*Article*

# The Role of Monotonicity in the Epistemic Analysis of Strategic Games

**Krzysztof R. Apt** [1,*] **and Jonathan A. Zvesper** [2]

[1] Centrum for Mathematics and Computer Science (CWI), University of Amsterdam, Science Park 123, 1098 XG Amsterdam, The Netherlands

[2] Oxford University, Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK

[*] Author to whom correspondence should be addressed; E-Mail: apt@cwi.nl; Tel: 31-20-592-4057; Fax: 31-20-592-4199.

**Abstract:** It is well-known that in finite strategic games true common belief (or common knowledge) of rationality implies that the players will choose only strategies that survive the iterated elimination of strictly dominated strategies. We establish a general theorem that deals with monotonic rationality notions and arbitrary strategic games and allows to strengthen the above result to arbitrary games, other rationality notions, and transfinite iterations of the elimination process. We also clarify what conclusions one can draw for the customary dominance notions that are not monotonic. The main tool is Tarski's Fixpoint Theorem.

**Keywords:** true common beliefs; arbitrary games; monotonicity; Tarski's Fixpoint Theorem

## 1. Introduction

### 1.1. Contributions

In this paper we provide an epistemic analysis of arbitrary strategic games based on possibility correspondences. We prove a general result that is concerned with monotonic program properties[1] used by the players to select optimal strategies.

---

[1]The concept of a monotonic property is introduced in Section 2.

More specifically, given a belief model for the initial strategic game, denote by **RAT**($\phi$) the property that each player $i$ uses a property $\phi_i$ to select his strategy ('each player $i$ is $\phi_i$-rational'). We establish in Section 3 the following general result:

Assume that each property $\phi_i$ is monotonic. The set of joint strategies that the players choose in the states in which **RAT**($\phi$) is a true common belief is included in the set of joint strategies that remain after the iterated elimination of the strategies that for player $i$ are not $\phi_i$-optimal.

In general, transfinite iterations of the strategy elimination are possible. For some belief models the inclusion can be reversed.

This general result covers the usual notion of rationalizability in finite games and a 'global' version of the iterated elimination of strictly dominated strategies used in [1] and studied for arbitrary games in [2]. It does not hold for the 'global' version of the iterated elimination of weakly dominated strategies. For the customary, 'local' version of the iterated elimination of strictly dominated strategies we justify in Section 4 the statement

> *true common belief (or common knowledge) of rationality implies that the players will choose only strategies that survive the iterated elimination of strictly dominated strategies*

for arbitrary games and transfinite iterations of the elimination process. Rationality refers here to the concept studied in [3]. We also show that the above general result yields a simple proof of the well-known version of the above result for finite games and strict dominance by a mixed strategy.

The customary, local, version of strict dominance is non-monotonic, so the use of monotonic properties has allowed us to provide epistemic foundations for a non-monotonic property. However, weak dominance, another non-monotonic property, remains beyond the reach of this approach. In fact, we show that in the above statement we cannot replace strict dominance by weak dominance. A mathematical reason is that its global version is also non-monotonic, in contrast to strict dominance, the global version of which is monotonic. To provide epistemic foundations of weak dominance the only currently known approaches are [4] based on lexicographic probability systems and [5] based on a version of the 'all I know' modality.

## 1.2. Connections

The relevance of monotonicity in the context of epistemic analysis of finite strategic games has already been pointed out in [6]. The distinction between local and global properties is from [7] and [8].

To show that for some belief models an equality holds between the set of joint strategies chosen in the states in which **RAT**($\phi$) is true common belief and the set of joint strategies that remain after the iterated elimination of the strategies that for player $i$ are not $\phi_i$-rational requires use of transfinite ordinals. This complements the findings of [9] in which transfinite ordinals are used in a study of limited rationality, and [10], where a two-player game is constructed for which the $\omega_0$ (the first infinite ordinal) and $\omega_0 + 1$ iterations of the rationalizability operator of [3] differ.

In turn, [11] show that arbitrary ordinals are necessary in the epistemic analysis of arbitrary strategic games based on partition spaces. Further, as shown in [2], the global version of the iterated elimination of strictly dominated strategies, when used for arbitrary games, also requires transfinite iterations of the underlying operator.

Finally, [12] invokes Tarski's Fixpoint Theorem, in the context of what the author calls "general systems", and uses this to prove that the set of rationalizable strategies in a finite non-cooperative game is the largest fixpoint of a certain operator. That operator coincides with the global version of the elimination of never-best-responses.

Some of the results presented here were initially reported in a different presentation, in [13].

## 2. Preliminaries

### 2.1. Strategic Games

Given $n$ players ($n > 1$) by a **strategic game** (in short, a **game**) we mean a sequence $(S_1, \ldots, S_n, p_1, \ldots, p_n)$, where for all $i \in \{1, \ldots, n\}$

- $S_i$ is the non-empty set of **strategies** available to player $i$,

- $p_i$ is the **payoff function** for the player $i$, so $p_i : S_1 \times \ldots \times S_n \to \mathcal{R}$, where $\mathcal{R}$ is the set of real numbers.

We denote the strategies of player $i$ by $s_i$, possibly with some superscripts. We call the elements of $S_1 \times \ldots \times S_n$ **joint strategies**. Given a joint strategy $s$ we denote the $i$th element of $s$ by $s_i$, write sometimes $s$ as $(s_i, s_{-i})$, and use the following standard notation:

- $s_{-i} := (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$,

- $S_{-i} := S_1 \times \ldots \times S_{i-1} \times S_{i+1} \times \ldots \times S_n$.

Given a finite non-empty set $A$ we denote by $\Delta A$ the set of probability distributions over $A$ and call any element of $\Delta S_i$ a **mixed strategy** of player $i$.

In the remainder of the paper we assume an initial strategic game

$$H := (H_1, \ldots, H_n, p_1, \ldots, p_n)$$

A **restriction** of $H$ is a sequence $(G_1, \ldots, G_n)$ such that $G_i \subseteq H_i$ for all $i \in \{1, \ldots, n\}$. Some of $G_i$s can be the empty set. We identify the restriction $(H_1, \ldots, H_n)$ with $H$. We shall focus on the complete lattice that consists of the set of all restrictions of the game $H$ ordered by the componentwise set inclusion:

$$(G_1, \ldots, G_n) \subseteq (G'_1, \ldots, G'_n) \text{ iff } G_i \subseteq G'_i \text{ for all } i \in \{1, \ldots, n\}$$

So in this lattice $H$ is the largest element in this lattice.

### 2.2. Possibility Correspondences

In this and the next subsection we essentially follow the survey of [14]. Fix a non-empty set $\Omega$ of **states**. By an **event** we mean a subset of $\Omega$.

A **possibility correspondence** is a mapping from $\Omega$ to the powerset $\mathcal{P}(\Omega)$ of $\Omega$. We consider three properties of a possibility correspondence $P$:

(i) for all $\omega$, $P(\omega) \neq \emptyset$,

(ii) for all $\omega$ and $\omega'$, $\omega' \in P(\omega)$ implies $P(\omega') = P(\omega)$,

(iii) for all $\omega$, $\omega \in P(\omega)$.

If the possibility correspondence satisfies properties (i) and (ii), we call it a ***belief correspondence*** and if it satisfies properties (i)–(iii), we call it a ***knowledge correspondence***.[2] Note that each knowledge correspondence $P$ yields a partition $\{P(\omega) \mid \omega \in \Omega\}$ of $\Omega$.

Assume now that each player $i$ has at its disposal a possibility correspondence $P_i$. Fix an event $E$. We define

$$\Box E := \Box^1 E := \{\omega \in \Omega \mid \forall i \in \{1, \ldots, n\} \; P_i(\omega) \subseteq E\}$$

by induction on $k \geq 1$

$$\Box^{k+1} E := \Box\Box^k E$$

and finally

$$\Box^* E := \bigcap_{k=1}^{\infty} \Box^k E$$

If all $P_i$s are belief correspondences, we usually write $B$ instead of $\Box$ and if all $P_i$s are knowledge correspondences, we usually write $K$ instead of $\Box$. When $\omega \in B^* E$, we say that the event $E$ is ***common belief in the state*** $\omega$ and when $\omega \in K^* E$, we say that the event $E$ is ***common knowledge in the state*** $\omega$.

An event $F$ is called ***evident*** if $F \subseteq \Box F$. That is, $F$ is evident if for all $\omega \in F$ we have $P_i(\omega) \subseteq F$ for all $i \in \{1, \ldots, n\}$. In what follows we shall use the following alternative characterizations of common belief and common knowledge based on evident events:

$$\omega \in \Box^* E \text{ iff for some evident event } F \text{ we have } \omega \in F \subseteq \Box E \tag{1}$$

where $\Box = B$ or $\Box = K$ (see [16], respectively Proposition 4 on page 180 and Proposition on page 174), and

$$\omega \in K^* E \text{ iff for some evident event } F \text{ we have } \omega \in F \subseteq E \tag{2}$$

([17], page 1237).

### 2.3. Models for Games

We now relate these considerations to strategic games. Given a restriction $G := (G_1, \ldots, G_n)$ of the initial game $H$, by a ***model*** for $G$ we mean a set of states $\Omega$ together with a sequence of functions $\overline{s_i} : \Omega \to G_i$, where $i \in \{1, \ldots, n\}$. We denote it by $(\Omega, \overline{s_1}, \ldots, \overline{s_n})$.

In what follows, given a function $f$ and a subset $E$ of its domain, we denote by $f(E)$ the range of $f$ on $E$ and by $f \,|\, E$ the restriction of $f$ to $E$.

By the ***standard model*** $\mathcal{M}$ for $G$ we mean the model in which

---

[2]Note that the notion of a belief has two meanings in the literature on epistemic analysis of strategic games, so also in this paper. From the context it is always clear which notion is used. In the modal logic terminology a belief correspondence is a frame for the modal logic KD45 and a knowledge correspondence is a frame for the modal logic S5, see, e.g. [15].

- $\Omega := G_1 \times \ldots \times G_n$

- $\overline{s_i}(\omega) := \omega_i$, where $\omega = (\omega_1, \ldots, \omega_n)$

So the states of the standard model for $G$ are exactly the joint strategies in $G$, and each $\overline{s_i}$ is a projection function. Since the initial game $H$ is given, we know the payoff functions $p_1, \ldots, p_n$. So in the context of $H$ the standard model is an alternative way of representing a restriction of $H$.

Given a (not necessarily standard) model $\mathcal{M} := (\Omega, \overline{s_1}, \ldots, \overline{s_n})$ for a restriction $G$ and a sequence of events $\overline{E} = (E_1, \ldots, E_n)$ in $\mathcal{M}$ (*i.e.*, of subsets of $\Omega$) we define

$$G_{\overline{E}} := (\overline{s_1}(E_1), \ldots, \overline{s_n}(E_n))$$

and call it the ***restriction of*** $G$ ***to*** $\overline{E}$. When each $E_i$ equals $E$ we write $G_E$ instead of $G_{\overline{E}}$.

Finally, we extend the notion of a model for a restriction $G$ to a ***belief model*** for $G$ by assuming that each player $i$ has a belief correspondence $P_i$ on $\Omega$. If each $P_i$ is a knowledge correspondence, we refer then to a ***knowledge model***. We write each belief model as

$$(\Omega, \overline{s_1}, \ldots, \overline{s_n}, P_1, \ldots, P_n)$$

## 2.4. Operators

Consider a fixed complete lattice $(D, \subseteq)$ with the largest element $\top$. In what follows we use ordinals and denote them by $\alpha, \beta, \gamma$. Given a, possibly transfinite, sequence $(G_\alpha)_{\alpha < \gamma}$ of elements of $D$ we denote their join and meet respectively by $\bigcup_{\alpha < \gamma} G_\alpha$ and $\bigcap_{\alpha < \gamma} G_\alpha$.

Let $T$ be an operator on $(D, \subseteq)$, *i.e.*, $T : D \to D$.

- We call $T$ ***monotonic*** if for all $G, G'$, $G \subseteq G'$ implies $T(G) \subseteq T(G')$, and ***contracting*** if for all $G$, $T(G) \subseteq G$.

- We say that an element $G$ is a ***fixpoint*** of $T$ if $G = T(G)$ and a ***post-fixpoint*** of $T$ if $G \subseteq T(G)$.

- We define by transfinite induction a sequence of elements $T^\alpha$ of $D$, where $\alpha$ is an ordinal, as follows:

    - $T^0 := \top$,

    - $T^{\alpha+1} := T(T^\alpha)$,

    - for all limit ordinals $\beta$, $T^\beta := \bigcap_{\alpha < \beta} T^\alpha$.

- We call the least $\alpha$ such that $T^{\alpha+1} = T^\alpha$ the ***closure ordinal*** of $T$ and denote it by $\alpha_T$. We call then $T^{\alpha_T}$ the ***outcome of*** (iterating) $T$ and write it alternatively as $T^\infty$.

So an outcome is a fixpoint reached by a transfinite iteration that starts with the largest element. In general, the outcome of an operator does not need to exist but we have the following classic result due to [18].[3]

---

[3] We use here its 'dual' version in which the iterations start at the largest and not at the least element of a complete lattice.

**Tarski's Fixpoint Theorem** Every monotonic operator $T$ on $(D, \subseteq)$ has an outcome, *i.e.*, $T^\infty$ is well-defined. Moreover,

$$T^\infty = \nu T = \cup\{G \mid G \subseteq T(G)\}$$

where $\nu T$ is the largest fixpoint of $T$.

In contrast, a contracting operator does not need to have a largest fixpoint. But we have the following obvious observation.

**Note 1.** *Every contracting operator $T$ on $(D, \subseteq)$ has an outcome,* i.e., *$T^\infty$ is well-defined.*                □

In Section 4 we shall need the following lemma, that modifies the corresponding lemma from [8] from finite to arbitrary complete lattices.

**Lemma 1.** *Consider two operators $T_1$ and $T_2$ on $(D, \subseteq)$ such that*

- *for all $G$, $T_1(G) \subseteq T_2(G)$,*

- *$T_1$ is monotonic,*

- *$T_2$ is contracting.*

*Then $T_1^\infty \subseteq T_2^\infty$.*

**Proof.** We first prove by transfinite induction that for all $\alpha$

$$T_1^\alpha \subseteq T_2^\alpha \tag{3}$$

By the definition of the iterations we only need to consider the induction step for a successor ordinal. So suppose the claim holds for some $\alpha$. Then by the first two assumptions and the induction hypothesis we have the following string of inclusions and equalities:

$$T_1^{\alpha+1} = T_1(T_1^\alpha) \subseteq T_1(T_2^\alpha) \subseteq T_2(T_2^\alpha) = T_2^{\alpha+1}$$

This shows that for all $\alpha$ (3) holds. By Tarski's Fixpoint Theorem and Note 1 the outcomes of $T_1$ and $T_2$ exist, which implies the claim.                □

### 2.5.  *Iterated Elimination of Non-Rational Strategies*

In this paper we are interested in analyzing situations in which each player pursues his own notion of rationality and this information is common knowledge or true common belief. As a special case we cover then the usually analyzed situation in which all players use the same notion of rationality.

Given player $i$ in the initial strategic game $H := (H_1, \ldots, H_n, p_1, \ldots, p_n)$ we formalize his notion of rationality using an ***optimality property*** $\phi(s_i, G_i, G_{-i})$ that holds between a strategy $s_i \in H_i$, a set $G_i$ of strategies of player $i$ and a set $G_{-i}$ of joint strategies of his opponents. Intuitively, $\phi_i(s_i, G_i, G_{-i})$ holds if $s_i$ is an 'optimal' strategy for player $i$ within the restriction $G := (G_i, G_{-i})$, assuming that he uses the property $\phi_i$ to select optimal strategies. In Section 4 we shall provide several natural examples of such properties.

We say that the property $\phi_i$ used by player $i$ is **monotonic** if for all $G_{-i}, G'_{-i} \subseteq H_{-i}$ and $s_i \in H_i$

$$G_{-i} \subseteq G'_{-i} \text{ and } \phi(s_i, H_i, G_{-i}) \text{ imply } \phi(s_i, H_i, G'_{-i})$$

So monotonicity refers to the situation in which the set of strategies of player $i$ is set to $H_i$ and the set of joint strategies of player $i$'s opponents is increased.

Each sequence of properties $\phi := (\phi_1, \ldots, \phi_n)$ determines an operator $T_\phi$ on the restrictions of $H$ defined by

$$T_\phi(G) := G'$$

where $G := (G_1, \ldots, G_n)$, $G' := (G'_1, \ldots, G'_n)$, and for all $i \in \{1, \ldots, n\}$

$$G'_i := \{s_i \in G_i \mid \phi_i(s_i, H_i, G_{-i})\}$$

Note that in defining the set of strategies $G'_i$ we use in the second argument of $\phi_i$ the set $H_i$ of player's $i$ strategies in the *initial* game $H$ and not in the *current* restriction $G$. This captures the idea that at every stage of the elimination process player $i$ analyzes the status of each strategy in the context of his initial set of strategies.

Since $T_\phi$ is contracting, by Note 1 it has an outcome, *i.e.*, $T_\phi^\infty$ is well-defined. Moreover, if each $\phi_i$ is monotonic, then $T_\phi$ is monotonic and by Tarski's Fixpoint Theorem its largest fixpoint $\nu T_\phi$ exists and equals $T_\phi^\infty$. Finally, $G$ is a fixpoint of $T_\phi$ iff for all $i \in \{1, \ldots, n\}$ and all $s_i \in G_i$, $\phi_i(s_i, H_i, G_{-i})$ holds.

Intuitively, $T_\phi(G)$ is the result of removing from $G$ all strategies that are not $\phi_i$-rational. So the outcome of $T_\phi$ is the result of the iterated elimination of strategies that for player $i$ are not $\phi_i$-rational.

## 3. Two Theorems

We now assume that each player $i$ employs some property $\phi_i$ to select his strategies, and we analyze the situation in which this information is true common belief or common knowledge. To determine which strategies are then selected by the players we shall use the $T_\phi$ operator.

We begin by fixing a belief model $(\Omega, \overline{s_1}, \ldots, \overline{s_n}, P_1, \ldots, P_n)$ for the initial game $H$. Given an optimality property $\phi_i$ of player $i$ we say that player $i$ is $\phi_i$-**rational in the state** $\omega$ if $\phi_i(\overline{s_i}(\omega), H_i, (G_{P_i(\omega)})_{-i})$ holds. Note that when player $i$ believes (respectively, knows) that the state is in $P_i(\omega)$, the set $(G_{P_i(\omega)})_{-i}$ represents his belief (respectively, his knowledge) about other players' strategies. That is, $(H_i, (G_{P_i(\omega)})_{-i})$ is the restriction he believes (respectively, knows) to be relevant to his choice.

Hence $\phi_i(\overline{s_i}(\omega), H_i, (G_{P_i(\omega)})_{-i})$ captures the idea that if player $i$ uses $\phi_i$ to select his strategy in the game he considers relevant, then in the state $\omega$ he indeed acts 'rationally'.

To reason about common knowledge and true common belief we introduce the event

$$\mathbf{RAT}(\phi) := \{\omega \in \Omega \mid \text{each player } i \text{ is } \phi_i\text{-rational in } \omega\}$$

and consider the following two events constructed out of it: $K^*\mathbf{RAT}(\phi)$ and $\mathbf{RAT}(\phi) \cap B^*\mathbf{RAT}(\phi)$. We then focus on the corresponding restrictions $G_{K^*\mathbf{RAT}(\phi)}$ and $G_{\mathbf{RAT}(\phi) \cap B^*\mathbf{RAT}(\phi)}$.

So strategy $s_i$ is an element of the $i$th component of $G_{K^*\mathbf{RAT}(\phi)}$ if $s_i = \overline{s_i}(\omega)$ for some $\omega \in K^*\mathbf{RAT}(\phi)$. That is, $s_i$ is a strategy that player $i$ chooses in a state in which it is common knowledge that each player $j$ is $\phi_j$-rational, and similarly for $G_{\mathbf{RAT}(\phi) \cap B^*\mathbf{RAT}(\phi)}$.

The following result then relates for arbitrary strategic games the restrictions $G_{\textbf{RAT}(\phi) \cap B^*\textbf{RAT}(\phi)}$ and $G_{K^*\textbf{RAT}(\phi)}$ to the outcome of the iteration of the operator $T_\phi$.

**Theorem 1.**

(i) *Suppose that each property $\phi_i$ is monotonic. Then for all belief models for $H$*

$$G_{\textbf{RAT}(\phi) \cap B^*\textbf{RAT}(\phi)} \subseteq T_\phi^\infty$$

(ii) *Suppose that each property $\phi_i$ is monotonic. Then for all knowledge models for $H$*

$$G_{K^*\textbf{RAT}(\phi)} \subseteq T_\phi^\infty$$

(iii) *For some standard knowledge model for $H$*

$$T_\phi^\infty \subseteq G_{K^*\textbf{RAT}(\phi)}$$

So part $(i)$ (respectively, $(ii)$) states that true common belief (respectively, common knowledge) of $\phi_i$-rationality of each player $i$ implies that the players will choose only strategies that survive the iterated elimination of non-$\phi$-rational strategies.

**Proof.**
$(i)$ Fix a belief model $(\Omega, \overline{s_1}, \ldots, \overline{s_n}, P_1, \ldots, P_n)$ for $H$. Take a strategy $s_i$ that is an element of the $i$th component of $G_{\textbf{RAT}(\phi) \cap B^*\textbf{RAT}(\phi)}$. Thus we have $s_i = \overline{s_i}(\omega)$ for some state $\omega$ such that $\omega \in \textbf{RAT}(\phi)$ and $\omega \in B^*\textbf{RAT}(\phi)$. The latter implies by (1) that for some evident event $F$

$$\omega \in F \subseteq \{\omega' \in \Omega \mid \forall i \in \{1, \ldots, n\}\ P_i(\omega') \subseteq \textbf{RAT}(\phi)\} \tag{4}$$

Take now an arbitrary $\omega' \in F \cap \textbf{RAT}(\phi)$ and $i \in \{1, \ldots, n\}$. Since $\omega' \in \textbf{RAT}(\phi)$, it holds that player $i$ is $\phi_i$-rational in $\omega'$, *i.e.*, $\phi_i(\overline{s_i}(\omega'), H_i, (G_{P_i(\omega')})_{-i})$ holds. But $F$ is evident, so $P_i(\omega') \subseteq F$. Moreover by (4) $P_i(\omega') \subseteq \textbf{RAT}(\phi)$, so $P_i(\omega') \subseteq F \cap \textbf{RAT}(\phi)$. Hence $(G_{P_i(\omega')})_{-i} \subseteq (G_{F \cap \textbf{RAT}(\phi)})_{-i}$ and by the monotonicity of $\phi_i$ we conclude that $\phi_i(\overline{s_i}(\omega'), H_i, (G_{F \cap \textbf{RAT}(\phi)})_{-i})$ holds.

By the definition of $T_\phi$ this means that $G_{F \cap \textbf{RAT}(\phi)} \subseteq T_\phi(G_{F \cap \textbf{RAT}(\phi)})$, *i.e.* $G_{F \cap \textbf{RAT}(\phi)}$ is a post-fixpoint of $T_\phi$. But $T_\phi$ is monotonic since each property $\phi_i$ is. Hence by Tarski's Fixpoint Theorem $G_{F \cap \textbf{RAT}(\phi)} \subseteq T_\phi^\infty$. But $s_i = \overline{s_i}(\omega)$ and $\omega \in F \cap \textbf{RAT}(\phi)$, so we conclude by the above inclusion that $s_i$ is an element of the $i$th component of $T_\phi^\infty$. This proves the claim.

$(ii)$ By the definition of common knowledge for all events $E$ we have $K^*E \subseteq E$. Hence for all $\phi$ we have $K^*\textbf{RAT}(\phi) \subseteq \textbf{RAT}(\phi) \cap K^*\textbf{RAT}(\phi)$ and consequently $G_{K^*\textbf{RAT}(\phi)} \subseteq G_{\textbf{RAT}(\phi) \cap K^*\textbf{RAT}(\phi)}$.

So part (ii) follows from part (i).

$(iii)$ Suppose $T_\phi^\infty = (G_1, \ldots, G_n)$. Consider the event $F := G_1 \times \ldots \times G_n$ in the standard model for $H$. Then $G_F = T_\phi^\infty$. Define each possibility correspondence $P_i$ by

$$P_i(\omega) := \begin{cases} F & \text{if } \omega \in F \\ \Omega \setminus F & \text{otherwise} \end{cases}$$

Each $P_i$ is a knowledge correspondence (also when $F = \emptyset$ or $F = \Omega$) and clearly $F$ is an evident event.

Take now an arbitrary $i \in \{1, \ldots, n\}$ and an arbitrary state $\omega \in F$. Since $T_\phi^\infty$ is a fixpoint of $T_\phi$ and $\overline{s_i}(\omega) \in G_i$ we have $\phi_i(\overline{s_i}(\omega), H_i, (T_\phi^\infty)_{-i})$, so by the definition of $P_i$ we have $\phi_i(\overline{s_i}(\omega), H_i, (G_{P_i(\omega)})_{-i})$. This shows that each player $i$ is $\phi_i$-rational in each state $\omega \in F$, *i.e.*, $F \subseteq \mathbf{RAT}(\phi)$.

Since $F$ is evident, we conclude by (2) that in each state $\omega \in F$ it is common knowledge that each player $i$ is $\phi_i$-rational, *i.e.*, $F \subseteq K^*\mathbf{RAT}(\phi)$. Consequently

$$T_\phi^\infty = G_F \subseteq G_{K^*\mathbf{RAT}(\phi)}$$

<div align="right">□</div>

Items $(i)$ and $(ii)$ show that when each property $\phi_i$ is monotonic, for all belief models of $H$ it holds that the joint strategies that the players choose in the states in which each player $i$ is $\phi_i$-rational and it is common belief that each player $i$ is $\phi_i$-rational (or in which it is common knowledge that each player $i$ is $\phi_i$-rational) are included in those that remain after the iterated elimination of the strategies that are not $\phi_i$-rational.

Note that monotonicity of the $\phi_i$ properties was not needed to establish item $(iii)$.

By instantiating the $\phi_i$'s with specific properties we get instances of the above result that refer to specific definitions of rationality. This will allow us to relate the above result to the ones established in the literature. Before we do this we establish a result that identifies a large class of properties $\phi_i$ for which Theorem 1 does not apply.

**Theorem 2.** *Suppose that a joint strategy $s \notin T_\phi^\infty$ exists such that*

$$\phi_i(s_i, H_i, (\{s_j\}_{j \neq i}))$$

*holds all $i \in \{1, \ldots, n\}$. Then for some knowledge model for $H$ the inclusion*

$$G_{K^*\mathbf{RAT}(\phi)} \subseteq T_\phi^\infty$$

*does not hold.*

**Proof.** We extend the standard model for $H$ by the knowledge correspondences $P_1, \ldots, P_n$ where for all $i \in \{1, \ldots, n\}$, $P_i(\omega) = \{\omega\}$. Then for all $\omega$ and all $i \in \{1, \ldots, n\}$

$$G_{P_i(\omega)} = (\{\overline{s_1}(\omega)\}, \ldots, \{\overline{s_n}(\omega)\})$$

Let $\omega' := s$. Then for all $i \in \{1, \ldots, n\}$, $G_{P_i(\omega')} = (\{s_1\}, \ldots, \{s_n\})$, so by the assumption each player $i$ is $\phi_i$-rational in $\omega'$, *i.e.*, $\omega' \in \mathbf{RAT}(\phi)$. By the definition of $P_i$s the event $\{\omega'\}$ is evident and $\omega' \in K\mathbf{RAT}(\phi)$. So by (1) $\omega' \in K^*\mathbf{RAT}(\phi)$. Consequently $s = (\overline{s_1}(\omega'), \ldots, \overline{s_n}(\omega')) \in G_{K^*\mathbf{RAT}(\phi)}$.

This yields the desired conclusion by the choice of $s$. <div align="right">□</div>

## 4. Applications

We now analyze to what customary game-theoretic properties the above two results apply. By a ***belief*** of player $i$ about the strategies his opponents play given the set $G_{-i}$ of their joint strategies we mean one of the following possibilities:

- a joint strategy of the opponents of player $i$, *i.e.*, $s_{-i} \in G_{-i}$, called a ***point belief***,

- or, in the case the game is finite, a joint mixed strategy of the opponents of player $i$ (*i.e.*, $(m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n)$, where $m_j \in \Delta G_j$ for all $j \neq i$), called an ***independent belief***,

- or, in the case the game is finite, an element of $\Delta G_{-i}$, called a ***correlated belief***.

In the second and third case the payoff function $p_i$ can be lifted in the standard way to an ***expected payoff*** function $p_i : H_i \times \mathcal{B}_i(G_{-i}) \to \mathcal{R}$, where $\mathcal{B}_i(G_{-i})$ is the corresponding set of beliefs of player $i$ held given $G_{-i}$.

We use below the following abbreviations, where $s_i, s_i' \in H_i$ and $G_{-i}$ is a set of the strategies of the opponents of player $i$:

- (***strict dominance***) $s_i' \succ_{G_{-i}} s_i$ for

  $$\forall s_{-i} \in G_{-i} \; p_i(s_i', s_{-i}) > p_i(s_i, s_{-i})$$

- (***weak dominance***) $s_i' \succ^w_{G_{-i}} s_i$ for

  $$\forall s_{-i} \in G_{-i} \; p_i(s_i', s_{-i}) \geq p_i(s_i, s_{-i}) \wedge \exists s_{-i} \in G_{-i} \; p_i(s_i', s_{-i}) > p_i(s_i, s_{-i})$$

In the case of finite games the relations $\succ_{G_{-i}}$ and $\succ^w_{G_{-i}}$ between a mixed strategy and a pure strategy are defined in the same way.

We now introduce natural examples of the optimality notion.

- $sd_i(s_i, G_i, G_{-i}) \equiv \neg \exists s_i' \in G_i \; s_i' \succ_{G_{-i}} s_i$

- (assuming $H$ is finite) $msd_i(s_i, G_i, G_{-i}) \equiv \neg \exists m_i' \in \Delta G_i \; m_i' \succ_{G_{-i}} s_i$

- $wd_i(s_i, G_i, G_{-i}) \equiv \neg \exists s_i' \in G_i \; s_i' \succ^w_{G_{-i}} s_i$

- (assuming $H$ is finite) $mwd_i(s_i, G_i, G_{-i}) \equiv \neg \exists m_i' \in \Delta G_i \; m_i' \succ^w_{G_{-i}} s_i$

- $br_i(s_i, G_i, G_{-i}) \equiv \exists \mu_i \in \mathcal{B}_i(G_{-i}) \; \forall s_i' \in G_i \; p_i(s_i, \mu_i) \geq p_i(s_i', \mu_i)$

So $sd_i$ and $wd_i$ are the customary notions of strict and weak dominance and $msd_i$ and $mwd_i$ are their counterparts for the case of dominance by a mixed strategy. Note that the notion $br_i$ of best response, comes in three 'flavours' depending on the choice of the set $\mathcal{B}_i(G_{-i})$ of beliefs.

Consider now the iterated elimination of strategies as defined in Subsection 2.5, so *with* the repeated reference by player $i$ to the strategy set $H_i$. For the optimality notion $sd_i$ such a version of iterated elimination was studied in [2], for $mwd_i$ it was used in [4], while for $br_i$ it corresponds to the rationalizability notion of [3].

In [10], [2] and [7] examples are provided showing that for the properties $sd_i$ and $br_i$ in general transfinite iterations (*i.e.*, iterations beyond $\omega_0$) of the corresponding operator are necessary to reach the outcome. So to establish for them part $(iii)$ of Theorem 1 transfinite iterations of the $T_\phi$ operator are necessary.

The following lemma holds.

**Lemma 2.** *The properties $sd_i$, $msd_i$ and $br_i$ are monotonic.*

**Proof.** Straightforward. $\square$

So Theorem 1 applies to the above three properties. In contrast, Theorem 1 does not apply to the remaining two properties $wd_i$ and $mwd_i$, since, as indicated in [8], the corresponding operators $T_{wd}$ and $T_{mwd}$ are not monotonic, and hence the properties $wd_i$ and $mwd_i$ are not monotonic.

In fact, the desired inclusion does not hold and Theorem 2 applies to these two optimality properties. Indeed, consider the following game:

|   | $L$ | $R$ |
|---|-----|-----|
| $U$ | $1,1$ | $0,1$ |
| $D$ | $1,0$ | $1,1$ |

Then the outcome of iterated elimination for both $wd_i$ and $mwd_i$ yields $G := (\{D\}, \{R\})$. Further, we have $wd_1(U, \{U, D\}, \{L\})$ and $wd_2(L, \{L, R\}, \{U\})$, and analogously for $mwd_1$ and $mwd_2$.

So the joint strategy $(U, L)$ satisfies the conditions of Theorem 2 for both $wd_i$ and $mwd_i$. Note that this game also furnishes an example for non-monotonicity of $wd_i$ since $wd_1(U, \{U, D\}, \{L, R\})$ does not hold.

This shows that the optimality notions $wd_i$ and $mwd_i$ cannot be justified in the used epistemic framework as 'stand alone' concepts of rationality.

## 5. Consequences of Common Knowledge of Rationality

In this section we show that common knowledge of rationality is sufficient to entail the customary iterated elimination of strictly dominated strategies. We also show that weak dominance is not amenable to such a treatment.

Given a sequence of properties $\phi := (\phi_1, \ldots, \phi_n)$, we introduce an operator $U_\phi$ on the restrictions of $H$ defined by

$$U_\phi(G) := G',$$

where $G := (G_1, \ldots, G_n)$, $G' := (G'_1, \ldots, G'_n)$, and for all $i \in \{1, \ldots, n\}$

$$G'_i := \{s_i \in G_i \mid \phi_i(s_i, G_i, G_{-i})\}.$$

So when defining the set of strategies $G'_i$ we use in the second argument of $\phi_i$ the set $G_i$ of player's $i$ strategies in the *current* restriction $G$. That is, $U_\phi(G)$ determines the 'locally' $\phi$-optimal strategies in $G$. In contrast, $T_\phi(G)$ determines the 'globally' $\phi$-optimal strategies in $G$, in that each player $i$ must consider all of his strategies $s'_i$ that occur in his strategy set $H_i$ in the *initial game $H$*.

So the 'global' form of optimality coincides with rationality, as introduced in Subsection 2.5, while the customary definition of iterated elimination of strictly (or weakly) dominated strategies refers to the iterations of the appropriate instantiation of the 'local' $U_\phi$ operator.

Note that the $U_\phi$ operator is non-monotonic for all non-trivial optimality notions $\phi_i$ such that $\phi_i(s_i, \{s_i\}, (\{s_j\}_{j \neq i}))$ for all joint strategies $s$, so in particular for $br_i, sd_i, msd_i, wd_i$ and $mwd_i$. Indeed, given $s$ let $G_s$ denote the corresponding restriction in which each player $i$ has a single strategy $s_i$.

Each restriction $G_s$ is a fixpoint of $U_\phi$. By non-triviality of $\phi_i$s we have $U_\phi(H) \neq H$, so for each restriction $G_s$ with $s$ including an eliminated strategy the inclusion $U_\phi(G_s) \subseteq U_\phi(H)$ does not hold, even though $G_s \subseteq H$. In contrast, as we saw, by virtue of Lemma 2 the $T_\phi$ operator is monotonic for $br_i, sd_i$ and $msd_i$.

First we establish the following consequence of Theorem 1. When each property $\phi_i$ equals $br_i$, we write here **RAT**$(br)$ and similarly with $U_{sd}$.

**Corollary 1.**

  (i) *For all belief models*

$$G_{\mathbf{RAT}(br) \cap B^* \mathbf{RAT}(br)} \subseteq U_{sd}^\infty$$

  (ii) *for all knowledge models*

$$G_{K^* \mathbf{RAT}(br)} \subseteq U_{sd}^\infty$$

*where in both situations we use in $br_i$ the set of poinr beliefs.*

**Proof.**
$(i)$ By Lemma 2 and Theorem 1$(i)$ $G_{\mathbf{RAT}(br) \cap B^* \mathbf{RAT}(br)} \subseteq T_{br}^\infty$ Each best response to a joint strategy of the opponents is not strictly dominated, so for all restrictions $G$

$$T_{br}(G) \subseteq T_{sd}(G)$$

Also, for all restrictions $G$, $T_{sd}(G) \subseteq U_{sd}(G)$. So by Lemma 1 $T_{br}^\infty \subseteq U_{sd}^\infty$, which concludes the proof.

$(ii)$ By part $(i)$ and the fact that $K^* \mathbf{RAT}(br) \subseteq \mathbf{RAT}(br)$. □

Part $(ii)$ formalizes and justifies in the epistemic framework used here the often used statement:

> common knowledge of rationality implies that the players will choose only strategies that survive the iterated elimination of strictly dominated strategies

for games with *arbitrary strategy sets* and *transfinite iterations* of the elimination process, and where best response means best response to a point belief.

In the case of finite games Theorem 1 implies the following result. For the case of independent beliefs it is implicitly stated in [19], explicitly formulated in [20] (see [14, page 181]) and proved using Harsanyi type spaces in [21].

**Corollary 2.** *Assume the initial game $H$ is finite.*

  (i) *For all belief models for $H$*

$$G_{\mathbf{RAT}(br) \cap B^* \mathbf{RAT}(br)} \subseteq U_{msd}^\infty,$$

  (ii) *for all knowledge models for $H$*

$$G_{K^* \mathbf{RAT}(br)} \subseteq U_{msd}^\infty,$$

*where in both situations we use in $br_i$ either the set of point beliefs or the set of independent beliefs or the set of correlated beliefs.*

**Proof.** The argument is analogous as in the previous proof but relies on a subsidiary result and runs as follows.

($i$) Denote respectively by $brp_i$, $bri_i$ and $brc_i$ the best response property w.r.t. *point*, *independent* and *correlated* beliefs of the opponents. Below $\phi$ stands for either $brp$, $bri$ or $brc$.

By Lemma 2 and Theorem 1 $G_{\mathbf{RAT}(\phi) \cap B^*\mathbf{RAT}(\phi)} \subseteq T_\phi^\infty$. Further, for all restrictions $G$ we have both $T_\phi(G) \subseteq U_\phi(G)$ and $U_{br}(G) \subseteq U_{bri}(G) \subseteq U_{brc}(G)$. So by Lemma 1 $T_\phi^\infty \subseteq U_{brc}^\infty$. But by the result of [22], (page 60) (that is a modification of the original result of [23]), for all restrictions $G$ we have $U_{brc}(G) = U_{msd}(G)$, so $U_{brc}^\infty = U_{msd}^\infty$, which yields the conclusion.

($ii$) By ($i$) and the fact that $K^*\mathbf{RAT}(br) \subseteq \mathbf{RAT}(br)$. □

Finally, let us clarify the situation for the remaining two optimality notions, $wd_i$ and $mwd_i$. For them the inclusions of Corollaries 1 and 2 do not hold. Indeed, it suffices to consider the following initial game $H$:

|   | $L$ | $R$ |
|---|-----|-----|
| $U$ | $1,0$ | $1,0$ |
| $D$ | $1,0$ | $0,0$ |

Here every strategy is a best response but $D$ is weakly dominated by $U$. So both $U_{wd}^\infty$ and $U_{mwd}^\infty$ are proper subsets of $T_{br}^\infty$. On the other hand by Theorem 1($iii$) for some standard knowledge model for $H$ we have $G_{K^*\mathbf{RAT}(br)} = T_{br}^\infty$. So for this knowledge model neither $G_{K^*\mathbf{RAT}(br)} \subseteq U_{wd}^\infty$ nor $G_{K^*\mathbf{RAT}(br)} \subseteq U_{mwd}^\infty$ holds.

## Acknowledgements

## References

1. Milgrom, P.; Roberts, J. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* **1990**, *58*, 1255–1278.

2. Chen, Y.C.; Long, N.V.; Luo, X. Iterated strict dominance in general games. *Games Econ. Behav.* **2007**, *61*, 299–315.

3. Bernheim, B.D. Rationalizable strategic behavior. *Econometrica* **1984**, *52*, 1007–1028.

4. Brandenburger, A.; Friedenberg, A.; Keisler, H. Admissibility in games. *Econometrica* **2008**, *76*, 307–352.

5. Halpern, J.; Pass, R. A logical characterization of iterated admissibility. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK XII)*; Stanford University,

Standord, CA, USA, 6–8 July 2009; Available online: http://arxiv.org/abs/0906.4326 (accessed on 30 September 2010).

6. van Benthem, J. Rational dynamics and epistemic logic in games. *Int. Game Theory Rev.* **2007**, *9*, 13–45.

7. Apt, K.R. The many faces of rationalizability. *The B.E. J. Theoretical Econ.* **2007**, 7, Article 18.; Available online: http://arxiv.org/abs/cs.GT/0608011 (accessed on 30th September 2010).

8. Apt, K.R. Relative Strength of Strategy Elimination Procedures. *Econ. Bull.* **2007**, *3*, 1–9; Available online: http://www.economicsbulletin.com/ (accessed on 30th September 2010).

9. Lipman, B.L. How to decide how to decide how to . . . : Modeling limited rationality. *Econometrica* **1991**, *59*, 1105–1125.

10. Lipman, B.L. A note on the implications of common knowledge of rationality. *Games Econ. Behav.* **1994**, *6*, 114–129.

11. Heifetz, A.; Samet, D. Knowledge spaces with arbitrarily high rank. *Games Econ. Behav.* **1998**, *22*, 260–273.

12. Luo, X. General systems and $\phi$-stable sets—A formal analysis of socioeconomic environments. *J. Math. Econ.* **2001**, *36*, 95–109.

13. Apt, K.R. Epistemic analysis of strategic games with arbitrary strategy sets. In *Proceedings of the 11th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK XI)*, Brussels, Belgium, June 25-27th, 2007; Available online: http://portal.acm.org (accessed on 30th September 2010).

14. Battigalli, P.; Bonanno, G. Recent results on belief, knowledge and the epistemic foundations of game theory. *Res. Econ.* **1999**, *53*, 149–225.

15. Blackburn, P.; de Rijke, M.; Venema, Y. *Modal Logic*; Cambridge University Press: Cambridge, UK, 2001.

16. Monderer, D.; Samet, D. Approximating common knowledge with common beliefs. *Games Econ. Behav.* **1989**, *1*, 170–190.

17. Aumann, R. Agreeing to disagree. *Ann. Stat.* **1976**, *4*, 1236–1239.

18. Tarski, A. A lattice-theoretic fixpoint theorem and its applications. *Pacific J. Math.* **1955**, *5*, 285–309.

19. Brandenburger, A.; Dekel, E. Rationalizability and correlated equilibria. *Econometrica* **1987**, *55*, 1391–1402.

20. Stalnaker, R. On the evaluation of solution concepts. *Theor. Decis.* **1994**, *37*, 49–73.

21. Brandenburger, A.; Friedenberg, A. Intrinsic correlation in games. *J. Econ. Theor.* **2008**, *141*, 28–67.

22. Osborne, M.J.; Rubinstein, A. *A Course in Game Theory*; The MIT Press: Cambridge, MA, USA, 1994.

23. Pearce, D.G. Rationalizable strategic behavior and the problem of perfection. *Econometrica* **1984**, *52*, 1029–1050.