A Syntactic Approach to Rationality in Games with Ordinal Payoffs

Giacomo Bonanno

Department of Economics
University of California
Davis CA 95616-8578, United States of America
gfbonanno@ucdavis.edu

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). Frame characterization results are also provided.

1 Introduction

The notion of rationalizability in games was introduced independently by Bernheim [2] and Pearce [16]. A strategy of player i is said to be rational if it maximizes player i's expected payoff, given her probabilistic 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

Giacomo Bonanno, Wiebe van der Hoek, Michael Wooldridge (eds.). *Logic and the Foundations of Game and Decision Theory (LOFT 7)*. Texts in Logic and Games 3, Amsterdam University Press 2008, pp. 59–86.

dominated strategies.[1] They captured 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 [18] using a universal type space, rather than Kripke structures (Kripke [13]). A characterization of common belief of rationality using probabilistic Kripke structures was first provided by Stalnaker [17], although it was implicit in Brandenburger and Dekel [8]. 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 [16] and Brandenburger and Dekel [8]) 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 [17], 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

[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 [8]).

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 common knowledge of rationality in the stronger sense. The characterization results proved in Section 5 (Propositions 5.4 and 5.8) 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 characterization. In Section 7 we discuss related literature, while Section 8 contains a summary and concluding remarks.

2 Multi-agent logics of belief and knowledge

We consider a multi-modal logic with $n+1$ operators $B_1, B_2, \ldots, B_n, B_*$ where, for $i = 1, \ldots, n$, the intended interpretation of $B_i\varphi$ is "player i believes that φ", while $B_*\varphi$ is interpreted as "it is common belief that φ". The formal language is built in the usual way (see [3] and [10]) 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 $\Box \in \{B_1, \ldots, B_n, B_*\}$,
$$\Box\varphi \wedge \Box(\varphi \rightarrow \psi) \rightarrow \Box\psi. \qquad (\mathbf{K})$$
3. Axioms **D**, **4** and **5** for individual beliefs: for $i = 1, \ldots, n$,
$$B_i\varphi \rightarrow \neg B_i \neg\varphi, \qquad (\mathbf{D}_i)$$
$$B_i\varphi \rightarrow B_i B_i\varphi, \qquad (\mathbf{4}_i)$$
$$\neg B_i\varphi \rightarrow B_i \neg B_i\varphi. \qquad (\mathbf{5}_i)$$
4. Axioms for common belief: for $i = 1, \ldots, n$,
$$B_*\varphi \rightarrow B_i\varphi, \qquad (\mathbf{CB1})$$
$$B_*\varphi \rightarrow B_i B_*\varphi, \qquad (\mathbf{CB2})$$
$$B_*(\varphi \rightarrow B_1\varphi \wedge \cdots \wedge B_n\varphi) \rightarrow (B_1\varphi \wedge \cdots \wedge B_n\varphi \rightarrow B_*\varphi). \qquad (\mathbf{CB3})$$

Rules of Inference:

1. Modus Ponens:
$$\text{From } \varphi \text{ and } (\varphi \to \psi) \text{ infer } \psi. \qquad (\mathbf{MP})$$

2. Necessitation for every modal operator: for $\Box \in \{B_1, \ldots, B_n, B_*\}$,
$$\text{From } \varphi \text{ infer } \Box\varphi. \qquad (\mathbf{Nec})$$

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 $i = 1, \ldots, n$,
$$B_i\varphi \to \varphi. \qquad (\mathbf{T}_i)$$

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, $\neg B_*\varphi \to B_*\neg B_*\varphi$, 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: $B_iB_*\varphi \to B_*\varphi$. Indeed, the logic $\mathbf{KD45}^*_n$ augmented with the axiom $B_iB_*\varphi \to B_*\varphi$ *coincides* with the logic $\mathbf{KD45}^*_n$ augmented with the axiom $\neg B_*\varphi \to B_*\neg B_*\varphi$ (see [6]).

On the semantic side we consider Kripke structures $\langle \Omega, \mathcal{B}_1, \ldots, \mathcal{B}_n, \mathcal{B}_* \rangle$ where Ω is a set of states or possible worlds and, for every $j \in \{1, \ldots, n, *\}$, $\mathcal{B}_j$ is a binary relation on Ω.[2] For every $\omega \in \Omega$ and for every $j \in \{1, \ldots, n, *\}$, let $\mathcal{B}_j(\omega) = \{\omega' \in \Omega : \omega \mathcal{B}_j \omega'\}$.

Definition 2.1. A $\mathrm{D45}^*_n$ *frame* is a Kripke structure $\langle \Omega, \mathcal{B}_1, \ldots, \mathcal{B}_n, \mathcal{B}_* \rangle$ that satisfies the following properties: for all $\omega, \omega' \in \Omega$ and $i = 1, \ldots, n$

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 $\mathcal{B}_i(\omega) \subseteq \mathcal{B}_i(\omega')$;

[2] Throughout the paper we shall use the Roman font for syntactic operators (e.g., B_i and K_i) and the Calligraphic font for the corresponding semantic relations (e.g., $\mathcal{B}_i$ and $\mathcal{K}_i$).

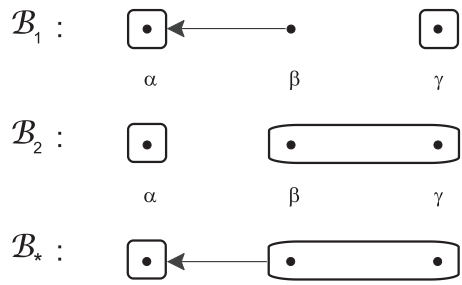

FIGURE 1. Illustration of a $D45_n^*$ frame.

4. $\mathcal{B}_*$ is the transitive closure of $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_n$, that is, $\omega' \in \mathcal{B}_*(\omega)$ if and only if there is a sequence $\langle \omega_1, \ldots, \omega_m \rangle$ in Ω such that (1) $\omega_1 = \omega$, (2) $\omega_m = \omega'$ and (3) for every $k = 1, \ldots, m-1$ there is an $i_k \in \{1, \ldots, n\}$ 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 $i = 1, \ldots, n$,

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 $j \in \{1, \ldots, n, *\}$, $\omega' \in \mathcal{B}_j(\omega)$ 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'.

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 $\omega \models \varphi$ the fact that φ is true at state ω. The truth set of φ is denoted by $\|\varphi\|$, that is, $\|\varphi\| = \{\omega \in \Omega : \omega \models \varphi\}$. Truth of a formula at a state is defined recursively as follows:

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

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

The following result is well-known:[3]

Proposition 2.2. The logic $\mathbf{KD45}^*_n$ is sound and complete with respect to the class of $\mathrm{D45}^*_n$ frames, that is, a formula is a theorem of $\mathbf{KD45}^*_n$ if and only if it is valid in every $\mathrm{D45}^*_n$ frame. Similarly, $\mathbf{S5}^*_n$ is sound and complete with respect to the class of $\mathrm{S5}^*_n$ frames.

3 Ordinal 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 3.1. A *finite strategic-form game with ordinal payoffs* is a quintuple $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$, where

- $N = \{1, \ldots, n\}$ 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,
- $\succeq_i$ is player i's ordering of O,[4]
- $z : S \to O$ (where $S = S_1 \times \cdots \times S_n$) is a function that associates with every strategy profile $s = (s_1, \ldots, s_n)$ 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, $S_{-i} = S_1 \times \cdots \times S_{i-1} \times S_{i+1} \times \cdots \times S_n$. 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 $s_{-i} \in S_{-i}$.

[3] See [4]. The same result has been provided for somewhat different axiomatizations of common belief by a number of authors (for example [14], [15] and [12]).

[4] That is, $\succeq_i$ is a binary relation on O that satisfies the following properties: for all $o, o', o'' \in O$, (1) either $o \succeq_i o'$ or $o' \succeq_i o$ (completeness or connectedness) and (2) if $o \succeq_i o'$ and $o' \succeq_i o''$ then $o \succeq_i o''$ (transitivity). The interpretation of $o \succeq_i o'$ 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 $o \succeq_i o'$ and not $o' \succeq_i o$. The interpretation of $o \succ_i o'$ is that player i strictly prefers outcome o to outcome o'.

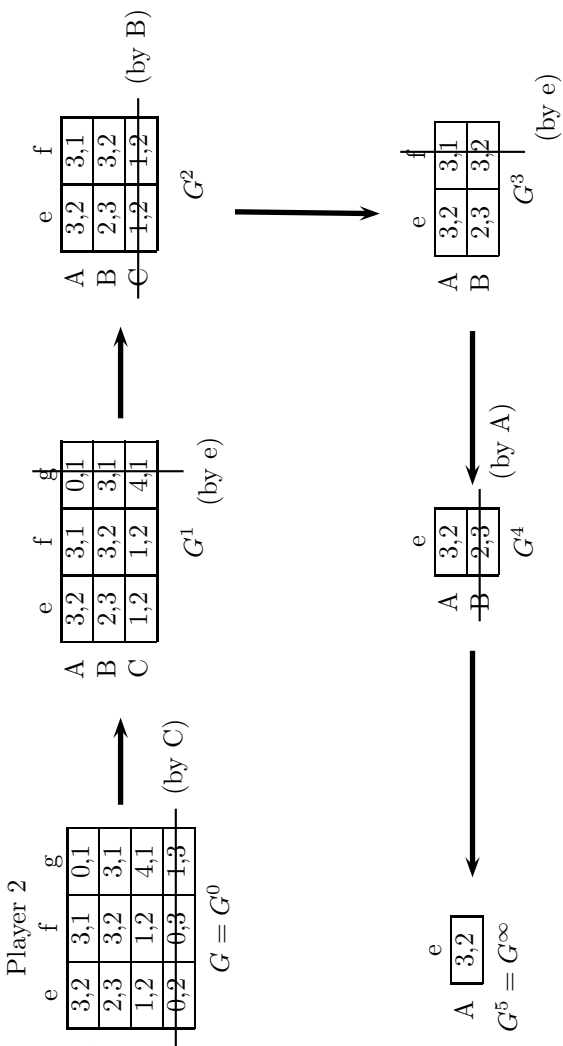

FIGURE 2. Illustration of the iterated deletion of strictly dominated strategies.

Definition 3.2. Given a game $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$ and $s_i \in S_i$, we say that, for player i, s_i is *strictly dominated in* G if there is another strategy $t_i \in S_i$ of player i 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 $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$ and G' be two games, where $G' = \langle N', \{S'_i\}_{i\in N'}, O', \{\succeq'_i\}_{i\in N'}, z'\rangle$. We say that G' is a *subgame* of G if $N' = N$, $O' = O$, for every $i \in N$: $\succeq'_i = \succeq_i$ and $S'_i \subseteq S_i$ (so that $S' \subseteq S$) and z' coincides with the restriction of z to S' (that is, for every $s' \in S'$, $z'(s') = z(s')$).

Definition 3.3 (IDSDS procedure)**.** The Iterated Deletion of Strictly Dominated Strategies is the following procedure. Given a game $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$ let $\langle G^0, G^1, \ldots, G^m, \ldots\rangle$ be the sequence of subgames of G defined recursively as follows. For all $i \in N$,

1. let $S_i^0 = S_i$ and let $D_i^0 \subseteq S_i^0$ be the set of strategies of player i that are strictly dominated in $G^0 = G$;
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 $S_i^\infty = \bigcap_{m\in\mathbb{N}} S_i^m$ (where $\mathbb{N}$ denotes the set of non-negative integers) and let G^∞ be the subgame of G with strategy sets S_i^∞. Let $S^\infty = S_1^\infty \times \cdots \times S_n^\infty$.[5]

The IDSDS procedure is illustrated in Figure 2, where:

$$
\begin{array}{llll}
S_1^0 = \{A, B, C, D\} & D_1^0 = \{D\} & S_2^0 = \{e, f, g\} & D_2^0 = \varnothing \\
S_1^1 = \{A, B, C\} & D_1^1 = \varnothing, & S_2^1 = \{e, f, g\} & D_2^1 = \{g\} \\
S_1^2 = \{A, B, C\} & D_1^2 = \{C\} & S_2^2 = \{e, f\} & D_2^2 = \varnothing \\
S_1^3 = \{A, B\} & D_1^3 = \varnothing & S_2^3 = \{e, f\} & D_2^3 = \{f\} \\
S_1^4 = \{A, B\} & D_1^4 = \{B\} & S_2^\infty = S_2^4 = \{e\} & D_2^4 = \varnothing \\
S_1^\infty = S_1^5 = \{A\} & & &
\end{array}
$$

Thus $S^\infty = \{(A, e)\}$.

In Figure 2 we have represented the ranking $\succeq_i$ 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

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

only if $z(s) \succeq_i z(s')$ (in each cell, the first number is the payoff of player 1 while the second number is the payoff of player 2).[6]

The next iterated deletion procedure differs from IDSDS in that at every round we delete strategy *profiles* rather than individual strategies. This procedure is the restriction to pure strategies of the algorithm introduced by Stalnaker [17].

Definition 3.4 (IDIP procedure)**.** Let $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$, be a game, together with 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 $T^\infty = \bigcap_{m\in\mathbb{N}} T^m$.[7]

The IDIP procedure is illustrated in Figure 3, where

$$S = T^0 = \{(A,d),(A,e),(A,f),(B,d),(B,e),(B,f),(C,d),(C,e),(C,f)\},$$
$$I^0 = \{(B,e),(C,f)\}$$

(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);

$$T^2 = \{(A,d),(A,e),(A,f),(C,d)\},$$
$$I^2 = \{(C,d)\}$$

[6] Note that the payoff function $u_i : S \to \mathbb{R}$ used in Figure 2 to represent the ranking $\succeq_i$ of player i is an *ordinal* function in the sense that it could be replaced by any other function v_i obtained by composing u_i with a strictly increasing function on the reals. That is, if $f_i : \mathbb{R} \to \mathbb{R}$ is such that $f_i(x) > f_i(y)$ whenever $x > y$, then $v_i : S \to \mathbb{R}$ defined by $v_i(s) = f_i(u_i(s))$ could be used as an alternative representation of $\succeq_i$ and the outcome of the IDSDS algorithm would be the same.

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

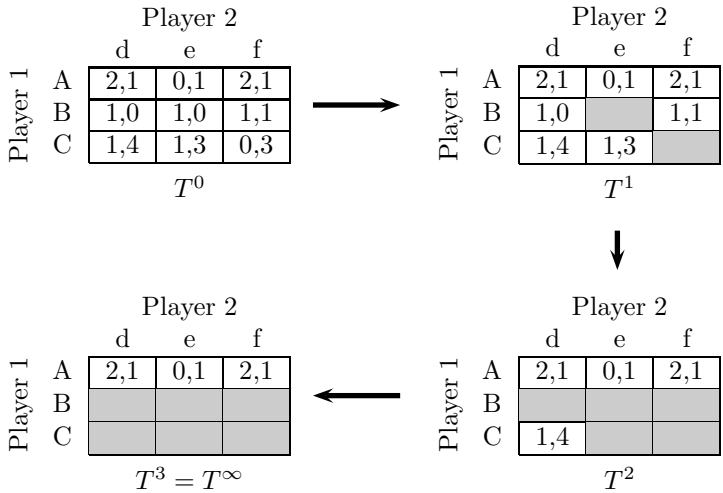

FIGURE 3. Illustration of the iterated deletion of inferior strategy profiles.

(the elimination of (C, d) is done through player 1 and strategy A);

$$T^3 = \{(A, d), (A, e), (A, f)\},$$
$$I^3 = \varnothing,$$

and thus $T^\infty = T^3$.

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 $s_i, t_i, \ldots$ 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 4.1. Fix a game $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$ with $S_i = \{s_i^1, s_i^2, \ldots, s_i^{m_i}\}$ (thus the cardinality of S_i is m_i). A game logic is called a *G-logic* if its set of strategy symbols is $\{s_i^k\}_{i=1,\ldots,n;k=1,\ldots,m_i}$ (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 $S_i = \{s_i^1, s_i^2, \ldots, s_i^{m_i}\}$, we denote by $\mathbf{L}_G^{\mathrm{D45}}$ (respectively, $\mathbf{L}_G^{\mathrm{S5}}$) the $\mathbf{KD45}_n^*$ (respectively, $\mathbf{S5}_n^*$) G-logic that satisfies the following additional axioms: for all $i = 1, \ldots, n$ and for all $k, \ell = 1, \ldots, m_i$, with $k \neq \ell$,

$$(s_i^1 \vee s_i^2 \vee \cdots \vee s_i^{m_i}), \tag{\textbf{G1}}$$

$$\neg(s_i^k \wedge s_i^\ell), \tag{\textbf{G2}}$$

$$s_i^k \to B_i s_i^k, \tag{\textbf{G3}}$$

$$(s_i^k \succeq_i s_i^\ell) \vee (s_i^\ell \succeq_i s_i^k), \tag{\textbf{G4}}$$

$$(s_i^\ell \succ_i s_i^k) \leftrightarrow ((s_i^\ell \succeq_i s_i^k) \wedge \neg(s_i^k \succeq_i s_i^\ell)). \tag{\textbf{G5}}$$

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

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

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

1.	$\varphi \to s_i^k$	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

[8] Note that, in general, logic $\mathbf{L}_G^{\mathrm{D45}}$ allows for incorrect beliefs. In particular, a player might have incorrect beliefs about the choices made by *other* players. By Proposition 4.2, however, a player cannot have mistaken beliefs about her own choice.

4.	$B_i s_i^k \to B_i \neg s_i^\ell$	3, rule RK[9]
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.	$s_i^1 \vee \cdots \vee s_i^{m_i}$	axiom **G1**
10.	$B_i s_i^k \to (s_i^1 \vee \cdots \vee s_i^{m_i})$	9, PL
11.	$B_i s_i^k \to \varphi$	8 (for every $\ell \neq k$), 10, PL
12.	$B_i s_i^k \to s_i^k$	1, 11, PL

Q.E.D.

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

Definition 4.3. Given a game $G = \langle N, \{S_i\}_{i\in N}, O, \{\succeq_i\}_{i\in N}, z\rangle$ 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 $\sigma(\omega) = (\sigma_1(\omega), \ldots, \sigma_n(\omega)) \in S$. 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 2.1).

Definition 4.4. Given a game G with $S_i = \{s_i^1, s_i^2, \ldots, s_i^{m_i}\}$, 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:

- $\omega \models s_i^h$ if and only if $\sigma_i(\omega) = s_i^h$,
- $\omega \models (s_i^k \succeq_i s_i^\ell)$ if and only if $z(s_i^k, \sigma_{-i}(\omega)) \succeq_i z(s_i^\ell, \sigma_{-i}(\omega))$.

Thus at state ω in a G-model it is true that player i chooses strategy s_i^h if and only if the strategy of player i associated with ω is s_i^h $(\sigma_i(\omega) = s_i^h)$ and it is true that strategy s_i^k is at least as good as strategy s_i^ℓ if and only if s_i^k in combination with $\sigma_{-i}(\omega)$ (the profile of strategies of players other than i associated with ω) yields an outcome which player i considers at least as good as the outcome yielded by s_i^ℓ in combination with $\sigma_{-i}(\omega)$.

[9] 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.

	d	e	f
a	2,1	0,1	2,1
b	1,0	1,0	1,1
c	1,4	1,3	0,3

FIGURE 4. A game: player 1 controls the rows (i.e., has strategies a, b and c), and player 2 the columns.

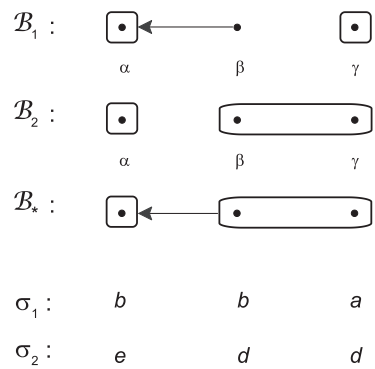

FIGURE 5. $\mathrm{D45}^*_n$ frame for the game of Figure 4.

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

Figure 4 illustrates a two-player game with strategy sets $S_1 = \{a, b, c\}$ and $S_2 = \{d, e, f\}$ and Figure 5 a $\mathrm{D45}^*_n$ frame for it. The corresponding model is given by the following valuation:

$$\begin{aligned}
\alpha \models\ & b \wedge e \wedge (b \succ_1 a) \wedge (c \succ_1 a) \wedge (b \succeq_1 c) \wedge (c \succeq_1 b) \\
& \wedge (f \succ_2 d) \wedge (f \succ_2 e) \wedge (e \succeq_2 d) \wedge (d \succeq_2 e), \\
\beta \models\ & b \wedge d \wedge (a \succ_1 b) \wedge (a \succ_1 c) \wedge (b \succeq_1 c) \wedge (c \succeq_1 b) \\
& \wedge (f \succ_2 d) \wedge (f \succ_2 e) \wedge (e \succeq_2 d) \wedge (d \succeq_2 e), \\
\gamma \models\ & a \wedge d \wedge (a \succ_1 b) \wedge (a \succ_1 c) \wedge (b \succeq_1 c) \wedge (c \succeq_1 b) \wedge (d \succeq_2 e) \\
& \wedge (e \succeq_2 d) \wedge (d \succeq_2 f) \wedge (f \succeq_2 d) \wedge (e \succeq_2 f) \wedge (f \succeq_2 e).
\end{aligned}$$

Proposition 4.5. Logic $\mathbf{L}^{\mathrm{D45}}_G$(respectively, $\mathbf{L}^{\mathrm{S5}}_G$) is sound with respect to the class of $\mathbb{M}^{\mathrm{D45}}_G$ (respectively, $\mathbb{M}^{\mathrm{S5}}_G$) models.

Proof. It follows from Proposition 2.2 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 4.4), $\omega \models s_i^k$ if and only if $\sigma_i(\omega) = s_i^k$; (2) axiom **G3** is an immediate consequence of the fact (see Definition 4.3) 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. Q.E.D.

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 $\mathbf{L}_G^{\mathrm{D45}}$ 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"):

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

Given a game G, let $\mathbf{L}_G^{\mathrm{D45}}$+**WR** (respectively, $\mathbf{L}_G^{\mathrm{S5}}$+**WR**) be the extension of $\mathbf{L}_G^{\mathrm{D45}}$ (respectively, $\mathbf{L}_G^{\mathrm{S5}}$) 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:

$$s_i^k \wedge B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k) \;\rightarrow\; \neg r_i. \qquad (\mathbf{SR})$$

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

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

Proposition 5.1. **WR** is a theorem of $\mathbf{L}_G^{\mathrm{D45}}$+**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.	$(r_i \wedge s_i^k) \rightarrow \neg(B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k))$	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
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.	$B_i(s_i^\ell \succ_i s_i^k) \rightarrow (B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg (s_i^\ell \succ_i s_i^k))$	5, 6, PL

8. $\neg(B_i(s_i^\ell \succeq_i s_i^k) \wedge \neg B_i \neg(s_i^\ell \succ_i s_i^k)) \rightarrow \neg B_i(s_i^\ell \succ_i s_i^k)$ 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

Q.E.D.

Definition 5.2. Given a game G, let $\mathbb{M}_G^{\mathrm{D45|WR}} \subseteq \mathbb{M}_G^{\mathrm{D45}}$ ($\mathbb{M}_G^{\mathrm{S5|WR}} \subseteq \mathbb{M}_G^{\mathrm{S5}}$) be the class of $\mathrm{D45}_n^*$ (respectively, $\mathrm{S5}_n^*$) G-models (see Definition 4.4) 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 $z(\sigma_i(\omega), \sigma_{-i}(\omega')) \succeq_i z(s_i, \sigma_{-i}(\omega'))$.[10]

Thus at state ω player i is rational if and only if, for every strategy s_i of hers, there is a state ω' that she considers possible at ω ($\omega' \in \mathcal{B}_i(\omega)$) where the strategy that she actually uses at ω ($\sigma_i(\omega)$) is at least as good as s_i against the strategies used by the other players at ω' ($\sigma_{-i}(\omega')$). For instance, in the model based on the frame of Figure 5 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 $z(a,d) \succeq_2 z(a,f)$, while for strategy e we have that $\beta \in \mathcal{B}_2(\beta)$, $\sigma_1(\beta) = b$ and $z(b,d) \succeq_2 z(b,e)$. Thus, in the model based on the frame of Figure 5, 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 5.3. Logic $\mathbf{L}_G^{\mathrm{D45}}$+**WR** (respectively, $\mathbf{L}_G^{\mathrm{S5}}$+**WR**) is sound with respect to the class of models $\mathbb{M}_G^{\mathrm{D45|WR}}$ (respectively, $\mathbb{M}_G^{\mathrm{S5|WR}}$).

Proof. By Proposition 4.5 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 $\mathcal{B}_i(\omega) \subseteq \|s_i^\ell \succ_i s_i^k\|$, that is (see Definition 4.4), $z(s_i^\ell, \sigma_{-i}(\omega')) \succ_i z(s_i^k, \sigma_{-i}(\omega'))$, for every $\omega' \in \mathcal{B}_i(\omega)$. It follows from Definition 5.2 that $\omega \models \neg r_i$. Q.E.D.

The following proposition says that common belief of the weak notion of rationality expressed by axiom **WR** characterizes the Iterated Deletion of Strictly Dominated Strategies (see Definition 3.3).[11]

[10] This could alternatively be written as $z(\sigma_i(\omega'), \sigma_{-i}(\omega')) \succeq_i z(s_i, \sigma_{-i}(\omega'))$, since, by definition of G-frame (see Definition 4.3), if $\omega' \in \mathcal{B}_i(\omega)$ then $\sigma_i(\omega') = \sigma_i(\omega)$.

[11] Proposition 5.4 is the syntactic-based, ordinal version of a semantic, probabilistic-based result of Stalnaker [17]. As noted in the Introduction, Stalnaker's result was, in turn, a reformulation of earlier results due to Bernheim [2], Pearce [16], Tan and Werlang [18] and Brandenburger and Dekel [8].

The characterization results given in Propositions 5.4 and 5.8 are not characteriza-

Proposition 5.4. Fix a finite strategic-form game with ordinal payoffs G. Then both (A) and (B) below hold.

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

(B) For every $s \in S^\infty$ there exists a model in $\mathbb{M}_G^{\mathrm{S5|WR}}$ and a state ω such that (1) $\sigma(\omega) = s$ and (2) $\omega \models K_*(r_1 \wedge \cdots \wedge r_n)$.[12]

Proof. (A) Fix a model in $\mathbb{M}_G^{\mathrm{D45|WR}}$ and a state α and suppose that $\alpha \models B_*(r_1 \wedge \cdots \wedge r_n)$. The proof is by induction. First we show that, for every player $i = 1, \ldots, n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_i(\omega) \notin D_i^0$ (see Definition 3.3). 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 $s_{-i} \in S_{-i}$, $z(\hat{s}_i, s_{-i}) \succ_i z(\sigma_i(\beta), s_{-i})$. Then, 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 5.2 that $\beta \models \neg r_i$, 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 $j = 1, \ldots, n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_j(\omega) \in S_j^m$. We want to show that, for every player $i = 1, \ldots, n$ 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 2.1)—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 5.2, $\beta \models \neg r_i$, contradicting the fact that $\beta \in \mathcal{B}_*(\alpha)$ and $\alpha \models B_* r_i$. Thus, for every player $i = 1, \ldots, n$ and for every $\omega \in \mathcal{B}_*(\alpha)$, $\sigma_i(\omega) \in \bigcap_{m \in \mathbb{N}} S_i^m = S_i^\infty$. 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 4.3, $\sigma_i(\beta) = \sigma_i(\alpha)$. Thus $\sigma_i(\alpha) \in S_i^\infty$.

(B) Let m be the cardinality of $S^\infty = S_1^\infty \times \cdots \times S_n^\infty$ and let $\Omega = \{\omega_1, \ldots, \omega_m\}$. Let $\sigma : \Omega \to S^\infty$ be a one-to-one function. For every player i,

tions 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 Section 6 we provide a reformulation of Propositions 5.4 and 5.8 in terms of frame characterization.

[12] 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.

Thus while part (A) says that if at a state there is common *belief* of rationality then the strategy profile played at that state belongs to the set of strategy profiles that are obtained by applying the IDSDS algorithm, part (B) says that any such strategy profile is realized at a state of some model where there is common *knowledge* of rationality (that is, common belief with the added property that individual beliefs satisfy the Truth Axiom $\mathbf{T}_i$).

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 ith coordinate of $\sigma(\omega)$. Let $\mathcal{K}_*$ be the transitive closure of $\bigcup_{i \in N} \mathcal{K}_i$ (then, for every $\omega \in \Omega$, $\mathcal{K}_*(\omega) = \Omega$). The structure so defined is clearly an $\mathrm{S5}^*_n$ G-frame. Consider the model corresponding to this frame (see Definition 4.4). 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 5.2). Hence, for every $\omega \in \Omega$, $\omega \models (r_1 \wedge \cdots \wedge r_n)$ and, therefore, for every $\omega \in \Omega$, $\omega \models K_*(r_1 \wedge \cdots \wedge r_n)$. Q.E.D.

Remark 5.5. Since $\mathbb{M}_G^{\mathrm{S5|WR}} \subseteq \mathbb{M}_G^{\mathrm{D45|WR}}$ it follows from part (B) of Proposition 5.4 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 5.6. Given a game G, let $\mathbb{M}_G^{\mathrm{D45|SR}} \subseteq \mathbb{M}_G^{\mathrm{D45}}$ ($\mathbb{M}_G^{\mathrm{S5|SR}} \subseteq \mathbb{M}_G^{\mathrm{S5}}$, respectively) 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 hers which is better than $\sigma_i(\omega)$ (the strategy she is actually using at ω) at some state ω' that she considers possible at ω, then $\sigma_i(\omega)$ is better than s_i at some other state ω'' that she considers possible at ω. For example, in the model based on the frame of Figure 5 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 hers, namely f, which is better than d at β and as good as d at γ, and $\mathcal{B}_2(\beta) = \{\beta, \gamma\}$. Thus she is irrational according to Definition 5.6.

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

Proposition 5.7. Logic $\mathbf{L}_G^{\mathrm{D45}}$+**SR** (respectively, $\mathbf{L}_G^{\mathrm{S5}}$+**SR**) is sound with respect to the class of models $\mathbb{M}_G^{\mathrm{D45|SR}}$ (respectively, $\mathbb{M}_G^{\mathrm{S5|SR}}$).

Proof. By Proposition 4.5 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 $\mathcal{B}_i(\omega) \subseteq \|s_i^\ell \succeq_i s_i^k\|$ [that is—see Definition 4.4—$z(s_i^\ell, \sigma_{-i}(\omega')) \succeq_i z(s_i^k, \sigma_{-i}(\omega'))$, 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 5.6 that $\omega \models \neg r_i$. Q.E.D.

The following proposition says that common *knowledge* of the stronger notion of rationality expressed by axiom **SR** characterizes the Iterated Deletion of Inferior Profiles (see Definition 3.4).[13]

Proposition 5.8. Fix a finite strategic-form game with ordinal payoffs G. Then both (A) and (B) below hold.

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

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

Proof. (A) As in the case of Proposition 5.4, the proof is by induction. Fix a model in $\mathbb{M}_G^{\text{S5}|\text{SR}}$ and a state α and suppose that $\alpha \models K_*(r_1 \wedge \cdots \wedge r_n)$. First we show that, for every $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \notin I^0$ (see Definition 3.4). 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 2.1), $\beta \in \mathcal{K}_i(\beta)$. It follows from Definition 5.6 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 $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \notin I^m$. 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 $z(\tilde{s}_i, s_{-i}) \succeq_i z(\sigma_i(\beta), s_{-i})$. By Definition 4.3, 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 5.6 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 $\omega \in \mathcal{K}_*(\alpha)$, $\sigma(\omega) \in \bigcap_{m \in \mathbb{N}} T^m = T^\infty$. It only remains to establish that $\sigma(\alpha) \in T^\infty$, but this follows from reflexivity of $\mathcal{K}_*$.

[13] Proposition 5.8 is the syntactic-based, ordinal version of a semantic, probabilistic-based result due to Stalnaker [17]. For a correction of that result see Bonanno and Nehring [5].

	c	d
a	1,1	1,0
b	1,1	0,1

FIGURE 6. A game where player 1 has strategies a and b, and player 2 has c and d.

(B) Let m be the cardinality of T^∞ and let $\Omega = \{\omega_1, \ldots, \omega_m\}$. 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 ith coordinate of $\sigma(\omega)$. Let $\mathcal{K}_*$ be the transitive closure of $\bigcup_{i \in N} \mathcal{K}_i$ (then, for every $\omega \in \Omega$, $\mathcal{K}_*(\omega) = \Omega$). The structure so defined is clearly an $\mathrm{S5}^*_n$ G-frame. Consider the model corresponding to this frame (see Definition 4.4). 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 if $z(s_i, \sigma_{-i}(\omega')) \succ_i z(\sigma_i(\omega), \sigma_{-i}(\omega'))$ 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 5.6). Hence, for every $\omega \in \Omega$, $\omega \models (r_1 \wedge \cdots \wedge r_n)$ and, therefore, for every $\omega \in \Omega$, $\omega \models K_*(r_1 \wedge \cdots \wedge r_n)$. Q.E.D.

Note that Proposition 5.8 is not true if one replaces knowledge with belief, as illustrated in the game of Figure 6 and corresponding frame in Figure 7. In the corresponding model we have that, according to the stronger notion of rationality expressed by Definition 5.6, $\alpha \models r_1 \wedge r_2$ 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).[14] 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 5.5), with axiom **SR** there is a crucial difference between the implications of common *belief* of rationality and those of common *knowledge* of rationality.

6 Frame characterization

The characterization results proved in the previous section (Propositions 5.4 and 5.8) 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 6.1. An axiom characterizes (or is characterized by) a class $\mathbb{F}$

[14] In the game of Figure 6 we have that $S^\infty = S = \{(a, c), (a, d), (b, c), (b, d)\}$ while $T^\infty = \{(a, c), (b, c)\}$.

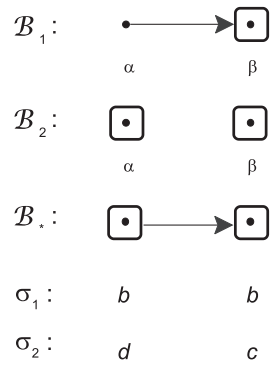

FIGURE 7. A frame for the game of Figure 6.

of Kripke frames if the axiom is valid in every model based on a frame that belongs to $\mathbb{F}$ and, conversely, 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.[15]

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 $\mathbb{M}_G^{\text{D45|WR}}$, $\mathbb{M}_G^{\text{S5|WR}}$, $\mathbb{M}_G^{\text{D45|SR}}$ and $\mathbb{M}_G^{\text{S5|SR}}$ (Definitions 5.2 and 5.6). Secondly we modify axioms **WR** and **SR** as follows:

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

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

One can derive axioms $\mathbf{WR}'$ and $\mathbf{SR}'$ from the logics considered previously by adding 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 5.4.

Proposition 6.2. Subject to the valuation rules specified in Definition 4.4 for the atomic propositions s_i^k and $(s_i^\ell \succeq_i s_i^k)$, axiom $\mathbf{WR}'$ is characterized by the class of D45_n^* game frames (see Definition 4.3) 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 $\alpha \models s_i^k$, that is,

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

$\sigma_i(\alpha) = s_i^k$. We want to show that $\alpha \models \neg B_i(s_i^\ell \succ_i s_i^k)$. Suppose not. Then $\mathcal{B}_i(\alpha) \subseteq \|s_i^\ell \succ_i s_i^k\|$, that is,

$$\text{for every } \omega \in \mathcal{B}_i(\alpha), \quad z(s_i^\ell, \sigma_{-i}(\omega)) \succ_i z(s_i^k, \sigma_{-i}(\omega)). \tag{6.1}$$

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 (6.1) 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, that is, there is a state $\omega \in \Omega$ and a player $i \in N$ such that $\sigma_i(\omega) \notin S_i^\infty$. 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 $\hat{m} = m(\alpha, i)$. Then

$$\sigma_i(\alpha) \in D_i^{\hat{m}} \tag{6.2}$$

and, since (see Definition 3.3), for every $j \in N$ and for every $p, q \in \mathbb{N} \cup \{\infty\}$, $S_j^{p+q} \subseteq S_j^p$,

$$\text{for every } j \in N \text{ and } \omega \in \Omega, \quad \sigma_j(\omega) \in S_j^{\hat{m}}. \tag{6.3}$$

Let $s_i^k = \sigma_i(\alpha)$. By (6.2) and (6.3) , there exists a $s_i^\ell \in S_i$ such that, for every $\omega \in \Omega$, $z(s_i^\ell, \sigma_{-i}(\omega)) \succ_i z(s_i^k, \sigma_{-i}(\omega))$. Thus $\mathcal{B}_i(\alpha) \subseteq \|s_i^\ell \succ_i s_i^k\|$ and thus $\alpha \models s_i^k \wedge B_i(s_i^\ell \succ_i s_i^k)$, so that axiom $\mathbf{WR}'$ is falsified at α. Q.E.D.

The next proposition is the counterpart of Proposition 5.8.

Proposition 6.3. Subject to the valuation rules specified in Definition 4.4 for the atomic propositions s_i^k and $(s_i^\ell \succeq_i s_i^k)$, axiom $\mathbf{SR}'$ is characterized by the class of S5_n^* game frames (see Definition 4.3) 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 $\alpha \models s_i^k \wedge K_i(s_i^\ell \succeq_i s_i^k)$. Then $\sigma_i(\alpha) = s_i^k$ and $\mathcal{K}_i(\alpha) \subseteq \|s_i^\ell \succeq_i s_i^k\|$, that is,

$$\text{for all } \omega \in \mathcal{K}_i(\alpha), \quad z(s_i^\ell, \sigma_{-i}(\omega)) \succeq_i z(s_i^k, \sigma_{-i}(\omega)). \tag{6.4}$$

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

$$z(s_i^\ell, \sigma_{-i}(\beta)) \succ_i z(s_i^k, \sigma_{-i}(\beta)). \tag{6.5}$$

It follows from (6.4) and (6.5) that $(s_i^k, \sigma_{-i}(\beta)) = (\sigma_i(\beta), \sigma_{-i}(\beta))$ is inferior relative to the set $\{s \in S : s = \sigma(\omega) \text{ for some } \omega \in \mathcal{K}_i(\alpha)\}$, contradicting the hypothesis that $\sigma(\omega) \in T^\infty$ for all $\omega \in \Omega$.

Conversely, fix an $S5_n^*$ frame not in the class, that is, there is a state $\omega \in \Omega$ such that $\sigma(\omega) \notin T^\infty$. For every $\omega \in \Omega$, let

$$m(\omega) = \begin{cases} \infty & \text{if } \sigma(\omega) \in T^\infty, \\ m & \text{if } \sigma(\omega) \in I^m = T^m \setminus T^{m+1}. \end{cases}$$

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

$$z(s_i^\ell, \sigma_{-i}(\alpha)) \succ_i z(\sigma_i(\alpha), \sigma_{-i}(\alpha)) \tag{6.6}$$

and

$$\forall \omega \in \Omega, \text{ if } (\sigma_i(\alpha), \sigma_{-i}(\omega)) \in T^{m_0} \text{ then } z(s_i^\ell, \sigma_{-i}(\omega)) \succeq_i z(\sigma_i(\alpha), \sigma_{-i}(\omega)). \tag{6.7}$$

By definition of m_0, since (see Definition 3.4) 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 (6.7) that $\mathcal{K}_i(\alpha) \subseteq \|s_i^\ell \succeq_i s_i^k\|$, that is, $\alpha \models K_i(s_i^\ell \succeq_i s_i^k)$. Since the frame is an S5 frame, $\mathcal{K}_i$ is reflexive and, therefore, $\alpha \in \mathcal{K}_i(\alpha)$. It follows from this and (6.6) that $\alpha \models \neg K_i \neg(s_i^\ell \succ_i s_i^k)$. Thus $\alpha \models s_i^k \wedge K_i(s_i^\ell \succeq_i s_i^k) \wedge \neg K_i \neg(s_i^\ell \succ_i s_i^k)$, so that axiom **SR′** is falsified at α. Q.E.D.

There appears to be an important difference between the results of Section 5 and those of this section, namely that, while Propositions 5.4 and 5.8 give a *local* result, Propositions 6.2 and 6.3 provide a *global* one. For example, Proposition 5.4 says 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 6.2, says 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 5.4. Fix a model and a state α and suppose that $\alpha \models B_*(r_1 \wedge \cdots \wedge r_n)$. Since, for every formula φ, $B_*\varphi \to B_*B_*\varphi$ is a theorem of $\mathbf{KD45}_n^*$, it follows that $\alpha \models B_*B_*(r_1 \wedge \cdots \wedge r_n)$, that is, for every $\omega \in \mathcal{B}_*(\alpha)$, $\omega \models B_*(r_1 \wedge \cdots \wedge r_n)$. Thus, it follows from Proposition 5.4 that $\sigma(\omega) \in S^\infty$, for every $\omega \in \mathcal{B}_*(\alpha)$.[16] 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

[16] This fact was proved directly in the proof of Proposition 5.4.

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 difference between the results of Section 5 and those of this section lies in the fact that Propositions 5.4 and 5.8 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 6.2 and 6.3 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 **WR′** (or **SR′**) and the rule of necessitation: from $s_1^k \rightarrow \neg B_1(s_1^\ell \succ_1 s_1^k)$ we get, by Necessitation, that $B_1(s_1^k \rightarrow \neg B_1(s_1^\ell \succ_1 s_1^k)) \wedge B_2(s_1^k \rightarrow \neg B_1(s_1^\ell \succ_1 s_1^k))$ and thus, whatever is implied by **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 Related literature

As noted in the introduction, the iterated elimination of strictly dominated strategies as a solution concept for strategic-form games goes back to Bernheim [2] and Pearce [16] and has been further studied and characterized by a number of authors. From the point of view of this paper, the most important contribution in this area is due to Stalnaker [17], who put forward the novel proposal of characterizing solution concepts for games in terms of classes of models. Stalnaker carried out his analysis within the standard framework of von Neumann-Morgenstern payoffs and defined dominance in terms of mixed strategies. Furthermore his analysis was semantic rather than syntactic. Our approach differs from Stalnaker's in that we formulate rationality syntactically within an axiomatic system and provide characterization results in line with the notion of frame characterization in modal logic. Furthermore, we do this in a purely ordinal framework that does not require probabilistic beliefs and von Neumann-Morgenstern payoffs. However, our intellectual debt towards Stalnaker is clear. In particular, the IDIP algorithm (see Definition 3.4) is the adaptation to ordinal games of the algorithm he introduced in [17].

A syntactic epistemic analysis of iterated strict dominance was also proposed by de Bruin [9, p. 86]. However his approach is substantially different from ours. First of all, his analysis is explicitly carried out only for two-person games, while we allowed for any number of players. Secondly,

de Bruin assumes von Neumann Morgenstern payoffs and his definition of strict dominance involves domination by mixed strategies [9, p. 51], while we considered ordinal payoffs and defined dominance in terms of pure strategies only (see Definition 3.2). Thirdly, de Bruin restricts attention to knowledge (that is, in his axiom system he imposes the Truth Axiom $\mathbf{T}_i$ on individual beliefs: [9, p. 51]) and thus does not investigate the difference between the implications of common belief of rationality and those of common knowledge of rationality (hence in his analysis there is no counterpart to the difference highlighted in Propositions 5.4 and 5.8 of Section 5). More importantly, however, de Bruin introduces the notion of strict dominance *directly into the syntax* by using atomic propositions of the form $nsd_i(A_i, A_j)$ whose intended interpretation is "player i uses a pure strategy in A_i which is not strictly dominated by a mixed strategy over A_i given that player j plays a pure strategy in A_j". Furthermore, his definition of rationality *incorporates* the notion of strict dominance. De Bruin's definition of rationality [9, p. 86] consists of two parts: a basis step without knowledge, $r_i \rightarrow nsd_i(A_i, A_j)$, and an inductive step with knowledge: $(r_i \wedge K_i X_i \wedge K_i X_j) \rightarrow nsd_i(X_i, X_j)$. According to de Bruin the advantage of his two-part definition of rationality is that

> Drawing a line between a basis case without beliefs, and an inductive step with beliefs makes it possible to mimic every single round of elimination of the solution concept by a step in the hierarchy of common belief in rationality. This becomes highly explicit in the inductive character of the proof. [9, p. 100]

However, as pointed out at the end of the previous section, this mimicking of the elimination steps occurs also in the proof of Proposition 5.4 without the need for a two-part definition of rationality and, more importantly, without incorporating the notion of dominance in the syntax.

The disadvantage of de Bruin's approach is that one loses the distinction between syntax and semantics and the ability to link the two by means of frame characterization results. In our analysis, the notion of strict dominance is purely a semantic notion, which has no syntactic counterpart. On the other hand the definition of rationality is expressed syntactically and it is epistemically based, in that it evaluates a player's rationality by comparing her action with her beliefs about the desirability of alternative actions. The characterization results then establish a correspondence between the output of an algorithm (such as the iterated deletion of strictly dominated strategies) and common belief of an independently formulated notion of rationality.

Börgers [7] provides a characterization of pure-strategy dominance that differs from ours. Like us, Börgers assumes that only the ordinal rankings of the players are commonly known; however—unlike us—he also assumes that

	d	e
a	0,0	0,0
b	1,1	0,0
c	0,0	1,1

	d	e
a	0,0	0,0
b	1,1	0,0

FIGURE 8. Two games in which player 2 has strategies d and e, whereas player 1 has either three strategies (left) or two (right).

each player has a von Neumann-Morgenstern utility function on the set of outcomes, forms probabilistic beliefs about the opponents' strategy choices and chooses a pure strategy that maximizes her expected utility, given those beliefs. He thus asks the question: what pure-strategy profiles are consistent with common belief of rationality, where the latter is defined as expected utility maximization with respect to *some* von Neumann-Morgenstern utility function and *some* beliefs about the opponents' strategies? Börgers shows that a pure strategy is rational in this sense if and only if it is not dominated by another *pure* strategy. Thus there is no need to consider dominance by a mixed strategy. However, he shows that the relevant notion of dominance in this case is *not* strict dominance but the following stronger notion: a strategy $s_i \in S_i$ of player i is dominated if and only if, for every subset of strategy profiles $X_{-i} \subseteq S_{-i}$ of the players other than i, there exists a strategy $t_i \in S_i$ (which can vary with X_{-i}) that *weakly* dominates s_i relative to X_{-i}.[17] For example, in the game illustrated in Figure 8 (left), strategy a of player 1 is dominated (by b relative to $\{d, e\}$ and also relative to $\{d\}$ and by c relative to $\{e\}$). Thus in the corresponding model shown in Figure 9, at state α, while player 1 is rational according to our axiom **WR** (since no strategy is strictly dominated; indeed at state α there is common knowledge of rationality), she is not rational according to Börgers' definition.[18]

On the other hand, while stronger than the notion expressed by our axiom **WR**, Börgers' notion of rationality is *weaker* than our axiom **SR**, as can be seen in the game of Figure 8 (right). Here strategy a of player 1 is dominated by b relative to $\{d, e\}$ and also relative to $\{d\}$ but not relative to $\{e\}$. Thus a is a rational strategy according to Börgers' definition. On the other hand, in the model of Figure 9 (viewed now as a model for the game of Figure 8 (right)) player 1 is not rational at state α (where her choice is

[17] We say that t_i weakly dominates s_i relative to X_{-i} if (1) $z(t_i, x_{-i}) \succeq_i z(s_i, x_{-i})$, for all $x_{-i} \in X_{-i}$, and (2) there exists an $\hat{x}_{-i} \in X_{-i}$ such that $z(t_i, \hat{x}_{-i}) \succ_i z(s_i, \hat{x}_{-i})$.

[18] The reason for this is as follows: if player 1 assigns positive probability to both α and β, then she would get a higher expected utility by switching to strategy b. The same is true if she assigns probability 1 to α. On the other hand, if she assigns probability 1 to β then she can increase her utility by switching to c.

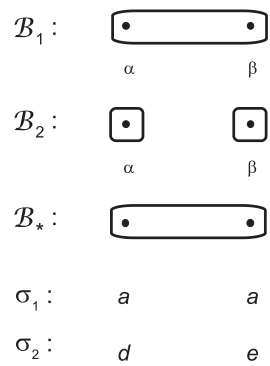

FIGURE 9. A model for the games of Figure 8.

a) according to the notion of rationality expressed by axiom **SR** (indeed, in this game, $T^\infty = \{(a,e),(b,d)\}$ and thus $(a,d) \notin T^\infty$).

8 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 [11]). Our focus has been on strategic-form games with ordinal payoffs and non-probabilistic beliefs. 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 [17].

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.[19]

We have provided two versions of our characterization results. The first (Propositions 5.4 and 5.8), which comes closer to the previous game-theoretic literature, is based on an explicit account of the role of common

[19] In the sense that whatever is incompatible with our notion of rationality is also incompatible with the stronger notion of Bayesian rationality.

belief of rationality and thus requires a syntax that contains atomic propositions that are interpreted as "player i is rational". The second characterization (Propositions 6.2 and 6.3) 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 this qualitative (that is, non probabilistic) analysis to extensive-form games with perfect information and the notion of backward induction.

References

[1] P. Battigalli & G. Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2):149–225, 1999.

[2] D. Bernheim. Rationalizable strategic behavior. *Econometrica*, 52(4):1002–1028, 1984.

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

[4] G. Bonanno. On the logic of common belief. *Mathematical Logic Quarterly*, 42(1):305–311, 1996.

[5] G. Bonanno & K. Nehring. On Stalnaker's notion of strong rationalizability and Nash equilibrium in perfect information games. *Theory and Decision*, 45(3):291–295, 1998.

[6] G. Bonanno & K. Nehring. Common belief with the logic of individual belief. *Mathematical Logic Quarterly*, 46(1):49–52, 2000.

[7] T. Börgers. Pure strategy dominance. *Econometrica*, 61(2):423–430, 1993.

[8] A. Brandenburger & E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55(6):1391–1402, 1987.

[9] B. de Bruin. *Explaining Games: On the Logic of Game Theoretic Explanations*. Ph.D. thesis, University of Amsterdam, 2004. *ILLC Publications* DS-2004-03.

[10] B. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1984.

[11] E. Dekel & F. Gul. Rationality and knowledge in game theory. In D. Kreps & K. Wallis, eds., *Advances in economics and econometrics*, pp. 87–172. Cambridge University Press, 1997.

[12] R. Fagin, J. Halpern, Y. Moses & M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

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

[14] L. Lismont. La connaissance commune en logique modale. *Mathematical Logic Quarterly*, 39(1):115–130, 1993.

[15] L. Lismont & P. Mongin. On the logic of common belief and common knowledge. *Theory and Decision*, 37(1):75–106, 1994.

[16] D. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 52(4):1029–1050, 1984.

[17] R. Stalnaker. On the evaluation of solution concepts. *Theory and Decision*, 37(1):49–74, 1994.

[18] T. Tan & S. Werlang. The Bayesian foundation of solution concepts of games. *Journal of Economic Theory*, 45(2):370–391, 1988.