Common Beliefs and Public Announcements in Strategic Games with Arbitrary Strategy Sets

Krzysztof R. Apt and Jonathan A. Zvesper*

September 26, 2007

Abstract

We provide an epistemic analysis of arbitrary strategic games based on possibility correspondences. Such an analysis calls for the use of transfinite iterations of the corresponding operators. In the case of common beliefs and common knowledge our approach is based on Tarski's Fixpoint Theorem and applies to 'monotonic' properties. In the case of an analysis based on the notion of a public announcement our approach applies to 'global properties'. Both classes of properties include the notions of rationalizability and the iterated elimination of strictly dominated strategies. We also provide an axiomatic presentation of the main results concerning common beliefs and monotonic properties.

Keywords: epistemic analysis, possibility correspondences, fixpoints, rationalizability, public announcements.

1 Introduction

1.1 Background

Epistemic analysis of strategic games (in short, games) aims at predicting the choices of rational players in the presence of (partial or common) knowledge

*Centrum for Mathematics and Computer Science (CWI), Kruislaan 413, 1098 SJ Amsterdam, the Netherlands, and University of Amsterdam

or belief about the behaviour of other players. Most often it focusses on the iterated elimination of never best responses (a notion termed as rationalizability), the iterated elimination of strictly dominated strategies (IESDS) and on justification of the strategies selected in Nash and correlated equilibria.

Starting with Aumann [1987], Brandenburger and Dekel [1987] and Tan and Werlang [1988] a large body of literature arose that investigates the epistemic foundations of rationalizability by modelling the reasoning employed by players in choosing their strategies. Such an analysis, based either on possibility correspondences and partition spaces, or Harsanyi type spaces, is limited either to finite or compact games with continuous payoffs, or to two-player games, see, e.g., Battigalli and Bonanno [1999] or Ely and Peski [2006].

In turn, in the case of IESDS the epistemic analysis has focussed on finite games (with an infinite hierarchy of beliefs) and strict dominance either by pure or by mixed strategies, see, e.g. Brandenburger, Friedenberg and Keisler [2004].

1.2 Contributions

In this paper we provide an epistemic analysis of arbitrary strategic games based on possibility correspondences. More specifically, denote by $\mathbf{RAT}(\overline{\phi})$ the property that each player i uses a monotonic property[1] ϕ_i to select his strategy ('each player i is ϕ_i-rational'). Then the following sets of strategy profiles coincide:

- those that the players choose in the states in which $\mathbf{RAT}(\overline{\phi})$ is common knowledge,
- those that the players choose in the states in which $\mathbf{RAT}(\overline{\phi})$ is true and is common belief,
- those that remain after the iterated elimination of the strategies that for player i are not ϕ_i-optimal.

This requires that transfinite iterations of the strategy elimination are allowed and covers the usual notion of rationalizability and a global version of the iterated elimination of strictly dominated strategies. For the customary, local version of the iterated elimination of strictly dominated strategies (that is defined using a non-monotonic property) we justify the statement

[1] The concepts of monotonic, global and local properties are introduced in Section 3.

> 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 Bernheim [1984].

We also provide an axiomatic presentation of some of these results. This clarifies the logical underpinnings of the epistemic analysis and shows that the use of transfinite iterations can be naturally captured by a single inference rule that involves greatest fixpoints. Also, it shows that the relevant monotonic properties can be defined using positive formulae.

Finally, inspired by van Benthem [2007], we provide an alternative characterization of the strategies that remain after iterated elimination of strategies that for player i are not ϕ_i-optimal, based on the concept of a public announcement due to Plaza [1989]. This yields a generalization of van Benthem's results to arbitrary strategic games and to other properties than rationalizability, and a global version of the iterated elimination of strictly dominated strategies.

Apart of the necessity of the use of transfinite iterations when studying arbitrary strategic games, our analysis shows the relevance of two concepts of the underlying properties ϕ_i used by the players to select their strategies. The first one is monotonicity, and it allows us to use Tarski's Fixpoint Theorem. The second is globality, which intuitively means that each subgame obtained by iterated elimination of strategies is analyzed *in the context* of the given initial game. While the epistemic analysis of arbitrary games based on possibility correspondences applies only to monotonic properties, the one based on public announcement applies to global properties.

1.3 Connections

Our results complement the findings of Lipman [1991] in which transfinite ordinals are used in a study of limited rationality, and Lipman [1994], where a two-player game is constructed for which the ω_0 (the first infinite ordinal) and $\omega_0 + 1$ iterations of the rationalizability operator of Bernheim [1984] differ. In turn, Heifetz and Samet [1998] show that in general arbitrary ordinals are necessary in the epistemic analysis of strategic games based on partition spaces. Further, as argued in Chen, Long and Luo [2005], the notion of IESDS à la Milgrom and Roberts [1990], when used for arbitrary games, also requires transfinite iterations of the underlying operator.

The relevance of monotonicity in the context of epistemic analysis of finite strategic games has already been pointed out in van Benthem [2007]. The distinction between local and global properties is from Apt [2007b] and Apt [2007c]. Some of the results presented here were initially reported in Apt [2007a].

2 Preliminaries

This paper connects three concepts, operators on a complete lattice, strategic games and possibility correspondences. In this section we introduce these concepts and recall basic results concerning them.

2.1 Operators

Consider a fixed complete lattice $(D, \subseteq)$ with the largest element $\top$. In what follows we use ordinals and denote them by α, β, γ. 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 $\bigcap_{\alpha<\gamma} G_\alpha$.

Definition 1. Let T be an operator on $(D, \subseteq)$, i.e., $T : D \rightarrow D$.

- We call T ***monotonic*** if for all G_1, G_2

$$G_1 \subseteq G_2 \text{ implies } T(G_1) \subseteq T(G_2).$$

- We call T ***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^α of D, where α is an ordinal, as follows:

  - $T^0 := \top$,
  - $T^{\alpha+1} := T(T^\alpha)$,

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

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

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 Tarski [1955].[2]

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

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

where νT 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^{∞} is well-defined.* □

In Section 5 we shall need the following lemma.

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 α

$$T_1^{\alpha} \subseteq T_2^{\alpha}. \tag{1}$$

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 α. Then by

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

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 α (1) holds. By Tarski's Fixpoint Theorem and Note 1 the outcomes of T_1 and T_2 exist, which implies the claim. □

2.2 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 each $i \in [1..n]$

- S_i is the non-empty set of ***strategies*** (sometimes called ***actions***) 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. Given $s \in S_1 \times \ldots \times S_n$ we denote the ith 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 ΔA the set of probability distributions over A and call any element of ΔS_i a ***mixed strategy*** of player i.

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

$$H := (T_1, \ldots, T_n, p_1, \ldots, p_n).$$

A ***restriction*** of H is a sequence $(S_1, \ldots, S_n)$ such that $S_i \subseteq T_i$ for $i \in [1..n]$. We identify the restriction $(T_1, \ldots, T_n)$ 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:

$$(S_1, \ldots, S_n) \subseteq (S'_1, \ldots, S'_n) \text{ iff } S_i \subseteq S'_i \text{ for all } i \in [1..n].$$

So H is the largest element in this lattice and $\bigcup_{\alpha<\gamma}$ and $\bigcap_{\alpha<\gamma}$ are the customary set-theoretic operations on the restrictions.

Consider now a restriction $G := (S_1, \ldots, S_n)$ of H and two strategies s_i, s'_i from T_i (so *not necessarily* from S_i). We say that s_i ***is strictly dominated on*** G by s'_i (and write $s'_i \succ_G s_i$) if

$$\forall s_{-i} \in S_{-i}\ p_i(s'_i, s_{-i}) > p_i(s_i, s_{-i}),$$

and that s_i ***is weakly dominated on*** G by s'_i (and write $s'_i \succ^w_G s_i$) if

$$\forall s_{-i} \in S_{-i}\ p_i(s'_i, s_{-i}) \geq p_i(s_i, s_{-i}) \wedge \exists s_{-i} \in S_{-i}\ p_i(s'_i, s_{-i}) > p_i(s_i, s_{-i}).$$

In the case of finite games, once the payoff function is extended in the expected way to mixed strategies, the relations $\succ_G$ and $\succ^w_G$ between a mixed strategy and a pure strategy are defined in the same way.

A ***belief*** of player i held in $G := (S_1, \ldots, S_n)$ can be

- a joint strategy of the opponents of player i in G (i.e., $s_{-i} \in S_{-i}$),
- 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 $m_j \in \Delta S_j$ for all j),
- or, in the case the game is finite, a ***correlated strategy*** of the opponents of player i (i.e., $m \in \Delta S_{-i}$).

Each payoff function p_i can be modified to an ***expected payoff*** function $p_i : S_i \times \mathcal{B}_i \rightarrow \mathcal{R}$, where $\mathcal{B}_i$ is one of the above three sets of beliefs of player i.

Further, given a restriction $G' := (S'_1, \ldots, S'_n)$ of H, we say that the strategy s_i from T_i is a ***best response in*** G' ***to some belief*** μ_i ***held in*** G if

$$\forall s'_i \in S'_i\ p_i(s_i, \mu_i) \geq p_i(s'_i, \mu_i).$$

2.3 Possibility correspondences

In this and the next subsection we essentially follow the exposition of Battigalli and Bonanno [1999]. Fix a non-empty set Ω of ***states***. By an ***event*** we mean a subset of Ω.

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

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

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

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

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***.[3] Note that each belief correspondence P yields a partition $\{P(\omega) \mid \omega \in \Omega\}$ of Ω.

Assume now that each player i has to its disposal a possibility correspondence P_i. Given an event E we define then

$$\Box E := \{\omega \in \Omega \mid \forall i \in [1..n]\ P_i(\omega) \subseteq E\}.$$

If all P_is are belief correspondences, we write BE instead of $\Box E$ and if all P_is are knowledge correspondences, we write KE instead of $\Box E$.

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 $i \in [1..n]$. Following Aumann [1976] if each P_i is a knowledge correspondence, we say that an event E is a ***common knowledge in the state*** $\omega \in \Omega$ if for some evident event F we have $\omega \in F \subseteq E$. We write then $\omega \in K^* E$.

Further, using a characterization of Monderer and Samet [1989], if each P_i is a belief correspondence, we say that an event E is a ***common belief in the state*** $\omega \in \Omega$ if for some evident event F we have $\omega \in F \subseteq BE$. We write then $\omega \in B^* E$.

These two definitions can be made even more similar: thanks to an observation of Monderer and Samet [1989] we can replace in the definition of common knowledge the condition $\omega \in F \subseteq E$ by $\omega \in F \subseteq KE$. This and the observation that $\Box(E \cap F) = \Box E \cap \Box F$ allows us to define common knowledge and common belief in a uniform way by putting

$$\Box^* E := \bigcup \{F \subseteq \Omega \mid F \subseteq \Box(E \cap F)\}.$$

Intuitively, an event E is a common knowledge in the state ω if in ω every player knows that every player knows that every player ... knows E, where we say that ***every player knows*** E in the state ω if $\omega \in KE$, and analogously with the common belief.

[3] 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. Blackburn, de Rijke and Venema [2001].

2.4 Models for games

We now link these considerations with the strategic games. Given a restriction $G := (S_1, \ldots, S_n)$ of the initial game H, by a ***model*** for G we mean a set Ω of states such that $|\Omega| \geq |S_i|$ for all $i \in [1..n]$ (where for a set A we denote its cardinality by $|A|$), together with a sequence of functions $s_i : \Omega \rightarrow S_i$, where $i \in [1..n]$. We denote it by $(\Omega, s_1, \ldots, s_n)$.

In what follows we use s_i interchangingly to denote a strategy of player i or a function that, when applied to a state, yields a strategy of player i. If a confusion arises, we write $s_i(\cdot)$ to denote a function. Further, 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

- $\Omega := S_1 \times \ldots \times S_n$ (which means that for $\omega \in \Omega$, ω_i is well-defined),
- $s_i(\omega) := \omega_i$.

So the states of the standard model for G are exactly the joint strategies in G, and each $s_i(\cdot)$ 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 a standard model is just an alternative way of representing a restriction of H.

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

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

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 Ω. If each P_i is a knowledge correspondence, we refer then to a ***knowledge model***.

3 Local and global properties

The assumption that each player is rational is one of the basic stipulations within the framework of strategic games. However, rationality can be differently interpreted by different players.[4] This may for example mean that a player

- does not choose a strategy weakly/strictly dominated by another pure/mixed strategy,
- chooses only best replies to the (beliefs about the) strategies of the opponents.

In this paper we are interested in analyzing situations in which each player pursues his own notion of rationality, more specifically those situations in which this information is common knowledge or 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 a strategic game $H := (T_1, \ldots, T_n, p_1, \ldots, p_n)$ we formalize his notion of rationality using a property $\phi_i(s_i, G, G')$ that holds between a state $s_i \in T_i$ and restrictions G and G' of H. Intuitively, $\phi_i(s_i, G, G')$ holds if s_i is an 'optimal' strategy for player i within the restriction G in the context of G', assuming that he uses the property ϕ to select optimal strategies.

Here are some examples of the property ϕ_i which show that the above-mentioned rationality notions can be formalized in a number of natural ways:

- $sd_i(s_i, G, G')$ that holds iff the strategy s_i of player i is not strictly dominated on G by any strategy from the restriction $G' := (S'_1, \ldots, S'_n)$ of H (i.e., $\neg \exists s'_i \in S'_i \; s'_i \succ_G s_i$),
- (assuming H is finite) $msd_i(s_i, G, G')$ that holds iff the strategy s_i of player i is not strictly dominated on G by any of its mixed strategy from the restriction $G' := (S'_1, \ldots, S'_n)$ of H, (i.e., $\neg \exists m'_i \in \Delta S'_i \; m'_i \succ_G s_i$),
- $wd_i(s_i, G, G')$ that holds iff the strategy s_i of player i is not weakly dominated on G by any strategy from the restriction $G' := (S'_1, \ldots, S'_n)$ of H (i.e., $\neg \exists s'_i \in S'_i \; s'_i \succ^w_G s_i$),
- (assuming H is finite) $mwd_i(s_i, G, G')$ that holds iff the strategy s_i of player i is not weakly dominated on G by any of its mixed strategy from the restriction $G' := (S'_1, \ldots, S'_n)$ of H, (i.e., $\neg \exists m'_i \in \Delta S'_i \; m'_i \succ^w_G s_i$),

[4]This matter is obfuscated by the fact that the etymologically related noun 'rationalizability' stands by now for the concept introduced in Bernheim [1984] and Pearce [1984] that refers to the outcome of iterated elimination of never best responses.

- $br_i(s_i, G, G')$ that holds iff the strategy s_i of player i is a best response in the restriction $G' := (S'_1, \ldots, S'_n)$ of H to some belief μ_i held in G (i.e., for some belief μ_i held in G, $\forall s'_i \in S'_i \; p_i(s_i, \mu_i) \geq p_i(s'_i, \mu_i)$).

Two natural possibilities for G' are $G' = H$ or $G' = G$. We then abbreviate $\phi_i(s_i, G, H)$ to $\phi_i^g(s_i, G)$ and $\phi_i(s_i, G, G)$ to $\phi_i^l(s_i, G)$ and henceforth focus on the binary properties $\phi_i(\cdot, \cdot)$. The superscript 'g' stands for ***global*** and 'l' for ***local***. Global properties are then those in which a player's strategy is evaluated with respect to all his strategies from the initial game, whereas local properties are concerned solely with a comparison of strategies available in the restriction G.

We say that the property $\phi_i(\cdot, \cdot)$ used by player i is ***monotonic*** if for all restrictions G and G' of H and $s_i \in T_i$

$$G \subseteq G' \text{ and } \phi(s_i, G) \text{ implies } \phi(s_i, G').$$

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

$$T_{\overline{\phi}}(G) := (S'_1, \ldots, S'_n),$$

where $G := (S_1, \ldots, S_n)$ and for all $i \in [1..n]$

$$S'_i := \{s_i \in S_i \mid \phi_i(s_i, G)\}.$$

Since $T_{\overline{\phi}}$ is contracting, by Note 1 it has an outcome, i.e., $T_{\overline{\phi}}^{\infty}$ is well-defined. Moreover, if each ϕ_i is monotonic, then $T_{\overline{\phi}}$ is monotonic and by Tarski's Fixpoint Theorem its largest fixpoint $\nu T_{\overline{\phi}}$ exists and equals $T_{\overline{\phi}}^{\infty}$.

Intuitively, $T_{\overline{\phi}}(G)$ is the result of removing from G all strategies that are not ϕ_i-optimal. So the outcome of $T_{\overline{\phi}}$ is the result of the iterated elimination of strategies that for player i are not ϕ_i-optimal, where $i \in [1..n]$.

When each property ϕ_i equals sd^l, we write T_{sd^l} instead of $T_{\overline{sd^l}}$ and similarly with other specific properties. The natural examples of such an iterated elimination of strategies that were discussed in the literature are:[5]

- iterated elimination of strategies that are strictly dominated by another strategy;

[5]The reader puzzled by the existence of multiple definitions for the apparently uniquely defined concepts is encouraged to consult Apt [2007b].

This corresponds to the iterations of the T_{sd^l} operator in the case of Dufwenberg and Stegeman [2002]) and of the T_{sd^g} operator in the case of Chen, Long and Luo [2005].

- iterated elimination of strategies that are weakly dominated by another strategy;
- (for finite games) iterated elimination of strategies that are weakly, respectively strictly, dominated by a mixed strategy;

  This are the customary situations studied starting with Luce and Raiffa [1957] that corresponds to the iterations of the T_{msd^l}, respectively T_{mwd^l}, operator.
- iterated elimination of strategies that are never best responses to some belief;

  This corresponds to the iterations of the T_{br^g} operator in the case of Bernheim [1984] and the T_{br^l} operator in the case of Pearce [1984], in each case for an appropriate set of beliefs.

Usually only the first ω_o iterations of the corresponding operator T are considered, i.e., one studies T^{ω_0}, that is $\bigcap_{i<\omega_0} T^i$, and not T^∞.

In the next section we assume that each player i employs some property ϕ_i to select his strategies, and we analyze the situation in which this information is common knowledge. To determine which strategies are then selected by the players we shall use the $T_{\overline{\phi}}$ operator. We shall also explain why in general transfinite iterations are necessary.

4 Two theorems

All belief and knowledge models considered in this section are models for the initial game H and use a fixed set of states Ω.

Given now a property $\phi_i(\cdot, G)$ that player i uses to select his strategies in the restriction G of H, we say that, given a belief model $\mathcal{M} := (\Omega, s_1, \ldots, s_n, P_1, \ldots, P_n)$ for H, player i is ***ϕ_i-rational in the state*** ω if $\phi_i(s_i(\omega), G_{P_i(\omega)})$ holds. Note that when player i knows (respectively, believes) that the state is in $P_i(\omega)$, the restriction $G_{P_i(\omega)}$ represents his knowledge (respectively, his belief) about the players' strategies. That is, $G_{P_i(\omega)}$ is the game he knows (respectively, believes) to be relevant to his choice. Hence

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

We are interested in the strategies selected by each player in the states in which it is common knowledge (or true and is common belief) that each player i is ϕ_i-rational. To this end, given a belief model $\mathcal{M} := (\Omega, s_1, \ldots, s_n, P_1, \ldots, P_n)$ for H we introduce the following event:

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

and focus on the following two events:

$$CK(\overline{\phi}) := \{\omega \in \Omega \mid \text{for some knowledge model } \omega \in K^*\mathbf{RAT}(\overline{\phi})\},$$

$$CB(\overline{\phi}) := \{\omega \in \Omega \mid \text{for some belief model } \omega \in \mathbf{RAT}(\overline{\phi}) \text{ and } \omega \in B^*\mathbf{RAT}(\overline{\phi})\},$$

and the corresponding restrictions $G_{CK(\overline{\phi})}$ and $G_{CB(\overline{\phi})}$ of H.

So a strategy s_i is an element of the ith component of $G_{CK(\overline{\phi})}$ if for some knowledge model $(\Omega, s_1(\cdot), \ldots, s_n(\cdot), P_1, \ldots, P_n)$ for H we have $s_i = s_i(\omega)$ for some $\omega \in K^*\mathbf{RAT}(\overline{\phi})$. That is, s_i is a strategy that player i chooses in a state in which for some knowledge model it is common knowledge that each player j is ϕ_j-rational, and similarly for $G_{CB(\overline{\phi})}$.

The following result then relates for arbitrary strategic games the restriction $G_{CB(\overline{\phi})}$ to the outcome of the iteration of the operator $T_{\overline{\phi}}$.

Theorem 1. *Suppose that each property ϕ_i is monotonic. Then*

$$G_{CK(\overline{\phi})} = G_{CB(\overline{\phi})} = T_{\overline{\phi}}^{\infty}.$$

Proof. We prove three inclusions.
(i) $G_{CK(\overline{\phi})} \subseteq G_{CB(\overline{\phi})}$.

This is an immediate consequence of the inclusion $CK(\overline{\phi}) \subseteq CB(\overline{\phi})$ that holds for an arbitrary $\overline{\phi}$ on the account of the alternative characterization of common knowledge mentioned in Subsection 2.3 and the fact that for all events E we have $K^*E \subseteq E$.

(ii) $G_{CB(\overline{\phi})} \subseteq T_{\overline{\phi}}^{\infty}$.

Take a strategy s_i that is an element of the ith component of $G_{CB(\overline{\phi})}$. Thus for some belief model $\mathcal{M} := (\Omega, s_1(\cdot), \ldots, s_n(\cdot), P_1, \ldots, P_n)$ for H we have $s_i = s_i(\omega)$ for some $\omega \in CB(\overline{\phi})$. So $\omega \in \mathbf{RAT}(\overline{\phi})$ and $\omega \in B^*\mathbf{RAT}(\overline{\phi})$. The latter implies that for some evident event F

$$\omega \in F \subseteq \{\omega' \in \Omega \mid \forall i \in [1..n]\ P_i(\omega') \subseteq \mathbf{RAT}(\overline{\phi})\}. \tag{2}$$

Take now an arbitrary $\omega' \in F \cap \mathbf{RAT}(\overline{\phi})$ and $i \in [1..n]$. Since $\omega' \in \mathbf{RAT}(\overline{\phi})$, player i is ϕ_i-rational in ω', i.e., $\phi_i(s_i(\omega'), G_{P_i(\omega')})$ holds. But F is evident, so $P_i(\omega') \subseteq F$. Moreover by (2) $P_i(\omega') \subseteq \mathbf{RAT}(\overline{\phi})$, so $P_i(\omega') \subseteq F \cap \mathbf{RAT}(\overline{\phi})$. Hence $G_{P_i(\omega')} \subseteq G_{F \cap \mathbf{RAT}(\overline{\phi})}$ and by the monotonicity of ϕ_i we conclude that $\phi_i(s_i(\omega'), G_{F \cap \mathbf{RAT}(\overline{\phi})})$ holds.

By the definition of $T_{\overline{\phi}}$ this means that $G_{F \cap \mathbf{RAT}(\overline{\phi})} \subseteq T_{\overline{\phi}}(G_{\mathbf{RAT}(\overline{\phi})})$, i.e. that $G_{F \cap \mathbf{RAT}(\overline{\phi})}$ is a post-fixpoint of $T_{\overline{\phi}}$. Hence by Tarski's Fixpoint Theorem $G_{F \cap \mathbf{RAT}(\overline{\phi})} \subseteq T_{\overline{\phi}}^{\infty}$.

But $s_i = s_i(\omega)$ and $\omega \in F \cap \mathbf{RAT}(\overline{\phi})$, so we conclude by the above inclusion that s_i is an element of the ith component of $T_{\overline{\phi}}^{\infty}$. This proves $G_{CB(\overline{\phi})} \subseteq T_{\overline{\phi}}^{\infty}$.

(iii) $T_{\overline{\phi}}^{\infty} \subseteq G_{CK(\overline{\phi})}$.

Recall that $H = (T_1, \ldots, T_n, p_1, \ldots, p_n)$. We first define

- the functions $s_1 : \Omega \to T_1, \ldots, s_n : \Omega \to T_n$,
- an event E,
- the knowledge correspondences $P_1, \ldots, P_n$.

Suppose $T_{\overline{\phi}}^{\infty} = (S_1, \ldots, S_n)$. Choose $j_0 \in [1..n]$ such that the set S_{j_0} has the largest cardinality among the sets $S_1, \ldots, S_n$. Define the function $s_{j_0} : \Omega \to T_{j_0}$ arbitrarily, but so that it is onto (note that this is possible since by assumption $|\Omega| \geq |T_{j_0}|$) and let $E := s_{j_0}^{-1}(S_{j_0})$.

Our aim is to ensure that

$$G_E = T_{\overline{\phi}}^{\infty}.$$

So we define each function $s_k : \Omega \to T_k$, where $k \neq j_0$, in such a way that $s_k^{-1}(S_k) = E$. Note that this is possible since $|E| \geq |S_{j_0}| \geq |S_k|$.

Next, we define each knowledge correspondence P_i arbitrarily but so that for all $\omega \in E$ we have $P_i(\omega) = E$. Then for all $i \in [1..n]$

$$G_{P_i(\omega)} = G_E.$$

We now show that in all $\omega \in E$, in the resulting knowledge model for H, each player i is ϕ_i-rational in ω. So take an arbitrary $\omega \in E$ and $i \in [1..n]$. By the definition of the function $s_i(\cdot)$ a strategy $s_i \in S_i$ exists such that $s_i = s_i(\omega)$. Now, $T_{\overline{\phi}}^{\infty}$ is a fixpoint of $T_{\overline{\phi}}$, so $\phi_i(s_i, T_{\overline{\phi}}^{\infty})$ holds. But $T_{\overline{\phi}}^{\infty} = G_E = G_{P_i(\omega)}$, so $\phi_i(s_i(\omega), G_{P_i(\omega)})$ holds, i.e. player i is indeed ϕ_i-rational in ω.

To complete the proof take now an arbitrary strategy $s_i \in S_i$. By the definition of the function $s_i(\cdot)$ a state $\omega \in E$ exists such that $s_i = s_i(\omega)$. Further, we just showed that each player j is ϕ_j-rational in ω. But by the definition of the knowledge correspondences E is an evident event, so it is common knowledge in ω that each player j is ϕ_j-rational in ω. Hence $\omega \in CK(\overline{\phi})$ and consequently s_i is an element of the ith component of $G_{CK(\overline{\phi})}$.

This proves that $T_{\overline{\phi}}^{\infty} \subseteq G_{CK(\overline{\phi})}$. □

The above theorem, through the definitions of the events $CK(\overline{\phi})$ and $CB(\overline{\phi})$, makes use of arbitrary models for the initial game H. But it also holds when we admit only standard models for H in these two definitions. Indeed, among the three inclusions that we established in its proof we only need to reconsider the last one, $T_{\overline{\phi}}^{\infty} \subseteq G_{CK(\overline{\phi})}$. The argument in the presence of only standard models runs as follows.

Suppose $T_{\overline{\phi}}^{\infty} = (S_1, \ldots, S_n)$. Consider the event $F := S_1 \times \ldots \times S_n$ in the standard model for H. Then $G_F = T_{\overline{\phi}}^{\infty}$. Define each knowledge 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 indeed a knowledge correspondence (also if $F = \emptyset$ or $F = \Omega$) and clearly F is an evident event.

Take now an arbitrary $i \in [1..n]$ and an arbitrary state $\omega \in F$. Since $T_{\overline{\phi}}^{\infty}$ is a fixpoint of $T_{\overline{\phi}}$ and $s_i(\omega) \in S_i$, $\phi_i(s_i(\omega), T_{\overline{\phi}}^{\infty})$ holds, so by the definition of P_i, $\phi_i(s_i(\omega), G_{P_i(\omega)})$ holds.

This shows that each player i is ϕ_i-rational in each state $\omega \in F$. Since F is evident, we conclude that in each state $\omega \in F$ it is common knowledge that each player i is ϕ_i-rational. So $F \subseteq CK(\overline{\phi})$ and consequently $T_{\overline{\phi}}^{\infty} = G_F \subseteq G_{CK(\overline{\phi})}$.

The above theorem shows that when each property ϕ_i is monotonic, the strategy profiles that the players choose in the states in which it is common knowledge that each player i is ϕ_i-rational (or in which each player i is ϕ_i-rational and it is common belief that each player i is ϕ_i-rational), are included in those that remain after the iterated elimination of the strategies that are not ϕ_i-optimal. It generalizes corresponding results established for finite strategic games (for their survey see Battigalli and Bonanno [1999]) to the case of arbitrary strategic games and arbitrary monotonic properties ϕ_i.

In Chen, Long and Luo [2005], Lipman [1994] and Apt [2007b] examples are provided showing that for the properties of strict dominance (namely sd^g) and best response (namely br^g) in general transfinite iterations (i.e., iterations beyond ω_0) of the corresponding operator are necessary to reach the outcome. So to achieve equalities in the above theorem transfinite iterations of the $T_{\overline{\phi}}$ operator are necessary.

By instantiating ϕ_is to specific properties we get instances of the above result that relate to specific definitions of rationality. Before we do this we establish another result that will apply to another class of properties ϕ_i.

Theorem 2. *Suppose that*

$$\phi_i(s_i, (\{s_1\}, \ldots, \{s_n\})) \text{ holds for all } i \in [1..n] \text{ and } s_i \in T_i. \tag{3}$$

Then

$$G_{CK(\overline{\phi})} = G_{CB(\overline{\phi})} = H.$$

Proof. As noted in the proof of Theorem 1, for all $\overline{\phi}$ we have $G_{CK(\overline{\phi})} \subseteq G_{CB(\overline{\phi})}$. Since $G_{CB(\overline{\phi})} \subseteq H$, it suffices to prove that $H \subseteq G_{CK(\phi)}$.

So take a strategy s_i of player i in H. We choose a knowledge model in stages. First we stipulate that each function $s_i(\cdot)$ is onto. This is possible since by assumption $|\Omega| \geq |T_i|$ for all $i \in [1..n]$. So a state ω exists such that $s_i = s_i(\omega)$. Next, we choose for each player j a knowledge correspondence P_j such that $P_j(\omega) = \{\omega\}$. Then

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

and, on the account of (3), each player j is ϕ_j-rational in ω.

By the choice of the knowledge correspondences $\{\omega\}$ is an evident event. Hence it is common knowledge in ω that each player j is ϕ_j-rational in ω. So by definition s_i is an element of the ith component of CK_ϕ. □

Note that any property ϕ_i that satisfies (3) and is not trivial (that is, for some strategy s_i, $\phi_i(s_i, H)$ does not hold) is not monotonic.

Also here the claim holds if in the definitions of $CK(\overline{\phi})$ and $CB(\overline{\phi})$ we admit only standard models for H. The reason is that all we needed was for each function $s_i(\cdot)$ to be onto; but this is the case for standard models.

5 ... and their consequences

Let us analyze now the consequences of the above two theorems. Consider first Theorem 1. The following lemma, in which we refer to the properties introduced in Section 3, clarifies the matters.

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

Proof. Straightforward. □

So Theorem 1 applies to the above three properties. (Note that br_i^g actually comes in three 'flavours' depending on the choice of beliefs.) Strict dominance in the sense of sd_i^g is studied in Chen, Long and Luo [2005], while br_i^g corresponds to the rationalizability notion of Bernheim [1984].

In contrast, Theorem 1 does not apply to the properties wd_i^g and mwd_i^g, since, as indicated in Apt [2007c], the corresponding operators T_{wd^g} and T_{mwd^g} are not monotonic, and hence the properties wd_i^g and mwd_i^g are not monotonic.

To see the consequences of Theorem 2 note that the properties sd_i^l, msd_i^l, wd_i^l, mwd_i^l and br_i^l satisfy (3). In particular, this theorem shows that the 'customary' concepts of strict dominance, sd_i^l and msd_i^l cannot be justified in the used epistemic framework as 'stand alone' concepts of rationality. Indeed, this theorem shows that common knowledge that each player is rational in one of these two senses does not exclude any strategy.

What *can* be done is to justify these two concepts as *consequences* of the common knowledge of rationality defined in terms of br_i^g, the 'global' version of the best response property, Namely, we have the following result. When each property ϕ_i equals br_i^g), we write here $CK(br^g)$ instead of $CK(\overline{br^g})$ and analogously for CB.

Theorem 3. *For all games H*

$$G_{CK(br^g)} = G_{CB(br^g)} \subseteq T^{\infty}_{sd^l},$$

where we take as the set of beliefs the set of joint strategies of the opponents.

Proof. By Lemma 2 and Theorem 1 $G_{CK(br^g)} = G_{CB(br^g)} = T^{\infty}_{br^g}$. Each best response to a joint strategy of the opponents is not strictly dominated, so for all restrictions G

$$T_{br^g}(G) \subseteq T_{sd^g}(G).$$

Also, for all restrictions G

$$T_{sd^g}(G) \subseteq T_{sd^l}(G).$$

So by Lemma 1 $T^{\infty}_{br^g} \subseteq T^{\infty}_{sd^l}$, which concludes the proof. □

The above result 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.

In the case of finite games we have the following known result. For a proof using Harsanyi type spaces see Brandenburger and Friedenberg [2006].

Theorem 4. *For all finite games H*

$$G_{CK(br^g)} = G_{CB(br^g)} \subseteq T^{\infty}_{msd^l},$$

where we take as the set of beliefs the set of joint mixed strategies of the opponents.

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

Again by Lemma 2 and Theorem 1 $G_{CK(br^g)} = G_{CB(br^g)} = T^{\infty}_{br^g}$. Further, for all restrictions G

$$T_{br^g}(G) \subseteq T_{br^l}(G)$$

and

$$T_{br^l}(G) \subseteq T_{brc^l}(G),$$

where brc_i^l stands for the best response property w.r.t. the correlated strategies of the opponents. So by Lemma 1 $T^{\infty}_{br^g} \subseteq T^{\infty}_{brc^l}$.

But by the result of Osborne and Rubinstein [1994, page 60] (that is a modification of the original result of Pearce [1984]) for all restrictions G we have $T_{brc^l}(G) = T_{msd^l}(G)$, so $T^{\infty}_{brc^l} = T^{\infty}_{msd^l}$, which yields the conclusion. $\Box$

6 Axiomatic presentation

The proof of Theorem 1 relied on a judicious manipulation of events. So it is natural to ask what are the proof-theoretic principles involved. In this section we present a formal language $\mathcal{L}_\nu$ that will be interpreted over belief models. We will then give syntactic proof rules for $\mathcal{L}_\nu$ that lead to an axiomatic proof of the main part of Theorem 1. Throughout the section we assume, as usual, the initial game H and monotonic properties $\phi_1, \ldots, \phi_n$. Later we shall introduce a language that allows us to define and analyze the relevant properties.

To start with, we consider the simpler language $\mathcal{L}$ the formulae of which are defined by the following recursive definition, where $i \in [1..n]$:

$$\psi ::= rat_i \mid \psi \wedge \psi \mid \neg\psi \mid \Box_i\psi \mid O_i\psi,$$

where each rat_i is a constant. We abbreviate the formula $\bigwedge_{i\in[1..n]} rat_i$ to rat, $\bigwedge_{i\in[1..n]} \Box_i\psi$ to $\Box\psi$ and $\bigwedge_{i\in[1..n]} O_i\psi$ to $O\psi$.

Formulae of $\mathcal{L}$ will be interpreted as events in belief models for H. Given a belief model $(\Omega, s_1, \ldots, s_n, P_1, \ldots, P_n)$ for H, we define the ***interpretation function*** $\mathcal{I}(\cdot) : \mathcal{L} \to \mathcal{P}(\Omega)$ as follows:

- $\mathcal{I}(rat_i) = \{\omega \in \Omega \mid \phi_i(s_i(\omega), G_{P_i(\omega)})\}$,
- $\mathcal{I}(\phi \wedge \psi) = \mathcal{I}(\phi) \cap \mathcal{I}(\psi)$,
- $\mathcal{I}(\neg\psi) = \Omega - \mathcal{I}(\psi)$,
- $\mathcal{I}(\Box_i\psi) = \{\omega \in \Omega \mid P_i(\omega) \subseteq \mathcal{I}(\psi)\}$,
- $\mathcal{I}(O_i\psi) = \{\omega \in \Omega \mid \phi_i(s_i(\omega), G_{\mathcal{I}(\psi)})\}$.

Note that $\mathcal{I}(rat)$ is the event that every player is rational, $\mathcal{I}(\Box\psi)$ is the event that every player believes the event $\mathcal{I}(\psi)$ and $\mathcal{I}(O\psi)$ is the event that every player's strategy is optimal in the context of the restriction $G_{\mathcal{I}(\psi)}$.

$\mathcal{L}$ is a modal language in the sense of Blackburn, de Rijke and Venema [2001]. Although $\mathcal{L}$ can express some connections between our formal definitions of optimality, rationality and beliefs, it is not a very expressive language. If our interest were to reason about particular games, we could extend the language with atoms s_i expressing the event that the strategy s_i is chosen. This choice is often made when defining modal languages for models of games, see, e.g., de Bruin [2004]. However, we are interested in a language to reason about games with arbitrary strategy sets, and in particular in a language that can express the non-trivial part of the left-to-right inclusion of Theorem 1 (part (ii) of its proof).

Specifically, we want a language that can express the following statement:

Imp If it is true common belief that every player is rational, then all players choose strategies that survive the iterated elimination of non-optimal strategies.

To this end we extend the vocabulary of $\mathcal{L}$ with a single set variable denoted by x and the largest fixpoint operator νx. (The corresponding extension of the first-order logic by the dual, least fixpoint operator μx was first studied in Gurevich [1984].) Modulo one caveat the resulting language $\mathcal{L}_\nu$ is defined as follows, where '...' stands for the already given definition of $\mathcal{L}$:

$$\psi ::= \ldots \mid x \mid \nu x.\phi.$$

The caveat is the following: ϕ must be

- ***positive in*** x, which means that each occurrence of x in ϕ is under the scope of an even number of negation signs ($\neg$),
- ν***-free***, which means that it does not contain any occurrences of the νx operator.

(The latter restriction is not necessary, but simplifies matters and is sufficient for our considerations.)

To define the interpretation function $\mathcal{I}(\cdot)$ for $\mathcal{L}_\nu$ we must keep track of the variable x. Therefore we first extend the function $\mathcal{I}(\cdot)$ on $\mathcal{L}$ to a function $\mathcal{I}(\cdot) : \mathcal{L}_\nu \times \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ by padding it with a dummy argument. We give one clause as an example:

- $\mathcal{I}(\Box_i\psi, E) = \{\omega \in \Omega \mid P_i(\omega) \subseteq \mathcal{I}(\psi, E)\}$.

Then we put

- $\mathcal{I}(x, E) = E$,

and finally define

- $\mathcal{I}(\nu x.\psi) = \bigcup\{E \subseteq \Omega \mid E \subseteq \mathcal{I}(\psi, E)\}$.

It is straightforward to see that the restriction to positive in x and ν-free formulae ψ ensures that $\mathcal{I}(\psi, \cdot)$ is a monotonic operator on the powerset $\mathcal{P}(\Omega)$ of Ω. Hence by Tarski's Fixpoint Theorem $\mathcal{I}(\nu x.\psi)$ is its largest fixpoint.

This language can express **Imp**. To see this, first notice that common belief is definable in $\mathcal{L}_\nu$ using the νx operator. The analogous characterization of common knowledge is given in Fagin et al. [1995, Section 11.5].

Note 2. *Let ψ be a formula of $\mathcal{L}$ and x a variable. Then $\mathcal{I}(\nu x.\Box(x \wedge \psi))$ is the event that the event $\mathcal{I}(\psi)$ is common belief.*

Proof. ψ is a formula of $\mathcal{L}$, so x does not occur in ψ. Note that for all $F \subseteq \Omega$ we have

- $\mathcal{I}(\Box(x \wedge \psi), F) = \mathcal{I}(\Box x, F) \cap \mathcal{I}(\Box\psi, F)$,
- $\mathcal{I}(\Box x, F) = \Box\mathcal{I}(x, F) = \Box F$,
- $\mathcal{I}(\Box\psi, F) = \Box\mathcal{I}(\psi, F) = \Box\mathcal{I}(\psi)$,

where the 'outer' $\Box$ is defined in Subsection 2.3. Hence

$$\begin{aligned}\mathcal{I}(\nu x.\Box(x \wedge \psi)) &= \bigcup\{F \subseteq \Omega \mid F \subseteq \mathcal{I}(\Box(x \wedge \psi), F)\} \\ &= \bigcup\{F \subseteq \Omega \mid \Box(F \cap \mathcal{I}(\psi))\} \\ &= \Box^*\mathcal{I}(\psi),\end{aligned}$$

where $\Box^*$ is defined in Subsection 2.3. $\Box$

From now on we abbreviate the (well-formed) formula $\nu x.\Box(x \wedge \psi)$ for ψ being a formula of $\mathcal{L}$ to $\Box^*\psi$. So $\Box^*$ is a new modality added to the language $\mathcal{L}_\nu$.

We can also define the iterated elimination of non-optimal strategies.

Note 3. *In the game determined by the event $\mathcal{I}(\nu x.Ox)$, every player selects a strategy which survives the iterated elimination of non-optimal strategies.*

Proof. We must show the following inclusion:

$$G_{\mathcal{I}(\nu x.Ox)} \subseteq T_{\overline{\phi}}^{\infty}.$$

Let $G' := (S'_1, \ldots, S'_n) = G_{\mathcal{I}(\nu x.Ox)}$. By Tarski's Fixpoint Theorem it suffices to show that $G' \subseteq T_{\overline{\phi}}(G')$. So take any $j \in [1..n]$ and any $s'_j \in S'_j$. We must show that $\phi_j(s'_j, G')$ holds. By definition for some $\omega \in \mathcal{I}(\nu x.Ox)$ we have $s_j(\omega) = s'_j$. Then there is some E such that $\omega \in E$ and $E \subseteq \mathcal{I}(Ox, E)$. Therefore for all $i \in [1..n]$, $\phi_i(s_i(\omega), G_E)$ holds, so in particular $\phi_j(s'_j, G_E)$ holds.

But $E \subseteq \mathcal{I}(Ox, E)$ implies $E \subseteq \mathcal{I}(\nu x.Ox)$ and therefore $G_E \subseteq G_{\mathcal{I}(\nu x.Ox)} = G'$. Hence by monotonicity of ϕ_j we get $\phi_j(s'_j, G')$ as desired. □

Now consider the following formula:

$$(rat \land \Box^* rat) \to \nu x.Ox. \tag{4}$$

By Notes 2 and 3, we can see that wherever the formula (4) holds, then if it is true common belief that every player is rational, then each player selects a strategy that survives the iterated elimination of non-optimal strategies.

We call an $\mathcal{L}_\nu$-formula ψ ***valid*** if for every belief model $(\Omega, s_1, \ldots, s_n, P_1, \ldots, P_n)$ for H we have $\mathcal{I}(\psi) = \Omega$.

We are now in a position to connect $\mathcal{L}_\nu$ to **Imp**: the statement **Imp** asserts that the formula (4) is valid.

In the rest of this section we will discuss a simple proof system in which we can derive (4). This will provide an alternative way of proving the corresponding inclusion in Theorem 1.

We will use an axiom and rule of inference for the fixpoint operator taken from Kozen [1983] and one axiom for rationality analogous to the one called in de Bruin [2004] an 'implicit definition' of rationality. We give these in Figure 1 denoting by $\psi[x \mapsto \chi]$ the formula obtained from ψ by substituting each occurence of the variable x with the formula χ.

First we establish the soundness of this proof system, that is that its axioms are valid and the proof rules preserve validity.

Lemma 3. *The proof system* ***P*** *is sound.*

Proof. We show first the validity of the axiom $ratDis$. Let $(\Omega, s_1, \ldots, s_n, P_i, \ldots, P_n)$ be a belief model for H. We must show that $\mathcal{I}(rat \to (\Box\psi \to O\psi)) = \Omega$.

Axiom schemata	
$rat \to (\Box\psi \to O\psi)$	$ratDis$
$\nu x.\psi \to \psi[x \mapsto \nu x.\psi]$	νDis

Rule of inference

$$\frac{\chi \to \psi[x \mapsto \chi]}{\chi \to \nu x.\psi} \ \nu Ind$$

Figure 1: Proof system **P**

That is, that for any ψ the inclusion $\mathcal{I}(rat) \cap \mathcal{I}(\Box\psi) \subseteq \mathcal{I}(O\psi)$ holds. So take some $\omega \in \mathcal{I}(rat) \cap \mathcal{I}(\Box\psi)$. Then for every $i \in [1..n]$, $\phi_i(s_i(\omega), G_{P_i(\omega)})$, and $P_i(\omega) \subseteq \mathcal{I}(\psi)$. So by monotonicity of ϕ_i, $\phi_i(s_i(\omega), G_{\mathcal{I}(\psi)})$, i.e. $\omega \in \mathcal{I}(O_i\psi)$ as required.

The axioms νDis and the rule νInd were introduced in Kozen [1983], and their soundness proof is standard. This axiom and the rule formalize, respectively, the following two consequences of Tarski's Fixpoint Theorem concerning a monotonic operator T:

- νT is a post-fixpoint of T, i.e., $\nu T \subseteq T(\nu T)$ holds,
- if Y is a post-fixpoint of T, i.e., $Y \subseteq T(Y)$ holds, then $Y \subseteq \nu T$. □

Next, we establish the already announced claim.

Theorem 5. *The formula (4) is a theorem of the proof system **P**.*

Proof. The following formula is an instance of the axiom $ratDis$ (with $\psi := \Box^* rat \wedge rat$):

$$rat \to (\Box(\Box^* rat \wedge rat) \to O(\Box^* rat \wedge rat)),$$

and the following is an instance of νDis (with $\psi := \Box(x \wedge rat)$):

$$\Box^* rat \to \Box(\Box^* rat \wedge rat)$$

Putting these two together via some simple propositional logic, we obtain:

$$(\Box^* rat \wedge rat) \rightarrow O(\Box^* rat \wedge rat).$$

This last formula is of the right shape to apply the rule νInd (with $\chi := \Box^* rat \wedge rat$ and $\psi := Ox$), to obtain:

$$(\Box^* rat \wedge rat) \rightarrow \nu x.Ox,$$

which is precisely the formula (4). □

The derivation of 4 has shown which proof-theoretic principles are required, or at least sufficient, to obtain the part of Theorem 1 on which we are concentrating. It is interesting to note that no axioms or rules for the modalities $\Box$ and O were needed in order to derive (4). This corresponds to the fact that in the proof of the corresponding inclusion in Theorem 1 we did not use the fact that the possibility correspondences were belief correspondences.

Corollary 1. *The formula (4) is valid.* □

In the language $\mathcal{L}_\nu$, $rat_1, \ldots, rat_n$ are propositional constants. We can define them in terms of the $\Box_i$ and O_i modalities but to this end we need to extend the language $\mathcal{L}_\nu$ to a second-order one by allowing quantifiers over set variables, so by allowing formulae of the form $\exists X \phi$. It is clear how to extend the semantics to this larger class of formulae. In the resulting language each rat_i constant is definable by a formula of the latter language:

$$rat_i \equiv \forall X(\Box_i X \rightarrow O_i X), \quad (5)$$

where $\forall X \phi$ is an abbreviation for $\neg\exists X \neg\phi$.

The following observation then shows correctness of this definition.

Note 4. *For all $i \in [1..n]$ the formula (5) is valid.* □

Let us mention that such second-order extensions of propositional modal logics were first considered in Fine [1970].

To further our syntactic analysis, we now give a language $\mathcal{L}_O$ which can be used to define and analyze the optimality properties $\phi_i(\cdot, \cdot)$. It is a first-order language formed from a family of n ternary relation symbols $x \geq_z^i y$, where $i \in [1..n]$, along with the binary relation $x \in X$ between a first-order variable and a set variable. $\mathcal{L}_O$ is given by the following recursive definition:

$$\phi ::= x \in X \mid x \geq^i_z y \mid \neg\phi \mid \phi \wedge \phi \mid \exists x\phi,$$

where $i \in [1..n]$.

We use the same abbreviations $\rightarrow$ and $\vee$ as above and further abbreviate $\neg y \geq^i_z x$ to $x >^i_z y$, $\exists x(x \in X \wedge \phi)$ to $\exists x \in X\,\phi$, $\forall x(x \in X \rightarrow \phi)$ to $\forall x \in X\,\phi$, and write $\forall x\phi$ for $\neg\exists x\neg\phi$.

By an ***optimality condition*** for player i we now mean a formula containing exactly one free first-order variable and the set variable X, and in which all the occurrences of the atomic formula $x \geq^j_z y$ are with j equal to i. In particular, we are interested in the following optimality conditions:

- $sd^l_i(x, X) := \forall y \in X \exists z \in X x \geq^i_z y$,
- $sd^g_i(x, X) := \forall y \exists z \in X x \geq^i_z y$,
- $wd^l_i(x, X) := \forall y \in X(\forall z \in X x \geq^i_z y \vee \exists z \in X x >^i_z y)$,
- $wd^g_i(x, X) := \forall y(\forall z \in X x \geq^i_z y \vee \exists z \in X x >^i_z y)$,
- $br^l_i(x, X) := \exists z \in X \forall y \in X x \geq^i_z y$,
- $br^g_i(x, X) := \exists z \in X \forall y\, x \geq^i_z y$.

We now give a semantics for $\mathcal{L}_O$-formulae in the context of a model $(\Omega, s_1, \ldots, s_n)$ for the initial game H. An ***assignment*** is a function α that maps each first-order variable to a state in Ω and each set variable to an event in (a subset of) Ω. The semantics is given by a satisfaction relation between an assignment α and a formula ϕ of $\mathcal{L}_O$, with $\models_\alpha \phi$ meaning that α satisfies ϕ. This relation is defined as follows:

- $\models_\alpha x \in X$ iff $\alpha(x) \in \alpha(X)$,
- $\models_\alpha x \geq^i_z y$ iff $p_i(s_i(\alpha(x)), s_{-i}(\alpha(z))) \geq p_i(s_i(\alpha(y)), s_{-i}(\alpha(z)))$,
- $\models_\alpha \neg\phi$ iff not $\models_\alpha \phi$,
- $\models_\alpha \phi \wedge \psi$ iff $\models_\alpha \phi$ and $\models_\alpha \psi$,
- $\models_\alpha \exists x\phi$ iff there is an $\omega \in \Omega$ such that $\models_{\alpha[x \mapsto \omega]} \phi$,

where:

$$\alpha[x \mapsto \omega](x_0) := \begin{cases} \alpha(x) & \text{if } x \neq x_0 \\ \omega & \text{otherwise.} \end{cases}$$

This semantics allows us to relate the above six optimality conditions to the corresponding optimality properties that are concerned solely with pure strategies.

Note 5. *For each optimality condition* ϕ_i*, where* $\phi \in \{sd^l, sd^g, wd^l, wd^g, br^l, br^g\}$

$$\models_\alpha \phi_i(x, X) \textit{ iff the property } \phi_i(s_i(\alpha(x)), G_{\alpha(X)}) \textit{ holds.}$$

□

To relate optimality conditions to monotonic optimality properties we need one more definition. We say that a formula ϕ of $\mathcal{L}_O$ is *positive* just when every occurrence of the set variable X occurs under a positive number of negation signs ($\neg$). So for example the formula $br_i^l(x, X)$, that is, $\exists z \in X \forall y \in X x \geq_z^i y$, is not positive, since the second occurrence of X is under one negation sign, while $br_i^g(x, X)$, that is, $\exists z \in X \forall y\, x \geq_z^i y$, is positive.

The following observation then links syntactic matters with monotonicity.

Note 6. *For every positive optimality condition* $\phi_i(x, X)$ *for player* i *the corresponding property* $\phi_i(s_i, G)$ *(used by player* i*) is monotonic.* □

Among the above six optimality conditions only $sd_i^g(x, X)$ and $br_i^g(x, X)$ are positive. The corresponding other four properties, as already mentioned earlier, are not monotonic. By the above observation they cannot be defined by positive formulae.

7 Public announcements

In this section, inspired by van Benthem [2007], we provide an alternative characterization of the iterated elimination of strategies that for player i are not ϕ_i-optimal, based on the concept of a public announcement.

Let us clarify first what we would like to achieve. Consider a model $\mathcal{M}$ for the initial game H. The process of iterated elimination of the strategies that are not ϕ_i-optimal, formalized by the iterated applications of the $T_{\overline{\phi}}$ operator, produces a sequence $T_{\overline{\phi}}^{\alpha}$, where α is an ordinal, of restrictions of

H. We would like to mimic it on the side of the models, so that we get a corresponding sequence $\mathcal{M}^\alpha$ of models of these restrictions.

To make this idea work we need to define an appropriate way of reducing models. We take care of it by letting the players repeatedly announce that they only select ϕ_i-optimal strategies. This brings us to the notions of public announcements and their effects on the models.

Given a model $\mathcal{M} = (\Omega, s_1, \ldots, s_n)$ we define

- a ***public announcement*** by player i in a model $\mathcal{M}$ as an event E in $\mathcal{M}$,
- given a vector $\overline{E} := (E_1, \ldots, E_n)$ of public announcements by players $1, \ldots, n$ we let

$$[\overline{E}](\mathcal{M}) := (\cap_{i=1}^n E_i, (s_i \,|\, \cap_{i=1}^n E_i)_{i \in [1..n]})$$

and call it ***the effect of the public announcements of*** $\overline{E}$ ***on*** $\mathcal{M}$.

Given a property $\phi_i(\cdot, G)$ that player i uses to select his strategies in the restriction G of H and a model $\mathcal{M} := (\Omega, s_1, \ldots, s_n)$ for G we define $[\![\phi_i]\!]$ as the event in $\mathcal{M}$ that player i selects optimally his strategies with respect to G. Formally:

$$[\![\phi_i]\!] := \{\omega \in \Omega \mid \phi_i(s_i(\omega), G)\}$$

(Note that in the notation of the previous section we have $[\![\overline{\phi}]\!] = \mathcal{I}(O_i\mathbf{T})$, where $\mathbf{T} := \psi \vee \neg\psi$.) We abbreviate the vector $([\![\phi_1]\!], \ldots, [\![\phi_n]\!])$ to $[\![\overline{\phi}]\!]$.

We want now to obtain the reduction of a model $\mathcal{M}$ of G to a model $\mathcal{M}$ of $T_{\overline{\phi}}(G)$ by means of the just defined vector $[\![\overline{\phi}]\!]$ of public announcements.

The effect of the public announcements of $\overline{E}$ on a model of G should ideally be a model of the restriction $G_{\overline{E}}$. Unfortunately, this does not hold in such generality. Indeed, let the two-player game G have the strategy sets $S_1 := \{U, D\}$, $S_2 := \{L, R\}$ and consider the model $\mathcal{M}$ for G with $\Omega := \{\omega_{ul}, \omega_{dr}\}$ and the functions s_1 and s_2 defined by

$$s_1(\omega_{ul}) = U, \ s_2(\omega_{ul}) = L, \ s_1(\omega_{dr}) = D, \ s_2(\omega_{dr}) = R.$$

This simple example is depicted in Figure 2.

Let $\overline{E} = (\{\omega_{ul}\}, \{\omega_{dr}\})$; then $[\overline{E}](\mathcal{M}) = \emptyset$, which is not a model of $G_{\overline{E}} = (\{U\}, \{R\})$.

	L	R
U	ω_{ul}	
D		ω_{dr}

Figure 2: A motivating example for the use of standard models

A remedy lies in restricting one's attention to the standard models. However, in order to find a faithful public announcement analogue to strategy elimination we must also narrow the concept of a public announcement as follows. A ***proper public announcement*** by player i in a standard model is a subset of $\Omega = S_1 \times \ldots \times S_n$ of the form $S_1 \times \ldots \times S_{i-1} \times S'_i \times S_{i+1} \times \ldots \times S_n$.

So a proper public announcement by a player is an event that amounts to a 'declaration' by the player that he will limit his attention to a subset of his strategies, that is, will discard the remaining strategies. So when each player makes a proper public announcement, their combined effect on the standard model is that the set of states (or equivalently, the set of joint strategies) becomes appropriately restricted. An example, which is crucial for us, of a proper public announcement in a standard model is of course $[\![\phi_i]\!]$.

The following note links in the desired way two notions we introduced. It states that the effect of the proper public announcements of $\overline{E}$ on the standard model for G is the standard model for the restriction of G to $\overline{E}$.

Note 7. *Let $\mathcal{M}$ be the standard model for G and $\overline{E}$ a vector of proper public announcements by players $1, \ldots, n$ in $\mathcal{M}$. Then $[\overline{E}](\mathcal{M})$ is the standard model for $G_{\overline{E}}$.*

Proof. We only need to check that $\cap_{i=1}^n E_i$ is the set of joint strategies of the restriction $G_{\overline{E}}$. But each E_i is a proper announcement, so it is of the form $S_1 \times \ldots \times S_{i-1} \times S'_i \times S_{i+1} \times \ldots \times S_n$, where $G = (S_1, \ldots, S_n)$. So $\cap_{i=1}^n E_i = S'_1 \times \ldots \times S'_n$.

Moreover, each function $s_i(\cdot)$ is a projection, so $G_{\overline{E}} = (s_1(E_1), \ldots, s_n(E_n)) = (S'_1, \ldots, S'_n)$. □

We also have the following observation that links the vector $[\![\overline{\phi}]\!]$ of public announcements with the operator $T_{\overline{\phi}}$ of Section 3.

Note 8. *Let $\mathcal{M} := (\Omega, s_1, \ldots, s_n)$ be the standard model for G. Then*

$$T_{\overline{\phi}}(G) = G_{[\![\overline{\phi}]\!]}.$$

Proof. Let $G = (S_1, \ldots, S_n)$, $T_{\overline{\phi}}(G) = (S'_1, \ldots, S'_n)$ and $G_{[\![\overline{\phi}]\!]} = (S''_1, \ldots, S''_n)$.

Fix $i \in [1..n]$. Then we have the following string of equivalences:

$$\begin{array}{lll} & & s_i \in S'_i \\ & \text{iff} & s_i \in S_i \wedge \phi_i(s_i, G) \\ (s_i(\cdot) \text{ is onto}) & \text{iff} & s_i \in S_i \wedge \exists \omega \in \Omega\, (s_i = s_i(\omega) \wedge \phi_i(s_i(\omega), G)) \\ & \text{iff} & s_i \in S_i \wedge \exists \omega \in [\![\phi_i]\!]\, (s_i = s_i(\omega)) \\ & \text{iff} & s_i \in S''_i. \end{array}$$

□

Denote now by $[\![\overline{\phi}]\!]^\infty$ the iterated effect of the public announcements of $[\![\overline{\phi}]\!]$ starting with the standard model for the initial game H. The following conclusion then relates the iterated elimination of the strategies that for player i are not ϕ_i-optimal to the iterated effects of the corresponding public announcements.

Corollary 2. $[\![\overline{\phi}]\!]^\infty$ *is the standard model for the restriction* $T^\infty_{\overline{\phi}}$.

Proof. By Notes 7 and 8. □

Note that in the above corollary each effect of the public announcements of $[\![\overline{\phi}]\!]$ is considered on a different standard model. Note also that the above result holds for arbitrary properties ϕ_i, not necessarily monotonic ones.

We already mentioned in Section 1 that for various natural properties $\overline{\phi}$ transfinite iterations of $T_{\overline{\phi}}$ may be needed to reach the outcome $T^\infty_{\overline{\phi}}$. So the same holds for the iterated effects of the corresponding public announcements. It is useful to point out that, as shown in Parikh [1992] a similar situation can arise in case of natural dialogues the aim of which is to reach common knowledge.

This analysis gives an account of public announcements of the *optimality* of players' strategies. We now extend this analysis to public announcements of *rationality*. To this end we additionally assume for each player a belief correspondence $P_i : \Omega \to \mathcal{P}(\Omega)$, that is we consider belief models.

We define then the event of player i being ϕ_i-rational in the restriction G as

$$\langle \phi_i \rangle := \{\omega \in \Omega \mid \phi_i(s_i(\omega), G_{P_i(\omega)})\}.$$

(Note that in the notation of the previous section we have $\langle \phi_i \rangle = \mathcal{I}(rat_i)$.) Again we abbreviate $(\langle \phi_1 \rangle, \ldots, \langle \phi_n \rangle)$ to $\langle \overline{\phi} \rangle$. Note that $\langle \overline{\phi} \rangle$ depends on the underlying belief model $(\mathcal{M}, s_1, \ldots, s_n, P_1, \ldots, P_n)$ and on G.

We extend the definition of the effect of the public announcements $\overline{E} := (E_1, \ldots, E_n)$ to belief models in the natural way, by restricting each possibility correspondence to the intersection of the events in $\overline{E}$:

$$[\overline{E}](\mathcal{M}, P_1, \ldots, P_n) = ([\overline{E}]\mathcal{M}, P_1 \,|\, \cap_{i=1}^{n} E_i, \ldots, P_n \,|\, \cap_{i=1}^{n} E_i).$$

This definition is in the same spirit as in Plaza [1989] and in Osborne and Rubinstein [1994, page 72], where it is used in the analysis of the puzzle of the hats.

We aim to find a class of belief models for which, under a mild restriction on the properties ϕ_i, $\langle\overline{\phi}\rangle^\infty$, the iterated effect of the public announcements of $\langle\overline{\phi}\rangle$ starting with the standard belief model for the initial game H, will be the standard belief model for $T_{\overline{\phi}}^{\infty}$. We will therefore use a natural choice of possibility correspondences, which we call the ***standard possibility correspondences***:

$$P_i(\omega) = \{\omega' \in \Omega \mid s_i(\omega) = s_i(\omega')\}.$$

By the ***standard knowledge model*** for a restriction G we now mean the standard model for G endowed with the standard possibility correspondences.

The following observation holds.

Note 9. *Consider the standard knowledge model* $(\Omega, s_1, \ldots, s_n, P_1, \ldots, P_n)$ *for a restriction* $G := (S_1, \ldots, S_n)$ *of* H *and a state* $\omega \in \Omega$. *Then*

$$G_{P_i(\omega)} = (S_1, \ldots, S_{i-1}, \{\omega_i\}, S_{i+1}, \ldots, S_n).$$

Proof. Immediate by the fact that in the standard knowledge model for each possibility correspondence we have $P_i(\omega) = \{\omega' \in \Omega \mid \omega'_i = \omega_i\}$. □

Intuitively, this observation states that in each state of a standard knowledge model each player knows his own choice of strategy but knows nothing about the strategies of the other players. So standard possibility correspondences represent the beliefs of each player after he has privately selected his strategy but no information between the players has been exchanged. It is in that sense that the standard knowledge models are natural. In van Benthem [2007] in effect only such models are considered.

A large class of properties ϕ_i satisfy the following restriction:

A For all $G := (S_1, \ldots, S_n)$ and $G' := (S'_1, \ldots, S'_n)$ such that $S_j = S'_j$ for all $j \neq i$,

$\phi_i(s_i, G) \leftrightarrow \phi_i(s_i, G')$.

That restriction on the properties ϕ_i is sufficient to obtain the following analogue of Corollary 2 for the case of public announcements of *rationality.*

Corollary 3. *Suppose that each property* $\phi_1, \ldots, \phi_n$ *satisfies* $\boldsymbol{A}$. *Then* $\langle\overline{\phi}\rangle^\infty$ *is the standard knowledge model for the restriction* $T_{\overline{\phi}}^\infty$.

Proof. Notice that it suffices to prove for each restriction G the following statement for each i:

$$\forall \omega \in \Omega \; \phi_i(\omega_i, G) \; \leftrightarrow \; \phi_i(\omega_i, G_{P_i(\omega)}). \tag{6}$$

Indeed, (6) entails that $\langle\phi_i\rangle = [\![\phi_i]\!]$, in which case the result follows from Corollary 2 and the observation that the possibility correspondences are restricted in the appropriate way.

But (6) is a direct consequence of the assumption of **A** and of Note 9. □

To see the consequences of the above result note that **A** holds for each global property sd_i^g, msd_i^g, wd_i^g, mwd_i^g and (all three forms of) br_i^g introduced in Section 3.

For each ϕ_i equal sd_i^g and finite games Corollary 3 boils down to Theorem 7 in van Benthem [2007]. The corresponding result for each ϕ_i equal to br_i^g, with the beliefs consisting of the joint strategies of the opponents, and finite games is mentioned at the end of Section 5.4 of that paper.

It is interesting to recall that the properties wd_i^g and mwd_i^g, in contrast to sd_i^g and msd_i^g and br_i^g, are not monotonic. So, in contrast to Theorem 1, we have now a characterization of $T_{\overline{\phi}}^\infty$ for both forms of weak dominance.

Also it is important to note that the above Corollary does not hold for the corresponding local properties sd_i^l, msd_i^l, wd_i^l, mwd_i^l and br_i^l introduced in Section 3. Indeed, for each such property ϕ_i by Note 9 $\phi_i(\omega_i, G_{P_i(\omega)})$ holds for each state ω and restriction G. Consequently $\langle\overline{\phi}\rangle = \Omega$. So when each ϕ_i is a local property listed above, $\langle\overline{\phi}\rangle$ is an identity operator on the standard knowledge models, that is $\langle\overline{\phi}\rangle^\infty$ is the standard knowledge model for the initial game H and not $T_{\overline{\phi}}^\infty$.

Still, as the following result shows, it is possible for finite games to draw conclusions about the outcome of the iterated elimination of strategies that are not optimal in a local sense.

Corollary 4. *Assume the initial game* H *is finite. Then for each* $\phi \in \{sd, msd, wd, mwd, br\}$, $\langle\phi^g\rangle^\infty$ *is the standard knowledge model for the restriction* $T_{\phi^l}^\infty$.

Proof. We rely on the following results that for finite games link the outcomes of the iterations of the corresponding local and global properties:

- (see Apt [2007b])

  $T^{\infty}_{br^{\,l}} = T^{\infty}_{br^{\,g}}$,

- (see Apt [2007c])

  $T^{\infty}_{\phi^l} = T^{\infty}_{\phi^g}$ for $\phi \in \{sd, wd\}$,

- (see Brandenburger, Friedenberg and Keisler [2006b])

  $T^{\infty}_{msd^l} = T^{\infty}_{msd^g}$,

- (see Brandenburger, Friedenberg and Keisler [2006a])

  $T^{\infty}_{mwd^l} = T^{\infty}_{mwd^g}$.

The conclusion now follows by Corollary 3. □

This corollary states that for finite games the outcome of, for example the customary iterated elimination of weakly dominated strategies, $T^{\infty}_{wd^l}$, can be obtained by iterating on the standard knowledge models the effect of the public announcements by all players of the corresponding *global version* of weak dominance, wd_i^g. So, yet again, we see an intimate interplay between the local and global notions of dominance.

Acknowledgements

We acknowledge helpful discussions with Adam Brandenburger, who suggested Theorems 3 and 4, and with Giacomo Bonanno who, together with a referee of Apt [2007b], suggested to incorporate common beliefs in the analysis. Joe Halpern pointed us to Monderer and Samet [1989]. Johan van Benthem made us aware of the alternative approach to epistemic analysis based on public announcements.

References

K. R. Apt

[2007a] Epistemic analysis of strategic games with arbitrary strategy sets, in: *Proceedings 11th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK07)*. Available from http://arxiv.org/abs/0706.1001. To appear at the ACM Portal.

[2007b] The many faces of rationalizability, *The B.E. Journal of Theoretical Economics, 7(1)*. (Topics), Article 18, 39 pages. Available from http://arxiv.org/abs/cs.GT/0608011.

[2007c] Relative strength of strategy elimination procedures, *Economics Bulletin*, 3, pp. 1–9. Available from http://economicsbulletin.vanderbilt.edu/Abstract.asp?PaperID=EB-07C7001%5.

R. Aumann

[1976] Agreeing to disagree, *The Annals of Statistics*, 4, pp. 1236–1239.

[1987] Correlated equilibrium as an expression of Bayesian rationality, *Econometrica*, 55, pp. 1–18.

P. Battigalli and G. Bonanno

[1999] Recent results on belief, knowledge and the epistemic foundations of game theory, *Research in Economics*, 53, pp. 149–225.

J. van Benthem

[2007] Rational dynamics and epistemic logic in games, *International Game Theory Review*, 9, pp. 13–45. To appear.

B. D. Bernheim

[1984] Rationalizable strategic behavior, *Econometrica*, 52, pp. 1007–1028.

P. Blackburn, M. de Rijke, and Y. Venema

[2001] *Modal Logic*, Cambridge University Press.

A. Brandenburger and E. Dekel

[1987] Rationalizability and correlated equilibria, *Econometrica*, 55, pp. 1391–1402.

A. Brandenburger and A. Friedenberg

[2006] Intrinsic correlation in games. Working paper. Available from http://pages.stern.nyu.edu/~abranden.

A. Brandenburger, A. Friedenberg, and H. Keisler

[2004] Admissibility in games. Working paper. Revised October 2006. Available from http://pages.stern.nyu.edu/~abranden.

[2006a] Admissibility in games. Working paper. Available from http://pages.stern.nyu.edu/~abranden.

[2006b] Fixed points for strong and weak dominance. Working paper. Available from http://pages.stern.nyu.edu/~abranden/.

B. de Bruin

[2004] *Explaining Games: On the logic of game theoretic explanations*, PhD thesis, ILLC, University of Amsterdam.

Y.-C. Chen, N. V. Long, and X. Luo

[2005] Iterated strict dominance in general games. Available from `http://www.sinica.edu.tw/~xluo/pa10.pdf`.

M. Dufwenberg and M. Stegeman

[2002] Existence and uniqueness of maximal reductions under iterated strict dominance, *Econometrica*, 70, pp. 2007–2023.

J. Ely and M. Peski

[2006] Hierarchies of belief and interim rationalizability, *Theoretical Economics*, 1, pp. 19–65. Available from `http://ideas.repec.org/a/the/publsh/163.html`.

R. Fagin, J. Halpern, M. Vardi, and Y. Moses

[1995] *Reasoning about knowledge*, MIT Press, Cambridge, MA, USA.

K. Fine

[1970] Propositional quantifiers in modal logic, *Theoria*, 36, pp. 336–346.

Y. Gurevich

[1984] Toward logic tailored for computational complexity, in: *Proceedings Logic Colloquium '83*, vol. 104 of Lecture Notes in Mathematics.

A. Heifetz and D. Samet

[1998] Knowledge spaces with arbitrarily high rank, *Games and Economic Behavior*, 22, pp. 260–273.

D. Kozen

[1983] Results on the propositional mu-calculus, *Theoretical Computer Science*, 27, pp. 333–354.

B. L. Lipman

[1991] How to decide how to decide how to . . . : Modeling limited rationality, *Econometrica*, 59, pp. 1105–1125.

[1994] A note on the implications of common knowledge of rationality, *Games and Economic Behavior*, 6, pp. 114–129.

R. D. Luce and H. Raiffa

[1957] *Games and Decisions*, John Wiley and Sons, New York.

P. Milgrom and J. Roberts

[1990] Rationalizability, learning, and equilibrium in games with strategic complementarities, *Econometrica*, 58, pp. 1255–1278.

D. Monderer and D. Samet

[1989] Approximating common knowledge with common beliefs, *Games and Economic Behaviour*, 1, pp. 170–190.

M. J. Osborne and A. Rubinstein

[1994] *A Course in Game Theory*, The MIT Press, Cambridge, Massachusetts.

R. Parikh

[1992] Finite and infinite dialogues, in: *Logic from Computer Science*, Y. N. Moschovakis, ed., Mathematical Sciences Research Institute Publications, 21, Springer, pp. 481–497.

D. G. Pearce

[1984] Rationalizable strategic behavior and the problem of perfection, *Econometrica*, 52, pp. 1029–1050.

J. Plaza

[1989] Logics of public communications, in: *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, eds., pp. 201–216.

T.-C. Tan and S. Werlang

[1988] The Bayesian foundations of solution concepts of games, *Journal Of Economic Theory*, 45, pp. 370–391.

A. Tarski

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