An Epistemic Characterization of Iterated Deletion of Inferior Strategy Profiles in Preference-Based Type Spaces[⋆]

Michael Trost

Institute for Research in Economic Evolution
University of Freiburg

`michael.trost@vwl.uni-freiburg.de`

Abstract. Bonanno (2008) provides an epistemic characterization for the solution concept of iterated deletion of inferior strategy profiles (IDIP) by embedding strategic games with ordinal payoffs in non-probabilistic epistemic models which are built on Kripke frames. In this paper we follow the event-based approach to epistemic game theory and supplement strategic games with type space models, where each type is associated with a preference relation on the state space. In such a framework IDIP can be characterized by the conditions that at least one player has correct beliefs about the state of the world and there is common belief that every player is rational, has correct beliefs about the state of the world, is aware about the own choice of strategy and has strictly monotone preferences. Moreover, we compare the epistemic motivations for IDIP and its mixed strategy variant known as strong rationalizability (SR). Presuppose the above conditions. Whenever there is also common belief of expected utility maximization IDIP still applies. But if there is common belief of expected payoff maximization, then SR results.

Key words: Iterated deletion of inferior strategy profiles, strong rationalizability, epistemic game theory.

1 Introduction

The solution concept of iterated deletion of inferior strategy profiles (IDIP) introduced by Bonanno (2008) is the pure strategy variant of the strong rationalizability concept (SR) proposed by Stalnaker (1994). Unlike standard deletion processes (e.g. iterated deletion of strictly dominated strategies or iterated deletion of weakly dominated strategies) both solution concepts are based on deletion processes which eliminate strategy profiles rather than strategies at each round. An epistemic motivation for SR is

---

[⋆] I would like to thank Thomas Gehrig and the participants of the economic theory seminar at the University of California, Davis, in particular Burkhard Schipper and Giacomo Bonanno. The main issue of this paper originates from a question raised by latter. Furthermore, I am indebted to the two anonymous referees for their valuable comments.

provided by Stalnaker (1994) and Bonanno and Nehring (1998) where latter amend some inconsistencies in the argumentation of the former. An epistemic motivation of IDIP is given in Bonanno (2008). While Stalnaker (1994) and Bonanno and Nehring (1998) follow standard epistemic game theory and frame strategic games by Harsanyi (or probabilistic) type spaces[1], Bonanno (2008) starts with strategic games with ordinal payoffs and embed them in non-probabilistic epistemic models which are built on Kripke frames. In this paper we also provide an epistemic motivation for IDIP, but obey to the standard approach and take the strategic game form as basis for our analysis. However, we encounter the problem that an epistemic characterization of IDIP satisfying the minimal requirement that every player acts rationally is impossible, if strategic games are framed by Harsanyi type spaces.

In order to realize that problem recall that in a Harsanyi type space any type of any player is associated with a probability measure on the state space. This measure quantifies the beliefs of a type about the players' choices and types. In such a framework rationality means expected payoff maximization. As Pearce (1984, Lemma 3) demonstrated a strategy is never a rational choice of an expected payoff maximizer if and only if it is strictly dominated by some mixture. In the following we present a game at which strategy profiles containing strategies strictly dominated by some mixtures survive iterated deletion of inferior strategies.

Let us apply the solution concepts of IDIP and SR on game $\mathit{\Gamma}_1$ depicted in figure 1. The solutions of IDIP and SR are in bold in the two game matrices below. While SR leads to the unique solution (u, r), IDIP takes the strategy profiles $(m, l), (m, c), (m, r), (u, r)$ as solutions. Although strategy m is strictly dominated by the $60 : 40$ mixture of the strategies u and d, any strategy profile containing strategy m survives the iterated deletion of inferior strategy profiles. Since this strategy is never a rational choice of an expected payoff maximizer it is impossible to construct a Harsanyi type space to game $\mathit{\Gamma}_1$ containing a state at which player R is rational and chooses strategy m.

To circumvent this impossibility result we generalize the Harsanyi type space to so-called preference-based type spaces. In such type spaces each type is associated with a preference relation on acts defined on the state space instead of a probability measure on the state space. Only in the case that the preference relation of every type is representable by an expected payoff function the preference-based type spaces can be transformed into some Harsanyi type space. The beliefs of the players in preference-based type spaces are formed according to Morris (1996) and an event is said to be believed by a player if its complement is (Savage-)null. Such belief formation is compatible with the probabilistic beliefs in Harsanyi type spaces.

Two epistemic characterizations for IDIP are given in this paper. The first one aims to set only weak requirements on the common beliefs in order to characterize IDIP. The second one is an tightening of the first one and

---

[1] Indeed, they embed games in so-called probabilistic epistemic models. But it is not difficult to see that these epistemic models and those type spaces are essentially the same. See e.g. Battigalli and Bonanno (1999, Remark 10).

Game Γ_1

		Player C		
		l	c	r
	u	$(1,1)$	$(3,2)$	$(4,3)$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d	$(4,3)$	$(1,4)$	$(1,2)$

Solutions under IDIP

		Player C		
		l	c	r
	u	$(1,1)$	$(3,2)$	$(\mathbf{4},\mathbf{3})$
Player R	m	$(\mathbf{2},\mathbf{2})$	$(\mathbf{2},\mathbf{2})$	$(\mathbf{2},\mathbf{2})$
	d	$(4,3)$	$(1,4)$	$(1,2)$

Solutions under SR

		Player C		
		l	c	r
	u	$(1,1)$	$(3,2)$	$(\mathbf{4},\mathbf{3})$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d	$(4,3)$	$(1,4)$	$(1,2)$

Fig. 1. IDIP and SR applied to game Γ_1

assumes that there is common belief that the players are Bayesian rational (i.e. that they are expected utility maximizers). Latter characterization reveals the difference in the epistemic motivation for SR and IDIP. Presuppose that at least one player has correct beliefs and there is common belief that every player has correct beliefs and is aware about the own choice of strategy. Whenever there is also common belief that all players are expected payoff maximizers then strong rationalizability is the appropriate solution concept. But if there is common belief that all players are expected utility maximizers then IDIP gives all possible outcomes of the game.

This paper is organized as follows. In the next section we reproduce the definition of IDIP and summarize important properties of this solution concept. In section 3 we supplement the strategic game with a preference-based type space. Such framework enables us to to decompose the strategic game into individual decision problems under subjective uncertainty. The two epistemic characterizations of IDIP are discussed in section 4. Finally, these characterizations are compared to those of SR in section 5.

2 Properties of IDIP

As mentioned above, unlike Bonanno (2008), we take the strategic game form as a basis of our epistemic analysis of IDIP. A *(finite) strategic game* Γ is a tuple $\Gamma := (S^i, z^i)_{i\in N}$, where N denotes the non-empty, finite set of players, S^i the non-empty, finite set of strategies available for player i, and $z^i : \times_{j\in N} S^j \rightarrow \mathbb{R}$ the player i's payoff function. As usual, a combination $s := (s^j)_{j\in N}$ of strategies is referred to as strategy profile and combination $s^{-i} := (s^j)_{j\in N\setminus\{i\}}$ denotes the profile listing strategies of players different to i. Henceforth, we denote the set of all strategy profiles by $S := \times_{i\in N} S^i$ and the set of all profiles listing strategies of players different to i by $S^{-i} := \times_{j\in N\setminus\{i\}} S^j$. Observe payoff function z^i assigns to every strategy profile $(s^j)_{j\in N}$ a real-valued payoff $z^i\left((s^j)_{j\in N}\right)$ usually interpreted as monetary payoff.[2] Let $\tilde{S}^i \subseteq S^i$ for any player i. The strategic game

$$\Gamma|_{\tilde{S}} := \left(\tilde{S}^i, z^i|_{\tilde{S}}\right)_{i\in N},$$

where $z^i|_{\tilde{S}}$ denotes the restriction of the payoff function on the domain $\tilde{S} := \times_{i\in N}\tilde{S}^i$, is called the *reduction of game Γ on strategy space $\tilde{S}$*.

Consider some strategic game $\Gamma := (S^i, z^i)_{i\in N}$ and pick some player $i \in N$. Let $\tilde{S}^i \subseteq S^i$ and $\tilde{S}^{-i} \subseteq S^{-i}$ be non-empty sets. A strategy $s^i \in \tilde{S}^i$ is called *weakly dominated in $\tilde{S}^i$ given $\tilde{S}^{-i}$*, whenever a strategy $\tilde{s}^i \in \tilde{S}^i$ exists such that $z^i(\tilde{s}^i, \tilde{s}^{-i}) \geq z^i(s^i, \tilde{s}^{-i})$ holds for any $\tilde{s}^{-i} \in \tilde{S}^{-i}$, where the inequality is strict for some $\hat{s}^{-i} \in \tilde{S}^{-i}$. A strategy $s^i \in \tilde{S}^i$ is called *strictly dominated in $\tilde{S}^i$ given $\tilde{S}^{-i}$* whenever a strategy $\tilde{s}^i \in \tilde{S}^i$ exists such that $z^i(\tilde{s}^i, \tilde{s}^{-i}) > z^i(s^i, \tilde{s}^{-i})$ holds for any $\tilde{s}^{-i} \in \tilde{S}^{-i}$. If strategy $s^i \in \tilde{S}^i$ is not strictly dominated (weakly dominated, respectively) $\tilde{S}^i$, then we say s^i is strictly undominated (weakly undominated, resp.) in $\tilde{S}^i$ given $\tilde{S}^{-i}$. In case that $\tilde{S}^i = S^i$ or $\tilde{S}^{-i} = S^{-i}$ holds, we usually omit the reference to these sets.

A *solution concept T for strategic games* is a mapping that assigns to each strategic game $\Gamma := (S^i, z^i)_{i\in N}$ a set $T(\Gamma) \subseteq S$ of strategy profiles. If there is no confusion about the underlying strategic game Γ we often just write T for the solution set $T(\Gamma)$. Let $WU^i(\Gamma)$ the set of strategies of player i that are weakly undominated in strategic game $\Gamma := (S^i, z^i)_{i\in N}$, then $WU(\Gamma) := \times_{i\in N}WU^i(\Gamma)$ gives the set of all weakly undominated strategy profiles (i.e. the set of all strategy profiles consisting only of weakly undominated strategies). We call this solution concept *weak undominance*. The process of iterated deletion of weakly dominated strategies is the sequence $(WU_k(\Gamma)_{k\in\mathbb{N}}$ inductively determined by $WU_1(\Gamma) := WU(\Gamma)$ and $WU_{k+1}(\Gamma) := WU(\Gamma|_{WU_k(\Gamma)})$ for all $k \geq 1$. A strategy profile $s \in WU_k(\Gamma)$ is said to survive k rounds of deletion of weakly dominated strategies. The solution concept of *iterated weak undominance WU_∞* is defined by $WU_\infty(\Gamma) := \cap_{k\in\mathbb{N}}WU_k(\Gamma)$ for any strategic game Γ. Verbally, it consists of all strategy profiles which only contain strategies surviving each round of deletion of weakly dominated strategies. The solution concepts of strict undominance (denoted by SU), of k rounds of deletion of strictly dominated strategies (SU_k) and of iterated strict undominance (SU_∞) are similarly defined.

In the following we reproduce the solution concept generated by *iterated deletion of inferior strategy profiles* (*IDIP*) and introduced by Bonanno (2008). This solution concept is a pure strategy variant of the *strong rationalizability* concept of Stalnaker (1994) which we will discuss in section 5. Both solution concepts differ considerably from conventional dele-

---

[2] In game theory payoff and utility are often viewed as interchangeable. A crucial feature of our analysis is that we distinguish between (observable) payoffs und (subjective) utility. The relationship between the strategy choices of the players and the payoffs they receive is captured by the payoff functions and is part of the strategic game. The players' rankings of their available strategies are not a feature of the strategic game and will be addressed in a separate item (see section 3). In such a setting, only in case that the players' preferences are representable by some expected payoff functions the (Bernoulli) utilities and the payoffs coincide.

tion processes like the iterated deletion of strictly dominated strategies or the iterated deletion of weakly dominated strategies. In latter processes strategies (more precisely, all strategy profiles containing at least one of those strategies) are successively deleted, while in the two solution concepts of Stalnaker (1994) and Bonanno (2008) specific strategy profiles are successively erased.

Definition 1. *Let* $\mathit{\Gamma} := (S^i, z^i)_{i \in N}$ *be a finite strategic game and* X *be a non-empty subset of* $S := \times_{i \in N} S^i$*. A strategy profile* $s \in X$ *is inferior relative to* X*, if there exists a player* $i \in N$ *and a strategy* $\tilde{s}^i \in S^i$ *such that*
(i) $z^i(\tilde{s}^i, s^{-i}) > z^i(s^i, s^{-i})$ *holds.*
(ii) for any $\tilde{s}^{-i} \in S^{-i}$*, if* $(s^i, \tilde{s}^{-i}) \in X$ *holds then* $z^i(\tilde{s}^i, \tilde{s}^{-i}) \geq z^i(s^i, \tilde{s}^{-i})$ *results.*

Let X be a non-empty subset of S, then $I(X)$ denotes the set of all strategy profiles inferior to X. A strategy profile belonging to $S \setminus I(S)$ is simply said to be inferior.

Definition 2. *Let* $\mathit{\Gamma} := (S^i, z^i)_{i \in N}$ *be a finite strategic game. Set* $R_0 := S$ *and* $R_k := R_{k-1} \setminus I(R_{k-1})$ *for all* $k \geq 1$*. A strategy profile* s *survives the iterated deletion of inferior strategy profiles, if* $s \in R_\infty := \cap_{k=1}^{\infty} R_k$ *holds.*

This solution concept is applied to strategic game $\mathit{\Gamma}_1$. The elimination process is depicted in figure 2 and contains the sets

$$\begin{aligned} R_1 &= \{(u,c),(u,r),(m,l),(m,c),(m,r),(d,c),(d,r)\} \\ R_2 &= \{(u,c),(u,r),(m,l),(m,c),(m,r)\} \\ R_3 &= \{(u,r),(m,l),(m,c),(m,r)\} \;=\; R_4 \;=\; \ldots\;. \end{aligned}$$

Consequently, strategy profiles of $R_\infty := \{(u,r),(m,l),(m,c),(m,r)\}$ survive this process. Note the outcome is not rectangular. The reason is that unlike conventional deletion processes IDIP refers to strategy profiles rather than strategies. The following remark summarizes properties of IDIP.

Remark 3. *Consider a finite strategic game* $\mathit{\Gamma}$ *and let* $(R_k)_{k \in N}$ *be the sequence of sets generated by iterated deletion of inferior strategy profiles. Then*
(a) $SU_k \supseteq R_k \supseteq WU_k$ *holds for every round* k*.*
(b) $SU_\infty \supseteq R_\infty \supseteq WU_\infty \neq \emptyset$ *holds.*
(c) *there exists a round* l *such that* $R_k = R_\infty$ *holds for any* $k \geq l$*.*
(d) *for any strategy profile* $s \in R_\infty$ *and for any player* $i \in N$ *strategy* s^i *is weakly undominated in* S^i *given* $\{\tilde{s}^{-i} \in S^{-i} : (s^i, \tilde{s}^{-i}) \in R_\infty\}$*.*

3 Preference-Based Type Spaces

The objective of epistemic game theory is to provide a decision-theoretic foundation for outcomes of games. Such foundation requires to supplement the strategic game which fixes the rules of the game with a type spaces which display the players' decision-makings. A *(finite) preference-based type space to a strategic game* $\mathit{\Gamma} := (S^i, z^i)_{i \in N}$ is a tuple $(T^i, \succsim^i)_{i \in N}$

R_0

Player C

		l	c	r
	u	$(1,1)$	$(3,2)$	$(4,3)$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d	$(4,3)$	$(1,4)$	$(1,2)$

R_1

Player C

		l	c	r
	u		$(3,2)$	$(4,3)$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d		$(1,4)$	$(1,2)$

R_2

Player C

		l	c	r
	u		$(3,2)$	$(4,3)$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d			

R_3

Player C

		l	c	r
	u			$(4,3)$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d			

Fig. 2. Process of iterated deletion of inferior strategy profiles

containing for each player a finite set T^i of types and a preference mapping $\succsim^i$ that assigns to every type $t^i \in T^i$ a complete, transitive and non-trivial preference relation $\succsim^i_{t^i}$ on $\mathbb{R}^\Omega$, where $\mathbb{R}^\Omega$ denotes the set of all acts from state space $\Omega := \times_{i \in N}(S^i \times T^i)$ to $\mathbb{R}$.[3]

Following the terminology of Savage (1954) a member of state space Ω is called a *state of the world* and a subset E of Ω an *event*. The ωth component of act x is denoted by x_ω and indicates the payoff the decision maker receives, when she has chosen act x and state ω occurs. For any event $E \subseteq \Omega$, x_E denotes the tuple $(x_\omega)_{\omega \in E}$ and $x_{\neg E}$ the tuple $x_{\Omega \setminus E}$. Let $x, y \in \mathbb{R}^\Omega$ and $E \subseteq \Omega$, then $(x_E, y_{\neg E})$ corresponds to the act that yields payoff x_ω if state $\omega \in E$ is realized and payoff y_ω if state $\omega \in \neg E$ is realized. For notational simplicity, we sometimes write ω for event $\{\omega\}$. Hereafter $t^i(\omega)$ and $s^i(\omega)$ denote the projections of state ω on T^i and S^i, respectively. Let $s^i \in S^i$ be an available strategy of player i. If the game is embedded in a type space, this strategy induces an act $x \in \mathbb{R}^\Omega$, where $x_\omega := z^i(s^i, s^{-i}(\omega))$ holds for any $\omega \in \Omega$. Note, sometimes we speak of a strategy, but actually mean the act induced by this strategy. This should not cause problems, because it should be clear from the context, whether the strategy or its induced act is meant. For example, let $s^i, \tilde{s}^i \in S^i$, then $s^i \succsim^i_{t^i} \tilde{s}^i$ means that type t^i of player i weakly prefers the act induced by s^i to the act induced by $\tilde{s}^i$. From now on, we call a strategic game framed by a preference-based type space a *framed strategic game*.

Preference-based type spaces are not new in epistemic game theory. For example, Epstein and Wang (1996) and Di Tillo (2008) construct universal preference-based type spaces, where universality means that any coherent hierarchy of preference relations can be generated by the type space.

---

[3] A preference relation $\succsim$ on $\mathbb{R}^\Omega$ is *complete*, if $x \succsim y$ or $y \succsim x$ holds for any acts x, y, is *transitive*, if $x \succsim y$ and $y \succsim z$ implies $x \succsim z$, and is *non-trivial*, if there are acts x, y such that $x \succ y$ holds.

To obtain such a type space Epstein and Wang (1996) postulate regularity axioms on types' preference relations, while Di Tillo (2008) imposes only mild restrictions on the types' preference relations, but assumes that the set of possible consequences (i.e. payoffs) is finite. Unlike these approaches our preference-based type space consists only of a finite number of types and therefore is not universal. Our assumptions with regard to types' preference relations are less restrictive than those of Epstein and Wang (1996) and with regard to the cardinality of the set of possible consequences less restrictive than those of Di Tillo (2008). A finite preference-based type space very similar to ours can be found in Athreya (2001). However, there the uncertainty of a player refers only to the other players' strategy-type profiles. Since in our framework the domain of the preference relations include the player's own strategies and types, our setting allows a preference-based description of the ability for introspection.

Standard epistemic game theory as surveyed in Battigalli and Bonanno (1999) embed strategic games in so-called Harsanyi type space. A *Harsanyi* (or *probabilistic*) *type space to a strategic game* $\Gamma := (S^i, z^i)_{i \in N}$ is a tuple $(T^i, p^i)_{i \in N}$ containing for each player i a finite set T^i of types and a mapping $p^i : T^i \rightarrow \Delta(\Omega)$ assigning to every type $t^i \in T^i$ a probability measure $p^i_{t^i}$ on state space $\Omega := \times_{i \in N}(S^i \times T^i)$ (see Harsanyi (1967/68)). Obviously, each Harsanyi type space corresponds essentially to the preference-based type space, for which each type t^i is associated with the preference relation that is representable by the corresponding expected payoff function $U : \mathbb{R}^{\Omega} \rightarrow \mathbb{R}$ with $x \mapsto \sum_{\omega \in \Omega} p^i_{t^i}(\omega) x_{\omega}$. Therefore the class of all Harsanyi type spaces to some fixed strategic game Γ can be considered as a subclass of the class of all preference-based type spaces to Γ.

While the strategic game form fixes the rules of the game, the type structure should describe the players' beliefs about the players' choices and types. In Harsanyi type spaces the formation of types' beliefs is straightforward. There an event is believed by a type, whenever she assigns probability one to this event. Because our preference-based construction of a type space does not imply that the types are endowed with a probabilistic belief, we are forced to choose another way for defining types' beliefs. We proceed like Morris (1996) and deduce directly the types' beliefs from their preference relations.

Consider a framed strategic game $(S^i, z^i, T^i, \succsim^i)_{i \in N}$. A state $\tilde{\omega}$ is said to be *considered as possible by the player i at state ω*, whenever the preferences of type t^i_{ω} are affected by outcomes at state $\tilde{\omega}$, that is, whenever there exists acts $x, y, z \in \mathbb{R}^{\Omega}$ such that $(y_{\tilde{\omega}}, x_{\neg\tilde{\omega}}) \succ^i_{t^i_{\omega}} (z_{\tilde{\omega}}, x_{\neg\tilde{\omega}})$ holds. The event consisting of all states which are considered as possible by i at ω is called the *possibility set of i at ω* and is denoted by $P^i(\omega)$ from now on. Player i *believes at state ω that event E occurs*, if every state that i considers as possible at ω belongs to E, or in formal terms, if $P^i(\omega) \subseteq E$ holds. It turns out that an event E is believed by i at ω, whenever $\neg E$ is (Savage-)null, that is, whenever $(x_E, y_{\neg E}) \sim^i_{t^i_{\omega}} (x_E, z_{\neg E})$ holds for any acts $x, y, z \in \mathbb{R}^{\Omega}$. Note, in the subclass of Harsanyi type spaces this belief concept agrees with the probabilistic belief concept. An event E is said to be *commonly believed at state ω*, if every player believes that E occurs, every player believes that every player believes that E occurs, every player believes that every player believes that every player believes that E occurs and so on. It turns out that event E is commonly believed, if and only if $E \subseteq P_*(\omega)$ holds where P^* is the transitive closure of the players' possibility correspondences $(P^i)_{i\in N}$.[4]

The objective of epistemic game theory is to examine how players' theories about the other players' choices and theories do affect the outcome of the game. Those theories are generally called *statements about the world*. In the following we describe specific statements which become crucial for our epistemic analysis of the solution concepts of IDIP and SR.

A player i is said to be *rational at* ω, if at this state her choice of strategy is a rational choice, i.e. if $s^i_\omega \succsim^i_{t^i_\omega} s^i$ holds for any $s^i \in S^i$. Clearly, whenever the preference relation of the player is representable by an expected utility (payoff, respectively) function, then rationality corresponds to expected utility (payoff, resp.) maximization. A player i is called *aware about the own choice of strategy at* ω, if at this state she deems only states as possible at which she chooses the strategy that she is actually choosing, or in symbols, if

$$\left(x_{\{s^i_\omega\}\times T^i\times\Omega^{-i}}, y_{\neg\{s^i_\omega\}\times T^i\times\Omega^{-i}}\right) \sim^i_{t^i_\omega} \left(x_{\{s^i_\omega\}\times T^i\times\Omega^{-i}}, z_{\neg\{s^i_\omega\}\times T^i\times\Omega^{-i}}\right)$$

holds for any acts $x, y, z \in \mathbb{R}^\Omega$. A player i is said to have *correct* or *true beliefs at state* ω, if she deems the actual state as possible, that means, she does not err at state ω. In formal terms, having correct beliefs means that there exists some acts $x, y, z \in \mathbb{R}^\Omega$ such that $(x_\omega, z_{\neg\omega}) \succ^i_{t^i_\omega} (y_\omega, z_{\neg\omega})$ holds. A preference relation of player i is *strictly monotone* at state ω, if for any acts $x, y \in \mathbb{R}^\Omega$ with x weakly dominating y given $P^i(\omega)$ the ranking $x \succ^i_{t^i_\omega} y$ results.[5] A preference relation of player i is *concave* at state ω, if for any acts x, y and any scalar $\lambda \in [0,1]$ the ranking $x \succsim y$ implies the ranking $x \succsim \lambda x + (1-\lambda)y$. We say, a preference relation of player i is *representable by an expected utility function*, if there exists a *Bernoulli utility function* $u^i : \mathbb{R} \to \mathbb{R}$ and a probability measure p^i on Ω such that for any acts $x, y \in \mathbb{R}^\Omega$ the ranking $x \succsim y$ holds, if and only if $\sum_{\omega\in\Omega} p^i(\omega)u^i(x_\omega) \geq \sum_{\omega\in\Omega} p^i(\omega)u^i(y_\omega)$ holds. In case that $u^i(\lambda) = \lambda$ is satisfied for any $\lambda \in \mathbb{R}$ we say that player i's preference relation $\succsim$ is representable by an expected payoff function. A preference-based axiomization of expected payoff representation suitable to our setting can be found in Jensen (1967) and a preference-based axiomization of expected utility representation in Wakker (1984).

Consider some strategic game $\Gamma := (S^i, z^i)_{i\in N}$. A preference-based type space to Γ is said to be *consistent* with a statement about the world, whenever it contains a state at which this statement is satisfied. A statement about the world *characterizes* a set $T \subseteq S := \times_{i\in N} S^i$ of strategy profiles, if the following two conditions are satisfied:

---

[4] That is, $\tilde{\omega} \in P_*(\omega)$ holds if and only if there is a finite sequence $(i_1, \ldots, i_m)$ in N and a finite sequence $(\omega_0, \omega_1, \ldots, \omega_m)$ in Ω such that $\omega_0 = \omega$, $\omega_m = \tilde{\omega}$ and, for every $k = 1, \ldots, m$, $\omega_k \in P^{i_k}(\omega_{k-1})$ hold.

[5] Consider an event $E \subseteq \Omega$. An *act* x *weakly dominates act* y *given* E, if $x_\omega \geq y_\omega$ holds for any $\omega \in E$ and this inequality is strict for some $\omega \in E$.

(i) (*Consistency*) If a preference-based type space to Γ is consistent with the statement, then at every state of the type space satisfying this statement a strategy profile $s \in T$ is realized.

(ii) (*Existence*) For every $s \in T$ there exists a preference-based type space to Γ which contains a state at which this statement is satisfied and at which the strategy profile s is realized.

An *epistemic statement* is a statement referring to the beliefs of the players. An *epistemic characterization for a solution concept* is given, whenever an epistemic statement is found that, for any strategic game, characterizes the set of strategy profiles resulting from this solution concept. In the following sections we provide epistemic characterizations for the solution concepts IDIP and SR.

4 Epistemic Characterizations of IDIP

In this section we provide epistemic characterizations of IDIP. At the outset we state epistemic conditions entailing common belief that a strategy profile is realized that survives IDIP. It turns out that such common belief is implied by common belief of rationality, awareness on the own choice of strategy, correct beliefs and strict monotonicity. As we will see, these epistemic conditions are a constituent part of the epistemic characterizations of IDIP derived here. Indeed, an epistemic characterization of IDIP is attained, if additional to these epistemic conditions we presuppose that at least one player has correct beliefs. Furthermore, we will show that IDIP is still characterizable, even if the epistemic condition of common belief of strict monotonicity is tightened. More precisely, if we presuppose that there common belief of preferences being representable by an expected utility function with some strictly increasing Bernoulli function IDIP still results.

Lemma 4. *Consider a framed strategic game $(S^i, z^i, T^i, \succsim^i)_{i \in N}$. If at state ω there is common belief that every player is rational, is aware on the own choice of strategy, has correct beliefs and has strictly monotone preferences, then at state ω there is common belief that only strategy profiles are chosen which survive the iterated deletion of inferior strategy profiles.*

Note above lemma states epistemic conditions such that there is common belief among the players that a strategy profile is realized that survives IDIP. However, these conditions do not imply that the actual strategies survive this process as the following example (figure 3) of a type space to strategic game Γ_1 demonstrates, where $u : \mathbb{R} \to \mathbb{R}$ denotes the strictly increasing mapping assigning $u(x) := 2x$ to any $x \leq 2$ and $u(x) := 3 + \frac{1}{2}x$ to any $x > 2$.

Suppose the actual state of the world is $\tilde{\omega} := (u, t_1^R, l, t_1^C)$ and consider event

$$E := \{(u, t_1^R, r, t_3^C), (m, t_2^R, l, t_1^C), (m, t_2^R, c, t_2^C), (m, t_2^R, r, t_3^C)\} .$$

Obviously, it holds $P^R(\tilde{\omega}) \subseteq E$ as well as $P^C(\tilde{\omega}) \subseteq E$. Since $P^R(\omega) \subseteq E$ and $P^C(\omega) \subseteq E$ are satisfied for any $\omega \in E$, we obtain $P_*(\tilde{\omega}) \subseteq E$.

Player	Type	Preference relation
R	t_1^R	$x \succsim_{t_1^R}^R y :\Leftrightarrow x_{(u,t_1^R,r,t_3^C)} \geq y_{(u,t_1^R,r,t_3^C)}$
	t_2^R	$x \succsim_{t_2^R}^R y :\Leftrightarrow u(x_{(m,t_2^R,l,t_1^C)}) + u(x_{(m,t_2^R,c,t_2^C)}) + u(x_{(m,t_2^R,r,t_3^C)}) \geq u(y_{(m,t_2^R,l,t_1^C)}) + u(y_{(m,t_2^R,c,t_2^C)}) + u(y_{(m,t_2^R,r,t_3^C)})$
C	t_1^C	$x \succsim_{t_1^C}^C y :\Leftrightarrow x_{(m,t_2^R,l,t_1^C)} \geq y_{(m,t_2^R,l,t_1^C)}$
	t_2^C	$x \succsim_{t_2^C}^C y :\Leftrightarrow x_{(m,t_2^R,c,t_2^C)} \geq y_{(m,t_2^R,c,t_2^C)}$
	t_3^C	$x \succsim_{t_3^C}^C y :\Leftrightarrow x_{(m,t_2^R,r,t_3^C)} \geq y_{(m,t_2^R,r,t_3^C)}$

Fig. 3. A type space to strategic game $\mathit{\Gamma}_1$

Furthermore, at any state belonging to event E each player is rational, is aware about her choice of strategy, has correct beliefs and has strictly monotone preferences. Hence, at state $\tilde{\omega}$ there is common belief that all players are rational, are aware about their own choices, have correct beliefs and have strictly monotone preference relations. Note that at any state of event E a strategy profile is realized that survives IDIP. Therefore, in accordance to lemma 4, at state $\tilde{\omega}$ there is common belief that a strategy profile is chosen which is iteratively non-inferior. However, the actual strategy profile (u, l) does not survive this process. Analyzing further this game, it turns out that at $\tilde{\omega}$ every player is rational, aware about her choice of strategy and has strictly monotone preferences, but does not deem the actual state as possible. Indeed, achieving that the strategy profile realized at the actual state survives the deletion process, the condition that at least one player considers the actual state as possible (or in other words, at least one player has correct beliefs) has to be added to the assumptions given in lemma 3. This additional assumption is satisfied at all states of event E and for this reason iteratively non-inferior strategy profiles are selected there.

Theorem 5. *The set of strategy profiles surviving the iterated elimination of inferior strategy profiles is characterized by correct beliefs of some player and common belief of rationality, awareness on the own choice of strategy, correct beliefs and strict monotonicity.*

Stricter epistemic characterizations of IDIP are possible. For example, we can tighten the epistemic condition of common belief of strict monotonicity without ruling out strategy profiles predicted by IDIP. Indeed, even if we presuppose that there is common belief of preferences being representable by an expected utility function with a strictly increasing Bernoulli function IDIP still results.

Corollary 6. *Iterated deletion of inferior strategy profiles is characterized by correct beliefs of some player and common belief of rationality, awareness on the own choice of strategy, correct beliefs and preferences representable by expected utility functions with a strictly increasing Bernoulli function.*

5 A Comparison with Characterizations of SR

In this section we compare the epistemic characterizations of IDIP derived in the preceding section with epistemic characterizations of SR. Stalnaker (1994) and Bonanno and Nehring (1998) already give an epistemic characterization for this solution concept in Harsanyi type spaces. Providing an epistemic characterization of SR in preference-based type spaces enables us to detect the crucial preference properties which should be commonly believed such that only outcomes predicted by SR are realized. It turns out that adding common belief of concavity to the epistemic conditions stated in theorem 5 suffices to characterize that solution concept.

Consider some strategic game $\Gamma := (S^i, z^i)_{i \in N}$ and pick some player $i \in N$. A probability measure σ^i on S^i is called a *mixed strategy of player* i, where $\sigma^i(s^i)$ gives the probability that pure strategy s^i is realized. The set of all mixed strategies of player i is labeled by $\Delta(S^i)$. Let $\tilde{S}^i$ be a non-empty subset of S^i, then $\Delta(\tilde{S}^i)$ denotes the set of all mixed strategies whose support is a subset of $\tilde{S}^i$. Note, each pure strategy $s^i \in S^i$ can be identified with the degenerated mixed strategy which assigns probability 1 to strategy s^i. For this reason we can view any $\tilde{S}^i \subseteq S^i$ as a subset of $\Delta(\tilde{S}^i)$. By a slight abuse of notation we denote the expected payoff of mixed strategy $\sigma^i \in \Delta(S^i)$ given the other players choose profile $s^{-i} \in S^{-i}$ by $z^i(\sigma^i, s^{-i})$. Let $\tilde{S}^{-i}$ be a non-empty subset of S^{-i}, then strategy $s^i \in \tilde{S}^i$ is called *weakly dominated by some mixture of* $\tilde{S}^i$ (or equivalently, *weakly dominated in* $\Delta(\tilde{S}^i)$) *given* $\tilde{S}^{-i}$, whenever a mixed strategy $\tilde{\sigma}^i \in \Delta(\tilde{S}^i)$ exists such that $z^i(\tilde{\sigma}^i, \tilde{s}^{-i}) \geq z^i(s^i, \tilde{s}^{-i})$ holds for any $\tilde{s}^{-i} \in \tilde{S}^{-i}$ and this inequality is strict for some $\tilde{s}^{-i} \in \tilde{S}^{-i}$. A strategy $s^i \in \tilde{S}^i$ is called *strictly dominated by some mixture of* $\tilde{S}^i$ (or equivalently, *strictly dominated in* $\Delta(\tilde{S}^i)$) *given* $\tilde{S}^{-i}$ whenever a strategy $\tilde{s}^i \in \tilde{S}^i$ exists such that $z^i(\tilde{\sigma}^i, \tilde{s}^{-i}) > z^i(s^i, \tilde{s}^{-i})$ holds for any $\tilde{s}^{-i} \in \tilde{S}^{-i}$. If strategy $s^i \in \tilde{S}^i$ is not strictly dominated (weakly dominated, respectively) by some mixture of $\tilde{S}^i$ given $\tilde{S}^{-i}$, then we say s^i is strictly undominated (weakly undominated, resp.) by mixtures of $\tilde{S}^i$ given $\tilde{S}^{-i}$. In case that $\tilde{S}^i = S^i$ or $\tilde{S}^{-i} = S^{-i}$ holds, we usually omit the reference to these sets.

The solution concept of *weak undominance in mixtures* assigns to each strategic game the set of all strategy profiles consisting only of strategies weakly undominated by mixtures. Henceforth, it is denoted by $\overline{WU}$. Fix some strategic game Γ. The process of iterated deletion of strategies weakly dominated by some mixture is the sequence $(\overline{WU}_k(\Gamma))_{k \in \mathbb{N}}$ inductively determined by $\overline{WU}_1(\Gamma) := \overline{WU}(\Gamma)$ and $\overline{WU}_{k+1}(\Gamma) := \overline{WU}(\Gamma|_{\overline{WU}_k(\Gamma)})$ for all $k \geq 1$. The solution concept of *iterated weak undominance in mixtures* $\overline{WU}_\infty$ is defined by $\overline{WU}_\infty(\Gamma) := \cap_{k \in \mathbb{N}} \overline{WU}_k(\Gamma)$ for any strategic game Γ. A strategy profile $s \in \overline{WU}_k(\Gamma)$ is said to survive k rounds of deletion of strategies weakly dominated by some mixture. The solution concepts of strict undominance in mixtures (denoted by $\overline{SU}$), of k rounds of deletion of strategies strictly dominated by some mixture ($\overline{SU}_k$) and of iterated strict undominance in mixtures ($\overline{SU}_\infty$) are similarly defined.

Next we reproduce the solution concept of strong rationalizability (SR) introduced by Stalnaker (1994). This concept is based on iterated deletion of strategy profiles being inferior to some mixture. Like IDIP this deletion process refers to strategy profiles and not to strategies.

Definition 7. *Let $\Gamma := (S^i, z^i)_{i\in N}$ be a finite strategic game and X be a non-empty subset of $S := \times_{i\in N} S^i$. A strategy profile $s \in X$ is called inferior to some mixture relative to X, if there exists a player $i \in N$ and a mixed strategy $\tilde{\sigma}^i \in \Delta(S^i)$ such that*
(i) $z^i(\tilde{\sigma}^i, s^{-i}) > z^i(s^i, s^{-i})$ *holds.*
(ii) for any $\tilde{s}^{-i} \in S^{-i}$, *if* $(s^i, \tilde{s}^{-i}) \in X$ *holds then* $z^i(\tilde{\sigma}^i, \tilde{s}^{-i}) \geq z^i(s^i, \tilde{s}^{-i})$ *results.*

Let X be a non-empty subset of S, then $\overline{I}(X)$ denotes the set of strategy profiles inferior to some mixture relative to X. A strategy profile belonging to set $S \setminus \overline{I}(S)$ is simply said to be inferior to mixtures.

Definition 8. *Let $\Gamma := (S^i, z^i)_{i\in N}$ be a finite strategic game. Set $\overline{R}_0 := S$ and $\overline{R}_k := \overline{R}_{k-1} \setminus \overline{I}(\overline{R}_{k-1})$ for all $k \geq 1$. A strategy profile s survives the iterated deletion of strategy profiles being inferior to mixtures, if $s \in \overline{R}_\infty := \cap_{k=1}^{\infty} \overline{R}_k$ holds.*

The solution concept $\overline{R}_\infty$ is called *strong rationalizability*. We apply it to strategic game Γ_1. The deletion process is depicted in figure 4 and contains the sets

$$\begin{aligned} \overline{R}_1 &= \{(u,c),(u,r),(d,c),(d,r)\} \\ \overline{R}_2 &= \{(u,c),(u,r)\} \\ \overline{R}_3 &= \{(u,r)\} \;=\; \overline{R}_4 \;=\; \ldots\;. \end{aligned}$$

Consequently, strong rationalizability applied to game Γ_1 gives the solution $\overline{R}_\infty := \{(u,r)\}$.

Game Γ_1 → $\overline{R}_1$

Player C

Player R		l	c	r
	u	$(1,1)$	$(3,2)$	$(4,3)$
Player R	m	$(2,2)$	$(2,2)$	$(2,2)$
	d	$(4,3)$	$(1,4)$	$(1,2)$

$\overline{R}_1$

Player C

		l	c	r
	u		$(3,2)$	$(4,3)$
Player R	m			
	d		$(1,4)$	$(1,2)$

$\overline{R}_2$

Player C

		l	c	r
	u		$(3,2)$	$(4,3)$
Player R	m			
	d			

$\overline{R}_3$

Player C

		l	c	r
	u			$(4,3)$
Player R	m			
	d			

Fig. 4. Process of iterated deletion of strategy profiles being inferior to mixtures

The following remark summarizes properties of strong rationalizability.

Remark 9. *Consider a finite strategic game Γ and let $(\overline{R}_k)_{k\in N}$ be the sequence of sets generated by iterated deletion of strategy profiles inferior to mixtures. Then*

(a) *$\overline{SU}_k \supseteq \overline{R}_k \supseteq \overline{WU}_k$ holds for every round k.*
(b) *$\overline{SU}_\infty \supseteq \overline{R}_\infty \supseteq \overline{WU}_\infty \neq \emptyset$ holds.*
(c) *there exists a round l such that $\overline{R}_k = \overline{R}_\infty$ holds for any $k \geq l$.*
(d) *for any strategy profile $s \in \overline{R}_\infty$ and for any player $i \in N$ strategy s^i is weakly undominated in $\Delta(S^i)$ given $\{\tilde{s}^{-i} \in S^{-i} : (s^i, \tilde{s}^{-i}) \in \overline{R}_\infty\}$.*

As mentioned above our objective is to figure out minimal requirements on common beliefs such that SR gives all possible solutions. Proceeding like in the case of IDIP we begin our epistemic analysis of SR with an epistemic motivation of common belief of strongly rationalizable strategy profiles.

Lemma 10. *Consider a framed strategic game $(S^i, z^i, T^i, \succsim^i)_{i\in N}$. If at state ω there is common belief that every player is rational, is aware about the own choice of strategy, has correct beliefs and has strictly monotone and concave preferences, then at state ω there is common belief that strongly rationalizable strategy profiles are realized.*

Next theorem gives an epistemic characterization of SR. It turns out that SR is the appropriate solution concept, if common belief of concave preferences is added to the epistemic conditions of theorem 5.

Theorem 11. *Strong rationalizability is characterized by correct beliefs of some player and common belief of rationality, awareness on the own choice of strategy, correct beliefs, strict monotonicity and concavity.*

The more strict epistemic characterization of SR provided by Stalnaker (1994) and Bonanno and Nehring (1998) can be derived from theorem 11. Corollary 12 reproduces, for the class of framed strategic games, their result.[6]

Corollary 12 (Stalnaker (1994), Bonanno and Nehring (1998)). *Strong rationalizability is characterized by correct beliefs of some player and common belief of rationality, awareness on the own choice of strategy, correct beliefs and preferences being representable by some expected payoff function.*

Reviewing corollaries 6 and 12 we discover a catchy rule highlighting the difference in probabilistic epistemic motivations for those solution concepts. Presuppose that at least one player has correct beliefs and there is common belief of rationality, awareness on the own choice of strategy and correct beliefs. Whenever there is also common belief that all players are expected utility maximizers with strict increasing Bernoulli functions

---

[6] In comparison to the original version of Stalnaker (1994) and Bonanno and Nehring (1998) we have to add the epistemic condition that there is common belief that every player is aware about the own strategy. This condition is due to our formulation of the preference-based type space. In our type space the preference relations of players' types are defined on the same state space $\Omega := \times_{i\in N}(S^i \times T^i)$ which includes the own strategy-type pairs. This formulation deviates from the standard formulation at which the state space of player j is given by $\times_{i\in N\setminus\{j\}}(S^i \times T^i)$ excluding the own strategy-type pairs $S^j \times T^j$.

then IDIP is the appropriate solution concept. But if this condition is tightened to common belief that all players are expected payoff maximizers then strong rationalizability gives all possible outcomes of the game.

This result is not self-evident. For example, consider the solution concepts of iterated deletion of strategies strictly dominated in mixtures and its pure strategy variant, iterated deletion of strategies strictly dominated by pure strategies. According to Pearce (1984), Bernheim (1984), Tan and Werlang (1988) and Stalnaker (1994) former solution concept is characterized by (i) rationality, (ii) preferences being representable by an expected payoff function and (iii) common belief of (i) and (ii). However, the modification of condition (ii) to the condition that preferences are representable by an expected utility function with a strictly increasing Bernoulli function does not lead to an epistemic characterization of the latter solution concept.[7] To see that consider the following strategic game Γ_2, where we only depict the payoffs of player R.

Game Γ_2

		Player C	
		l	r
	u	1	1
Player R	m	2	0
	d	1	2

Fig. 5. Expected utility maximization applied to game Γ_2

If the preference relation of rational player R is representable by an expected utility function with a strictly increasing Bernoulli function, then strategy u is never a rational choice, although this strategy is not strictly dominated by some pure strategy. In the case that player R attaches positive probability to the event that player C selects strategy r strategy d yields higher expected utility than u, otherwise strategy m yields higher expected utility than u.

---

[7] Indeed, one can show that these assumptions characterize iterated deletion of dominated strategies in sense of Börgers (1993).

Bibliography

Athreya, R. (2001). Modelling Beliefs in Games with Generalized Preferences. In van Benthem, J., Ed., *Proceedings of TARK VIII (2001)*, 123–136, Siena.

Battigalli, P. and Bonanno, G. (1999). Recent Results on Belief, Knowledge and the Epistemic Foundations of Game Theory. *Research in Economics*, 53:149–225.

Bernheim, D. (1984). Rationalizable Strategic Behavior. *Econometrica*, 52(4):1007–1028.

Bonanno, G. (2008). A Syntactic Approach to Rationality in Games with Ordinal Payoffs. In Bonnano, G., van der Hoek, W., and Wooldridge, M., Eds., *Logics and the Foundations of Game and Decision Theory*, 59–86. Amsterdam University Press, Amsterdam.

Bonanno, G. and Nehring, K. (1998). On Stalnaker's Notion of Strong Rationalizability and Nash Equilibrium in Perfect Information Games. *Theory and Decision*, 45(3):291–295.

Börgers, T. (1993). Pure Strategy Dominance. *Econometrica*, 61(2):423–430.

Di Tillo, A. (2008). Subjective Expected Utility in Games. *Theoretical Economics*, 3:287–323.

Epstein, L. and Wang, T. (1996). ”Beliefs about Beliefs” without Probabilities. *Econometrica*, 64(6):1343–1373.

Jensen, N. (1967). An Introduction to Bernoullian Utility Theory: I: Utility Functions. *Swedish Journal of Economics*, 69:163–183.

Morris, S. (1996). The Logic of Belief and Belief Change: A Decision Theoretic Approach. *Journal of Economic Theory*, 69(1):1–23.

Pearce, D. (1984). Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica*, 52(4):1029–1050.

Savage, L. (1954). *The Foundations of Statistics*. Wiley, New York.

Stalnaker, R. (1994). On the Evaluation of Solution Concepts. *Theory and Decision*, 37(1):49–74.

Tan, T. and Werlang, S. (1988). The Bayesian Foundation of Solution Concepts of Games. *Journal of Economic Theory*, 45(2):173–195.

Wakker, P. (1984). Cardinal Coordinate Independence for Expected Utility. *Journal of Mathematical Psychology*, 28:110–117.