-robust

[1]See van Damme [12] for a survey of the literature on forward induction.

[2]Detailed discussions of the definition of consistent pairs have been given by Balkenborg [1] and Squires [10]. Both authors also discuss in detail the relation between consistent pairs and other solution concepts.

normal forms may correspond to robust extensive forms. Here, however, we shall not commit ourselves to Kohlberg and Mertens' view. Instead, we shall consider directly the extensive form. It seems that intuitions are more easily developed if the extensive form is considered. If the resulting concepts turn out not to be normal form invariant, we shall regard the extensive form intuition as more important than the lack of normal form invariance.[3]

We shall apply consistent pairs to *perturbed versions* of extensive games, i.e. to versions in which players are liable to tremble in the sense of Selten [9]. This will provide rudimentary foundations for statements about players' beliefs concerning information sets which cannot be reached under any rational play.

To apply consistent pairs to perturbed versions of extensive games we shall make a change in the definition of consistent pairs. We shall strengthen the requirement that a player's belief has full support on the other player's set of rational strategies, and assume in addition that each rational strategy is regarded "much more likely" than any tremble. This seems to be in line with the basic idea of consistent pairs. Without this adjustment the support requirement in the definition of consistent pairs would, in fact, be void.

Once this adjustment is made, it is easy to see that consistent pairs can be expected to exhibit some form of forward induction. Suppose that a player's belief about another player's behaviour is formed in accordance with the new requirement. A player's observations about the other player's behaviour will induce a revision of this player's belief about the other player's strategy using Bayes' rule. As long as the observed behaviour is consistent with some rational strategy, the posterior belief will give the rationality hypothesis much more probability than the hypothesis that the other player has trembled. Thus, players meet the requirement "that the inferences players draw about a player's future behavior should be consistent with rational behavior of this player in the past" which is how van Damme ([12], p.56) recently described the notion of forward induction.

In this paper, we shall consider two types of trembles, extensive form trembles, i.e. trembles which are uncorrelated across information sets, as in Selten [9], and normal form trembles, similar to those in Fudenberg et. al. [6]. Intuitively, normal form trembles correspond to the case in which trembles are completely correlated across information sets. It will be shown that consistent pairs may be very different depending on which of these two types of trembles are considered.

Extensive form trembles generate solution concepts that satisfy "backward

[3]We shall not return to the subject of normal form invariance in this paper. However, it will be immediate that only the second of the two concepts defined in the paper is normal form invariant.

induction" whereas solution concepts based on normal form trembles may violate "backward induction". Thus, one can say that our investigation of consistent pairs in perturbed extensive games will uncover interactions between forward and backward induction.

The intuition of forward induction has long been known to be both appealing but also problematic. In our approach this is reflected by the fact that consistent pairs need not exist. Formally, consistent pairs are fixed points in the space of nonempty subsets of the strategy sets. The favourable case is that this problem has a unique solution. However, as in [5] and [8], in some games either none, or multiple fixed points exist. Among these problems the non-existence problem is the most worrying one. The examples that we shall give to illustrate these problems will differ from related normal form examples in [5] and [8] in that they are robust under small changes in the players' payoffs.

This paper is structured as follows: In Section 2 the formal definition of consistent pairs in extensive games is given. Section 3 uses examples to illustrate the relation between consistent pairs and forward induction. Section 4 deals with non-existence and multiplicity of consistent pairs. The examples of Sections 3 and 4 suggest a general conjecture about consistent pairs in games of perfect information. This conjecture will be proved in Section 5. The result of that section will be used to relate the non-existence of consistent pairs in certain games of perfect information to an impossibility result of Basu [2].

2 Definitions

Attention will be restricted to finite, perfect recall extensive games with two players: $i = 1, 2$. In games with more than two players the issue would arise whether one player's beliefs about two other players' behaviour should take the form of a product measure. We restrict attention to two player games to avoid this point.

It will be unnecessary to have explicit notation for the various parts of the game tree and the payoffs. The only notation that we shall need is the symbol "S_i" for the set of player i's pure strategies.

As was explained in the Introduction, we shall consider perturbed versions of the given game in which both players have a small probability of making a mistake, i.e. of "trembling". We shall consider two forms of trembles: extensive form trembles as in Selten [9], and normal form trembles, similar to those in Fudenberg et. al. [6].

We begin with the case of extensive form trembles. An extensive form tremble of a player i, t_i, is a function which assigns to every information set of player i a measure defined on the set of actions available at that set. The measure indicates for every action the minimum probability with which it is chosen. At every information set, the minimum probabilities have to be strictly positive for all available actions, and their sum has to be less than one.

Player i's beliefs about how the game is played will be described by a triple (μ^i, t^i_i, t^i_j) whereby μ^i reflects player i's beliefs about player j's behavior [4] in the case that j does not tremble, and is hence a probability measure on S_j, and t^i_i and t^i_j are extensive form trembles, one for player i, and one for player j, indicating how i expects himself and j to tremble.

It will be useful to compare the probabilities which player i attaches to deliberate choices of player j with those which he attaches to trembles of player j. Therefore, we define the "ratio of player i's belief about player j's tremble, t^i_j, and a belief μ^i of player i" to be the quotient of the largest probability which t^i_j assigns to any action of player j and the minimum *of all positive probabilities* which μ^i assigns to strategies of player j. We denote this ratio by t^i_j/μ^i. If the ratio is close to zero, then any tremble is much less likely than even the least likely of all possible strategies.

Given a triple (μ^i, t^i_i, t^i_j) we can determine for every strategy s_i of player i the expected utility which i receives if he plays s_i [5]. A strategy which maximises expected utility will be called a "best reply to (μ^i, t^i_i, t^i_j)".

The goal is to define "consistent pairs" $(\hat{S}_1, \hat{S}_2)$ of sets of pure strategies of the two players, whereby for every player i the set $\hat{S}_i$ is the set of player i's "rational" choices. We shall postulate that these sets should solve Problem 1 defined below. In the definition of Problem 1 two parameters ε and δ appear. Both are assumed to be elements of (0,1). "Consistent pairs" will be the solutions of Problem 1 (or Problem 2 below) in the case that ε and δ are close to zero. Proposition 1 below will show that this is well-defined.

Problem 1 *Find nonempty sets $\hat{S}_i \subseteq S_i$ ($i = 1, 2$) such that for every $i = 1, 2$ the set $\hat{S}_i$ is the set of all strategies which are best replies to triples (μ^i, t^i_i, t^i_j) with the properties that μ^i has support $\hat{S}_j$, that the minimum probabilities determined by t^i_i and t^i_j are not greater than ε, and that the ratio t^i_j/μ^i is not greater than δ.*

Hence, a strategy of a player is "rational" if it is a best response to a belief which reflects that the other player is rational, that both players tremble with small probabilities, and that any rational choice of the other player is much more likely than a tremble of that player. This last condition is related to the condition in [5] and [8] that the support of each player's beliefs is the other player's set of rational strategies. It is modified so as to take the existence

[4]Throughout this paper when "i" denotes one player, "j" denotes the other player, i.e. $j \neq i$.

[5]Assuming that trembles at different information sets are uncorrelated, and that trembles are not correlated with strategy choices.

of trembles into account. Below, we shall sometimes refer to the modified condition as the "strengthened full support condition".

Observe that in Problem 1 it is not assumed that the trembles are "common knowledge". In fact, the players are permitted to hold arbitrary beliefs about their own and the other player's trembles, provided that they expect these trembles to be "small". Intuitively, Problem 1 refers thus to an environment in which "rational behaviour" is understood to be behaviour from some limited set of possibilities, but "irrational behaviour" is without commonly recognized structure.

Another important feature of Problem 1 is that the sets $\hat{S}_i$ are supposed to contain *all* best replies to beliefs with the required properties, not just some of them. This reflects that the notion of rationality formalised in Problem 1 is an "eductive" [6] one. If there is a justification for an action in terms of admissible beliefs, then this action will always be called "rational".

Next, we define a problem that is analogous to Problem 1, but that involves normal form rather than extensive form trembles. A normal form tremble of a player i, τ_i, is a function which assigns to every pure strategy of player i a minimum probability with which that strategy has to be chosen. The minimum probabilities have to be strictly positive, and their sum has to be less than one.

We can now proceed in analogy to what we did above, and define player i's beliefs as a triple $(\mu^i, \tau_i^i, \tau_j^i)$. We can also define the notion of a best reply to such a triple. The details of these definitions are omitted. We also need to define the "ratio of a tremble τ_j^i and a belief μ^i", τ_j^i/μ^i. It is the quotient of the largest probability assigned by τ_j^i to any strategy, and the smallest positive probability assigned by μ^i to a strategy.

The analog of Problem 1 for normal form trembles is:

Problem 2 *Find nonempty sets $\hat{S}_i \subseteq S_i$ $(i = 1, 2)$ such that for every $i = 1, 2$ the set $\hat{S}_i$ is the set of all strategies which are best replies to triples $(\mu^i, \tau_i^i, \tau_j^i)$ with the properties that μ^i has support $\hat{S}_j$, that the minimum probabilities determined by τ_i^i and τ_j^i are not greater than ε, and that the ratio τ_j^i/μ^i is not greater than δ.*

As was said above we shall focus on solutions of Problems 1 and 2 in the case that ε and δ are "small". The following result shows that this is well-defined:

Proposition 1 *For any game there exist $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ such that the set of solutions of Problem 1 is the same for all $(\varepsilon, \delta) \in (0, \bar{\varepsilon}) \times (0, \bar{\delta})$. The same is true for Problem 2.*

[6] In the sense of Binmore [4].

Proof: We give the argument only for Problem 1. For Problem 2 the proof is completely analogous. For any pair $(\hat{S}_1, \hat{S}_2)$ we consider the set of all $(\varepsilon, \delta) \in (0,1)^2$ for which $(\hat{S}_1, \hat{S}_2)$ solves Problem 1. The proof will be built on the following property of this set: Suppose $\varepsilon^1, \varepsilon^2, \varepsilon^3 \in (0,1)$ satisfy $\varepsilon^1 < \varepsilon^2 < \varepsilon^3$, and $\delta^1, \delta^2, \delta^3 \in (0,1)$ satisfy $\delta^1 < \delta^2 < \delta^3$. If $(\hat{S}_1, \hat{S}_2)$ solves Problem 1 for both $(\varepsilon^1, \delta^1)$ and for $(\varepsilon^3, \delta^3)$ then this pair solves Problem 1 also for $(\varepsilon^2, \delta^2)$.

To show this we denote for any $(\varepsilon, \delta) \in (0,1)^2$ by $B^{(\varepsilon,\delta)}(\hat{S}_j)$ the set of all strategies of player 2 that are best responses to a triple $(\mu^i,\ t^i_i,\ t^i_j)$ for which μ^i has support $\hat{S}_j$, the minimum probabilities determined by t^i_i and t^i_j are not greater than ε, and the ratio t^i_j/μ^i is not greater than δ. Note that $B^{(\varepsilon,\delta)}(\hat{S}_j)$ is weakly decreasing (in terms of set-inclusion) as ε or δ decreases.

By this last observation we now have: $B^{(\varepsilon^1,\delta^1)}(\hat{S}_j) \subseteq B^{(\varepsilon^2,\delta^2)}(\hat{S}_j) \subseteq B^{(\varepsilon^3,\delta^3)}$ $(\hat{S}_j)$. Since the first and the last of these sets are both equal to $\hat{S}_i$ the same must be true for the set in the middle. Hence: $\hat{S}_i = B^{(\varepsilon^2,\delta^2)}(\hat{S}_j)$ for both i. Thus $(\hat{S}_1, \hat{S}_2)$ solves Problem 1 for parameters $(\varepsilon^2, \delta^2)$. This completes the first step of the proof.

We now distinguish three groups of pairs $(\hat{S}_1, \hat{S}_2)$. A pair belongs to the first group if there are no values of (ε, δ) for which the pair solves Problem 1. A pair belongs to the second group if there are values of (ε, δ) for which it solves Problem 1, but if there is no sequence $(\varepsilon^n, \delta^n)_{n \in \mathbb{N}}$ such that the pair solves Problem 1 for all pairs of parameter values contained in this sequence and such that $(\varepsilon^n, \delta^n) \to (0,0)$ as $n \to \infty$. Finally, the third group contains all remaining pairs $(\hat{S}_1, \hat{S}_2)$.

Now, clearly, we can choose $\bar{\varepsilon}$ and $\bar{\delta}$ such that $\varepsilon < \bar{\varepsilon}$ and $\delta < \bar{\delta}$ implies that no pair $(\hat{S}_1, \hat{S}_2)$ in the first or the second of the above groups solves Problem 1 for parameter values ε and δ. We can moreover choose $\bar{\varepsilon}$ and $\bar{\delta}$ also such that for every pair $(\hat{S}_1, \hat{S}_2)$ in the third group there exists some $\varepsilon > \bar{\varepsilon}$ and some $\delta > \bar{\delta}$ such that the pair $(\hat{S}_1, \hat{S}_2)$ solves Problem 2 for the parameter values ε and δ.

With this choice of $\bar{\varepsilon}$ and $\bar{\delta}$ Proposition 1 is true. To see this note that by construction for $\varepsilon < \bar{\varepsilon}$ and $\delta < \bar{\delta}$ no pair of sets belonging to the first two of the above groups can solve Problem 1. Hence it suffices to show that all pairs of sets that belong to the third of the above groups will solve Problem 1 for all $\varepsilon < \bar{\varepsilon}$ and all $\delta < \bar{\delta}$. But this follows from the first step of this proof, together with the fact that by construction there exists for any pair of sets in the third group and for all $\varepsilon < \bar{\varepsilon}$ and all $\delta < \bar{\delta}$ some $\varepsilon^1, \varepsilon^2, \delta^1, \delta^2$ such that $\varepsilon^1 < \varepsilon < \varepsilon^2$, $\delta^1 < \delta < \delta^2$, and such that the given pair solves Problem 1 both for $(\varepsilon^1, \delta^1)$ and for $(\varepsilon^2, \delta^2)$.

Q.E.D.

We can now extend the definition of "consistent pairs" to extensive games. As we have considered two types of trembles, we define two types of "consistent pairs" for extensive games.

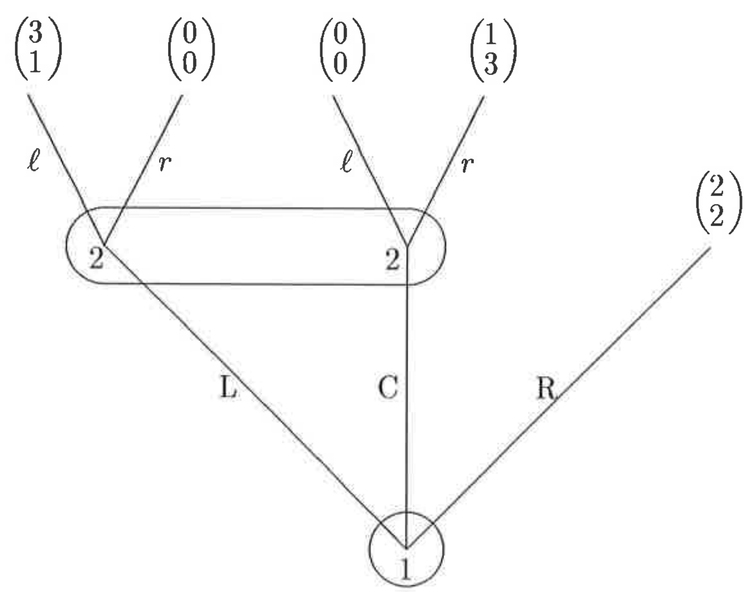
Definition 1 *A pair of strategy sets* $(\hat{S}_1, \hat{S}_2)$ *is called a "consistent pair of type 1" (resp. "consistent pair of type 2") if it solves Problem 1 (resp. Problem 2) for* $(\varepsilon, \delta) \in (0, \bar{\varepsilon}) \times (0, \bar{\delta})$ *whereby* $\bar{\varepsilon}$ *and* $\bar{\delta}$ *are the parameter values referred to in Proposition 1.*

3 Consistent Pairs and Forward Induction

This section contains examples which illustrate that consistent pairs satisfy a version of forward induction. The intuitive reason why consistent pairs and forward induction are related was already indicated in the Introduction. We shall also describe in this section some interaction between backward and forward induction. For this we shall compare consistent pairs of type 1 and of type 2.

The first two examples are well-known from the literature on forward induction. We shall show that in these examples consistent pairs support the conclusions obtained by previous analyses that were based on forward induction.

EXAMPLE 1: The Battle of the Sexes with an outside option for player 1.



Example 1

Example 1 is well-known from van Damme's work ([11], [12]). It is often used to illustrate the notion of forward induction. In Example 1 every player has just one information set, and therefore there is no distinction between consistent pairs of types 1 and 2. In the following, we shall therefore refer to "consistent pairs" without specification of the type.

CLAIM 1: In Example 1 there is a unique consistent pair: $\hat{S}_1 = \{\mathrm{L}\}$, $\hat{S}_2 = \{\ell\}$.

According to the solution described in Claim 1, player 1 does not take up the outside option, and, in the Battle of the Sexes, player 1's most preferred outcome is played. This is what also previous discussions of forward induction in this example ([11]) have concluded. The intuition for this example is that player 1, by not choosing the outside option, indicates that he expects to get at least 2 from the Battle of the Sexes. Therefore, player 2 should think that player 1 played L, and she should respond optimally by choosing ℓ. Hence, it is indeed rational for player 1 not to choose the outside option, and to choose L.

To see that the indicated pair of sets is indeed a consistent pair note first that $\hat{S}_1$ is clearly the set of all best responses of player 1 to beliefs that are admissible[7] given $\hat{S}_2$. As regards player 2, the strengthened full support requirement implies that when her information set is reached she must believe that it is most likely that player 1 chose L. Hence her unique best response is ℓ.

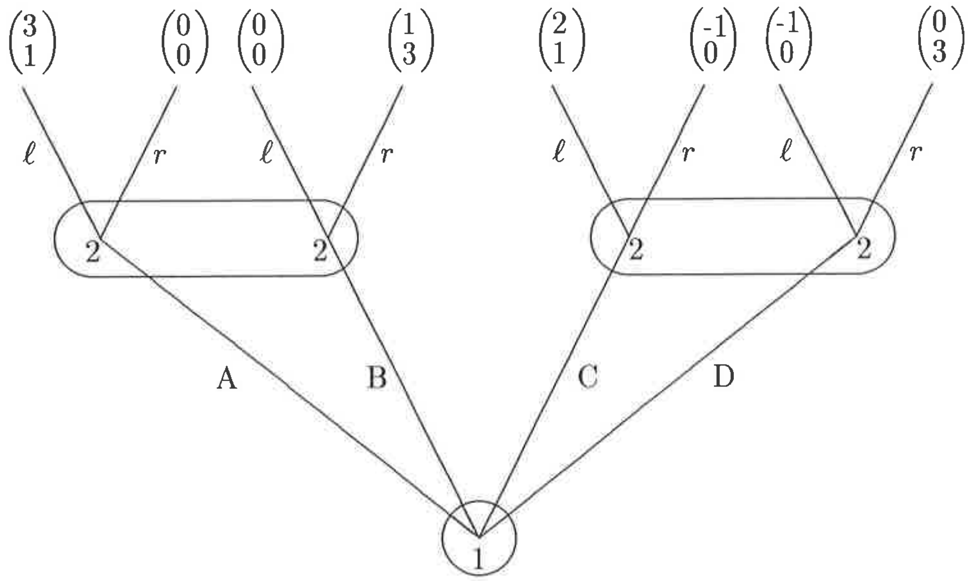
To see that there is no other consistent pair observe first that C is strictly dominated by R, so that we must have $C \notin \hat{S}_1$. If $\hat{S}_1 = \{L\}$ then the argument of the preceding paragraph shows $\hat{S}_2 = \{\ell\}$, so that we have the consistent pair of Claim 1.

If $\hat{S}_1 = \{R\}$ then player 2's information set can be reached only if player 1 trembles. Hence, if player 2 has to move, she can hold any arbitrary belief about player 1's preceding choice. Since $\hat{S}_2$ must contain *all* best responses of player 2 it must contain ℓ as well as r. Since player 1 may hold any arbitrary belief about $\hat{S}_2$, his strategy L might be a best response for him, and hence $L \in \hat{S}_1$, contradicting our initial hypothesis.

Finally, consider the case $\hat{S}_1 = \{L, R\}$. Then the strengthened full support requirement implies that player 2, if her information set is reached, must conclude that the most likely explanation for this is that player 1 chose L. Her best response is hence ℓ. So we must have $\hat{S}_2 = \{\ell\}$. Hence R is not a best response for player 1, contradicting our initial hypothesis that $R \in \hat{S}_1$.

EXAMPLE 2: A version of the Battle of the Sexes in which player 1, while choosing his strategy, can also choose to "sacrifice" 1 unit of utility.

[7] I.e. that satisfy the conditions of Problems 1/2 for the values (ε, δ) referred to in Definition 1.



Example 2

This example is due to Ben-Porath and Dekel [3] and van Damme [11]. Observe that player 2 has two information sets in Example 2. Therefore, extensive form trembles are different from normal form trembles, and we have to consider both definitions of consistent pairs.

CLAIM 2: In Example 2 there is a unique consistent pair of type 1: $\hat{S}_1 = \{A\}$, $\hat{S}_2 = \{(\ell,\ell),(\ell,r)\}$[8]. This is also the unique consistent pair of type 2.

According to this consistent pair, player 1 does not sacrifice utility, and the only outcome that is possible under rational behaviour gives player 1 his maximal utility in the Battle of the Sexes. This is the conclusion that was also obtained in previous investigations of this example that invoked forward induction ([3], [11]).

Although in Example 2 the definitions of consistent pairs of type 1 and of type 2 do not coincide all arguments that will be given below to prove Claim 2 apply to both types of consistent pairs. Therefore, we shall refer below to "consistent pairs" without specifying the type.

It is easy to see that the pair described in Claim 2 is a consistent pair. Therefore, we only show that no other consistent pair exists. Observe first that A and B will always yield strictly higher expected utility than D, so that $D \notin \hat{S}_1$.

[8]Strategies for player 2 indicate first her choice at her left information set, and then her choice at her right information set. The same convention applies in the context of similar examples below.

Next, we wish to prove that $C \notin \hat{S}_1$. The proof is indirect. Suppose $C \in \hat{S}_1$. As $D \notin \hat{S}_1$ the strengthened full support requirement implies that every element of $\hat{S}_2$ must have the second component ℓ. Therefore, C yields expected utility close to 2, and hence $B \notin \hat{S}_1$. This leaves two possible cases. The first case, $\hat{S}_1 = \{C\}$, can be excluded as we excluded the case $\hat{S}_1 = \{R\}$ in the discussion of Example 1. In the second case, $\hat{S}_1 = \{A, C\}$, the strengthened full support requirement implies that player 2 must choose ℓ at her left information set, and therefore C is not a best response, contradicting our hypothesis. We conclude $C \notin \hat{S}_1$.

So far, we have shown: $\hat{S}_1 \subseteq \{A, B\}$. Hence if her right information set is reached, player 2 must attribute this to a tremble of player 1. She may hence have any belief about what happened. Since $\hat{S}_2$ must contain all her best responses, she may hence play either ℓ or r at her second information set. Thus, player 1 can hold any beliefs about what would happen if he were to choose C. In particular, he may expect a utility close to 2. As $C \notin \hat{S}_1$ player 1 must expect for all admissible beliefs that his best response gives an expected utility of at least 2. Therefore, B can never be a best response, and we must have $\hat{S}_1 = \{A\}$.

But, if $\hat{S}_1 = \{A\}$, then the strengthened full support requirement implies that player 2 must choose ℓ at her left information set. As argued before, she can make any choice at her right information set, and hence we obtain the pair given in Claim 2.

In the previous two examples we obtained the same results as previous studies based on forward induction. Thus, our concept does capture some of the flavour of forward induction. The next example shows that sometimes consistent pairs yield results that differ from previous investigations of forward induction.

EXAMPLE 3: The twice repeated Battle of the Sexes with no discounting.

Here, we refer to the specification of the Battle of the Sexes which is obtained from Figure 1 by dropping the outside option for player 1.

CLAIM 3: In Example 3 the pair (S_1, S_2)[9] is a consistent pair of type 1 and of type 2.

Claim 3 is easy to verify, and therefore we omit the proof. Observe that Claim 3 is not a uniqueness assertion.

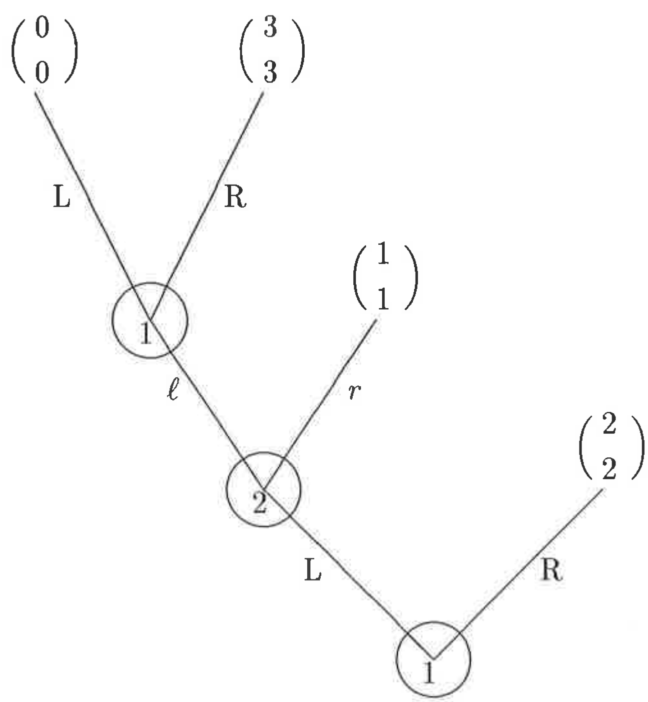
Thus, in Example 3, we have a consistent pair (of both types) in which the outcome of the game is completely indeterminate. This conflicts with

[9] Recall that S_i is player i's set of pure strategies.

van Damme's [11] result that forward induction excludes in the twice repeated Battle of the Sexes outcomes in which the same pure strategy Nash equilibrium of the one shot game is played twice. Van Damme used "stable equilibria" (in the sense of Kohlberg and Mertens [7]) as a formalisation of forward induction. The fact that stable equilibria differ so much from consistent pairs in this example appears to be due to the fact that stable equilibria, by construction, assume that the outcome that results from rational play is uniquely determined, whereas consistent pairs admit the possibility of multiple outcomes.

The next example illustrates the interaction between forward and backward induction.

EXAMPLE 4: A coordination game of perfect information.



Example 4

CLAIM 4: In Example 4 there is a unique consistent pair of type 1: $\hat{S}_1 = \{(L,R)\}$ [10], $\hat{S}_2 = \{\ell\}$. This is also the unique consistent pair of type 2.

[10]The notation for strategies of player 1 is that I first indicate his choice at his first information set, and then his choice at his second information set. The same notational

Observe that the strategies that are predicted by this consistent pair for the two players are also the unique subgame-perfect equilibrium strategies in this example.

The above result is not surprising if consistent pairs of type 1 are considered, and hence it is assumed that trembles are not correlated. Even if the strengthened full support requirement is not imposed[11], the subgame-perfect equilibrium strategies are the only candidates for rational behaviour. Adding this requirement does not affect the conclusion.

The above result is, however, not so obvious for consistent pairs of type 2. Suppose, for the moment, that we drop from Problem 2 the strengthened full support requirement. Then, one consistent pair of type 2 consists of the set $\{(L,R),(R,L)(R,R)\}$ for player 1, and the set $\{\ell,r\}$ for player 2. [12] The intuition for this is as follows: Take first the solution for player 2 as given, and consider player 1. Since both ℓ and r are rational choices of player 2, both L and R are rational choices at player 1's initial information set. If he begins with L he must continue with R. If, however, he begins with R, his second choice will be meaningless, since his second information set will never be reached. Therefore, both (R,L) and (R,R) are rational.

Consider next player 2. If she observes player 1 choosing L, she can interpret this either as a rational choice, since one of player 1's rational strategies prescribes the initial move L, or she can think that player 1 trembled. In the first case player 2 will expect player 1 to behave rationally also in future, i.e. to choose R, and hence she will choose ℓ. In the second case, since trembles are correlated, player 2 may expect player 1 to tremble also at his second information set. Hence she may prefer to choose r. Hence both ℓ and r are rational choices for player 2.

If we now take the strengthened full support requirement into account, then the sets just described no longer constitute a consistent pair. The strengthened full support requirement implies that player 2 gives priority to those explanations of her observations that don't question the rationality of player 1. Hence, when observing player 1 choosing L initially, player 2 should conclude that it is most likely that player 1 is playing his rational strategy (L,R). Then ℓ is her only rational choice.

It is easy to verify that the sets described in Claim 4 constitute a consistent pair of type 2. To see that they form *the only* consistent pair of type 2 note

convention applies in the context of similar examples below.

[11]Formally, we drop here from Problem 1 the requirement that a player's belief must have full support on the other player's set of rational strategies, and give each of these strategies a certain minimum probability, and instead we only require that the support of a player's belief must be *contained in* the other player's set of rational strategies. Proposition 1 remains true for the altered problem.

[12]Other solutions exist as well.

that any solution which makes L one of player 1's rational choices at his first information set must prescribe that player 1 continues with R, and, by the strengthened full support requirement, that player 2 chooses ℓ. This implies that (L, R) is player 1's *only* rational choice, and hence that we have the consistent pair of Claim 4.

It remains to check that there cannot be any solution which makes R player 1's *only* rational choice at his first information set. Since player 2's information set would not be reached, and since we have assumed normal form trembles, we would have $\hat{S}_2 = \{\ell, r\}$, and hence also (L, R) would be a rational choice of player 1, which contradicts the assumption that R is the only rational initial choice of player 1.

The interaction between forward and backward induction will be further illustrated in Example 7 and Proposition 3 below.

4 Non-Existence and Multiplicity of Consistent Pairs

In the previous section we have seen that consistent pairs in extensive games do reflect certain intuitions about rational behaviour. In this section it is shown that they suffer from non-existence and multiplicity problems. To some extent, these problems are reflections of intrinsic problems associated with the intuitive notions, specifically forward induction, captured by consistent pairs.

Proposition 2 *There are robust examples of games for which there is a unique consistent pair of type 1. There are also robust examples for which there is no consistent pair of type 1. There are also robust examples for which there are several consistent pairs of type 1, and none of them is "largest" in terms of set-inclusion. The same assertions are true for consistent pairs of type 2.*

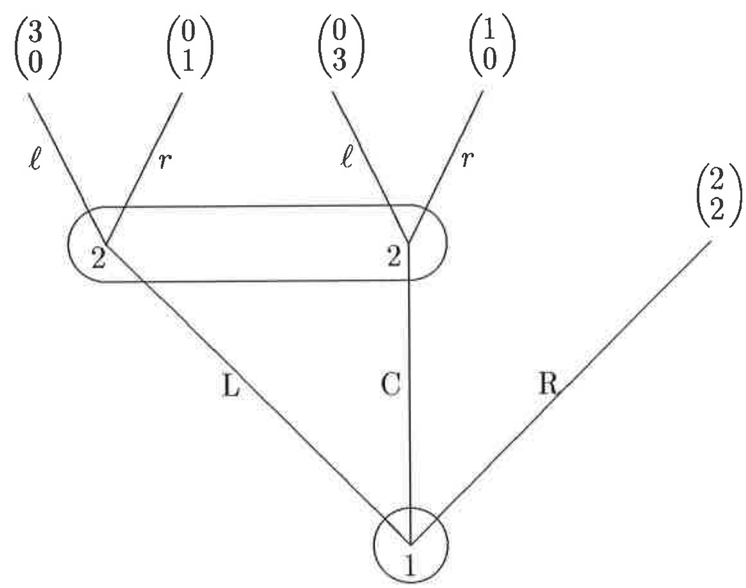
In this result, an example is called "robust" if the properties of this example, as referred to in the proposition, remain unchanged if the payoffs in the example are changed a little bit.

Proposition 2 is analogous to "Proposition 2" of [5]. However, the examples given in [5] to prove the non-existence and multiplicity parts of Proposition 2 are not robust in the sense just described. By contrast, the examples that we shall use here to prove Proposition 2 are robust.

Before proving Proposition 2 we explain why Proposition 2 asserts not only that there may be multiple consistent pairs, but in addition that it may happen that none of them is largest. This point is explained in more detail in the analogous context in [5]. There, it is shown that for normal form games the multiplicity of consistent pairs may persist, even if one drops the full support requirement for players' beliefs, but that there will always be a largest consistent pair. An analogous argument applies in the current context. Thus, the interesting point is not really the possible multiplicity of consistent pairs, but the fact that it may be that none of several consistent pairs is largest.[13]

As regards the proof of Proposition 2, the possibility that there may be a unique consistent pair was already demonstrated by Example 1 in the preceding section. It is immediate that this example is robust. We therefore begin with an example to demonstrate the possibility of non-existence.

EXAMPLE 5: The Battle of the Sexes with outside option, but with inverted payoffs for player 2: player 2 enjoys an outing only if player 1 is *not* present.



Example 5

In Example 5, as in Example 1, there is no difference between consistent pairs of type 1 and consistent pairs of type 2. In the following, we shall therefore refer, without qualification, to "consistent pairs".

CLAIM 5: There are no consistent pairs in Example 5.

For the proof of Claim 5 suppose first there were a solution such that $L \notin \hat{S}_1$. Note that C is strictly dominated by R, hence also $C \notin \hat{S}_1$. Hence $\hat{S}_1 = \{R\}$. But then player 2, when her information set is reached, must attribute

[13] Balkenborg [1] explains how the existence of largest set-valued solutions is related to the applicability of Tarski's fixed point theorem.

this to a tremble. Since she can hold arbitrary beliefs about trembles, and since every best response must be included in her set of rational strategies, both ℓ and r are rational choices for her. But then player 1 can hold arbitrary beliefs about player 2's behaviour, and, since all best responses must be included in his set of rational choices, it must be that $L \in \hat{S}_1$. This contradicts our assumption $L \notin \hat{S}_1$.

Now suppose $L \in \hat{S}_1$. Recall that C is strictly dominated, and hence will never be in $\hat{S}_1$. Therefore the strengthened full support requirement implies that player 2, when called upon to move, must conclude that it is most likely that player 1 played L, and hence she must play r. But this implies that player 1's only rational strategy is R, which contradicts our initial assumption $L \in \hat{S}_1$.

This completes the proof of Claim 5. It remains to observe that all arguments remain valid if payoffs are changed a little bit.

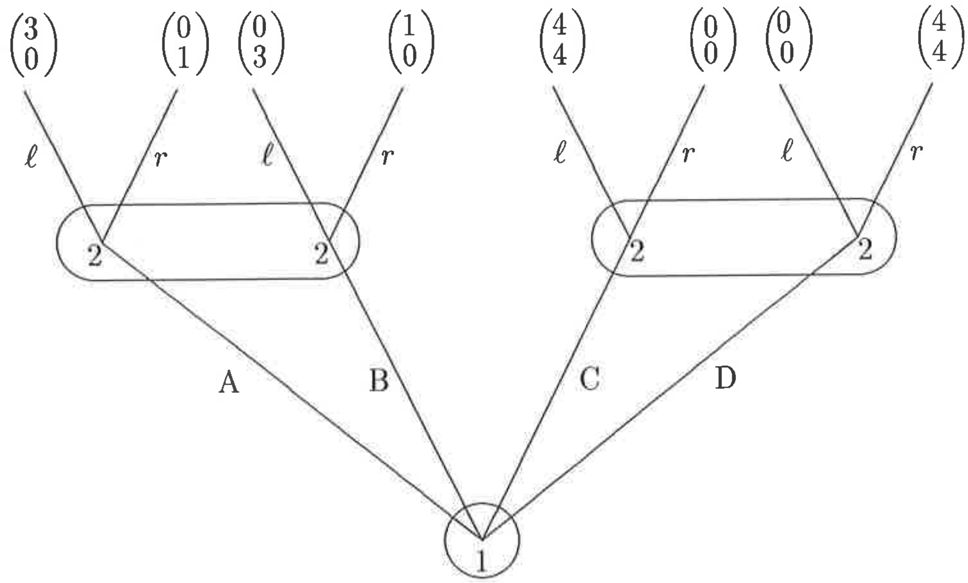
Intuitively, in this example, as in Example 1, forward induction suggests that player 2, when her information set is reached, should conclude that player 1 has chosen L. Player 2 should then choose r, and therefore player 1 should choose R. This argument is problematic, because the final conclusion is that there is no rational move of player 1 which reaches player 2's information set, and therefore there does not seem to be any foundation for applying a forward induction argument to player 2's beliefs at that information set.

To complete the proof of Proposition 2 we need to give a robust example in which there are several, but no largest consistent pairs.

EXAMPLE 6: As Example 5, but the outside option is replaced by a coordination game. (See next page.)

CLAIM 6: In Example 6 there are exactly two consistent pairs of type 1: $\hat{S}_1 = \{C\}$, $\hat{S}_2 = \{(\ell, \ell), (r, \ell)\}$, and $\hat{S}_1 = \{D\}$, $\hat{S}_2 = \{(\ell, r), (r, r)\}$. These two pairs are also the only consistent pairs of type 2.

In Example 6, because player 2 has two information sets, it is not immediate that there is no difference between consistent pairs of type 1 and consistent pairs of type 2. However, all arguments that are given below apply to both types of consistent pairs, and so we don't distinguish between them, referring instead to "consistent pairs" without specifying the type.



Example 6

It is easy to verify that the two pairs described in the claim are indeed consistent pairs. To show that no other consistent pairs exist, we consider first the possibility that $C \in \hat{S}_1$ and that $D \notin \hat{S}_1$. Then the strengthened full support requirement implies that player 2 should choose ℓ at her right information set. Therefore, by choosing C player 1 can guarantee himself almost 4. If he plays D he will get approximately 0, and if he plays A or B he will get less than 4, independent of what player 2 does. Hence: $\hat{S}_1 = \{C\}$. From this we can deduce that $\hat{S}_2 = \{(\ell, \ell), (r, \ell)\}$. Thus we have obtained the first of the two pairs described in Claim 6. An analogous argument shows that $D \in \hat{S}_1$ and $C \notin \hat{S}_1$ implies $\hat{S}_1 = \{D\}$ and $\hat{S}_2 = \{(\ell, r), (r, r)\}$ which is the other of the two pairs in Claim 6. We can thus complete the proof by showing that there is no consistent pair with the property that neither C nor D are in $\hat{S}_1$, or with the property that both C and D are in $\hat{S}_1$.

Suppose first neither C nor D were in $\hat{S}_1$. Then both ℓ and r must be rational for player 2 at her right information set. But this implies that player 1 may believe that he gets almost 4 when he chooses C or D, hence both must be rational choices. Thus there is a contradiction.

Consider next the case that both, C and D, are in $\hat{S}_1$. Then both ℓ and r must be rational choices of player 2 at her right information set. Notice next that player 1's choice B cannot be a rational choice since it yields at most payoff of 1, and, for any belief about player 2's behaviour, either C or D yields at least expected utility 2. We complete the proof by showing that both the hypothesis $A \in \hat{S}_1$ and the hypothesis $A \notin \hat{S}_1$ lead to contradictions. Consider first $A \in \hat{S}_1$. As $B \notin \hat{S}_1$, the strengthened full support requirement implies that player 2 must choose r at her left information set. Thus A yields approximately payoff 1. Choosing A can then not be rational, since, for any belief, either C or D will yield higher expected utility. This contradicts the hypothesis with which we started. Now suppose $A \notin \hat{S}_1$. Then, at player 2's left information set, both ℓ and r are rational choices. Suppose that player 1's belief attaches probability 1 to player 2's choice of ℓ at her left information set, but that player 1 attaches probability .5 to player 2's choice of ℓ at her right information set. Player 1's best response to this belief is A. Hence A is a rational choice of player 1. This contradicts again the hypothesis with which we started.

This completes the proof of Claim 6. It remains to observe that all arguments remain valid if payoffs are changed a little bit.

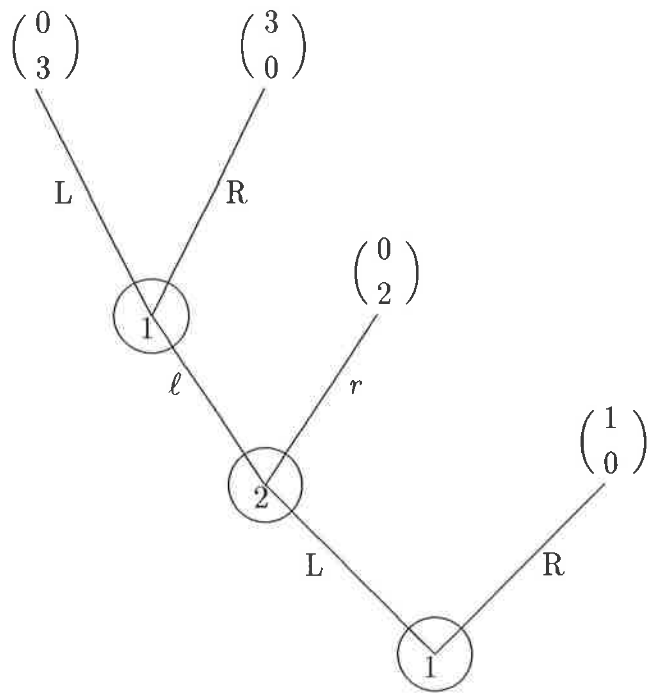
We conclude this section with a further example of non-existence of consistent pairs. The example is interesting because only consistent pairs of type 2, but not consistent pairs of type 1 fail to exist. The example is also relevant because it helps to prepare a more general analysis of consistent pairs in games of perfect information which we shall give in the next section.

EXAMPLE 7: A short "centipede" game. (See next page.)

CLAIM 7: In Example 7 there is a unique consistent pair of type 1: $\hat{S}_1 = \{(R, R)\}, \hat{S}_2 = \{r\}$. There is no consistent pair of type 2.

The unique consistent pair of type 1 consists of the singleton sets containing the subgame-perfect equilibrium strategies. It is to be expected that this solution is found if uncorrelated trembles are considered. We don't prove this part of the claim.

If we consider normal form trembles, and ignore for the moment the strengthened full support requirement for beliefs, then the pair of sets that was obtained in this context for Example 4 is also a solution for Example 7. The reason for this is roughly as above, and is therefore omitted.



Example 7

If one starts from this solution, and then introduces the strengthened full support requirement for beliefs, one is tempted to conclude that player 2 should interpret an initial choice of L by player 1 as a rational choice, and hence expect player 1 to choose R at his second information set, and that therefore player 2 should choose r. However, this would imply that player 1 should choose R initially, and, since player 1's second choice is then irrelevant, his rational strategies would appear to be (R, L) and (R, R). But note that we now have a problem analogous to that in Example 5. There is no rational strategy of player 1 that reaches player 2's information set. Hence the forward induction argument cannot be applied to player 2's choice, and both ℓ and r should be rational.

As in Example 5 this leads to a circular chain of arguments, and it there is no consistent pair of type 2. To see this formally note first that the assumption of a solution which makes L one of player 1's rational choices at his initial information set is contradictory because the argument of the preceding paragraph implies that player 2 chooses r, and hence that the only rational initial choice of player 1 is R. On the other hand, a solution which makes R the only rational choice of player 1 at his initial information set leads to the conclusion that player 2 can choose ℓ or r, and hence (L, R) should also be a rational choice of player 1.

5 Consistent Pairs in Games of Perfect Information

This section deals with particularly simple games, finite games of perfect information, and describes general properties of consistent pairs for these games. The result that we obtain generalises observations that were made for Examples 4 and 7 in the previous sections. For simplicity, the result refers only to those finite games of perfect information in which no player is indifferent between two different outcomes of the game. This condition is convenient because it ensures uniqueness of the subgame-perfect equilibrium, and it is generically satisfied.

Proposition 3 *For all finite, 2 player games of perfect information in which no player is indifferent between two different outcomes of the game: (i) There is a unique consistent pair of type 1. It consists of singleton sets, containing the subgame-perfect equilibrium strategies. (ii) There is either a unique consistent pair of type 2, or there is no consistent pair of type 2. If there is a unique consistent pair* $(\hat{S}_1, \hat{S}_2)$ *of type 2 then all strategy combinations in* $\hat{S}_1 \times \hat{S}_2$ *lead to the subgame perfect equilibrium outcome.*

Our earlier analysis of Examples 4 and 7 illustrates the assertions of Proposition 3. In particular, Example 4 is a case in which there is a unique consistent pair of type 2. As predicted by Proposition 3, all strategy combinations compatible with that pair generate the subgame-perfect equilibrium outcome. Example 7 is a case in which no consistent pair of type 2 exists. Proposition 3 asserts that no other cases can exist.

Proof: The proof of part (i) of the Proposition is elementary. Therefore, we prove part (ii) only. We describe the proof only informally. The first step is to show that if there is any consistent pair of type 2, $(\hat{S}_1, \hat{S}_2)$, then all strategy combinations in $\hat{S}_1 \times \hat{S}_2$ lead to the same outcome. To see this consider the set of all paths induced by strategy combinations $\hat{S}_1 \times \hat{S}_2$. Suppose this set contained more than one element. Consider the set of all decision nodes through which more than one path passes. There must be some "latest" nodes in this set. We focus on some such node, and on the player making a decision at this node. This player has several "rational" (according to the given consistent pair) choices. For every of his rational choices there is a unique continuation provided that the other player plays rationally. Moreover, the strengthened full support requirement for beliefs implies that, when this node is reached, the player expects with very high probability that the other player will behave rationally in the future. Hence, for each of his rational choices, he is almost certain about the associated final outcome of the game. Since he is not indifferent between any two outcomes of the game, one of the rational choices must yield a higher expected utility than the others. This contradicts the assumption that also the other choices are rational.

We conclude that for any consistent pair of type 2 there must be a unique path corresponding to it. Now suppose we knew the path associated with some consistent pair of type 2. We could then construct the corresponding solution. In fact, for every player i a strategy must then be rational if and only if it satisfies the following two conditions: Along the given path, it prescribes the moves that ensure that we stay on the path. In subgames that are reached if player j deviates from the path player i plays a strategy that is a best reply to a belief that attaches positive probabilities to all conceivable strategies of player j in that subgame. In subgames that are reached if player i deviates from the path his strategy is arbitrary. It is immediate from the definition of type 2 consistent pairs that the set of rational strategies of player i must be the set of strategies that satisfy these conditions.

To find all consistent pairs of type 2, it suffices to determine for each path the associated pair of sets in the way just described, and to check whether it is a type 2 consistent pair. No other consistent pairs of type 2 can exist. The proof can now be completed by showing that for all paths other than the subgame-perfect equilibrium path the associated pair of sets will not be a type 2 consistent pair. For this we proceed inductively. First, it is obvious that the last choice made along the path must be the one prescribed by the subgame-perfect equilibrium strategy. Next, assume that we had shown that the last n choices made along the path were those prescribed by the subgame-perfect equilibrium strategy. We must show that the same is true for the (n+1)-th choice. Suppose it were not. Let i be the player who makes the relevant choice. If he makes the choice suggested by the path he will obtain the same outcome as he would in the subgame-perfect equilibrium. If he deviated and made the subgame-perfect equilibrium choice then the subgame-perfect equilibrium continuations would not necessarily be the only, but still one of the possible continuations. But, by the definition of subgame-perfect equilibria this choice yields strictly higher utility than the path that we are considering. Therefore, the definition of type 2 consistent pairs would require that also the subgame-perfect equilibrium choice is a rational choice. This would contradict what was shown above, namely that every consistent pair of type 2 must have a unique corresponding path.

Q.E.D.

For generic games of perfect information this proposition simplifies the determination of consistent pairs. One first determines the subgame-perfect equilibrium. This immediately yields the unique type 1 consistent pair. Then one constructs the pair of sets associated with the subgame-perfect equilibrium path in the way described in the second paragraph of the above proof. Finally one checks whether the resulting pair is a type 2 consistent pair. If the answer is positive, then the constructed pair is in fact the unique consistent pair of type 2. Otherwise, there is no consistent pair of type 2.

The construction of the candidate solution that was described in the above proof suggests, in addition, that there is an intuitive condition that characterizes games in which a unique type 2 consistent pair exists. It is that no player can hope to gain by deviating from the subgame-perfect equilibrium path and pretending to be irrational. To avoid heavy formalism we do not give a formal statement of this result, but, roughly speaking, the relevant test is the following: When considering to deviate from the subgame-perfect equilibrium path, player i must anticipate that player j may continue in the ensuing subgame with any one of his best responses to a full support belief over i's strategies in this subgame. If there is no belief of player i that respects this restriction, and that gives higher expected utility than the subgame-perfect equilibrium path, then a unique type 2 consistent pair exists. Otherwise, no such pair exists.

This discussion makes clear that consistent pairs of type 2 fail to exist in games of perfect information whenever Basu's [2] impossibility result applies to a game. The above discussion thus integrates Basu's result into the framework of this paper, and shows that the impossibility result stems from the combination of the assumption of correlated trembles with a forward induction argument.

References

[1] Balkenborg, D., The Properties of Persistent Retracts and Related Concepts, PhD-thesis, Universität Bonn, 1992.

[2] Basu, K., On the Non-Existence of a Rationality Definition for Extensive Games, *International Journal of Game Theory* 19 (1990), 33-44.

[3] Ben-Porath, E. and E. Dekel, Signaling Future Actions and the Potential for Sacrifice, *Journal of Economic Theory* 57 (1992), 36-51.

[4] Binmore, K., Modelling Rational Players I, *Economics and Philosophy* 3 (1987), 179-214.

[5] Börgers, T. and L. Samuelson, "Cautious" Utility Maximization and Iterated Weak Dominance, *International Journal of Game Theory* 21 (1992), 13-25.

[6] Fudenberg, D., D. Kreps and D. Levine, On the Robustness of Equilibrium Refinements, *Journal of Economic Theory* 44 (1988), 354-380.

[7] Kohlberg, E. and J.F. Mertens, On the Strategic Stability of Equilibria, *Econometrica* 54 (1986), 1003-1037.

[8] Samuelson, L., Dominated Strategies and Common Knowledge, *Games and Economic Behavior* 4 (1992), 284-313.

[9] Selten, R., Re-Examination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory* 4 (1975), 25-55.

[10] Squires, D., Impossibility Theorems for Normal Form Games, mimeo., G.R.E.Q.E., Marseille, 1993.

[11] van Damme, E., Stable Equilibria and Forward Induction, *Journal of Economic Theory* 48 (1989), 476-496.

[12] van Damme, E., Refinements of Nash Equilibrium, in: J.J. Laffont (ed.), *Advances in Economic Theory, Sixth World Congress, Volume 1*, Cambridge: Cambridge University Press, 1992.