Refinements of Rationalizability for Normal-Form Games[*]

Jean Jacques Herings
CentER and Department of Econometrics
Tilburg University
e-mail: P.J.J.Herings@kub.nl

Vincent J. Vannetelbosch[†]
CORE and IRES
University of Louvain
e-mail: vannetel@core.ucl.ac.be

April 22, 1997

Abstract

In normal-form games, rationalizability (Bernheim [3], Pearce [11]) on its own fails to exclude some very implausible strategy choices. Three main refinements of rationalizability have been proposed in the literature: cautious, perfect, and proper rationalizability. Nevertheless, some of these refinements also fail to eliminate unreasonable outcomes and suffer from several drawbacks. Therefore, we introduce the trembling-hand rationalizability concept, where the players' actions have to be best responses also against perturbed conjectures. We also propose another refinement: weakly perfect rationalizability, where players' actions that are not best responses are only played with a very small probability.

We show the relationship between perfect rationalizability and weakly perfect rationalizability as well as the relationship between proper rationalizability and weakly perfect rationalizability : weakly perfect rationalizability is a weaker refinement than both perfect and proper rationalizability. Moreover, in two-player games it holds that weakly perfect rationalizability is a weaker refinement than trembling-hand rationalizability. The other relationships between the various refinements are illustrated by means of examples. For the relationship between any other two refinements we give examples showing that the remaining set of strategies corresponding to the first refinement can be either smaller or larger than the one corresponding to the second refinement.

JEL Classification: C72; Keywords: rationalizability, refinements.

[*] We would like to thank Eric van Damme, Claude d'Aspremont, Pierre Dehez, Françoise Forges and Hans Peters for helpful comments and suggestions. The usual disclaimer applies. This paper is an extended version of Vannetelbosch [16]. P.J.J. Herings is financially supported by the Netherlands Organization for Scientific Research (NWO), grant PPS-96-010.

[†] Corresponding author: CORE, 34 voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium.

1 Introduction

A notion like the Nash equilibrium assumes common expectations of the players' behaviour. That is, each player holds a correct conjecture about her opponents' strategy choice. But once we admit the possibility that a player may have several strategies that she could reasonably use, conjectures and strategies actually played may be mismatched. This is what distinguishes rationalizability (Bernheim [3], Pearce [11]) from equilibrium concepts.

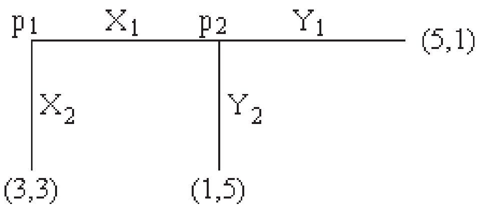
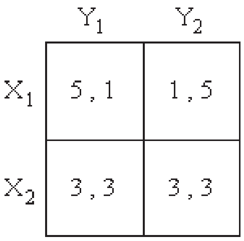


Figure 1: The extensive form of G1

But rationalizability for normal-form games on its own fails to exclude some implausible strategy choices. Consider the following game in extensive-form given in Figure 1. At the beginning of the game G1, player 1 chooses between her action X_2 or letting player 2 decide on one of the feasible outcomes: $5-1$ and $1-5$ (action X_1). Then, if the game has not ended, player 2 chooses between the outcome $5-1$ (action Y_1) and the outcome $1-5$ (action Y_2). The normal-form of G1 is given in Figure 2. It can be shown that $\{(X_1, Y_1), (X_1, Y_2), (X_2, Y_1), (X_2, Y_2)\}$ are rationalizable; in other words, all pure strategies are possible best responses and rationalizable in G1. But once we look at the extensive-form of G1 (see Figure 1), player 2's action Y_2 is an optimal action for him whenever the subgame is reached, while Y_1 is not a credible choice (Y_1 is strictly dominated in the subgame which starts with player 2's move). Therefore, (X_2, Y_2) is the only plausible rationalizable choice.

To avoid unreasonable outcomes, three main refinements of rationalizability have been proposed in the literature: perfect rationalizability (Bernheim [3]), proper rationalizability (Schuhmacher [12]), and cautious rationalizability (Pearce [11]). We also propose another refinement: weakly perfect rationalizability, where players' actions that are not best responses are only played with a very small probability. Nevertheless, these refinements may also fail to exclude implausible outcomes and suffer from some drawbacks: adding



	Y_1	Y_2
X_1	5 , 1	1 , 5
X_2	3 , 3	3 , 3

Figure 2: A two-player game: G1

dominated strategies may enlarge the set of rationalizable strategies, while adding a pure strategy which was already available as a mixed strategy may reduce the set of rationalizable strategies. To remedy these drawbacks, we introduce another refinement, the trembling-hand rationalizability concept, where players' actions have to be best responses also against perturbed conjectures.

The main results of the paper are as follows. We show the relationship between perfect rationalizability and weakly perfect rationalizability as well as the relationship between proper rationalizability and weakly perfect rationalizability : weakly perfect rationalizability is a weaker refinement than both perfect and proper rationalizability. Moreover, in two-player games it holds that weakly perfect rationalizability is a weaker refinement than trembling-hand rationalizability. The other relationships between the various refinements are illustrated by means of examples. For the relationship between any other two refinements we give examples showing that the remaining set of strategies corresponding to the first refinement can be either smaller or larger than the one corresponding to the second refinement. Finally, we also show that all these refinements of rationalizability possess the so-called pure strategy property.

The paper is organized as follows. In Section 2 the rationalizability concept is presented. Section 3 is devoted to the refinements. We derive some generally holding relationships there. Section 4 shows by means of examples that there are no other relationships between the refinements of rationalizability for normal-form games than the ones derived in Section 3.

2 Rationalizability

We consider a normal-form game $\Gamma(I, S, U)$, where I is a finite set of players. Each player $i \in I$ has a finite pure-strategy set S_i. We denote by $S \equiv \prod_{i \in I} S_i$ the Cartesian product set of strategy profiles. Let $U = (U_i)_{i \in I}$ be a list of all players' payoff functions $U_i : S \to \mathbb{R}$ that give player i's vN-M utility $U_i(s)$ for each strategy profile $s \in S$. Let M_i be the set of player i's mixed strategies; M_i is the set of all possible probability distributions over S_i. A mixed strategy $c_i \in M_i$ is a probability distribution over pure strategies. We denote by $c_i\left(s_i\right)$ the probability that c_i assigns to s_i. Let $M \equiv \prod_{i \in I} M_i$ be the set of mixed strategy profiles; where $c \in M$ is a mixed strategy profile. The support of a mixed strategy c_i is the set of pure strategies to which c_i assigns positive probability. A mixed strategy profile c gives rise to an expected payoff for each player. Let $U_i\left(c\right)$ be player i's expected payoff to strategy profile c, which is equal to $\sum_{s \in S}\left(\prod_{j \in I} c_j\left(s_j\right)\right) U_i\left(s\right)$. Player i's opponents in the game $\Gamma\left(I, S, U\right)$ are denoted by $-i$.

As general notation, given any set X, we denote by $\mathrm{ch}(X)$ the convex hull of the set X, i.e. the smallest convex set containing X.

Rationalizability (Bernheim [3], Pearce [11]) for normal-form games is based on the following assumptions: (**A1**) the players are rational, (**A2**) **A1** is common knowledge among the players, and (**A3**) the structure of the game (strategy sets, payoff functions) is common knowledge. Our formulation of rationality is based on expected utility maximization given uncorrelated[1] conjectures about the opponents' strategies.

Definition 1 *A strategy $c_i \in M_i$ is rational if there exists a conjecture $c_{-i} = \left(c_j\right)_{j \in I \setminus \{i\}} \in \prod_{j \neq i} M_j$ such that $\forall\ c_i^{'} \in M_i : U_i\left(c_i, c_{-i}\right) \geq U_i\left(c_i^{'}, c_{-i}\right)$.*

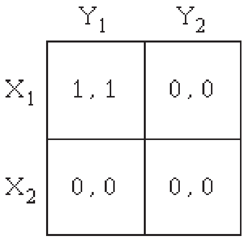
Formally, rationalizability for normal-form games is defined by the following iterative process.

Definition 2 *Let $R^0 \equiv M$. Then $R^k \equiv \prod_{i \in I} R_i^k$ $(k \geq 1)$ is inductively defined as follows: for $i \in I$, $c_i \in R_i^k$ if: (i) $c_i \in M_i$; (ii) $\exists\ c_{-i} \in \prod_{j \neq i} ch\left(R_j^{k-1}\right)$ such that $\forall\ c_i^{'} \in M_i$: $U_i\left(c_i, c_{-i}\right) \geq U_i\left(c_i^{'}, c_{-i}\right)$. The set of rationalizable mixed strategy profiles is the limit set $R^{\infty} \equiv \lim_{k \to \infty} R^k = \bigcap_{k=0}^{\infty} R^k$.*

For the purposes of expected utility calculations[2], Pearce [11] has shown that a conjecture over R_j^{k-1} can be regarded as an element of $\mathrm{ch}\left(R_j^{k-1}\right)$. Part (ii) in Definition 2 means that player i holds uncorrelated conjectures (or beliefs) about her opponents' strategies. Remark that in Definition 2, $\left\{R_i^k; k \geq 0\right\}$ is a weakly decreasing sequence, i.e. $R_i^{k+1} \subseteq R_i^k$, $\forall\ k \in \mathbb{N}$, $\forall\ i \in I$. Bernheim [3] and Pearce [11] have shown that, $\forall\ i \in I$, the limit set $R_i^{\infty} \equiv \lim_{k \to \infty} R_i^k = \bigcap_{k=0}^{\infty} R_i^k$ is nonempty and closed, and that the sequence converges in

[1] Correlated rationalizability, introduced by Brandenburger and Dekel [6], weakens rationalizability because allowing correlated conjectures about the strategies of the opponents makes more strategies rationalizable. In the paper, we only consider the case where the players hold uncorrelated conjectures.

[2] The convex hull operator is used in Definition 2 because, when player i holds a conjecture about which strategies belonging to R_j^{k-1} player j will use, it may be that, although both mixed strategies $c_j^{'}$ and $c_j^{''}$ are in R_j^{k-1}, the mixture $c_j^{'''} = \frac{1}{2}c_j^{'} + \frac{1}{2}c_j^{''}$ is not.



	Y_1	Y_2
X_1	1, 1	0, 0
X_2	0, 0	0, 0

Figure 3: A two-player game: G2

a finite number of steps. Moreover, Pearce [11] has shown that, $\forall\ k \in \mathbb{N}$, $\forall\ i \in I$, the set R_i^k has the pure strategy property. Let $SR_i^k \equiv \left\{ s_i \in S_i \mid c_i\left(s_i\right) > 0 \text{ for some } c_i \in R_i^k \right\}$.

Definition 3 *R_i^k has the pure strategy property if $SR_i^k = \{s_i \in S_i \mid c_i(s_i) = 1$ for some $c_i \in R_i^k \big\}$.*

That is, R_i^k has the pure strategy property if $c_i \in R_i^k$ implies that every pure strategy given positive weight by c_i is also in R_i^k; the set SR_i^k coincides with the set of pure strategies in R_i^k. By definition, $\forall\ i \in I$, M_i has the pure strategy property. Since, $\forall\ k \in \mathbb{N}$, $\forall\ i \in I$, the set R_i^k has the pure strategy property, it follows that the set of rationalizable strategies, R_i^∞, contains at least one pure strategy for each player. Note that all Nash equilibrium strategies are rationalizable; therefore, every strategy which is used with positive probability in some Nash equilibrium must be rationalizable.

Theorem 1 *For every game in normal-form we have, $\forall\ i \in I$, $R_i^\infty = \bigcap_{k=0}^\infty R_i^k \neq \emptyset$, the limit set R_i^∞ is closed and satisfies the pure strategy property $\forall i \in I$, and there exists $n \in \mathbb{N}$ such that: $R_i^{k+1} = R_i^k, \quad \forall k \geq n, \forall\ i \in I$.*

We denote by R^k the set of k-step rationalizable mixed strategy profiles, i.e. the set of mixed strategy profiles which survive k rounds of iteration.

Consider the two-player normal-form game G2 (see Figure 3) from Pearce [11]. This game possesses two pure Nash equilibria: $\{(X_1, Y_1), (X_2, Y_2)\}$. Then, it is straightforward that $SR^\infty = \{(X_1, Y_1), (X_1, Y_2), (X_2, Y_1), (X_2, Y_2)\}$; i.e. all pure strategy profiles are rationalizable. Nonetheless, the pure strategy profiles (X_1, Y_2), (X_2, Y_1) and (X_2, Y_2) seem unreasonable: these profiles are weakly dominated by the profile (X_1, Y_1). Moreover, the outcomes associated to these profiles are risk dominated and Pareto dominated by the outcome associated to (X_1, Y_1).

To exclude these unreasonable outcomes, we consider various refinements of the rationalizability concept, all of which require a rationalizable strategy profile to satisfy some particular robustness condition.

3 Refinements of Rationalizability

3.1 Perfect Rationalizability

Perfect rationalizability is due to Bernheim [3]. The idea behind the perfectness notion is that each player with a small probability makes mistakes, which has the consequence that every pure strategy is chosen with a positive (although possibly small) probability. Börgers [4, p.274] has given the following informal definition[3].

> Consider any finite normal-form game. Assume that every player has to choose each of his pure strategies with a certain strictly positive minimum probability. Assume that the minimum probabilities are common knowledge. Then apply *rationalizability* to this perturbed game. Strategies are *perfectly rationalizable* if they are the limit of rationalizable strategies in perturbed games as the minimum probabilities in these perturbed games converge to zero.

Let $\mathrm{int}\,(M_i)$ denote the interior of M_i. The mixed strategies in the subset $\mathrm{int}\,(M_i) \subseteq M_i$ are called interior or completely mixed strategies of player i. These mixed strategies assign positive probabilities to all pure strategies of player i. We denote by $M_i\,(\mu)$ the set of strategies of player i that assign probabilities of at least $\mu > 0$ to all pure strategies of player i; $M_i\,(\mu) \equiv \{c_i \in \mathrm{int}\,(M_i) \mid c_i\,(s_i) \geq \mu,\ \forall\ s_i \in S_i\}$. That is, $M_i\,(\mu) \subseteq \mathrm{int}\,(M_i) \subseteq M_i$. Formally, perfect rationalizability for normal-form games is defined by the following iterative procedure[4].

Definition 4 *Let $B^0\,(\mu) \equiv \prod_{i \in I} M_i\,(\mu)$. Then $B^k\,(\mu) \equiv \prod_{i \in I} B_i^k\,(\mu)$ $(k \geq 1)$ is inductively defined as follows: for $i \in I$, $c_i \in B_i^k\,(\mu)$ if: (i) $c_i \in M_i\,(\mu)$; (ii) $\exists\ c_{-i} \in \prod_{j \neq i} ch\left(B_j^{k-1}\,(\mu)\right)$ such that $\forall\ c_i^{'} \in M_i\,(\mu)$: $U_i\,(c_i, c_{-i}) \geq U_i\left(c_i^{'}, c_{-i}\right)$. The set of perfectly rationalizable strategy profiles is the limit set $B^\infty \equiv \lim_{\mu \to 0^+} B^\infty\,(\mu)$ where $B^\infty\,(\mu) \equiv \lim_{k \to \infty} B^k\,(\mu) = \bigcap_{k=0}^{\infty} B^k\,(\mu)$.*

In Definition 4, the limit set B^∞ is given by

$$\lim_{\mu \to 0^+} B^\infty\,(\mu) = \left\{c \in M \mid \exists \left\{\mu^t\right\}_{t=0}^{\infty} \to 0^+,\ \exists \left\{c^t\right\}_{t=0}^{\infty} \to c,\ c^t \in B^\infty\left(\mu^t\right)\right\}.$$

[3] Börgers [4] has shown that it is approximate common knowledge that the players maximize expected utility using full support conjectures (with also correlated conjectures allowed) if and only if they play strategies that survive the procedure which begins with one round of elimination of weakly dominated strategies and continues with iterated elimination of strictly dominated strategies.

[4] Our definition of perfect rationalizability is slightly different from Bernheim's definition: we assume that the minimum probabilities are the same for all pure strategies and for all players. It is straightforward that every perfectly rationalizable strategy profile $c \in B^\infty$ is perfectly rationalizable in the sense of Bernheim's definition. And one can verify that all the relationships we derive and counterexamples that we give, would still be valid if one uses Bernheim's definition.

It is easy to show that for every $\mu > 0$ the set $B^{\infty}(\mu)$ is nonempty and compact, from which it follows easily that the set B^{∞} is nonempty and compact. Moreover, it is not difficult to show that, $\forall\ k \in \mathbb{N}$, $\forall\ i \in I$, the set $B_i^k(\mu)$ has the pure strategy property, defined as follows. Let

$$M_i^k(\mu) \equiv \left\{c_i \in M_i(\mu) \mid \text{for some } c_i' \in B_i^k(\mu),\ c_i'(s_i) > \mu,\ \text{and } c_i\left(s_i\right) = 1 - \left(\#S_i - 1\right)\mu\right\}.$$

Definition 5 *$B_i^k(\mu)$ has the pure strategy property if $M_i^k(\mu) \subseteq B_i^k(\mu)$.*

That is, $B_i^k(\mu)$ has the pure strategy property if $c_i' \in B_i^k(\mu)$ and s_i is a pure strategy that is given weight exceeding μ by c_i' implies that the strategy c_i where c_i gives the minimum probability μ to all pure strategies $s_i' \neq s_i$ and the maximal probability $1 - (\#S_i - 1)\,\mu$ to pure strategy s_i is also in $B_i^k(\mu)$. We chose to retain the name pure strategy property since, for μ small, the strategy c_i is indeed close to the pure strategy s_i. For limit sets we can always employ the Definition of the pure strategy property as given in Definition 3. This also applies to limit sets of refinements to be discussed later in the paper.

Theorem 2 *For every game in normal-form we have, $\forall\ i \in I$, $B_i^{\infty} = \lim_{\mu \to 0^+} B_i^{\infty}(\mu) \neq \emptyset$. Moreover, the limit set B_i^{∞} is closed and satisfies the pure strategy property $\forall i \in I$.*

We will denote the pure strategies that are approximately in $B_i^k(\mu)$ by $SB_i^k(\mu)$, so $SB_i^k(\mu) = \{s_i \in S_i \mid \exists\ c_i \in B_i^k(\mu) \text{ with } c_i(s_i) = 1 - (\#S_i - 1)\,\mu\}$.

Note that all pure strategies that are played with positive probability in a uniformly perfect equilibrium, the equilibrium concept used for instance in Harsanyi and Selten [8], are perfectly rationalizable.

3.2 Weakly Perfect Rationalizability

Unlike the perfect rationalizability concept, in the weakly perfect rationalizability concept, a player is not required to optimize against her conjecture subject to an explicit constraint on minimum weights, but her conjecture must put less than ε weight on strategies that are not best responses[5]. Formally, we define weakly perfect rationalizability by the following iterative procedure.

Definition 6 *Let $D^0(\varepsilon) \equiv \prod_{i\in I} \operatorname{int}(M_i)$. Then $D^k(\varepsilon) \equiv \prod_{i\in I} D_i^k(\varepsilon)$ ($k \geq 1$) is inductively defined as follows: for $i \in I$, $c_i \in D_i^k(\varepsilon)$ if: (i) $c_i \in \operatorname{int}(M_i)$; (ii) $\exists\ c_{-i} \in \prod_{j\neq i} ch\left(D_j^{k-1}(\varepsilon)\right)$ such that: $\forall\ s_i, s_i^{'} \in S_i$: $U_i\left(s_i^{'}, c_{-i}\right) < U_i\left(s_i, c_{-i}\right) \Rightarrow c_i\left(s_i^{'}\right) \leq \varepsilon$.*

[5] This non-conventional way to optimize has been introduced by Myerson [10].

The set of weakly perfectly rationalizable strategy profiles is $D^{\infty} \equiv \lim_{\varepsilon \to 0^+} D^{\infty}(\varepsilon)$ *where* $D^{\infty}(\varepsilon) \equiv \lim_{k \to \infty} D^{k}(\varepsilon) = \bigcap_{k=0}^{\infty} D^{k}(\varepsilon)$.

The set D^{∞} could equivalently have been defined as $D^{\infty} \equiv \lim_{\varepsilon \to 0^+} \mathrm{cl}\left(D^{\infty}(\varepsilon)\right)$. Since the sets $\mathrm{cl}\left(D^{\infty}(\varepsilon)\right)$ are easily seen to be nonempty and compact, it follows that the set D^{∞} is nonempty and closed. Again, it is not difficult to show that, $\forall\ k \in \mathbb{N}$, $\forall\ i \in I$, the set $D_i^k(\varepsilon)$ has the pure strategy property. Let

$$\tilde{M}_i^k(\varepsilon) \equiv \left\{ c_i \in \mathrm{int}(M_i) \mid \text{ for some } c_i' \in D_i^k(\varepsilon),\ c_i'(s_i) > \varepsilon, \text{ and } c_i(s_i') \leq \varepsilon,\ \forall s_i' \neq s_i \right\}.$$

Definition 7 $D_i^k(\varepsilon)$ *has the pure strategy property if* $\tilde{M}_i^k(\varepsilon) \subseteq D_i^k(\varepsilon)$.

That is, $D_i^k(\varepsilon)$ has the pure strategy property if $c_i' \in D_i^k(\varepsilon)$ and s_i is a strategy that is given weight exceeding ε by c_i' implies that every strategy c_i where c_i gives weight less than or equal to ε to all pure strategies $s_i' \neq s_i$ and the remaining probability to pure strategy s_i is also in $D_i^k(\varepsilon)$. Summarizing, we have the following result.

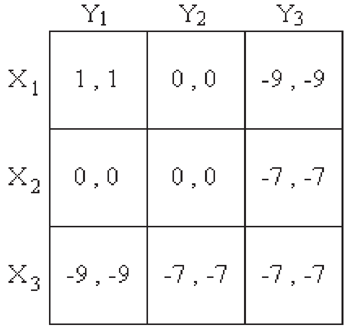
Theorem 3 *For every game in normal-form we have,* $\forall\ i \in I$, $D_i^{\infty} = \lim_{\varepsilon \to 0^+} D_i^{\infty}(\varepsilon) \neq \emptyset$. *Moreover, the limit set* D_i^{∞} *is closed and satisfies the pure strategy property* $\forall i \in I$.

We will denote the pure strategies that are approximately in $D_i^k(\varepsilon)$ by $SD_i^k(\varepsilon)$, so $SD_i^k(\varepsilon) = \left\{ s_i \in S_i \mid \exists\ c_i \in D_i^k(\varepsilon) \text{ with } c_i(s_i') = \varepsilon,\ \forall s_i' \neq s_i \right\}$. Clearly, all pure strategies that are played with positive probability in a perfect equilibrium as defined in van Damme [15] are weakly perfectly rationalizable.

Reconsider the two-player normal-form game G2 (see Figure 3). Remember that the pure strategy profiles (X_1, Y_2), (X_2, Y_1) and (X_2, Y_2), which seem unreasonable, are rationalizable. Nevertheless, none of these pure strategy profiles are perfectly or weakly perfectly rationalizable. Indeed, it is obvious that $D^1(\varepsilon) = D^k(\varepsilon)$ for all $k > 1$ and that $D^1(\varepsilon)$ is such that for all $(c_1, c_2) \in D_1^1(\varepsilon) \times D_2^1(\varepsilon) : c_1(X_2) \leq \varepsilon$ and $c_2(Y_2) \leq \varepsilon$. Then, the set of weakly perfectly rationalizable strategy profiles is $D^{\infty} \equiv \lim_{\varepsilon \to 0^+} D^{\infty}(\varepsilon) = \{(X_1, Y_1)\}$. Therefore, there is a unique strategy profile which survives this refinement (or perfect rationalizability) and it is (X_1, Y_1), i.e. the weakly dominant profile.

Theorem 4 *Every perfectly rationalizable strategy profile is weakly perfectly rationalizable.*

Proof. We have to show that, $\forall\ i \in I$, $c_i \in B_i^{\infty} \Rightarrow c_i \in D_i^{\infty}$. Remark that $B_i^0(\mu) \equiv \{c_i \in \mathrm{int}(M_i) \mid c_i(s_i) \geq \mu \quad \forall\ s_i \in S_i\}$ and $D_i^0(\varepsilon) = \mathrm{int}(M_i)$; therefore, $B_i^0(\mu) \subseteq D_i^0(\varepsilon)\ \forall\ i \in I$. Suppose $\mu = \varepsilon$ in Definitions 4 and 6, and let $\overline{\varepsilon} = \left(\max_{i \in I}(\#S_i)\right)^{-1}$. It is quite easy to show by induction on k that, $\forall\ k \in \mathbb{N}$, we have



	Y1	Y2	Y3
X1	1 , 1	0 , 0	-9 , -9
X2	0 , 0	0 , 0	-7 , -7
X3	-9 , -9	-7 , -7	-7 , -7

Figure 4: A two-player game: G3

that $B_i^k(\varepsilon) \subseteq D_i^k(\varepsilon) \quad \forall\ i \in I, \forall\ \varepsilon \in (0, \overline{\varepsilon})$. Assume $B_i^{k-1}(\varepsilon) \subseteq D_i^{k-1}(\varepsilon) \quad \forall\ i \in I$, $\forall\ \varepsilon \in (0, \overline{\varepsilon})$. If $c_i \in B_i^k(\varepsilon)$ [it means that (i) and (ii) in Definition 4 are satisfied, and that all pure strategies $s_i \in S_i$ which are not best responses against any $c_{-i} \in \prod_{j \neq i} ch\left(B_j^{k-1}(\varepsilon)\right)$ are assigned a probability of ε by all $c_i \in B_i^k(\varepsilon)$; i.e. $c_i(s_i) = \varepsilon$] then, since $\prod_{j \neq i} ch\left(D_j^{k-1}(\varepsilon)\right) \supseteq \prod_{j \neq i} ch\left(B_j^{k-1}(\varepsilon)\right)$, it is straightforward that (i), (ii) and (iii) are also satisfied in Definition 6; therefore $c_i \in D_i^k(\varepsilon)$. Thus, $\forall\ k \in \mathbb{N}$, we have that $B_i^k(\varepsilon) \subseteq D_i^k(\varepsilon) \quad \forall\ i \in I, \forall\ \varepsilon \in (0, \overline{\varepsilon})$; it implies that $B_i^\infty(\varepsilon) = \bigcap_{k=0}^\infty B_i^k(\varepsilon) \subseteq \bigcap_{k=0}^\infty D_i^k(\varepsilon) = D_i^\infty(\varepsilon) \quad \forall\ i \in I, \forall\ \varepsilon \in (0, \overline{\varepsilon})$ [note that since each set S_i is finite, the limits $B_i^\infty(\varepsilon)$ and $D_i^\infty(\varepsilon)$ are reached after a finite number of iterations, $\forall\ i \in I$]. Taking the limit $\varepsilon \to 0^+$, we have $B_i^\infty \subseteq D_i^\infty \quad \forall\ i \in I$. ■

Theorem 4 would still be true if Bernheim's definition of perfect rationalizability would have been used. In Section 4 we will give an example showing that the converse of Theorem 4 is not necessarily true. There exist games where the set of perfectly rationalizable strategy profiles is a proper subset of the set of weakly perfectly rationalizable ones, even if Bernheim's weaker definition of rationalizability approach is used. This is in contrast with the equilibrium approach, where perfect equilibrium can be defined in either way.

Consider now the example G3, taken from Myerson [10], which highlights how the perfectness notion (perfect or weakly perfect rationalizability) fails to eliminate all intuitively unreasonable outcomes. As in G2, (X_1, Y_1) would seem like the obvious outcome for the game G3. There are three Nash equilibria, and all are in pure strategies; these equilibria are (X_1, Y_1), (X_2, Y_2), and (X_3, Y_3). Of these three Nash equilibria, (X_3, Y_3) is not perfect nor proper, (X_2, Y_2) is perfect but not proper, and (X_1, Y_1) is both perfect and proper. The strategy profile (X_2, Y_2) is also weakly perfectly rationalizable. Indeed, $D^1(\varepsilon)$ is such that for all $(c_1, c_2) \in D_1^1(\varepsilon) \times D_2^1(\varepsilon) : c_1(X_3) \leq \varepsilon$ and $c_2(Y_3) \leq \varepsilon$. Then, player 1 may hold the following conjecture c_2 such that $c_2(Y_1) = \varepsilon$, $c_2(Y_2) = 1 - 2\varepsilon$, $c_2(Y_3) = \varepsilon$. Given this conjecture, her best response is to play X_2; indeed, $\forall\ \varepsilon \in (0,1)$ we have that $U_1(X_1, c_2) = -8\varepsilon < U_1(X_2, c_2) = -7\varepsilon$. Player 2 may hold a conjecture c_1 such that $c_1(X_1) = \varepsilon$, $c_1(X_2) = 1-2\varepsilon$, $c_1(X_3) = \varepsilon$. Given this conjecture, his best response is to play Y_2; indeed, $\forall\ \varepsilon \in (0,1)$ we have that $U_2(c_1, Y_1) = -8\varepsilon < U_2(c_1, Y_2) = -7\varepsilon$. Then, it is quite straightforward that $D^\infty \equiv \lim_{\varepsilon\to 0^+} D^\infty(\varepsilon) \supset \{(X_1,Y_1), (X_1,Y_2), (X_2,Y_1), (X_2,Y_2)\}$, $X_3 \notin D_1^\infty$, and $Y_3 \notin D_2^\infty$. In fact, adding the row X_3 and the column Y_3 to the game G2 has converted (X_2, Y_2) into an weakly perfectly rationalizable strategy profile, even though X_3 and Y_3 are weakly dominated strategies. Proper rationalizability is a refinement which deletes such unreasonable outcomes, like (X_1, Y_2), (X_2, Y_1), and (X_2, Y_2).

3.3 Proper Rationalizability

Schuhmacher [12] has developed the proper rationalizability concept which assumes that it is common knowledge that every player satisfies the ε-proper trembling condition, but the players have still no common expectations about the strategies of the opponents[6]. The ε-proper trembling condition requires that every player trembles in a more or less rational way[7]. That is, the players make more costly mistakes with a much smaller probability than less costly ones. Formally, given some $\varepsilon > 0$, a player i satisfies the ε-proper trembling condition if, given her conjecture $c_{-i} \in \prod_{j\neq i}\text{int}(M_j)$, she plays a completely mixed strategy $c_i \in \text{int}(M_i)$, that satisfies

$$\forall\ s_i, s_i^{'} \in S_i : U_i\left(s_i^{'}, c_{-i}\right) < U_i(s_i, c_{-i}) \Rightarrow c_i\left(s_i^{'}\right) \leq \varepsilon\ c_i(s_i)$$

Schuhmacher has shown that the common knowledge among the players of the ε-proper trembling condition implies that every player plays a strategy which survives the following procedure.

Definition 8 *Let $A^0(\varepsilon) \equiv \prod_{i\in I}\text{int}(M_i)$. Then $A^k(\varepsilon) \equiv \prod_{i\in I} A_i^k(\varepsilon)$ $(k \geq 1)$ is inductively defined as follows: for $i \in I$, $c_i \in A_i^k(\varepsilon)$ if: (i) $c_i \in \text{int}(M_i)$; (ii) $\exists\ c_{-i} \in \prod_{j\neq i} ch\left(A_j^{k-1}(\varepsilon)\right)$ such that $\forall\ s_i, s_i^{'} \in S_i : U_i\left(s_i^{'}, c_{-i}\right) < U_i(s_i, c_{-i}) \Rightarrow c_i\left(s_i^{'}\right) \leq \varepsilon\ c_i(s_i)$. The set of properly rationalizable strategy profiles is $A^\infty \equiv \lim_{\varepsilon\to 0^+} A^\infty(\varepsilon)$ where $A^\infty(\varepsilon) \equiv \lim_{k\to\infty} A^k(\varepsilon) = \bigcap_{k=0}^\infty A^k(\varepsilon)$.*

Part (iii) in the definition is the ε-proper trembling condition: if strategy $s_i^{'}$ is worse than strategy s_i against her conjecture c_{-i} about the behaviour of her opponents, then the

[6] The properness notion has been first introduced by Myerson [10], in the equilibrium approach, to refine the perfect equilibrium concept due to Selten [13]. Schuhmacher [12] has shown that proper rationalizability implies the backward induction outcome for generic extensive-form games with perfect information.

[7] The basic idea underlying the properness notion is that a player, although making mistakes, will try much harder to prevent the more costly mistakes than she will try to prevent the less costly ones; i.e. there is an element of rationality in the *mistake technology*.

probability of strategy $s_i^{'}$ is at most ε times the probability of strategy s_i. Schuhmacher [12] has shown that the limit set $\bigcap_{k=0}^{\infty} A_i^k(\varepsilon)$ contains at least the ε-proper equilibria. From Myerson [10], we have that for every normal-form game $\Gamma(I, S, U)$ there exists an ε-proper equilibrium. Therefore, the set of properly rationalizable strategy profiles is nonempty. Again, it is possible to define a pure-strategy property, although the definition becomes a bit more artificial in the case of proper rationalizability.

Definition 9 *$A_i^k(\varepsilon)$ has the pure strategy property if $c_i' \in A_i^k(\varepsilon)$ and $c_i'(s_i) > \varepsilon$ implies there exists $c_i \in A_i^k(\varepsilon)$ such that $c_i(s_i') \leq \varepsilon$, $\forall s_i' \neq s_i$.*

That is, $A_i^k(\varepsilon)$ has the pure strategy property if $c_i' \in A_i^k(\varepsilon)$ and s_i is a strategy that is given weight exceeding ε by c_i' implies that some strategy c_i where c_i gives weight less than or equal to ε to all pure strategies $s_i' \neq s_i$ and the remaining probability to pure strategy s_i is also in $A_i^k(\varepsilon)$. Notice that, opposite to the definition of the pure strategy property as defined for weakly perfect rationalizability, it can now no longer be required that every strategy c_i that gives weight less than or equal to ε to all pure strategies $s_i' \neq s_i$ and the remaining probability to pure strategy s_i also belongs to $A_i^k(\varepsilon)$.

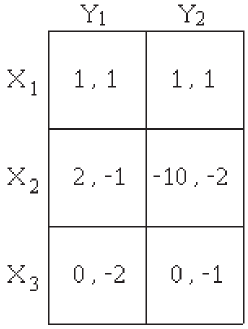
Theorem 5 *For every game in normal-form we have, $\forall\ i \in I$, $A_i^{\infty} = \lim_{\varepsilon \to 0^+} A_i^{\infty}(\varepsilon) \neq \emptyset$. Moreover, the limit set A_i^{∞} is closed and satisfies the pure strategy property $\forall i \in I$.*

We will denote the pure strategies that are approximately in $A_i^k(\varepsilon)$ by $SA_i^k(\varepsilon)$, so $SA_i^k(\varepsilon) = \left\{ s_i \in S_i \mid \exists\ c_i \in A_i^k(\varepsilon) \text{ with } c_i(s_i') \leq \varepsilon,\ \forall s_i' \neq s_i \right\}$.

Theorem 6 *Every properly rationalizable strategy profile is weakly perfectly rationalizable.*

> **Proof.** We have to show that, $\forall\ i \in I$, $c_i \in A_i^{\infty} \Rightarrow c_i \in D_i^{\infty}$. It is quite easy to show by induction on k that, $\forall\ k \in \mathbb{N}$, we have that $A_i^k(\varepsilon) \subseteq D_i^k(\varepsilon) \quad \forall\ i \in I$, $\forall\ \varepsilon \in (0,1)$. Remark that $A_i^0(\varepsilon) = D_i^0(\varepsilon) \quad \forall\ i \in I$, $\forall\ \varepsilon \in (0,1)$. Assume $A_i^{k-1}(\varepsilon) \subseteq D_i^{k-1}(\varepsilon) \quad \forall\ i \in I$, $\forall\ \varepsilon \in (0,1)$. If $c_i \in A_i^k(\varepsilon)$ [it means that (i), (ii) and (iii) in Definition 8 are satisfied] then, since $\varepsilon\, c_i(s_i) \leq \varepsilon$ and $\prod_{j \neq i} ch\left(D_j^{k-1}(\varepsilon)\right) \supseteq \prod_{j \neq i} ch\left(A_j^{k-1}(\varepsilon)\right)$, it is straightforward that (i), (ii) and (iii) are also satisfied in Definition 6; therefore $c_i \in D_i^k(\varepsilon)$. Thus, $\forall\ k \in \mathbb{N}$, we have that $A_i^k(\varepsilon) \subseteq D_i^k(\varepsilon) \quad \forall\ i \in I$, $\forall\ \varepsilon \in (0,1)$; it implies that $A_i^{\infty}(\varepsilon) = \bigcap_{k=0}^{\infty} A_i^k(\varepsilon) \subseteq \bigcap_{k=0}^{\infty} D_i^k(\varepsilon) = D_i^{\infty}(\varepsilon) \quad \forall\ i \in I$, $\forall\ \varepsilon \in (0,1)$ [note that since each set S_i is finite, the limits $A_i^{\infty}(\varepsilon)$ and $D_i^{\infty}(\varepsilon)$ are reached after a finite number of iterations, $\forall\ i \in I$]. Taking the limit $\varepsilon \to 0^+$, we have $A_i^{\infty} \subseteq D_i^{\infty} \quad \forall\ i \in I$. ■

Reconsider the two-player normal-form game G3 (see Figure 4). The strategy profile (X_1, Y_1) is the unique properly rationalizable strategy profile of G3. Indeed, $A^1(\varepsilon)$ is such that for all $(c_1, c_2) \in A_1^1(\varepsilon) \times A_2^1(\varepsilon) : c_1(X_3) \leq \varepsilon\, c_1(X_2)$ and $c_2(Y_3) \leq \varepsilon\, c_2(Y_2)$. Therefore,

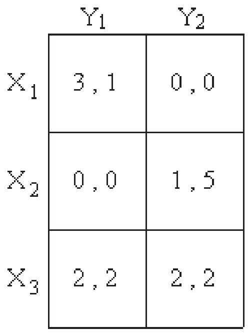


	Y_1	Y_2
X_1	1 , 1	1 , 1
X_2	2 , -1	-10 , -2
X_3	0 , -2	0 , -1

Figure 5: A two-player game: G4

for all $c_1 \in A_1^1(\varepsilon)$ we have that $U_2(c_1, Y_1) > U_2(c_1, Y_3)$; and for all $c_2 \in A_2^1(\varepsilon)$ we have that $U_1(X_1, c_2) > U_1(X_3, c_2)$. This implies that for all $(c_1, c_2) \in A_1^2(\varepsilon) \times A_2^2(\varepsilon) : c_1(X_3) \leq \varepsilon\, c_1(X_2)$, $c_1(X_3) \leq \varepsilon\, c_1(X_1)$, $c_2(Y_3) \leq \varepsilon\, c_2(Y_1)$, and $c_2(Y_3) \leq \varepsilon\, c_2(Y_2)$. Therefore, for all $c_1 \in A_1^2(\varepsilon)$ and $\varepsilon \in \left(0, \frac{1}{2}\right)$ we have that $U_2(c_1, Y_1) > U_2(c_1, Y_2)$; and for all $c_2 \in A_2^2(\varepsilon)$ and $\varepsilon \in \left(0, \frac{1}{2}\right)$ we have that $U_1(X_1, c_2) > U_1(X_2, c_2)$. This implies that for all $(c_1, c_2) \in A_1^3(\varepsilon) \times A_2^3(\varepsilon) : c_1(X_3) \leq \varepsilon\, c_1(X_2)$, $c_1(X_3) \leq \varepsilon\, c_1(X_1)$, $c_1(X_2) \leq \varepsilon\, c_1(X_1)$, $c_2(Y_2) \leq \varepsilon\, c_2(Y_1)$, $c_2(Y_3) \leq \varepsilon\, c_2(Y_1)$, and $c_2(Y_3) \leq \varepsilon\, c_2(Y_2)$. So $c_1(X_2) \leq \varepsilon\, c_1(X_1) \leq \varepsilon$, $c_1(X_3) \leq \varepsilon\, c_1(X_2) \leq \varepsilon^2$, $c_2(Y_2) \leq \varepsilon\, c_2(Y_1) \leq \varepsilon$, $c_2(Y_3) \leq \varepsilon\, c_2(Y_2) \leq \varepsilon^2$. Since the probabilities must sum to one, $c_1(X_1) \geq 1 - \varepsilon - \varepsilon^2$ and $c_2(Y_1) \geq 1 - \varepsilon - \varepsilon^2$. Therefore, we have that $A^\infty \equiv \lim_{\varepsilon \to 0^+} A^\infty(\varepsilon) = \{(X_1, Y_1)\}$; i.e. there is a unique properly rationalizable strategy profile, namely (X_1, Y_1). Thus, although (X_2, Y_2) is perfectly rationalizable for this game G3, it is not properly rationalizable.

Myerson's [10] properness notion was motivated by the fact that the perfectness notion has the drawback that adding dominated strategies may enlarge the set of perfect equilibria. In the non-equilibrium approach, we have shown that perfect and weakly perfect rationalizability may also suffer from this drawback. Nevertheless, van Damme [15] has shown that, for the equilibrium approach, the properness notion may suffer from the same drawback as well. The game G4 is such an example where both the perfectness notion and the properness notion fail to eliminate all intuitively unreasonable outcomes. In Figure 5, we have the normal-form of G4 taken from Pearce [11]. The game G4 has two pure Nash equilibria: $\{(X_2, Y_1), (X_1, Y_2)\}$. In fact, these two Nash equilibria are also trembling-hand perfect equilibria and proper equilibria. It can easily be shown that among the pure strategies, only player 1's action X_3 is not a properly rationalizable one;



	Y1	Y2
X1	3 , 1	0 , 0
X2	0 , 0	1 , 5
X3	2 , 2	2 , 2

Figure 6: A two-player game: G5

$X_3 \notin A_1^\infty$, $\{(X_1, Y_1), (X_1, Y_2), (X_2, Y_1), (X_2, Y_2)\} \subset A^\infty$.

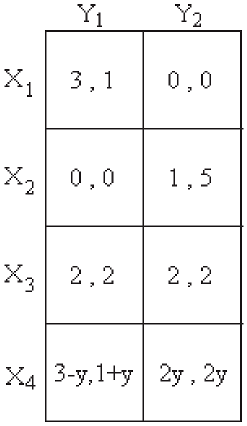
Van Damme [15] has mentioned a second drawback of the properness notion: in the equilibrium approach, the set of proper equilibria may change when a strategy that is already available as a mixed strategy is explicitly added as a pure strategy. This second drawback is illustrated by the games G5 and G6 (see Figures 6 and 7) taken from van Damme [15][8]. The game G6 results from G5 by adding the mixture $X_4 = (1 - y)\, X_1 + y\, X_3$ with $0 < y < 1$. We have that (X_3, Y_2) is a proper equilibrium of G5, but it is not proper in G6 when $y \geq \frac{1}{2}$. This second drawback also applies to proper rationalizability. Indeed, we have that among the pure strategies, only player 1's action X_2 is not a properly rationalizable one in G5; $X_2 \notin A_1^\infty$ and $\{(X_1, Y_1), (X_1, Y_2), (X_3, Y_1), (X_3, Y_2)\} \subset A^\infty$. Nevertheless, proper rationalizability singles out a unique outcome for G6, namely the strategy profile (X_1, Y_1), when $y \geq \frac{1}{2}$. These two drawbacks motivate us to introduce a further refinement: trembling-hand rationalizability for normal-form games.

3.4 Trembling-Hand Rationalizability

The starting point of trembling-hand rationalizability (THR) is that, in the definition of rationalizability, the rationality concept is strengthened by asking that a player's strategy be optimal not only given her conjecture but also given perturbed conjectures[9]. In the definition of THR, the perturbed conjecture puts weight on each pure strategy which

[8] This second drawback of the properness notion matters if both games G5 and G6 are considered as equivalent games (Kohlberg and Mertens [9] have studied the equivalence of games; see also van Damme [15, pp.259-265]).

[9] This restriction on the best-response correspondence may be interpreted as if the players have some doubt about the strategies played by their opponents.



	Y_1	Y_2
X_1	3 , 1	0 , 0
X_2	0 , 0	1 , 5
X_3	2 , 2	2 , 2
X_4	3-y,1+y	2y , 2y

Figure 7: A two-player game: G6

hasn't yet been deleted. Formally, THR is defined by modifying the iterative procedure of Definition 2.

Definition 10 *Let $T^0 \equiv M$. Then $T^k \equiv \prod_{i \in I} T_i^k$ $(k \geq 1)$ is inductively defined as follows: for $i \in I$, $c_i \in T_i^k$ if: (i) $c_i \in T_i^{k-1}$; (ii) $\exists$ $c_{-i} \in \prod_{j \neq i} ch\left(T_j^{k-1}\right)$ such that: c_j gives positive weight to each pure strategy in T_j^{k-1}, and $\forall$ $c_i^{'} \in T_i^{k-1} : U_i\left(c_i, c_{-i}\right) \geq U_i\left(c_i^{'}, c_{-i}\right)$. The set of trembling-hand rationalizable strategy profiles is $T^\infty \equiv \lim_{k \to \infty} T^k = \bigcap_{k=0}^{\infty} T^k$.*

In Definition 10, the set T_i^1 is the set of player i's *trembling-hand rational* strategies. At each step of the iteration, a strategy c_i of player i has to be a best response against some perturbed conjecture. That is, at step k of the iteration, to belong to T_i^k, a strategy c_i of player i has to be a best response against some perturbed conjecture $c_{-i} = \left(c_j\right)_{j \neq i} \in \prod_{j \neq i} \text{ch}\left(T_j^{k-1}\right)$ where c_j gives positive weight to each pure strategy in T_j^{k-1}. At step k of the iteration, no pure strategy in the set T_j^{k-1} is regarded as completely impossible. In Definition 10, $\left\{T_i^k; k \geq 0\right\}$ is a weakly decreasing sequence, i.e. $T_i^{k+1} \subseteq T_i^k \quad \forall\ k \in \mathbb{N}$, $\forall\ i \in I$. We denote by T^k the set of k-step trembling-hand rationalizable strategy profiles, i.e. the set of mixed strategy profiles which survive k rounds of iteration. The limit set is given by $T_i^\infty \equiv \lim_{k \to \infty} T_i^k = \bigcap_{k=0}^{\infty} T_i^k$, $\forall\ i \in I$. The pure strategy property is defined in the same way as for rationalizability, Definition 3. We will denote the pure strategies in T_i^k by ST_i^k. The proof of the following result goes along the same lines as the proof of Theorem 1 and is therefore omitted.

Theorem 7 *For every game in normal-form we have, $\forall\ i \in I$, $T_i^\infty = \bigcap_{k=0}^{\infty} T_i^k \neq \emptyset$, the limit set T_i^∞ is closed and satisfies the pure strategy property $\forall\ i \in I$, and there exists*

$n \in \mathbb{N}$ such that: $T_i^{k+1} = T_i^k \quad \forall\ k \geq n, \forall\ i \in I$.

In Theorem 4 we have shown that perfectly rationalizable strategy profiles are weakly perfectly rationalizable and in Theorem 6 that properly rationalizable strategy profiles are weakly perfectly rationalizable. Since the ideas lying underneath the trembling-hand rationalizability concept are closely related to those of the weakly perfect and proper rationalizability concepts, we might expect that trembling-hand rationalizable strategy profiles are also weakly perfectly rationalizable, or even that they are properly rationalizable. In fact, the trembling-hand rationalizability concept does not only require bad strategies to be expected with low probability, but even with zero probability. Therefore, Theorem 8 does not come as a surprise.

Theorem 8 *For any finite two-player game in normal-form, every trembling-hand rationalizable strategy profile is weakly perfectly rationalizable.*

Proof. We have to show that, $\forall i \in I$, $c_i \in T_i^\infty \Rightarrow c_i \in D_i^\infty$. Notice that $c_i \in T_i^1$ implies that there is no mixed strategy in M_i which weakly dominates c_i, using Lemma 4 of Pearce [11]. First we will show by induction on k that, $\forall k \in \mathbb{N}$, $ST_i^k \subseteq SD_i^k(\varepsilon)$, $\forall i \in I$, $\forall \varepsilon \in (0, \frac{1}{\max_{i \in I} \# S_i})$. Remark that $ST_i^0 = SD_i^0 = S_i$. Assume $ST_i^{k-1} \subseteq SD_i^{k-1}(\varepsilon)$. If $s_i^1 \in ST_i^k$, then there is $c_j^1 \in \text{ch}(T_j^{k-1})$ such that c_j^1 gives positive weight to each $s_j \in ST_j^{k-1}$, $j \neq i$, and $\forall c_i \in T_i^{k-1}$, $U_i(s_i^1, c_j^1) \geq U_i(c_i, c_j^1)$. Suppose there is $s_i^* \in S_i \setminus ST_i^{k-1}$ such that $U_i(s_i^*, c_j^1) > U_i(s_i^1, c_j^1)$. Without loss of generality s_i^* can be assumed to be a best response to c_j^1. Since $c_j^1 \in \text{ch}(T_j^{k-1}) \subseteq \text{ch}(T_j^{l-1})$, $\forall l \leq k$, it follows that $s_i^* \in ST_i^{k-1}$, a contradiction. Consequently, $U_i(s_i^1, c_j^1) \geq U_i(s_i, c_j^1)$, $\forall s_i \in S_i$. Since $s_i^1 \in ST_i^1$, there is $c_j^2 \in \text{int}(M_j)$ such that $U_i(s_i^1, c_j^2) \geq U_i(s_i, c_j^2)$, $\forall s_i \in S_i$. It follows that $U_i(s_i^1, (1-\varepsilon)c_j^1 + \varepsilon c_j^2) \geq U_i(s_i, (1-\varepsilon)c_j^1 + \varepsilon c_j^2)$, $\forall s_i \in S_i$. Moreover, $(1-\varepsilon)c_j^1 + \varepsilon c_j^2$ is a completely mixed strategy putting weight less than ε on each pure strategy in $S_j \setminus ST_j^{k-1} \supseteq S_j \setminus SD_j^{k-1}(\varepsilon)$, $j \neq i$, where the induction hypothesis is used for the inclusion. Therefore, using that $D_j^{k-1}(\varepsilon)$, $j \neq i$, satisfies the pure strategy property, $(1-\varepsilon)c_j^1 + \varepsilon c_j^2 \in \text{ch}(D_j^{k-1}(\varepsilon))$. So, $c_i^1 \in D_i^k(\varepsilon)$ where $c_i^1(s_i) = \varepsilon$, $\forall s_i \neq s_i^1$, and hence $s_i^1 \in SD_i^k(\varepsilon)$. We have shown that $ST_i^k \subseteq SD_i^k(\varepsilon)$.

Since the sets T_i^k and $D_i^k(\varepsilon)$, $\forall i \in I$, $\forall \varepsilon \in (0, \frac{1}{\max_{i \in I} \# S_i})$, can only change if the sets ST_j^k and $SD_i^k(\varepsilon)$ change, it follows that $\forall k, l \geq m$, where $m = \sum_{i \in I}(\# S_i - 1)$, $\forall i \in I$, $\forall \varepsilon \in (0, \frac{1}{\max_{i \in I} \# S_i})$, $T_i^k = T_i^l$ and $D_i^k(\varepsilon) = D_i^l(\varepsilon)$.

If $c_i' \in T_i^\infty$, then $c_i' \in T_i^k$ with $k \geq m+1$, so there is $c_j^3 \in \text{ch}(T_j^k)$ such that c_j^3 gives positive weight to each $s_j \in ST_j^k$ and $\forall c_i \in T_i^k$, $U_i(c_i', c_j^3) \geq U_i(c_i, c_j^3)$, from which it follows as above that $U_i(c_i', c_j^3) \geq U_i(s_i, c_j^3)$, $\forall s_i \in S_i$. Since $c_i' \in T_i^1$, there is $c_j^4 \in \text{int}(M_j)$ such that $U_i(c_i', c_j^4) \geq U_i(s_i, c_j^4)$, $\forall s_i \in S_i$. It follows that $U_i(c_i', (1-\varepsilon)c_j^3 + \varepsilon c_j^4) \geq U_i(s_i, (1-\varepsilon)c_j^3 + \varepsilon c_j^4)$, $\forall s_i \in S_i$. Moreover, $(1-\varepsilon)c_j^3 + \varepsilon c_j^4$ is a completely mixed strategy putting weight less than ε on each pure strategy in $S_j \setminus ST_j^k \supseteq S_j \setminus SD_j^k(\varepsilon)$. Therefore, using that $D_j^k(\varepsilon)$, $j \neq i$, satisfies the pure strategy property, $(1-\varepsilon)c_j^3 +$

	Y_1	Y_2	Y_3
X_1	2, 1, 1	1, 1, 1	0, 0, 1
X_2	0, 1, 1	1, 1, 1	0, 0, 1
X_3	2, 1, 1	0, 1, 1	0, 0, 1

Z_1

	Y_1	Y_2	Y_3
X_1	1, 1, 1	0, 1, 1	0, 0, 1
X_2	1, 1, 1	2, 1, 1	0, 0, 1
X_3	0, 1, 1	2, 1, 1	0, 0, 1

Z_2

	Y_1	Y_2	Y_3
X_1	1, 1, 0	1, 1, 0	0, 0, 0
X_2	1, 1, 0	1, 1, 0	0, 0, 0
X_3	0, 1, 0	0, 1, 0	2, 0, 0

Z_3

Figure 8: A three-player game: G7

$\varepsilon c_j^4 \in \mathrm{ch}(D_j^k(\varepsilon))$. So, $c_i''(\varepsilon) \in D_i^k(\varepsilon) = D_i^\infty(\varepsilon)$ where $c_i''(\varepsilon)(s_i) = \varepsilon$, if $c'(s_i) = 0$, and $c_i''(\varepsilon)(s_i) = c_i'(s_i)(1 - \#\{s_i' \mid c_i'(s_i') = 0\}\varepsilon)$ if $c'(s_i) \neq 0$. If $\varepsilon \to 0^+$, then $c_i''(\varepsilon) \to c_i'$, so $c_i' \in D_i^\infty$. ■

The proof of Theorem 8 is only valid for the two-player case since it relies on the linearity of $U_i\,(s_i, \cdot)$. Surprisingly, Theorem 8 cannot be generalized to three or more player games as is shown by Game G7 (see Figure 8). It is easily seen that $ST_1^1 = \{X_1, X_2, X_3\}$, $ST_2^1 = \{Y_1, Y_2\}$, and $ST_3^1 = \{Z_1, Z_2\}$. It is not possible in the first iteration to eliminate any pure strategy of player 1, since all strategies of player 1 are equally good against $(c_2, c_3) = ((1/3, 1/3, 1/3), (1/3, 1/3, 1/3))$. In the second iteration it is clearly impossible to eliminate any other pure strategy of player 2 or 3. Against $(c_2, c_3) = ((1/2, 1/2, 0), (1/2, 1/2, 0))$ all pure strategies of player 1 are equally good, so no further eliminations are possible. Consequently, for every $k \geq 1$, $ST_1^k = \{X_1, X_2, X_3\}$, $ST_2^k = \{Y_1, Y_2\}$, and $ST_3^k = \{Z_1, Z_2\}$.

Now we consider the weakly perfect rationalizability concept. Let any ε smaller than 1/3 be given. Obviously, in the first iteration again only the pure strategies Y_3 and Z_3 are eliminated, so $SD_1^1(\varepsilon) = \{X_1, X_2, X_3\}$, $SD_2^1(\varepsilon) = \{Y_1, Y_2\}$, and $SD_3^1(\varepsilon) = \{Z_1, Z_2\}$. In the second iteration, it is again impossible to eliminate any other pure strategy of player 2 or 3. Next we show that pure strategy X_3 of player 1 is eliminated in the second iteration, although it is easily seen that X_3 is not weakly dominated by any mixed strategy. Intuitively, compared to strategies X_1 and X_2, strategy X_3 is good against the conjectures (Y_1, Z_1), (Y_2, Z_2), and (Y_3, Z_3), but bad against all other pure strategy combinations. If every pure strategy is played with at least a small probability, then the pure strategy combinations against which strategy X_3 is bad will necessarily arise with positive probability. It will turn out that against any such conjecture at least one of the pure strategies X_1 and X_2 performs better. Let any $c_2 \in \mathrm{ch}(D_2^1(\varepsilon))$ and any $c_3 \in \mathrm{ch}(D_3^1(\varepsilon))$ be given. To simplify notation, let s and t denote the probability of the first action of player 2 and player 3, respectively, and β and γ the probability of the third action of player 2 and player 3, respectively, so $s = c_2(Y_1)$, $t = c_3(Z_1)$, $\beta = c_2(Y_3) \leq \varepsilon$, and $\gamma = c_3(Z_3) \leq \varepsilon$. Let us consider the payoffs of the pure strategies of player 1.

Strategy	Payoff
X_1	$2st + (1-s-\beta)t + s(1-t-\gamma) + \gamma(1-\beta)$
X_2	$(1-s-\beta)t + s(1-t-\gamma) + 2(1-s-\beta)(1-t-\gamma) + \gamma(1-\beta)$
X_3	$2st + 2(1-s-\beta)(1-t-\gamma) + 2\beta\gamma$

Table 1: The payoffs of the pure strategies of player 1

Pure strategy X_3 is at least as good as pure strategy X_1 if $t(3-3\beta)+s(3-3\gamma)+3\gamma \leq 4st + 5\beta\gamma + 2 - 2\beta$. So, if,

$$3 - 3\gamma - 4t > 0 \text{ and } s \leq \frac{(3\beta-3)t - 3\gamma - 2\beta + 2 + 5\beta\gamma}{3 - 3\gamma - 4t} \tag{1}$$

or

$$3 - 3\gamma - 4t < 0 \text{ and } s \geq \frac{(3\beta-3)t - 3\gamma - 2\beta + 2 + 5\beta\gamma}{3 - 3\gamma - 4t}.$$

If $3 - 3\gamma - 4t = 0$, then X_3 is strictly worse than X_1. Consider the case $3 - 3\gamma - 4t < 0$. It only holds that the right-hand side, i.e. the minimum probability to be put on strategy Y_1, is less than $1-\beta$ if $t > 1 - 2\beta\gamma/(1-\beta)$. But then $t+\gamma > (1-\beta+\gamma-3\beta\gamma)/(1-\beta) > 1$ since $\beta < 1/3$, a contradiction since the sum of t and γ should be strictly less than 1. So only case (1) remains.

Pure strategy X_3 is at least as good as X_2 if $t(1-\beta)+s(1-\gamma)+\gamma \leq 4st+3\beta\gamma$. So, if,

$$1 - \gamma - 4t > 0 \text{ and } s \leq \frac{3\beta\gamma - \gamma - (1-\beta)t}{1 - \gamma - 4t}$$

or

$$1 - \gamma - 4t < 0 \text{ and } s \geq \frac{3\beta\gamma - \gamma - (1-\beta)t}{1 - \gamma - 4t}. \tag{2}$$

If $1 - \gamma - 4t = 0$, then X_3 is strictly worse than action X_2. Consider the case where $1 - \gamma - 4t > 0$. It holds that the numerator of the right-hand side is negative (use that $\beta < 1/3$), a contradiction since s should be positive. So only case (2) remains.

Concluding, X_3 may be a best response of player 1 if

$$1 - \gamma - 4t < 0 < 3 - 3\gamma - 4t$$

and

$$\frac{3\beta\gamma - \gamma - (1-\beta)t}{1 - \gamma - 4t} \leq s \leq \frac{(3\beta-3)t - 3\gamma - 2\beta + 2 + 5\beta\gamma}{3 - 3\gamma - 4t}.$$

Next it is shown that the latter inequality can never be satisfied since the first term is always bigger than the third. Now, $1-\gamma-4t<0<3-3\gamma-4t$ and $(3\beta\gamma-\gamma-(1-\beta)t)/(1-\gamma-4t) \leq ((3\beta-3)t-3\gamma-2\beta+2+5\beta\gamma)/(3-3\gamma-4t)$ implies

$$t^2(4-4\beta)+t(4\beta-4+4\gamma-4\beta\gamma)+1-\beta-\gamma-\beta\gamma+2\beta\gamma^2 \leq 0. \tag{3}$$

The left-hand side of (3) is a quadratic function in t. Computing "b^2-4ac" to find the zero points of this function yields $16\gamma(\gamma-1)(1-4\beta+3\beta^2)$ which is smaller than 0 (use $\beta<1/3$). Therefore, the quadratic function in t has no zero points. By trying any value of the parameters, one sees that the left-hand side of (3) is actually positive everywhere, leading to a contradiction. Consequently, there are no values of s and t, given any $\beta,\gamma<1/3$, for which X_3 is the best response. So, X_3 can be eliminated. After this no further eliminations are possible. Consequently, for every $k\geq 2$, $SD_1^k(\varepsilon)=\{X_1,X_2\}\subset ST_1^k=\{X_1,X_2,X_3\}$, $SD_2^k(\varepsilon)=ST_2^k=\{Y_1,Y_2\}$, and $SD_3^k(\varepsilon)=ST_3^k=\{Z_1,Z_2\}$.

In many games, trembling-hand rationalizability can rule out implausible strategies that cannot be excluded by proper rationalizability (although in Section 4 we show that even for two-player normal-form games trembling-hand rationalizability is not a refinement of proper rationalizability). In Game G4, once we apply our concept THR, we obtain the following iterative deletion of pure strategy profiles: $ST^1=\{(X_2,Y_1),(X_1,Y_2),(X_2,Y_2),(X_1,Y_1)\}$; $ST^2=\{(X_2,Y_1),(X_1,Y_1)\}$; $T^3=\{(X_2,Y_1)\}$. Once player 1 will never play X_3, player 2's action Y_2 is never a best response against any trembling conjecture which puts weight on X_1 and X_2. Therefore, Y_1 is the unique trembling-hand rationalizable strategy of player 2. Knowing that player 2's choice is Y_1, player 1's best response is to play X_2 which is player 1's unique trembling-hand rationalizable strategy. In both G2 and G3, the strategy profile (X_1,Y_1) is the unique trembling-hand rationalizable strategy profile. Game G4 shows that sometimes it is possible to eliminate unreasonable strategies by means of trembling-hand rationalizability which cannot be eliminated by weakly perfect rationalizability, proper rationalizability, or even the proper equilibrium concept since the strategy profile (X_1,Y_2) constitutes a proper equilibrium in Game G4.

For some games the commonality of the knowledge that players are *trembling-hand* rational runs into problems. The following example[10] illustrates this inconsistency. In G8 (see Figure 9), trembling-hand rationalizability singles out the unique strategy profile (X_1,Y_1). Player 1 has two pure strategies; $S_1=\{X_1,X_2\}$. Player 2 has also two pure strategies; $S_2=\{Y_1,Y_2\}$. It is quite obvious that $T_1^1=T_1^\infty=\{X_1\}$ and $T_2^1=T_2^\infty=\{Y_1\}$.

[10] This inconsistency problem has been studied by Börgers and Samuelson [5]. To resolve such an inconsistency of common knowledge of *cautious rationality*, Asheim and Dufwenberg [1] have changed the object for the common knowledge: instead of common knowledge of rational choice, they assume common knowledge of rational reasoning.

	Y_1	Y_2
X_1	10 , 10	10 , 10
X_2	10 , 10	0 , 0

Figure 9: A two-player game: G8

Mutual knowledge of order 1 of *trembling-hand rationality* means that player 1 knows that player 2 will play Y_1. Therefore, player 1 is indifferent in playing X_1 or X_2. Nevertheless, the action X_2 is not trembling-hand rationalizable. The logical problem is: why should player 1 play a trembling-hand rationalizable strategy if player 1 knows that player 2 will play a trembling-hand rationalizable strategy?

Let $\overline{T}^0 \equiv M$. Then $\overline{T}^k \equiv \prod_{i \in I} \overline{T}_i^k$ $(k \geq 1)$ is inductively defined as follows: for $i \in I$, $c_i \in \overline{T}_i^k$ if: (i) $c_i \in M_i$; (ii) $\exists$ $c_{-i} \in \prod_{j \neq i} \mathrm{ch}\left(\overline{T}_j^{k-1}\right)$ such that: (iii) c_j gives positive weight to each pure strategy in $\overline{T}_j^{k-1}$; (iv) $\forall$ $c_i^{'} \in M_i$, $U_i\left(c_i, c_{-i}\right) \geq U_i\left(c_i^{'}, c_{-i}\right)$. Thus each set $\overline{T}_i^k$ consists of unconstrained best responses, while each set T_i^k consists of constrained best responses. The difference between the two iterative procedures that yield the sets $\overline{T}_i^k$ and T_i^k is that, in the kth step of the construction of the sets T_i^k, only trembling-hand best responses among the sets T_i^{k-1} of surviving strategies are considered. Strategies which have already been eliminated at a previous step as not trembling-hand rationalizable should never be readmitted into the set of trembling-hand rationalizable strategies. But a refinement of rationalizability admitting unconstrained best responses runs into problems since the limit set $\overline{T}^\infty \equiv \lim_{k \to \infty} \overline{T}^k$ may not exist. As an example, reconsider briefly Game G8. For player 1, we have that $\overline{T}_1^k = M_1$ if k even and $\overline{T}_1^k = \{X_1\}$ if k odd, whereas, $\forall$ $k > 0$, $T_1^k = \{X_1\}$. For player 2, we have that $\overline{T}_2^k = M_2$ if k even and $\overline{T}_2^k = \{Y_1\}$ if k odd, whereas, $\forall$ $k > 0$, $T_2^k = \{Y_1\}$.

Also the perfect, weakly perfect and proper rationalizability concepts suffer from a similar drawback albeit in a somewhat more disguised form. As long as μ and ε are positive, no problems arise as can be easily seen from the definitions of the sets $B^k(\mu)$, $D^k(\varepsilon)$, and $A^k(\varepsilon)$ (Definitions 4, 6, and 8, respectively). However, for the limit sets B^∞, D^∞, and A^∞ exactly the same problem arises as can be verified by means of Game G8, since for all concepts only the strategies X_1 and Y_1 remain. Again: why should player 1 stick to strategy X_1 if player 1 knows that player 2 will play Y_1?

3.5 Cautious Rationalizability

Cautious rationalizability, due to Pearce [11], imposes the condition on the set of rationalizable strategy profiles that the players do not take unnecessary risks. This condition requires that the players' conjectures give positive weight to all rationalizable alternatives, whereas the strategy profiles that are not rationalizable should be given zero weight. Formally, cautious rationalizability is defined by the following iterative procedure.

Definition 11 *Let $C^0 \equiv M$. Then $C^k \equiv \prod_{i \in I} C_i^k$ $(k \geq 1)$ is inductively defined as follows: for $i \in I$, $c_i \in C_i^k$ if: (i) $c_i \in R_i^\infty \left(C^{k-1}\right)$; (ii) $\exists$ $c_{-i} \in \prod_{j \neq i} ch \left(R_j^\infty \left(C^{k-1}\right)\right)$ such that: c_j gives positive weight to each pure strategy in $R_j^\infty \left(C^{k-1}\right)$, and $\forall$ $c_i^{'} \in R_i^\infty \left(C^{k-1}\right)$: $U_i \left(c_i, c_{-i}\right) \geq U_i \left(c_i^{'}, c_{-i}\right)$. The set of cautiously rationalizable strategy profiles is $C^\infty \equiv \lim_{k \to \infty} C^k = \bigcap_{k=0}^\infty C^k$.*

In Definition 11, $R_i^\infty \left(C^{k-1}\right)$ is player i's set of rationalizable strategies given that the set of players' strategy profiles is C^{k-1}. That is, $\forall$ $i \in I$, the set $R_i^\infty \left(C^{k-1}\right)$ is the limit set R_i^∞ of the iterative procedure of Definition 2 starting with $R^0 \equiv C^{k-1}$. In Definition 11, at each step of the iterative procedure, strategies that are not best responses are eliminated first, and then those that are not *cautious* responses are removed. We will denote the pure strategies in C_i^k by SC_i^k.

Consider the following example taken from Pearce [11]. Figure 10 gives us the payoff matrix of the normal-form game G9. In G9, action X_2 of player 1 is not perfectly nor weakly perfectly nor properly nor trembling-hand rationalizable. However, this action is cautiously rationalizable: $SC_1^\infty = \{X_1, X_2\}$ and $C_2^\infty = \{Y_1\}$. Pearce [11] has shown that, $\forall$ $i \in I$, the limit set C_i^∞ is nonempty, closed, and satisfies the pure strategy property.

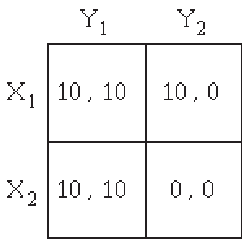
Theorem 9 *For every game in normal-form we have, $\forall$ $i \in I$, $C_i^\infty = \bigcap_{k=0}^\infty C_i^k \neq \emptyset$, the limit set C_i^∞ is closed and satisfies the pure strategy property $\forall i \in I$, and there exists $n \in \mathbb{N}$ such that: $C_i^{k+1} = C_i^k \quad \forall k \geq n$, $\forall i \in I$.*

The next section will make clear that the set of cautiously rationable strategy profiles can be either smaller or bigger than the set of strategy profiles obtained by any other refinement of rationalizability.

4 The Remaining Relationships

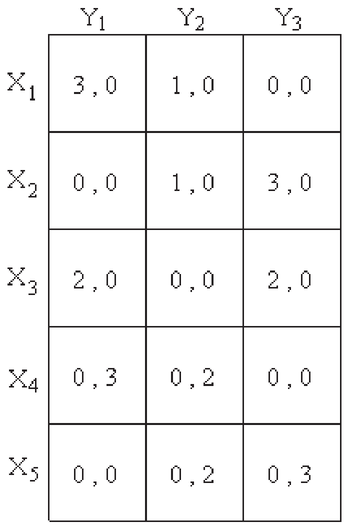
4.1 Two More Examples

The first example, G10, is due to Börgers [4]. Figure 11 gives us the payoff matrix of this two-player normal-form game. In G10, player 1's pure strategies or actions X_1, X_2, X_3 and



	Y_1	Y_2
X_1	10, 10	10, 0
X_2	10, 10	0, 0

Figure 10: A two-player game: G9

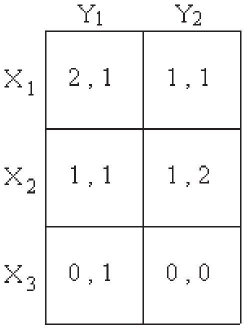


	Y_1	Y_2	Y_3
X_1	3, 0	1, 0	0, 0
X_2	0, 0	1, 0	3, 0
X_3	2, 0	0, 0	2, 0
X_4	0, 3	0, 2	0, 0
X_5	0, 0	0, 2	0, 3

Figure 11: A two-player game: G10

player 2's actions Y_1, Y_2, Y_3 are properly, trembling-hand, and cautiously rationalizable. Meanwhile, only player 1's actions X_1, X_2 and player 2's action Y_2 are perfectly rationalizable in G10. Given both examples G10 and G3, we conclude that there is no relationship between perfect rationalizability and these other refinements (proper, trembling-hand, and cautious rationalizability): perfect rationalizability may be weaker (example G3) or even stronger (example G10). Even if Bernheim's weaker definition of perfect rationalizability would be used, it would still be possible to eliminate pure strategy X_3 in G10.

The second example is the two-player normal-form game G11. Figure 12 gives us the payoff matrix of G11. In G11, proper and cautious rationalizability single out a unique strategy profile: (X_1, Y_2). Nevertheless, player 2's action Y_1 is trembling-hand rationalizable: $T_1^\infty = \{X_1\}$ and $T_2^\infty = M_2$. Therefore, there is no relationship between trembling-hand rationalizability and proper or cautious rationalizability: trembling-hand rationalizability may be weaker (Example G11) or stronger (Examples G9 and G4).



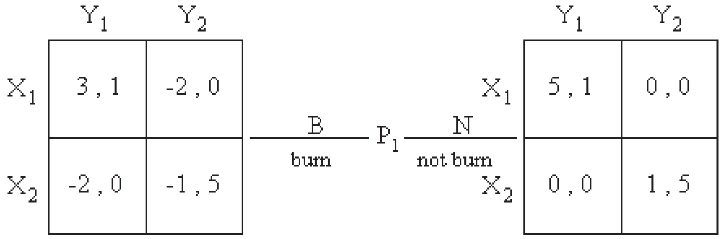
	Y_1	Y_2
X_1	2 , 1	1 , 1
X_2	1 , 1	1 , 2
X_3	0 , 1	0 , 0

Figure 12: A two-player game: G11

4.2 The Burning Money Game

Before concluding we briefly consider Ben-Porath and Dekel's [2] burning money game to get more insight into the consequences of using a particular refinement. This two-stage game is based on an idea of van Damme [14]. In the first stage, player 1 has a choice between her action B (burn money) that leads to a loss of 2 units of utility for her, and her action N (not burn money). After this choice is observed, player 1 and player 2 play a simultaneous-move game of coordination (see Figure 13). After the action N the payoffs are given by the right-hand matrix. After the action B the payoffs are given by the left-hand matrix; compared with the right-hand matrix, player 1's payoffs have all been reduced by 2 units, but player 2's payoffs are exactly the same. In the corresponding normal-form of this game (see Figure 14), player 1 has four pure strategies; $S_1 = \{BX_1, BX_2, NX_1, NX_2\}$. Player 2 has also four pure strategies; $S_2 = \{Y_1Y_1, Y_1Y_2, Y_2Y_1, Y_2Y_2\}$. The pure strategy Y_1Y_2 of player 2 means that he plays Y_1 if player 1 has burned money while he plays Y_2 otherwise.

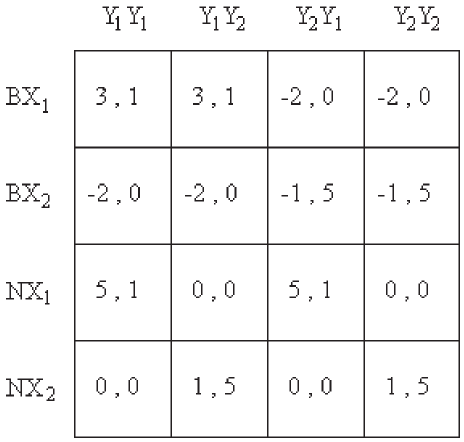
For this burning money game, trembling-hand rationalizability singles out a unique strategy profile: (NX_1, Y_1Y_1); that is, the fact that player 1 could have chosen to burn utility but did not do so ensures that she obtains her most preferred outcome. Indeed, in the game G12, once we apply our concept THR, we obtain the following iterative deletion of pure strategies: $BX_2 \notin ST_1^1$; $Y_2Y_1, Y_2Y_2 \notin ST_2^2$; $BX_2, NX_2 \notin ST_1^3$; $Y_2Y_1, Y_2Y_2, Y_1Y_2 \notin ST_2^4 = \{Y_1Y_1\}$; $BX_1, BX_2, NX_2 \notin ST_1^5 = \{NX_1\}$; $T^5 = \{(NX_1, Y_1Y_1)\}$. Nevertheless, player 1's action BX_1 (where player 1 burns money) is properly rationalizable. Indeed, $A^1(\varepsilon)$ is such that for all $(c_1, c_2) \in A_1^1(\varepsilon) \times A_2^1(\varepsilon) : c_1(BX_2) \leq \varepsilon\, c_1(NX_1)$ and $c_1(BX_2) \leq \varepsilon\, c_1(NX_2)$. Given these restrictions, for each pure strategy of player 2 there exists a conjec-



	Y_1	Y_2
X_1	3 , 1	-2 , 0
X_2	-2 , 0	-1 , 5

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

Figure 13: The burning money game



	Y_1Y_1	Y_1Y_2	Y_2Y_1	Y_2Y_2
BX_1	3 , 1	3 , 1	-2 , 0	-2 , 0
BX_2	-2 , 0	-2 , 0	-1 , 5	-1 , 5
NX_1	5 , 1	0 , 0	5 , 1	0 , 0
NX_2	0 , 0	1 , 5	0 , 0	1 , 5

Figure 14: Ben-Porath and Dekel's burning money game: G12

ture $c_1 \in A_1^1(\varepsilon)$ such that it is a best response against c_1. Indeed, for all $c_1 \in A_1^1(\varepsilon)$, player 2's expected payoffs are: $U_2(c_1, Y_1Y_1) = c_1(BX_1) + c_1(NX_1)$; $U_2(c_1, Y_1Y_2) = c_1(BX_1) + 5\,c_1(NX_2)$; $U_2(c_1, Y_2Y_1) = 5\,c_1(BX_2) + c_1(NX_1) \leq (1 + 5\varepsilon)\,c_1(NX_1)$; $U_2(c_1, Y_2Y_2) = 5\,c_1(BX_2) + 5\,c_1(NX_2) \leq (5 + 5\varepsilon)\,c_1(NX_2)$. For example, for all $\varepsilon \in (0,1)$, each pure strategy of player 2 is a best response against the conjecture $c_1 \in A_1^1(\varepsilon)$ defined by $c_1(BX_1) = \frac{5\varepsilon}{6(1+\varepsilon)}, c_1(BX_2) = \frac{\varepsilon}{6(1+\varepsilon)}, c_1(NX_1) = \frac{5}{6(1+\varepsilon)}, c_1(NX_2) = \frac{1}{6(1+\varepsilon)}$. For each pure strategy of player 1 belonging to $\{BX_1, NX_1, NX_2\}$ there exists a conjecture $c_2 \in A_2^2(\varepsilon)$ such that it is a best response against c_2. For example, each pure strategy belonging to $\{BX_1, NX_1, NX_2\}$ is a best response against the conjecture $c_2 \in A_2^1(\varepsilon)$ defined by $c_2(Y_1Y_1) = \frac{1}{12}, c_2(Y_1Y_2) = \frac{29}{60}, c_2(Y_2Y_1) = \frac{1}{12}, c_2(Y_2Y_2) = \frac{7}{20}$. Then, the sets of properly rationalizable strategies are the limit sets $A_1^\infty \supset \{BX_1, NX_1, NX_2\}$ and $A_2^\infty \supset \{Y_1Y_1, Y_1Y_2, Y_2Y_1, Y_2Y_2\}$; only player 1's pure strategy BX_2 does not belong to A_1^∞. Note that (NX_1, Y_1Y_1) is also the unique cautiously rationalizable strategy profile, with $(5,1)$ as the resulting payoffs. Therefore, trembling-hand and cautious rationalizability single out the outcome of forward induction (see Ben-Porath and Dekel [2], Hammond [7], van Damme [14]), while proper rationalizability (or weakly perfect rationalizability or perfect rationalizability) does not.

4.3 Conclusion

We conclude by summarizing the relationships between the refinements of rationalizability for normal-form games (see Table 2).

B^∞ Perfect rationalizability	$\subseteq$	D^∞ Weakly perfect rationalizability	Theorem 4
A^∞ Proper rationalizability	$\subseteq$	D^∞ Weakly perfect rationalizability	Theorem 6
T^∞ Trembling-hand rationalizability 2-person games	$\subseteq$	D^∞ Weakly perfect rationalizability	Theorem 8

Table 2: The relationships between the refinements

We have shown the relationship between perfect rationalizability and weakly perfect rationalizability (see Theorem 4) as well as the relationship between proper rationalizability and weakly perfect rationalizability (see Theorem 6): weakly perfect rationalizability is a weaker refinement than both perfect and proper rationalizability. Moreover, for 2-player normal-form games it holds that weakly perfect rationalizability is a weaker refinement than trembling-hand rationalizability (see Theorem 8). Unfortunately, there is no relationship between the other refinements (see Table 3).

A^∞	$\subset$	T^∞	Example G11, Figure 12
A^∞	$\supset$	T^∞	Example G12, Figure 14
B^∞	$\subset$	T^∞	Example G10, Figure 11
B^∞	$\supset$	T^∞	Example G3, Figure 4
T^∞	$\subset$	C^∞	Example G9, Figure 10
T^∞	$\supset$	C^∞	Example G11, Figure 12
D^∞	$\subset$	T^∞	Example G7, Figure 8
D^∞	$\supset$	T^∞	Example G3, Figure 4
C^∞	$\subset$	A^∞	Example G12, Figure 14
C^∞	$\supset$	A^∞	Example G9, Figure 10
D^∞	$\subset$	C^∞	Example G9, Figure 10
D^∞	$\supset$	C^∞	Example G3, Figure 4
C^∞	$\subset$	B^∞	Example G3, Figure 4
C^∞	$\supset$	B^∞	Example G9, Figure 10
A^∞	$\subset$	B^∞	Example G3, Figure 4
A^∞	$\supset$	B^∞	Example G10, Figure 11

A^∞ : set of properly rationalizable strategy profiles

B^∞ : set of perfectly rationalizable strategy profiles

C^∞ : set of cautiously rationalizable strategy profiles

D^∞ : set of weakly perfectly rationalizable strategy profiles

T^∞ : set of trembling-hand rationalizable strategy profiles

Table 3: No relationship between most refinements

References

[1] Asheim, G.B. and M. Dufwenberg, 1996, "Admissibility and Common Knowledge," CentER Discussion Paper 9616, (CentER, Tilburg University, Tilburg).

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

[3] Bernheim, D., 1984, "Rationalizable Strategic Behavior," *Econometrica* 52, 1007-1028.

[4] Börgers, T., 1994, "Weak Dominance and Approximate Common Knowledge," *Journal of Economic Theory* 64, 265-276.

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

[6] Brandenburger, A. and E. Dekel, 1987, "Rationalizability and Correlated Equilibria," *Econometrica* 55, 1391-1402.

[7] Hammond, P., 1993, "Aspects of Rationalizable Behavior," in K. Binmore, A. Kirman, P. Tani (eds), *Frontiers of Game Theory*, pp.277-305, MIT Press.

[8] Harsanyi, J.C., and R. Selten, 1988, *A General Theory of Equilibrium Selection in Games*, MIT Press.

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

[10] Myerson, R.B., 1978, "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory* 7, 73-80.

[11] Pearce, D.G., 1984, "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52, 1029-1050.

[12] Schuhmacher, F., 1995, "Proper Rationalizability and Backward Induction," mimeo, (University of Bonn, Bonn).

[13] Selten, R., 1975, "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4, 25-55.

[14] Van Damme, E., 1989, "Stable Equilibria and Forward Induction," *Journal of Economic Theory* 48, 476-496.

[15] Van Damme, E., 1991, *Stability and Perfection of Nash Equilibria*, Springer-Verlag (Second Edition).

[16] Vannetelbosch, V.J., 1996, "Refinements of Rationalizability for Normal-Form Games: The Main Ideas," IRES Discussion Paper 9612, (University of Louvain, Louvain-la-Neuve).