A Syntactic Approach to Rationality in Games

Giacomo Bonanno
Department of Economics,
University of California,
Davis, CA 95616-8578 - USA
e-mail: gfbonanno@ucdavis.edu

January 2007

Abstract

We consider strategic-form games with ordinal payoffs and provide a syntactic analysis of common belief/knowledge of rationality, which we define axiomatically. Two axioms are considered. The first says that a player is *irrational* if she chooses a particular strategy while believing that another strategy is better. We show that common belief of this weak notion of rationality characterizes the iterated deletion of pure strategies that are strictly dominated by *pure* strategies. The second axiom says that a player is *irrational* if she chooses a particular strategy while believing that a different strategy is at least as good and she considers it possible that this alternative strategy is actually better than the chosen one. We show that common *knowledge* of this stronger notion of rationality characterizes the restriction to pure strategies of the iterated deletion procedure introduced by Stalnaker (1994).

Keywords: rationality, common belief, rationalizability, dominated strategies, game logic, frame characterization.

1 Introduction

The notion of rationalizability in games was introduced independently by Bernheim [2] and Pearce [13]. A strategy of player i is said to be rational if it maximizes player i's expected payoff, given her beliefs about the strategies used by her opponents, that is, if it can be justified by some beliefs about her opponents' strategies. If player i, besides being rational, also attributes rationality to her opponents, then she must only consider as possible strategies of her opponents that are themselves justifiable. If, furthermore, player i believes that her opponents believe that she is rational, then she must believe that her opponents justify their own choices by only considering those strategies of player i that are justifiable, and so on. The strategies of player i that can be justified in this way are called rationalizable. Rationalizability was intended to capture the notion of common belief of rationality. Bernheim and Pearce showed that a strategy is rationalizable if and only if it survives the iterated deletion of strictly dominated strategies.[1] They expressed the notion of common belief of rationality only informally, that is, without making use of an epistemic framework. The first epistemic characterization of rationalizability was provided by Tan and Werlang [15] using a universal type space, rather than Kripke structures (Kripke [10]). A characterization of common belief of rationality using probabilistic Kripke structures was first provided by Stalnaker [14], although it was implicit in Brandenburger and Dekel [7]. Stalnaker also introduced a new, stronger, notion of rationalizability – which he called strong rationalizability – and showed that it corresponds to an iterated deletion procedure which is stronger than the iterated deletion of strictly dominated strategies. Stalnaker's approach is entirely semantic and uses the same notion of Bayesian rationality as Bernheim and Pearce, namely expected payoff maximization. This notion presupposes that the players' payoffs are von Neumann-Morgenstern payoffs. In contrast, in this paper we consider the larger class of strategic-form games with *ordinal* payoffs. Furthermore, we take a syntactic approach and define rationality axiomatically. We consider two axioms.

The first axiom says that a player is *irrational* if she chooses a particular strategy while believing that another strategy of hers is better. We show that common belief of this weak notion of rationality characterizes the iterated deletion of strictly dominated pure strategies. Note that, in the Bayesian approach based on von Neumann-Morgenstern payoffs, it can be shown (see Pearce [13] and Brandenburger and Dekel [7]) that a pure strategy s_i of player i is a best reply to some (possibly correlated) beliefs about the strategies of her opponents if and only if there is no *mixed* strategy of player i that strictly dominates s_i. The iterated deletion of strictly dominated strategies in the Bayesian approach thus allows the deletion of a pure strategy that is dominated by a mixed strategy, even though it may not be dominated by another pure strategy. Since we take a purely ordinal approach, the iterated deletion procedure that we consider only allows the removal of strategies that are dominated by *pure* strategies.

The second axiom that we consider says that a player is *irrational* if she chooses a particular strategy while believing that a different strategy is at least as good and she considers it possible that this alternative strategy is actually better than the chosen one. We show that common *knowledge* of this stronger notion of rationality characterizes the iterated deletion procedure introduced by Stalnaker [14], restricted – once again – to pure strategies.

The paper is organized as follows. In the next section we review the KD45 multi-agent logic for belief and common belief and the S5 logic for knowledge and common knowledge. In Section 3 we review the definition of strategic-form game with ordinal payoffs and the iterated deletion procedures mentioned above. In Section 4 we define game logics and introduce two axioms of rationality. In Section 5 we characterize common belief of rationality in the weaker sense and

[1] This characterization of rationalizability is true for two-player games and extends to n-player games only if correlated beliefs are allowed (see Brandenburger and Dekel [7]).

common knowledge of rationality in the stronger sense.

The characterization results proved in Section 5 (Propositions 15 and 19) are not characterizations in the sense in which this expression is used in modal logic, namely characterization of axioms in terms of classes of frames (see [3] p. 125). Thus in Section 6 we provide a reformulation of our results in terms of frame characterizations. Section 7 concludes.

2 Multi-agent logics of belief and knowledge

We consider a multi-modal logic with $n+1$ operators $B_1, B_2, ..., B_n, B_*$ where, for $i = 1, ..., n$, the intended interpretation of $B_i\phi$ is "player i believes that ϕ", while $B_*\phi$ is interpreted as "it is common belief that ϕ". The formal language is built in the usual way (see [3] and [8]) from a countable set A of atomic propositions, the connectives $\neg$ and $\vee$ (from which the connectives $\wedge$, $\rightarrow$ and $\leftrightarrow$ are defined as usual) and the modal operators.

We denote by $\mathbf{KD45}_n^*$ the logic defined by the following axioms and rules of inference.

AXIOMS:

1. All propositional tautologies.
2. Axiom **K** for every modal operator: for $\Box \in \{B_1, ..., B_n, B_*\}$,
$$\Box\phi \wedge \Box(\phi \rightarrow \psi) \rightarrow \Box\psi \quad (\mathbf{K})$$
3. Axioms **D**, **4** and **5** for individual beliefs: for $i = 1, ..., n$,
$$\begin{array}{ll} B_i\phi \rightarrow \neg B_i\neg\phi & (\mathbf{D}_i) \\ B_i\phi \rightarrow B_iB_i\phi & (\mathbf{4}_i) \\ \neg B_i\phi \rightarrow B_i\neg B_i\phi & (\mathbf{5}_i) \end{array}$$
4. Axioms for common belief: for $i = 1, ..., n$,
$$\begin{array}{ll} B_*\phi \rightarrow B_i\phi & (\mathbf{CB1}) \\ B_*\phi \rightarrow B_iB_*\phi & (\mathbf{CB2}) \\ B_*(\phi \rightarrow B_1\phi \wedge ... \wedge B_n\phi) \rightarrow (B_1\phi \wedge ... \wedge B_n\phi \rightarrow B_*\phi) & (\mathbf{CB3}) \end{array}$$

RULES OF INFERENCE:

1. Modus Ponens: from ϕ and $(\phi \rightarrow \psi)$ infer ψ **(MP)**
2. Necessitation for every modal operator: for $\Box \in \{B_1, ..., B_n, B_*\}$,
$$\text{from } \phi \text{ infer } \Box\phi \quad (\mathbf{NEC})$$

We denote by $\mathbf{S5}_n^*$ the logic obtained by adding to $\mathbf{KD45}_n^*$ the following axiom:

5. Axiom **T** for individual beliefs: for $i = 1, ..., n$,

$$B_i\phi \to \phi. \qquad (\mathbf{T}_i)$$

While $\mathbf{KD45}_n^*$ is a logic for individual and common beliefs, $\mathbf{S5}_n^*$ is the logic for (individual and common) knowledge. To stress the difference between the two, when we deal with $\mathbf{S5}_n^*$ we shall denote the modal operators by K_i and K_* rather than B_i and B_*, respectively.

Note that the common belief operator does not inherit all the properties of the individual belief operators. In particular, the negative introspection axiom for common belief, $\neg B_*\phi \to B_*\neg B_*\phi$, is *not* a theorem of $\mathbf{KD45}_n^*$. In order to obtain it as a theorem, one needs to strengthen the logic by adding the axiom that individuals are correct in their beliefs about what is commonly believed: $B_iB_*\phi \to B_*\phi$. Indeed, the logic $\mathbf{KD45}_n^*$ augmented with the axiom $B_iB_*\phi \to B_*\phi$ *coincides* with the logic $\mathbf{KD45}_n^*$ augmented with the axiom $\neg B_*\phi \to B_*\neg B_*\phi$ (see [6]).

On the semantic side we consider Kripke structures (see [10]) $\langle \Omega, \mathcal{B}_1, ..., \mathcal{B}_n, \mathcal{B}_* \rangle$ where Ω is a set of states or possible worlds and, for every $j \in \{1, ..., n, *\}$, $\mathcal{B}_j$ is a binary relation on Ω. For every $\omega \in \Omega$ and for every $j \in \{1, ..., n, *\}$, let $\mathcal{B}_j(\omega) = \{\omega' \in \Omega : \omega\mathcal{B}_j\omega'\}$.

Definition 1 *A* $D45_n^*$ frame *is a Kripke structure* $\langle \Omega, \mathcal{B}_1, ..., \mathcal{B}_n, \mathcal{B}_* \rangle$ *that satisfies the following properties: for all* $\omega, \omega' \in \Omega$ *and* $i = 1, ..., n$,

1. *Seriality:* $\mathcal{B}_i(\omega) \neq \varnothing$,
2. *Transitivity: if* $\omega' \in \mathcal{B}_i(\omega)$ *then* $\mathcal{B}_i(\omega') \subseteq \mathcal{B}_i(\omega)$,
3. *Euclideanness: if* $\omega' \in \mathcal{B}_i(\omega)$ *then* $\mathcal{B}_i(\omega) \subseteq \mathcal{B}_i(\omega')$,
4. $\mathcal{B}_*$ *is the transitive closure of* $\mathcal{B}_1 \cup ... \cup \mathcal{B}_n$*, that is,* $\omega' \in \mathcal{B}_*(\omega)$ *if and only if there is a sequence* $\langle \omega_1, ..., \omega_m \rangle$ *in* Ω *such that (1)* $\omega_1 = \omega$*, (2)* $\omega_n = \omega'$ *and (3) for every* $k = 1, ..., m-1$ *there is an* $i_k \in \{1, ..., n\}$ *such that* $\omega_{k+1} \in \mathcal{B}_{i_k}(\omega_k)$.

An $S5_n^*$ frame *is a* $D45_n^*$ *frame that satisfies the following additional property: for all* $\omega \in \Omega$ *and* $i = 1, ..., n$,

5. *Reflexivity:* $\omega \in \mathcal{B}_i(\omega)$.

Figure 1 illustrates the following $D45_n^*$ frame: $n = 2$, $\Omega = \{\alpha, \beta, \gamma\}$, $\mathcal{B}_1(\alpha) = \mathcal{B}_1(\beta) = \{\alpha\}$, $\mathcal{B}_1(\gamma) = \{\gamma\}$, $\mathcal{B}_2(\alpha) = \{\alpha\}$ and $\mathcal{B}_2(\beta) = \mathcal{B}_2(\gamma) = \{\beta, \gamma\}$. Thus $\mathcal{B}_*(\alpha) = \{\alpha\}$ and $\mathcal{B}_*(\beta) = \mathcal{B}_*(\gamma) = \{\alpha, \beta, \gamma\}$. We shall use the following convention when representing frames graphically: states are represented by points and for every two states ω and ω' and for every $j \in \{1, ..., n, *\}$, $\omega' \in \mathcal{B}_j(\omega)$ if and only if either (i) ω and ω' are enclosed in the same cell (denoted by a rounded rectangle), or (ii) there is an arrow from ω to the cell containing ω', or (iii) there is an arrow from the cell containing ω to the cell containing ω'.

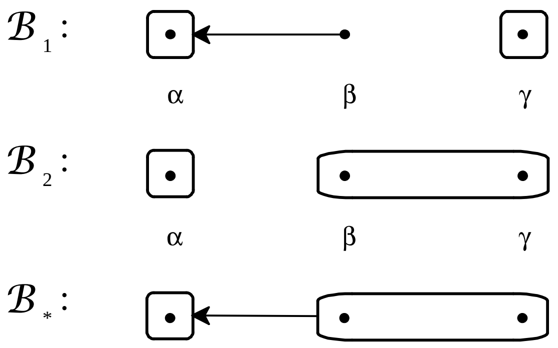


Figure 1

The link between syntax and semantics is given by the notions of valuation and model. A $D45_n^*$ *model* (respectively, $S5_n^*$ model) is obtained by adding to a $D45_n^*$ frame (respectively, $S5_n^*$ frame) a valuation $V : A \to 2^{\Omega}$, where A is the set of atomic propositions and 2^{Ω} denotes the set of subsets of Ω. Thus a valuation assigns to every atomic proposition p the set of states where p is true. Given a model and a formula ϕ, we denote by $\omega \models \phi$ the fact that ϕ is true at state ω. The truth set of ϕ is denoted by $\|\phi\|$, that is, $\|\phi\| = \{\omega \in \Omega : \omega \models \phi\}$. Truth of a formula at a state is defined recursively as follows.

if $p \in A$, $\omega \models p$ if and only if $\omega \in V(p)$,
$\omega \models \neg\phi$ if and only if $\omega \not\models \phi$,
$\omega \models \phi \vee \psi$ if and only if either $\omega \models \phi$ or $\omega \models \psi$ (or both),
$\omega \models B_i\phi \quad (i = 1, ..., n)$ if and only if $\mathcal{B}_i(\omega) \subseteq \|\phi\|$, that is,
if $\omega' \models \phi$ for all $\omega' \in \mathcal{B}_i(\omega)$,
$\omega \models B_*\phi$ if and only if $\mathcal{B}_*(\omega) \subseteq \|\phi\|$.

A formula ϕ is valid in a model if it is true at every state, that is, if $\|\phi\| = \Omega$. It is valid in a frame if it is valid in every model based on that frame.

The following result is well-known (see, for example [4], [11] and [12]).

Proposition 2 *Logic* $\mathbf{KD45}_n^*$ *is sound and complete with respect to the class of* $D45_n^*$ *frames, that is, a formula is a theorem of* $\mathbf{KD45}_n^*$ *if and only if it is valid in every* $D45_n^*$ *frame. Similarly,* $\mathbf{S5}_n^*$ *is sound and complete with respect to the class of* $S5_n^*$ *frames.*

3 Games and dominance

In this paper we restrict attention to finite strategic-form (or normal-form) games with ordinal payoffs, which are defined as follows.

Definition 3 *A* finite strategic-form game with ordinal payoffs *is a quintuple* $G = \left\langle N, \{S_i\}_{i \in N}, O, \{\succsim_i\}_{i \in N}, z \right\rangle$*, where*
$N = \{1, ..., n\}$ *is a set of players,*
S_i *is a finite set of strategies of player* $i \in N$*,*
O *is a finite set of outcomes,*
$\succsim_i$ *is player* i*'s ordering of* O*,*[2]
$z : S \to O$ *(where* $S = S_1 \times ... \times S_n$*) is a function that associates with every strategy profile* $s = (s_1, ..., s_n)$ *an outcome* $z(s) \in O$*.*

Given a player i we denote by S_{-i} the set of strategy profiles of the players other than i, that is, $S_{-i} = S_1 \times ... \times S_{i-1} \times S_{i+1} \times ... \times S_n$. When we want to focus on player i we shall denote the strategy profile $s \in S$ by (s_i, s_{-i}) where $s_i \in S_i$ and $s_{-i} \in S_{-i}$.

Definition 4 *Given a game* $G = \left\langle N, \{S_i\}_{i \in N}, O, \{\succsim_i\}_{i \in N}, z \right\rangle$ *and* $s_i \in S_i$*, we say that, for player* i*,* s_i *is* strictly dominated *in* G *if there is another strategy* $t_i \in S$ *such that – no matter what strategies the other players choose – player* i *prefers the outcome associated with* t_i *to the outcome associated with* s_i*, that is, if* $z(t_i, s_{-i}) \succ_i z(s_i, s_{-i})$*, for all* $s_{-i} \in S_{-i}$*.*

Let $G = \left\langle N, \{S_i\}_{i \in N}, O, \{\succsim_i\}_{i \in N}, z \right\rangle$ and $G' = \left\langle N', \{S_i'\}_{i \in N'}, O', \{\succsim_i'\}_{i \in N'}, z' \right\rangle$ be two games. We say that G' is a *subgame* of G if $N' = N$, $O' = O$, $\succsim_i' = \succsim_i$, $z' = z$ and, for every $i \in N$, $S_i' \subseteq S_i$.

Definition 5 *(IDSDS procedure). The Iterated Deletion of Strictly Dominated Strategies (IDSDS) is the following procedure. Given a game* $G = \left\langle N, \{S_i\}_{i \in N}, O, \{\succsim_i\}_{i \in N}, z \right\rangle$ *let* $\left\langle G^0, G^1, ..., G^m, ... \right\rangle$ *be the sequence of subgames of* G *defined recursively as follows. For all* $i \in N$*,*

1. let $S_i^0 = S_i$ *(thus* $G^0 = G$*) and let* $D_i^0 \subseteq S_i^0$ *be the set of strategies of player* i *that are strictly dominated in* G^0*;*

2. for $m \geq 1$*, let* $S_i^m = S_i^{m-1} \backslash D_i^{m-1}$ *and let* G^m *be the subgame of* G *with strategy sets* S_i^m*. Let* $D_i^m \subseteq S_i^m$ *be the set of strategies of player* i *that are strictly dominated in* G^m*.*

Let $S_i^\infty = \bigcap\limits_{m \in \mathbb{N}} S_i^m$ *(where* $\mathbb{N}$ *denotes the set of non-negative integers) and let* G^∞ *be the subgame of* G *with strategy sets* S_i^∞*. Let* $S^\infty = S_1^\infty \times ... \times S_n^\infty$*.*[3]

[2] That is, $\succsim_i$ is a binary relation on Ω that satisfies the following properties: for all $o, o', o'' \in O$, (1) either $o \succsim_i o'$ or $o' \succsim_i o$ (completeness) and (2) if $o \succsim_i o'$ and $o' \succsim_i o''$ then $o \succsim_i o''$ (transitivity). The interpretation of $o \succsim_i o'$ is that, according to player i, outcome o is at least as good as outcome o'. The strict ordering $\succ_i$ is defined as usual: $o \succ_i o'$ if and only if $o \succsim_i o'$ and not $o' \succsim_i o$. The interpretation of $o \succ_i o'$ is that player i prefers outcome o to outcome o'.

[3] Note that, since the strategy sets are finite, there exists an integer r such that $G^\infty = G^r = G^{r+k}$ for every $k \geq 1$.

The IDSDS procedure is illustrated in Figure 2, where
$S_1^0 = \{A, B, C, D\}$, $D_1^0 = \{D\}$, $S_2^0 = \{e, f, g\}$, $D_2^0 = \varnothing$;
$S_1^1 = \{A, B, C\}$, $D_1^1 = \varnothing$, $S_2^1 = \{e, f, g\}$, $D_2^1 = \{g\}$;
$S_1^2 = \{A, B, C\}$, $D_1^2 = \{C\}$, $S_2^2 = \{e, f\}$, $D_2^2 = \varnothing$;
$S_1^3 = \{A, B\}$, $D_1^3 = \varnothing$, $S_2^3 = \{e, f\}$, $D_2^3 = \{f\}$;
$S_1^4 = \{A, B\}$, $D_1^4 = \{B\}$, $S_2^\infty = S_2^4 = \{e\}$, $D_2^4 = \varnothing$;
$S_1^\infty = S_1^5 = \{A\}$. Thus $S^\infty = \{(A, e)\}$.

In Figure 2 we have represented the ranking $\succsim_i$ by a utility (or payoff) function $u_i : S \to \mathbb{R}$ satisfying the following property: $u_i(s) \geq u_i(s')$ if and only if $z(s) \succsim_i z(s')$ (in each cell, the first number is the payoff of player 1 while the second number is the payoff of player 2).

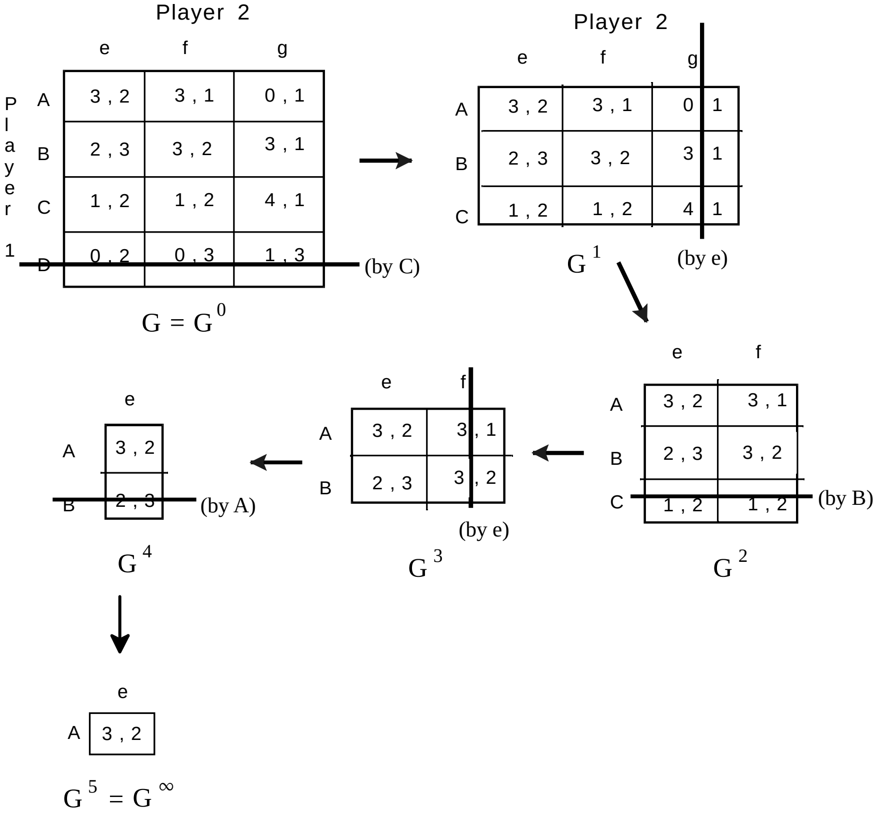


Figure 2

The next iterated deletion procedure differs from IDSDS in that at every round we delete strategy profiles rather than individual strategies.[4]

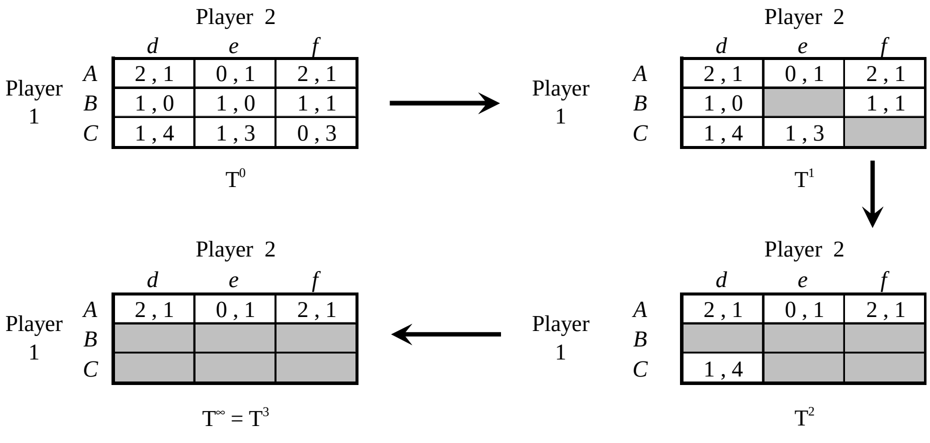
Definition 6 *(IDIP procedure) Given a game $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$, a subset of strategy profiles $X \subseteq S$ and a strategy profile $x \in X$, we say that x is* inferior relative to X *if there exists a player i and a strategy $s_i \in S_i$ of player i (thus s_i need not belong to the projection of X onto S_i) such that:*

1. $z(s_i, x_{-i}) \succ_i z(x_i, x_{-i})$, and

2. for all $s_{-i} \in S_{-i}$, if $(x_i, s_{-i}) \in X$ then $z(s_i, s_{-i}) \succeq_i z(x_i, s_{-i})$.

The Iterated Deletion of Inferior Profiles *(IDIP) is defined as follows. For $m \in \mathbb{N}$ define $T^m \subseteq S$ recursively as follows: $T^0 = S$ and, for $m \geq 1$, $T^m = T^{m-1} \backslash I^{m-1}$, where $I^{m-1} \subseteq T^{m-1}$ is the set of strategy profiles that are inferior relative to T^{m-1}. Let $T^\infty = \bigcap\limits_{m\in\mathbb{N}} T^m$.*[5]

The IDIP procedure is illustrated in Figure 3, where
$T^0 = S = \{(A,d),(A,e),(A,f),(B,d),(B,e),(B,f),(C,d),(C,e),(C,f)\}$,
$I^0 = \{(B,e),(C,f)\}$ (the elimination of (B,e) is done through player 2 and strategy f, while the elimination of (C,f) is done through player 1 and strategy B);
$T^1 = \{(A,d),(A,e),(A,f),(B,d),(B,f),(C,d),(C,e)\}$, $I^1 = \{(B,d),(B,f),(C,e)\}$ (the elimination of (B,d) and (B,f) is done through player 1 and strategy A, while the elimination of (C,e) is done through player 2 and strategy d);
$T^2 = \{(A,d),(A,e),(A,f),(C,d)\}$, $I^2 = \{(C,d)\}$ (the elimination of (C,d) is done through player 1 and strategy A);
$T^3 = \{(A,d),(A,e),(A,f)\}$, $I^3 = \varnothing$; thus $T^\infty = T^3$.



	d	e	f
A	2 , 1	0 , 1	2 , 1
B	1 , 0	1 , 0	1 , 1
C	1 , 4	1 , 3	0 , 3

T^0

	d	e	f
A	2 , 1	0 , 1	2 , 1
B	1 , 0	▨	1 , 1
C	1 , 4	1 , 3	▨

T^1

	d	e	f
A	2 , 1	0 , 1	2 , 1
B	▨	▨	▨
C	1 , 4	▨	▨

T^2

	d	e	f
A	2 , 1	0 , 1	2 , 1
B	▨	▨	▨
C	▨	▨	▨

$T^\infty = T^3$

Figure 3

[4] This procedure is the restriction to pure strategies of the algorithm introduced by Stalnaker [14].

[5] Since the strategy sets are finite, there exists an integer r such that $T^\infty = T^r = T^{r+k}$ for every $k \geq 1$.

4 Game logics

A logic is called a *game logic* if the set of atomic propositions upon which it is built contains atomic propositions of the following form:

- Strategy symbols s_i, t_i, ... The intended interpretation of s_i is "player i chooses strategy s_i".
- The symbols r_i whose intended interpretation is "player i is rational".
- Atomic propositions of the form $t_i \succeq_i s_i$, whose intended interpretation is "strategy t_i of player i is at least as good, for player i, as his strategy s_i", and atomic propositions of the form $t_i \succ_i s_i$, whose intended interpretation is "for player i strategy t_i is better than strategy s_i".

From now on we shall restrict attention to game logics.

Definition 7 *Fix a game* $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$ *with* $S_i = \{s_i^1, s_i^2, ..., s_i^{m_i}\}$ *(thus the cardinality of* S_i *is* m_i*). A game logic is called a* G-logic *if its set of strategy symbols is* $\{s_i^k\}_{i=1,...,n;k=1,...,m_i}$ *(with slight abuse of notation we use the symbol* s_i^k *to denote both an element of* S_i*, that is, a strategy of player* i*, and an element of* A*, that is, an atomic proposition whose intended interpretation is "player* i *chooses strategy* s_i^k*").*

Given a game G with $S_i = \{s_i^1, s_i^2, ..., s_i^{m_i}\}$, we denote by $\mathbf{L}_G^{D45}$(respectively, $\mathbf{L}_G^{S5}$) the $\mathbf{KD45}_n^*$ (respectively, $\mathbf{S5}_n^*$) G-logic that satisfies the following additional axioms: for all $i = 1, ..., n$ and for all $k, \ell = 1, ..., m_i$, with $k \neq \ell$,

$$\begin{array}{ll}
\left(s_i^1 \vee s_i^2 \vee ... \vee s_i^{m_i}\right) & (\mathbf{G1}) \\
\neg(s_i^k \wedge s_i^\ell) & (\mathbf{G2}) \\
s_i^k \rightarrow B_i s_i^k & (\mathbf{G3}) \\
\left(s_i^k \succeq_i s_i^\ell\right) \vee \left(s_i^\ell \succeq_i s_i^k\right) & (\mathbf{G4}) \\
\left(s_i^\ell \succ_i s_i^k\right) \leftrightarrow \left(\left(s_i^\ell \succeq_i s_i^k\right) \wedge \neg\left(s_i^k \succeq_i s_i^\ell\right)\right) & (\mathbf{G5})
\end{array}$$

Axiom **G1** says that player i chooses at least one strategy, while axiom **G2** says that player i cannot choose more than one strategy. Thus **G1** and **G2** together imply that each player chooses exactly one strategy. Axiom **G3**, on the other hand, says that player i is aware of his own choice: if he chooses strategy s_i^k then he believes that he chooses s_i^k. The remaining axioms state that the ordering of strategies is complete (**G4**) and that the corresponding strict ordering is defined as usual (**G5**).

Proposition 8 *Fix an arbitrary game* G*. The following is a theorem of logic* $\mathbf{L}_G^{D45}$*:* $B_i s_i^k \rightarrow s_i^k$*. That is, every player has correct beliefs about her own choice of strategy.*[6]

[6] Note that, in general, logic $\mathbf{L}_G^{D45}$ allows for incorrect beliefs. In particular, a player might have incorrect beliefs about the choices made by other players. By Proposition 8, however, a player cannot have mistaken beliefs about her own choice.

Proof. In the following 'PL' stands for Propositional Logic. Fix a player i and $k, \ell \in \{1, ..., m_i\}$ with $k \neq \ell$. Let ϕ denote the formula

$$(s_i^1 \vee ... \vee s_i^{m_i}) \wedge \neg s_i^1 \wedge ... \wedge \neg s_i^{k-1} \wedge \neg s_i^{k+1} \wedge ... \wedge \neg s_i^{m_i}.$$

1.	$\phi \to s_i^k$	tautology
2.	$\neg(s_i^k \wedge s_i^\ell)$	axiom **G2** (for $\ell \neq k$)
3.	$s_i^k \to \neg s_i^\ell$	2, PL
4.	$B_i s_i^k \to B_i \neg s_i^\ell$	3, rule RK [7]
5.	$B_i \neg s_i^\ell \to \neg B_i s_i^\ell$	axiom $\mathbf{D}_i$
6.	$s_i^\ell \to B_i s_i^\ell$	axiom **G3**
7.	$\neg B_i s_i^\ell \to \neg s_i^\ell$	6, PL
8.	$B_i s_i^k \to \neg s_i^\ell$	4, 5, 7, PL (for $\ell \neq k$)
9.	$s_i^1 \vee ... \vee s_i^{m_i}$	axiom **G1**
10.	$B_i s_i^k \to (s_i^1 \vee ... \vee s_i^{m_i})$	9, PL
11.	$B_i s_i^k \to \phi$	8 (for every $\ell \neq k$), 10, PL
12	$B_i s_i^k \to s_i^k$	1, 11, PL. ■

On the semantic side we consider models of games, which are defined as follows.

Definition 9 *Given a game $G = \langle N, \{S_i\}_{i \in N}, O, \{\succsim_i\}_{i \in N}, z \rangle$ and a Kripke frame $F = \langle \Omega, \{\mathcal{B}_i\}_{i \in N}, \mathcal{B}_* \rangle$, a* frame for G, *or* G-frame, *is obtained by adding to F n functions $\sigma_i : \Omega \to S_i$ ($i \in N$) satisfying the following property: if $\omega' \in \mathcal{B}_i(\omega)$ then $\sigma_i(\omega') = \sigma_i(\omega)$.*

Thus a G-frame adds to a Kripke frame a function that associates with every state ω a strategy profile $\sigma(\omega) = (\sigma_1(\omega), ..., \sigma_n(\omega)) \in S$. The restriction that if $\omega' \in \mathcal{B}_i(\omega)$ then $\sigma_i(\omega') = \sigma_i(\omega)$ is the semantic counterpart to axiom **G3**. Given a player i, as before we will denote $\sigma(\omega)$ by $(\sigma_i(\omega), \sigma_{-i}(\omega))$, where $\sigma_{-i}(\omega) \in S_{-i}$ is the profile of strategies of the players other than i.

We say that the G-frame $\langle \Omega, \{\mathcal{B}_i\}_{i \in N}, \mathcal{B}_*, \{\sigma_i\}_{i \in N}\} \rangle$ is a $D45_n^*$ G-frame (respectively, $S5_n^*$ G-frame) if the underlying Kripke frame $\langle \Omega, \{\mathcal{B}_i\}_{i \in N}, \mathcal{B}_* \rangle$ is a $D45_n^*$ frame (respectively, $S5_n^*$ frame: see Definition 1).

Definition 10 *Given a game G with $S_i = \{s_i^1, s_i^2, ..., s_i^{m_i}\}$, and a G-frame $F_G = \langle \Omega, \{\mathcal{B}_i\}_{i \in N}, \mathcal{B}_*, \{\sigma_i\}_{i \in N}\} \rangle$, a* model of G, *or* G-model, *is obtained by adding to F_G the following valuation:*

- $\omega \models s_i^k$ *if and only if* $\sigma_i(\omega) = s_i^k$,
- $\omega \models (s_i^k \succsim_i s_i^\ell)$ *if and only if* $z(s_i^k, \sigma_{-i}(\omega)) \succsim_i z(s_i^\ell, \sigma_{-i}(\omega))$ *and, similarly,* $\omega \models (s_i^k \succ_i s_i^\ell)$ *if and only if* $z(s_i^k, \sigma_{-i}(\omega)) \succ_i z(s_i^\ell, \sigma_{-i}(\omega))$.

Let $\mathbb{F}_G^{D45}$ (respectively, $\mathbb{F}_G^{S5}$) denote the set of $D45_n^*$ (respectively, $S5_n^*$) G-frames and $\mathbb{M}_G^{D45}$(respectively, $\mathbb{M}_G^{S5}$) the corresponding set of G-models.

[7] RK denotes the inference rule "from $\psi \to \chi$ infer $\Box\psi \to \Box\chi$", which is a derived rule of inference that applies to every modal operator $\Box$ that satisfies axiom **K** and the rule of Necessitation.

Figure 4 illustrates a game and a $D45_n^*$ frame for it. The corresponding model is given by the following valuation:

$\alpha \models B \wedge e \wedge (B \succ_1 A) \wedge (C \succ_1 A) \wedge (B \succeq_1 C) \wedge (C \succeq_1 B) \wedge (f \succ_2 d) \wedge (f \succ_2 e) \wedge (e \succeq_2 d) \wedge (d \succeq_2 e)$

$\beta \models B \wedge d \wedge (A \succ_1 B) \wedge (A \succ_1 C) \wedge (B \succeq_1 C) \wedge (C \succeq_1 B) \wedge (f \succ_2 d) \wedge (f \succ_2 e) \wedge (e \succeq_2 d) \wedge (d \succeq_2 e)$

$\gamma \models A \wedge d \wedge (A \succ_1 B) \wedge (A \succ_1 C) \wedge (B \succeq_1 C) \wedge (C \succeq_1 B) \wedge (d \succeq_2 e) \wedge (e \succeq_2 d) \wedge (d \succeq_2 f) \wedge (f \succeq_2 d) \wedge (e \succeq_2 f) \wedge (f \succeq_2 e)$.

		Player 2		
		d	*e*	*f*
Player 1	*A*	2 , 1	0 , 1	2 , 1
	B	1 , 0	1 , 0	1 , 1
	C	1 , 4	1 , 3	0 , 3

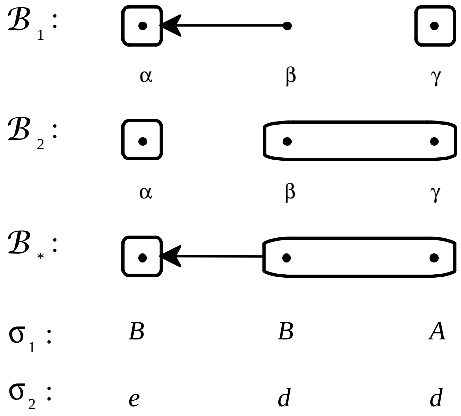


Figure 4

Proposition 11 *Logic* $\mathbf{L}_G^{D45}$ *(respectively,* $\mathbf{L}_G^{S5}$*) is sound with respect to the class of* $\mathbb{M}_G^{D45}$ *(respectively,* $\mathbb{M}_G^{S5}$*) models.*

Proof. It follows from Proposition 2 and the following observations: (1) axioms **G1** and **G2** are valid in every model because, for every state ω, there is a unique strategy $s_i^k \in S_i$ such that $\sigma_i(\omega) = s_i^k$ and, by the validation rules (see Definition 10), $\omega \models s_i^k$ if and only if $\sigma_i(\omega) = s_i^k$; (2) axiom **G3** is an immediate consequence of the fact (see Definition 9) that if $\omega' \in \mathcal{B}_i(\omega)$ then $\sigma_i(\omega') = \sigma_i(\omega)$; (3) axioms **G4** and **G5** are valid because, for every state ω, there is a unique profile of strategies $\sigma_{-i}(\omega)$ of the players other than i and the ordering $\succeq_i$ on O restricted to $z(S_i \times \sigma_{-i}(\omega))$ induces an ordering of S_i. ■

5 Rationality and common belief of rationality

So far we have not specified what it means for a player to be rational. The first extension of $\mathbf{L}_G^{D45}$ that we consider captures a very weak notion of rationality. The following axiom – called **WR** for 'Weak Rationality' – says that a player is *irrational* if she chooses a particular strategy while believing that a different strategy is better for her (recall that r_i is an atomic proposition whose intended interpretation is "player i is rational"):

$$s_i^k \wedge B_i(s_i^\ell \succ_i s_i^k) \rightarrow \neg r_i. \tag{\textbf{WR}}$$

Given a game G, let $\mathbf{L}_G^{D45}$+**WR** (respectively, $\mathbf{L}_G^{S5}$+**WR**) be the extension of $\mathbf{L}_G^{D45}$ (respectively, $\mathbf{L}_G^{S5}$) obtained by adding axiom **WR** to it.

The next axiom – called **SR** for 'Strong Rationality' – expresses a slightly stronger notion of rationality: it says that a player is irrational if she chooses a strategy while believing that a different strategy is at least as good and she considers it possible that this alternative strategy is actually better than the chosen one.

$$s_i^k \wedge B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k) \ \rightarrow \ \neg r_i. \tag{\textbf{SR}}$$

Given a game G, let $\mathbf{L}_G^{D45}$+**SR** (respectively, $\mathbf{L}_G^{S5}$+**SR**) be the extension of $\mathbf{L}_G^{D45}$(respectively, $\mathbf{L}_G^{S5}$) obtained by adding to axiom **SR**.

The following proposition shows that $\mathbf{L}_G^{D45}$+**SR** is an extension of $\mathbf{L}_G^{D45}$+**WR**.

Proposition 12 ***WR*** *is a theorem of* $\mathbf{L}_G^{D45}$+**SR**.

Proof. As before, 'PL stands for Propositional Logic'.

1.	$s_i^k \wedge B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k) \ \rightarrow \ \neg r_i$	Axiom **SR**
2.	$(r_i \wedge s_i^k) \rightarrow \neg \big(B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k) \big)$	1, PL
3.	$(s_i^\ell \succ_i s_i^k) \leftrightarrow (s_i^\ell \succeq_i s_i^k) \wedge \neg (s_i^k \succeq_i s_i^\ell)$	Axiom **G5**
4.	$(s_i^\ell \succ_i s_i^k) \rightarrow (s_i^\ell \succeq_i s_i^k)$	3, PL
5.	$B_i(s_i^\ell \succ_i s_i^k) \rightarrow B_i(s_i^\ell \succeq_i s_i^k)$	4, RK (see footnote 7)
6.	$B_i(s_i^\ell \succ_i s_i^k) \rightarrow \neg B_i \neg (s_i^\ell \succ_i s_i^k)$	Axiom $\mathbf{D}_i$
7.	$B_i(s_i^\ell \succ_i s_i^k) \rightarrow \big(B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k) \big)$	5, 6, PL
8.	$\neg \big(B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k) \big) \rightarrow \neg B_i(s_i^\ell \succ_i s_i^k)$	7, PL
9.	$(r_i \wedge s_i^k) \rightarrow \neg B_i(s_i^\ell \succ_i s_i^k)$	2, 8, PL
10.	$s_i^k \wedge B_i(s_i^\ell \succ_i s_i^k) \rightarrow \neg r_i.$	9, PL.

■

Definition 13 *Given a game G, let $\mathbb{M}_G^{D45|WR} \subseteq \mathbb{M}_G^{D45}$ (respectively, $\mathbb{M}_G^{S5|WR} \subseteq \mathbb{M}_G^{S5}$) be the class of $D45_n^*$ (respectively, $S5_n^*$) G-models (see Definition 10) where the valuation function satisfies the following additional condition:*

- *$\omega \models r_i$ if and only if, for every $s_i \in S_i$ there exists an $\omega' \in \mathcal{B}_i(\omega)$ such that $z(\sigma_i(\omega), \sigma_{-i}(\omega')) \succsim_i z(s_i, \sigma_{-i}(\omega'))$.*

For instance, in the model based on the frame of Figure 4 we have that $\alpha \models (r_1 \wedge \neg r_2)$, $\beta \models (r_1 \wedge r_2)$ and $\gamma \models (r_1 \wedge r_2)$. To see, for example, that $\beta \models r_2$ note that $\sigma_2(\beta) = d$ and for strategy f we have that $\gamma \in \mathcal{B}_2(\beta)$, $\sigma_1(\gamma) = A$ and $z(A, d) \succsim_2 z(A, f)$, while for strategy e we have that $\beta \in \mathcal{B}_2(\beta)$, $\sigma_1(\beta) = B$ and $z(B, d) \succsim_2 z(B, e)$.

Thus, in the model based on the frame of Figure 4, we have that at state β both players are rational, player 2 believes that player 1 is rational, but player 1 mistakenly believes that player 2 is irrational: $\beta \models r_1 \wedge r_2 \wedge B_2 r_1 \wedge B_1 \neg r_2$.

Proposition 14 *Logic $\mathbf{L}_G^{D45}$+**WR** (respectively, $\mathbf{L}_G^{S5}$+**WR**) is sound with respect to the class of models $\mathbb{M}_G^{D45|WR}$ (respectively, $\mathbb{M}_G^{S5|WR}$).*

Proof. By Proposition 11 it is sufficient to show that axiom **WR** is valid in an arbitrary such model. Suppose that $\omega \models s_i^k \wedge B_i(s_i^\ell \succ_i s_i^k)$. Then $\sigma_i(\omega) = s_i^k$ and $\mathcal{B}_i(\omega) \subseteq \left\| s_i^\ell \succ_i s_i^k \right\|$, that is (see Definition 10), $z(s_i^\ell, \sigma_{-i}(\omega')) \succ_i z(s_i^k, \sigma_{-i}(\omega'))$, for every $\omega' \in \mathcal{B}_i(\omega)$. It follows from Definition 13 that $\omega \models \neg r_i$. (Recall that, by Definition 9, $\sigma_i(\omega') = \sigma_i(\omega)$, for all $\omega' \in \mathcal{B}_i(\omega)$.) ■

The following proposition shows that common belief of the weak notion of rationality expressed by axiom **WR** characterizes the Iterated Deletion of Strictly Dominated Strategies (see Definition 5).[8]

Proposition 15 *Fix a finite strategic-form game with ordinal payoffs G. Then both (A) and (B) below hold.*

(A) Fix an arbitrary model in $\mathbb{M}_G^{D45|WR}$ and an arbitrary state α. If $\alpha \models B_*(r_1 \wedge ... \wedge r_n)$ then $\sigma(\alpha) \in S^\infty$.

(B) For every $s \in S^\infty$ there exists a model in $\mathbb{M}_G^{S5|WR}$ and a state α such that (1) $\sigma(\alpha) = s$ and (2) $\alpha \models K_*(r_1 \wedge ... \wedge r_n)$.[9]

Proof. (A) Fix a model in $\mathbb{M}_G^{D45|WR}$ and a state α and suppose that $\alpha \models B_*(r_1 \wedge ... \wedge r_n)$. The proof is by induction. First we show that, for every player $i = 1, ..., n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_i(\omega) \notin D_i^0$ (see Definition 5). Suppose not. Then there exist a player i and a $\beta \in \mathcal{B}_*(\alpha)$ such that $\sigma_i(\beta) \in D_i^0$, that is, strategy $\sigma_i(\beta)$ of player i is strictly dominated in G by some other strategy $\hat{s}_i \in S_i$: for every $s_{-i} \in S_{-i}$, $z(\hat{s}_i, s_{-i}) \succ_i z(\sigma_i(\beta), s_{-i})$. Thus, for every $\omega \in \mathcal{B}_i(\beta)$, $z(\hat{s}_i, \sigma_{-i}(\omega)) \succ_i z(\sigma_i(\beta), \sigma_{-i}(\omega))$. It follows from Definition 13 that

[8] Proposition 15 is the syntactic-based, ordinal version of a semantic, probabilistic-based result of Stalnaker [14]. As noted in the Introduction, Stalnaker's result was, in turn, a reformulation of earlier results due to Bernheim [2], Pearce [13], Tan and Werlang [15] and Brandenburger and Dekel [7].

[9] Recall that, in order to emphasize the distinction between belief and knowledge, when dealing with the latter we denote the modal operators by K_i and K_* rather than B_i and B_*, respectively. Similarly, we shall denote the accessibility relations by $\mathcal{K}_i$ and $\mathcal{K}_*$ rather than $\mathcal{B}_i$ and $\mathcal{B}_*$, respectively.

$\beta \models \neg r_i$, contradicting the hypothesis that $\beta \in \mathcal{B}_*(\alpha)$ and $\alpha \models B_* r_i$. Since, for every $\omega \in \Omega$, $\sigma_i(\omega) \in S_i^0 = S_i$, it follows that, for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_i(\omega) \in S_i^0 \backslash D_i^0 = S_i^1$. Next we prove the inductive step. Fix an integer $m \geq 1$ and suppose that, for every player $j = 1, ..., n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_j(\omega) \in S_j^m$. We want to show that, for every player $i = 1, ..., n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_i(\omega) \notin D_i^m$. Suppose not. Then there exist a player i and a $\beta \in \mathcal{B}_*(\alpha)$ such that $\sigma_i(\beta) \in D_i^m$, that is, strategy $\sigma_i(\beta)$ is strictly dominated in G^m by some other strategy $\tilde{s}_i \in S_i^m$. Since, by hypothesis, for every player j and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_j(\omega) \in S_j^m$, it follows – since $\mathcal{B}_i(\beta) \subseteq \mathcal{B}_*(\beta) \subseteq \mathcal{B}_*(\alpha)$ (see Definition 1) – that for every $\omega \in \mathcal{B}_i(\beta)$, $z(\tilde{s}_i, \sigma_{-i}(\omega)) \succ_i z(\sigma_i(\beta), \sigma_{-i}(\omega))$. Thus, by Definition 13 , $\beta \models \neg r_i$, contradicting the fact that $\beta \in \mathcal{B}_*(\alpha)$ and $\alpha \models B_* r_i$. Thus, for every player $i = 1, ..., n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_i(\omega) \in \bigcap\limits_{m \in \mathbb{N}} S_i^m = S_i^\infty$. It only remains to show that $\sigma_i(\alpha) \in S_i^\infty$. Fix an arbitrary $\beta \in \mathcal{B}_i(\alpha)$. Since $\mathcal{B}_i(\alpha) \subseteq \mathcal{B}_*(\alpha)$, $\beta \in \mathcal{B}_*(\alpha)$. Thus $\sigma_i(\beta) \in S_i^\infty$. By Definition 9, $\sigma_i(\beta) = \sigma_i(\alpha)$. Thus $\sigma_i(\alpha) \in S_i^\infty$.

(B) Let m be the cardinality of $S^\infty = S_1^\infty \times ... \times S_n^\infty$ and let $\Omega = \{\omega_1, ..., \omega_m\}$. Let $\sigma : \Omega \to S^\infty$ be a one-to-one function. For every player i, define the following equivalence relation on Ω: $\omega \mathcal{K}_i \omega'$ if and only if $\sigma_i(\omega) = \sigma_i(\omega')$, where $\sigma_i(\omega)$ is the i^{th} coordinate of $\sigma(\omega)$. Let $\mathcal{K}_*$ be the transitive closure of $\bigcup\limits_{i \in N} \mathcal{K}_i$ (then, for every $\omega \in \Omega$, $\mathcal{K}_*(\omega) = \Omega$). The structure so defined is clearly an $S5_n^*$ G-frame. Consider the model corresponding to this frame (see Definition 10). Fix an arbitrary state ω and an arbitrary player i. By definition of S^∞, for every $s_i \in S_i$ there exists an $\omega' \in \mathcal{K}_i(\omega)$ such that $z(\sigma_i(\omega), \sigma_{-i}(\omega')) \succeq_i z(s_i, \sigma_{-i}(\omega'))$. Thus $\omega \models r_i$ (see Definition 13). Hence, for every $\omega \in \Omega$, $\omega \models (r_1 \wedge ... \wedge r_n)$ and, therefore, for every $\omega \in \Omega$, $\omega \models K_*(r_1 \wedge ... \wedge r_n)$. ■

Remark 16 *Since $\mathbb{M}_G^{S5|WR} \subseteq \mathbb{M}_G^{D45|WR}$ it follows from part (B) of Proposition 15 that the implications of common* belief *of rationality (as implicitly defined by axiom **WR**) are the same as the implications of common* knowledge *of rationality.*

The above observation is not true for the stronger notion of rationality expressed by axiom **SR**, to which we now turn.

Definition 17 *Given a game G, let $\mathbb{M}_G^{D45|SR} \subseteq \mathbb{M}_G^{D45}$ (respectively, $\mathbb{M}_G^{S5|SR} \subseteq \mathbb{M}_G^{S5}$) be the class of $D45$ (respectively, $S5$) G-models where the valuation function satisfies the following condition:*

- *$\omega \models r_i$ if and only if, for every $s_i \in S_i$, whenever there exists an $\omega' \in \mathcal{B}_i(\omega)$ such that $z(s_i, \sigma_{-i}(\omega')) \succ_i z(\sigma_i(\omega), \sigma_{-i}(\omega'))$ then there exists an $\omega'' \in \mathcal{B}_i(\omega)$ such that $z(\sigma_i(\omega), \sigma_{-i}(\omega'')) \succ_i z(s_i, \sigma_{-i}(\omega''))$.*

Thus, at state ω, player i is rational if, whenever there is a strategy s_i of his which is better than $\sigma_i(\omega)$ (the strategy he is actually using at ω) at some state ω' that he considers possible at ω, then $\sigma_i(\omega)$ is better than s_i at some other state ω'' that he considers possible at ω. For example, in the model based on the frame of Figure 4 we have that $\omega \models (r_1 \wedge \neg r_2)$ for every $\omega \in \{\alpha, \beta, \gamma\}$. At state β, for instance, player 2 is choosing strategy d when there is another strategy of his, namely f, which is better than d at β and as good as d at γ and $\mathcal{B}_2(\beta) = \{\beta, \gamma\}$. Thus he is irrational according to Definition 17.

It is easily verified that $\mathbb{M}_G^{D45|SR} \subseteq \mathbb{M}_G^{D45|WR}$ and, similarly, $\mathbb{M}_G^{S5|SR} \subseteq \mathbb{M}_G^{S5|WR}$.

Proposition 18 *Logic* $\mathbf{L}_G^{D45}$+**SR** *(respectively,* $\mathbf{L}_G^{S5}$+**SR***) is sound with respect to the class of models* $\mathbb{M}_G^{D45|SR}$ *(respectively,* $\mathbb{M}_G^{S5|SR}$*).*

Proof. By Proposition 11 it is sufficient to show that axiom **SR** is valid in an arbitrary such model. Suppose that $\omega \models s_i^k \wedge B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k)$. Then $\sigma_i(\omega) = s_i^k$ and $\mathcal{B}_i(\omega) \subseteq \left\| s_i^\ell \succeq_i s_i^k \right\|$ [that is – see Definition 10 – $z(s_i^\ell, \sigma_{-i}(\omega')) \succeq_i z(s_i^k, \sigma_{-i}(\omega'))$, for every $\omega' \in \mathcal{B}_i(\omega)$] and there is an $\omega'' \in \mathcal{B}_i(\omega)$ such that $\omega'' \models s_i^\ell \succ_i s_i^k$, that is, $z(s_i^\ell, \sigma_{-i}(\omega'')) \succ_i z(s_i^k, \sigma_{-i}(\omega''))$. It follows from Definition 17 that $\omega \models \neg r_i$. ■

The following proposition shows that common *knowledge* of the stronger notion of rationality expressed by axiom **SR** characterizes the Iterated Deletion of Inferior Profiles (see Definition 6).[10]

Proposition 19 *Fix a finite strategic-form game with ordinal payoffs G. Then both (A) and (B) below hold.*

(A) Fix an arbitrary model in $\mathbb{M}_G^{S5|SR}$ *and an arbitrary state* α*. If* $\alpha \models K_*(r_1 \wedge ... \wedge r_n)$ *then* $\sigma(\alpha) \in T^\infty$*.*

(B) For every $s \in T^\infty$ *there exists a model in* $\mathbb{M}_G^{S5|SR}$ *and a state* α *such that (1)* $\sigma(\alpha) = s$ *and (2)* $\alpha \models K_*(r_1 \wedge ... \wedge r_n)$*.*

Proof. (A) As in the case of Proposition 15, the proof is by induction. Fix a model in $\mathbb{M}_G^{S5|SR}$ and a state α and suppose that $\alpha \models K_*(r_1 \wedge ... \wedge r_n)$. First we show that, for every $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \notin I^0$ (see Definition 6). Suppose, by contradiction, that there exists a $\beta \in \mathcal{K}_*(\alpha)$ such that $\sigma(\beta) \in I^0$, that is, $\sigma(\beta)$ is inferior relative to the entire set of strategy profiles S. Then there exists a player i and a strategy $\hat{s}_i \in S_i$ such that $z(\hat{s}_i, \sigma_{-i}(\beta)) \succ_i z(\sigma_i(\beta), \sigma_{-i}(\beta))$, and, for every $s_{-i} \in S_{-i}$, $z(\hat{s}_i, s_{-i}) \succeq_i z(\sigma_i(\beta), s_{-i})$. Thus $z(\hat{s}_i, \sigma_{-i}(\omega)) \succeq_i z(\sigma_i(\beta), \sigma_{-i}(\omega))$, for every $\omega \in \mathcal{K}_i(\beta)$; furthermore, by reflexivity of $\mathcal{K}_i$ (see Definition 1), $\beta \in \mathcal{K}_i(\beta)$. It follows from Definition 17 that $\beta \models \neg r_i$. Since $\beta \in \mathcal{K}_*(\alpha)$, this contradicts the hypothesis that $\alpha \models K_* r_i$. Thus, since, for every $\omega \in \Omega$, $\sigma(\omega) \in S = T^0$ we have shown that, for every $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \in T^0 \backslash I^0 = T^1$.

Now we prove the inductive step. Fix an integer $m \geq 1$ and suppose that, for every $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \in T^m$. We want to show that, for every $\omega \in \mathcal{K}_*(\alpha)$,

[10] Proposition 19 is the syntactic-based, ordinal version of a semantic, probabilistic-based result due to Stalnaker [14] . For a correction of that result see Bonanno and Nehring [5].

$\sigma(\omega) \notin I^m$. Suppose, by contradiction, that there exists a $\beta \in \mathcal{K}_*(\alpha)$ such that $\sigma(\beta) \in I^m$, that is, $\sigma(\beta)$ is inferior relative to T^m Then there exists a player i and a strategy $\tilde{s}_i \in S_i$ such that $z(\tilde{s}_i, \sigma_{-i}(\beta)) \succ_i z(\sigma_i(\beta), \sigma_{-i}(\beta))$, and, for every $s_{-i} \in S_{-i}$, if $(\tilde{s}_i, s_{-i}) \in T^m$ then $z(\tilde{s}_i, s_{-i}) \succsim_i z(\sigma_i(\beta), s_{-i})$. By Definition 9, for every $\omega \in \mathcal{K}_i(\beta)$, $\sigma_i(\omega) = \sigma_i(\beta)$ and by the induction hypothesis, for every $\omega \in \mathcal{K}_*(\alpha)$, $(\sigma_i(\omega), \sigma_{-i}(\omega)) \in T^m$. Thus, since $\mathcal{K}_i(\beta) \subseteq \mathcal{K}_*(\beta) \subseteq \mathcal{K}_*(\alpha)$, we have that, for every $\omega \in \mathcal{K}_i(\beta)$, $(\sigma_i(\beta), \sigma_{-i}(\omega)) \in T^m$. By reflexivity of $\mathcal{K}_i$, $\beta \in \mathcal{K}_i(\beta)$. It follows from Definition 17 that $\beta \models \neg r_i$. Since $\beta \in \mathcal{K}_*(\alpha)$, this contradicts the hypothesis that $\alpha \models K_* r_i$.
Thus, we have shown by induction that, for every $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \in \bigcap\limits_{m \in \mathbb{N}} T^m = T^\infty$. It only remains to establish that $\sigma(\alpha) \in T^\infty$, but this follows from reflexivity of $\mathcal{K}_*$.

(B) Let m be the cardinality of T^∞ and let $\Omega = \{\omega_1, ..., \omega_m\}$. Let $\sigma : \Omega \to T^\infty$ be a one-to-one function. For every player i, define the following equivalence relation on Ω: $\omega \mathcal{K}_i \omega'$ if and only if $\sigma_i(\omega) = \sigma_i(\omega')$, where $\sigma_i(\omega)$ is the i^{th} coordinate of $\sigma(\omega)$. Let $\mathcal{K}_*$ be the transitive closure of $\bigcup\limits_{i \in N} \mathcal{K}_i$ (then, for every $\omega \in \Omega$, $\mathcal{K}_*(\omega) = \Omega$). The structure so defined is clearly an $S5_n^*$ G-frame. Consider the model corresponding to this frame (see Definition 10). Fix an arbitrary state ω and an arbitrary player i. By definition of T^∞, for every player i and every $s_i \in S_i$ if there exists an $\omega' \in \mathcal{K}_i(\omega)$ such that $z(s_i, \sigma_{-i}(\omega')) \succ_i z(\sigma_i(\omega), \sigma_{-i}(\omega'))$ then there exists an $\omega'' \in \mathcal{K}_i(\omega)$ such that $z(\sigma_i(\omega), \sigma_{-i}(\omega'')) \succ_i z(s_i, \sigma_{-i}(\omega''))$. Thus $\omega \models r_i$ (see Definition 17). Hence, for every $\omega \in \Omega$, $\omega \models (r_1 \wedge ... \wedge r_n)$ and, therefore, for every $\omega \in \Omega$, $\omega \models K_*(r_1 \wedge ... \wedge r_n)$. ■

Note that Proposition 19 is not true if one replaces knowledge with belief, as illustrated in the game and frame of Figure 5. In the corresponding model we have that, according to the stronger notion of rationality expressed by Definition 17, $\alpha \models r_1 \wedge r_2$ and $\beta \models r_1 \wedge r_2$, so that $\alpha \models B_*(r_1 \wedge r_2)$, despite the fact that $\sigma(\alpha) = (B, d)$, which is an inferior strategy profile (relative to the entire game).[11] In other words, common belief of rationality, as expressed by axiom **SR**, is compatible with the players collectively choosing an inferior strategy profile. Thus, unlike the weaker notion expressed by axiom **WR** (see Remark 16), with axiom **SR** there is a crucial difference between the implications of common belief of rationality and those of common knowledge of rationality.

[11] In the game of Figure 5 we have that $S^\infty = S = \{(A, c), (A, d), (B, c), (B, d)\}$ while $T^\infty = \{(A, c), (B, c)\}$.

		Player 2	
		c	d
Player 1	A	1 , 1	1 , 0
	B	1 , 1	0 , 1

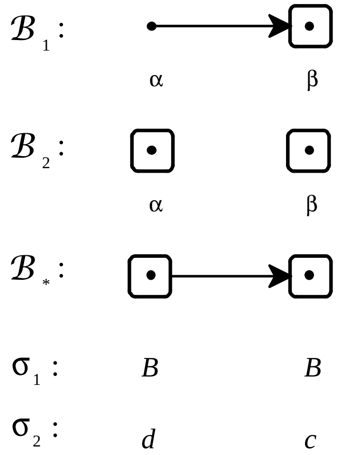


Figure 5

6 Frame characterization

The characterization results proved in the previous Section (Propositions 15 and 19) are not characterizations in the sense in which this expression is used in modal logic, namely characterization of axioms in terms of classes of frames (see [3] p. 125). In this Section we provide a reformulation of our results in terms of frame characterizations.

Definition 20 *An axiom characterizes (or is characterized by) a class $\mathbb{F}$ of Kripke frames if (1) the axiom is valid in every model based on a frame that belongs to $\mathbb{F}$ and, conversely, (2) if a frame does not belong to $\mathbb{F}$ then there is a model based on that frame and a state in that model at which an instance of the axiom is falsified.*[12]

We now modify the previous analysis as follows. First of all, we drop the symbols r_i from the set of atomic propositions and correspondingly drop the definitions of the classes of models $\mathbb{M}_G^{D45|WR}$, $\mathbb{M}_G^{S5|WR}$, $\mathbb{M}_G^{D45|SR}$ and $\mathbb{M}_G^{S5|SR}$ (Definitions 13 and 17). Secondly we modify axioms **WR** and **SR** as follows:

[12] For example, as is well known, the axiom $B_i\phi \to B_iB_i\phi$ is characterized by the class of frames where the relation $\mathcal{B}_i$ is transitive.

$$s_i^k \to \neg B_i(s_i^\ell \succ_i s_i^k) \tag{$\mathbf{WR'}$}$$

$$s_i^k \to \neg\left(B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg(s_i^\ell \succ_i s_i^k)\right). \tag{$\mathbf{SR'}$}$$

Axioms $\mathbf{WR'}$ and $\mathbf{SR'}$ correspond to adding to the logics considered previously the axiom that players are rational. In fact, from r_i and **WR** one obtains $\mathbf{WR'}$ (using Modus Ponens) and similarly for $\mathbf{SR'}$.

The next proposition is the counterpart of Proposition 15.

Proposition 21 *Subject to the valuation of the atomic propositions s_i^k, $\left(s_i^\ell \succeq_i s_i^k\right)$ and $\left(s_i^\ell \succ_i s_i^k\right)$ specified in Definition 10, axiom* $\boldsymbol{WR'}$ *is characterized by the class of $D45_n^*$ game frames (see Definition 9) that satisfy the following property: for all $i \in N$ and for all $\omega \in \Omega$, $\sigma_i(\omega) \in S_i^\infty$.*

Proof. Fix a model based on a frame in this class, a state α, a player i and two strategies s_i^k and s_i^ℓ of player i. Suppose that $a \models s_i^k$, that is, $\sigma_i(\alpha) = s_i^k$. We want to show that $a \models \neg B_i(s_i^\ell \succ_i s_i^k)$. Suppose not. Then $\mathcal{B}_i(\alpha) \subseteq \left\|s_i^\ell \succ_i s_i^k\right\|$, that is,

$$\text{for every } \omega \in \mathcal{B}_i(\alpha), \quad z(s_i^\ell, \sigma_{-i}(\omega)) \succ_i z(s_i^k, \sigma_{-i}(\omega)). \tag{1}$$

By hypothesis, for every player $j \neq i$ and for every $\omega \in \Omega$, $\sigma_j(\omega) \in S_j^\infty$. Thus it follows from this and (1) that $s_i^k \notin S_i^\infty$, contradicting the hypotheses that $\sigma_i(\alpha) = s_i^k$ and $\sigma_i(\omega) \in S_i^\infty$ for all $\omega \in \Omega$.

Conversely, fix a $D45_n^*$ frame not in the class. For every state ω and every player j let $m(\omega, j) = \begin{cases} \infty & \text{if } \sigma_j(\omega) \in S_j^\infty \\ m & \text{if } \sigma_j(\omega) \in D_j^m \end{cases}$. Let $\hat{m} = \min\{m(\omega, j) : j \in N, \omega \in \Omega\}$. By our hypothesis about the frame, $\hat{m} \in \mathbb{N}$. Let $i \in N$ and $\alpha \in \Omega$ be such that $\hat{m} = m(i, \alpha)$. Then

$$\sigma_i(\alpha) \in D_i^{\hat{m}} \tag{2}$$

and, since (see Definition 5), for every $j \in N$ and for every $p, q \in \mathbb{N} \cup \{\infty\}$, $S_j^{p+q} \subseteq S_j^p$,

$$\text{for every } j \in N \text{ and } \omega \in \Omega, \quad \sigma_j(\omega) \in S_j^{\hat{m}}. \tag{3}$$

Let $s_i^k = \sigma_i(\alpha)$. By (2) and (3) , there exists a $s_i^\ell \in S_i$ such that, for every $\omega \in \Omega$, $\quad z(s_i^\ell, \sigma_{-i}(\omega)) \succ_i z(s_i^k, \sigma_{-i}(\omega))$. Thus $\mathcal{B}_i(\alpha) \subseteq \left\|s_i^\ell \succ_i s_i^k\right\|$ and thus $\alpha \models s_i^k \wedge B_i \left\|s_i^\ell \succ_i s_i^k\right\|$, so that axiom $\mathbf{WR'}$ is falsified at α. ■

The next proposition is the counterpart of Proposition 19.

Proposition 22 *Subject to the valuation of the atomic propositions s_i^k, $\left(s_i^\ell \succeq_i s_i^k\right)$ and $\left(s_i^\ell \succ_i s_i^k\right)$ specified in Definition 10, axiom* ***SR′*** *is characterized by the class of $S5_n^*$ game frames (see Definition 9) that satisfy the following property: for all $\omega \in \Omega$, $\sigma(\omega) \in T^\infty$.*

Proof. Fix a model based on a frame in this class, a state α, a player i and two strategies s_i^k and s_i^ℓ of player i. Suppose that $a \models s_i^k \wedge B_i\left(s_i^\ell \succeq_i s_i^k\right)$, that is, $\sigma_i(\alpha) = s_i^k$ and $\mathcal{B}_i(\alpha) \subseteq \left\|s_i^\ell \succeq_i s_i^k\right\|$, that is,

$$\text{for all } \omega \in \mathcal{B}_i(\alpha), \quad z(s_i^\ell, \sigma_{-i}(\omega)) \succeq_i z(s_i^k, \sigma_{-i}(\omega)). \tag{4}$$

We want to show that $a \models B_i\neg(s_i^\ell \succ_i s_i^k)$. Suppose not. Then there exists a $\beta \in \mathcal{B}_i(\alpha)$ such that $\beta \models \left(s_i^\ell \succ_i s_i^k\right)$, that is,

$$z(s_i^\ell, \sigma_{-i}(\beta)) \succ_i z(s_i^k, \sigma_{-i}(\beta)). \tag{5}$$

It follows from (4) and (5) that $(s_i^k, \sigma_{-i}(\beta)) = (\sigma_i(\beta), \sigma_{-i}(\beta))$ is inferior relative to the set $\{s \in S : s = \sigma(\omega) \text{ for some } \omega \in \mathcal{B}_i(\alpha)\}$, contradicting the hypothesis that $\sigma(\omega) \in T^\infty$ for all $\omega \in \Omega$.

Conversely, fix an $S5_n^*$ frame not in the class. For every state ω, let $m(\omega) = \begin{cases} \infty & \text{if } \sigma(\omega) \in T^\infty \\ m & \text{if } \sigma(\omega) \in I^m = T^m \backslash T^{m+1} \end{cases}$. Let $m_0 = \min\{m(\omega) : \omega \in \Omega\}$. By our hypothesis about the frame, $m_0 \in \mathbb{N}$. Let $\alpha \in \Omega$ be such that $m_0 = m(\alpha)$. Then $\sigma(\alpha) \in I^{m_0}$, that is, there is a player i and a strategy $s_i^\ell \in S_i$ such that

$$z(s_i^\ell, \sigma_{-i}(\alpha)) \succ_i z(\sigma_i(\alpha), \sigma_{-i}(\alpha)) \tag{6}$$

and

$$\forall \omega \in \Omega, \text{ if } (\sigma_i(\alpha), \sigma_{-i}(\omega)) \in T^{m_0} \text{ then } z(s_i^\ell, \sigma_{-i}(\omega)) \succeq_i z(\sigma_i(\alpha), \sigma_{-i}(\omega)). \tag{7}$$

By definition of m_0, since (see Definition 6) for every $p, q \in \mathbb{N} \cup \{\infty\}$, $T^{p+q} \subseteq T^p$, for every $\omega \in \Omega$, $\sigma(\omega) \in T^{m_0}$. Thus, letting $s_i^k = \sigma_i(\alpha)$, it follows from (7) that $\mathcal{B}_i(\alpha) \subseteq \left\|s_i^\ell \succeq_i s_i^k\right\|$, that is, $\alpha \models B_i(s_i^\ell \succeq_i s_i^k)$. Since the frame is an S5 frame, $\mathcal{B}_i$ is reflexive and, therefore, $\alpha \in \mathcal{B}_i(\alpha)$. It follows from this and (6) that $\alpha \models \neg B_i\neg(s_i^\ell \succ_i s_i^k)$. Thus $\alpha \models s_i^k \wedge B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i\neg(s_i^\ell \succ_i s_i^k)$, so that axiom **SR′** is falsified at α. ■

There appears to be an important difference between the results of Section 5 and those of this section, namely that, while Propositions 15 and 19 give a *local* result, Propositions 21 and 22 provide a global one. For example, Proposition 15 states that if *at a state* there is common belief of rationality, then the strategy profile played *at that state* belongs to S^∞, while its counterpart in this Section, namely Proposition 21, states that the strategy profile played *at every state* belongs to S^∞. As a matter of fact, the results of Section 5 are also global in nature. Consider, for example, Proposition 15. Fix a model and a state α and suppose that $\alpha \models B_*(r_1 \wedge ... \wedge r_n)$. Since, for every formula ϕ, $B_*\phi \rightarrow B_*B_*\phi$ is a theorem of $\mathbf{KD45}_n^*$, it follows that $\alpha \models B_* B_* (r_1 \wedge ... \wedge r_n)$, that is, for every $\omega \in \mathcal{B}_*(\alpha)$, $\omega \models B_*(r_1 \wedge ... \wedge r_n)$. Thus, it follows from Proposition 15 that $\sigma(\omega) \in S^\infty$, for every $\omega \in \mathcal{B}_*(\alpha)$.[13] That is, if at a state there is common belief of rationality, then at that state, *as well as at all states reachable from it by the common belief relation* $\mathcal{B}_*$, it is true that the strategy profile played belongs to S^∞. This is essentially a global result, since from the point of view of a state α, the "global" space is precisely the set $\mathcal{B}_*(\alpha)$.

Thus the only real difference between the results of Section 5 and those of this section lies in the fact that Propositions 15 and 19 bring out the role of common belief by mimicking the informal argument that if player 1 is rational then she won't choose a strategy $s_1 \in D_1^0$ and if player 2 believes that player 1 is rational then he believes that $s_1 \notin D_1^0$ and therefore will not choose a strategy $s_2 \in D_2^1$, and if player 1 believes that player 2 believes that player 1 is rational, then player 1 believes that $s_2 \notin D_2^1$ and will, therefore, not choose a strategy $s_1 \in D_1^2$, and so on. Beliefs about beliefs about beliefs ... are explicitly modeled through the common belief operator. In contrast, Propositions 21 and 22 do not make use of the common belief operator. However, the logic is essentially the same. In particular, common belief of rationality is generated by the axiom $\mathbf{WR}'$ (or $\mathbf{SR}'$) and the rule of necessitation: from $s_1^k \to \neg B_1(s_1^\ell \succ_1 s_1^k)$ we get, by Necessitation, that $B_1\left(s_1^k \to \neg B_1(s_1^\ell \succ_1 s_1^k)\right) \wedge B_2\left(s_1^k \to \neg B_1(s_1^\ell \succ_1 s_1^k)\right)$ and thus, whatever is implied by $\mathbf{WR}'$ is believed by both players. Further iterations of the Necessitation rule yields beliefs about beliefs about beliefs ... about the rationality of every player.

7 Conclusion

We have examined the implications of common belief and common knowledge of two, rather weak, notions of rationality. Most of the literature on the epistemic foundations of game theory have dealt with the Bayesian approach, which identifies rationality with expected payoff maximization, given probabilistic beliefs (for surveys of this literature see [1] and [9]). Our focus has been on strategic-form games with ordinal payoffs and non-probabilistic Kripke structures. While most of the literature has been developed within the semantic approach, we have used a syntactic framework and expressed rationality in terms of syntactic axioms. We showed that the first, weaker, axiom of rationality characterizes the iterated deletion of strictly dominated strategies, while the stronger axiom characterizes the pure-strategy version of the algorithm introduced by Stalnaker [14].

The two notions of rationality used in this paper can, of course, be used also in the subclass of games with von Neumann-Morgenstern payoffs and the results would be the same. Furthermore, the standard notion of Bayesian rationality as expected payoff maximization is stronger than (that is, implies) both notions of rationality considered in this paper. Thus our results apply also to Bayesian rationality.

[13] This fact was proved directly in the proof of Proposition 15.

We have provided two version of our characterization results. The first (Propositions 15 and 19), which comes closer to the previous game-theoretic literature, is based on an explicit account of the role of common belief of rationality and thus requires a syntax that contains atomic propositions that are interpreted as "player i is rational". The second characterization (Propositions 21 and 22) is closer to the modal logic literature, where axioms are characterized in terms of properties of frames. However, we argued that the two characterizations are essentially identical.

We have restricted attention to strategic-form games. In future work we intend to extend the analysis to extensive-form games with perfect information and the notion of backward induction, which also does not require a probabilistic framework.

References

[1] Battigalli, Pierpaolo and Giacomo Bonanno, Recent results on belief, knowledge and the epistemic foundations of game theory, *Research in Economics*, 1999, **53**, 149-225.

[2] Bernheim, Douglas, Rationalizable strategic behavior, *Econometrica,* 1984, **52**, 1002-1028.

[3] Blackburn, P., M. de Rijke and Y. Venema, *Modal logic,* Cambridge University Press, 2001.

[4] Bonanno, Giacomo, On the logic of common belief, *Mathematical Logic Quarterly*, 1996, **42**, 305-311.

[5] Bonanno, Giacomo and Klaus Nehring, On Stalnaker's notion of strong rationalizability and Nash equilibrium in perfect information games, *Theory and Decision,* 1998, **45**, 291-295.

[6] Bonanno, Giacomo and Klaus Nehring, Common belief with the logic of individual belief, *Mathematical Logic Quarterly*, 2000, **46**, 49-52.

[7] Brandenburger, Adam and Eddie Dekel, Rationalizability and correlated equilibria, *Econometrica,* 1987, **55,** 1391-1402.

[8] Chellas, Brian, *Modal logic: an introduction,* Cambridge University Press, 1984.

[9] Dekel, Eddie and Faruk Gul, Rationality and knowledge in game theory, in: D. Kreps and K. Wallis (Eds.), *Advances in Economics and Econometrics*, Cambridge University Press, 1997, 87-172.

[10] Kripke, S., A semantical analysis of modal logic I: normal propositional calculi, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 1963, 9: 67-96.

[11] Lismont, Luc, La connaissance commune en logique modale, *Mathematical Logic Quarterly*, 1993, **39**, 115-130.

[12] Lismont, Luc and Philippe Mongin, On the logic of common belief and common knowledge, *Theory and Decision*, 1994, **37**, 75-106.

[13] Pearce, David, Rationalizable strategic behavior and the problem of perfection, *Econometrica,* 1984, **52**, 1029-1050.

[14] Stalnaker, Robert, On the evaluation of solution concepts, *Theory and Decision*, 1994, **37**, 49-74.

[15] Tan, T. and S. Werlang, The Bayesian foundation of solution concepts of games, *Journal of Economic Theory,* 1988, **45**, 370-391.