Common knowledge of payoff uncertainty in games

de Bruin, B. P.

Published in:
Synthese

DOI:
10.1007/s11229-007-9275-5

Publication date:
2008

Synthese (2008) 163:79–97
Knowledge, Rationality & Action 79–97
DOI 10.1007/s11229-007-9275-5

Common knowledge of payoff uncertainty in games

Boudewijn de Bruin

Received: 16 February 2007 / Accepted: 26 November 2007 / Published online: 15 December 2007

Abstract Using epistemic logic, we provide a non-probabilistic way to formalise payoff uncertainty, that is, statements such as 'player i has approximate knowledge about the utility functions of player j.' We show that on the basis of this formalisation common knowledge of payoff uncertainty and rationality (in the sense of excluding weakly dominated strategies, due to Dekel and Fudenberg (1990)) characterises a new solution concept we have called 'mixed iterated strict weak dominance.'

Keywords Common knowledge · Epistemic characterisation theorem · Payoff uncertainty · Rationality

1 Introduction

Interactive epistemology (Aumann 1999; Bonanno 2002; Kaneko 2002) deals with the beliefs and the knowledge of players of games. It comes in two versions. The semantic approach represents knowledge by means of possible worlds structures, identifying the knowledge of a player i with the set of propositions true at all worlds which i cannot distinguish from the actual world. The syntactic approach represents knowledge by sentences that are provable in extensions of various epistemic logics.

The semantic approach to interactive epistemology has been most commonly adopted (Board 2004; Heifetz and Mongin 2001; Stalnaker 1996), but interesting syntactic investigations have found their way into the economic and game theoretic literature (Baltag 2002; van Benthem 2001; Bonanno 2003; Clausing 2004; Pauly 2002; Rabinowicz 1998). Most such applications are concerned with extensive games

B. de Bruin (✉)
Oude Boteringestraat 52, Groningen 9712 GL, The Netherlands
e-mail: b.p.de.bruin@rug.nl

or cooperative game theory. This paper, by contrast, applies syntactic methods to the study of the interactive epistemology of normal form games.

In this paper we develop a syntactic framework to shed light on payoff uncertainty, that is, on game playing situations in which the players do not have exact knowledge concerning the utility functions of their opponents, but only approximate knowledge. Dekel and Fudenberg (1990) were among the first to study payoff uncertainty, and their pioneering paper gave rise to a literature relating, stated informally, common knowledge of payoff uncertainty and rationality to the iterated elimination of strictly dominated strategies preceded by one round of elimination of weakly dominated strategies, the Dekel-Fudenberg procedure. A key assumption guiding this literature is that payoff uncertainty be cast in probabilistic terms.

By contrast to this literature, using epistemic logic we develop a non-probabilistic model of payoff uncertainty, and show that under such a conception common knowledge of payoff uncertainty and rationality characterises a new solution concept that we call 'mixed iterated strict weak dominance.'

The structure of the paper is as follows. Section 2 introduces game theoretic and logical notation. Section 3 develops the formalisation of payoff uncertainty, rationality, and knowledge about payoff uncertainty and rationality. Section 4 contains the epistemic characterisation theorem and a proof. Section 5 compares our approach with the literature and motivates the axioms in terms of lexicographic beliefs. Section 6 concludes.

2 Notation

2.1 Game theory

Let $\Gamma = (\mathcal{I}, (\mathcal{A}_i)_i, (u_i)_i)$ be an N-person normal form game, and let X_1, X_2, and so on, be sets of strategies satisfying $X_i \subseteq \mathcal{A}_i$ for all $i \in \mathcal{I}$. The subgame of Γ 'spanned' by $\prod_i X_i$ is the game $(\mathcal{I}, (X_i)_i, (u_i|_{X_i})_i)$ resulting from Γ by removing for all i the strategies in the complement of X_i (with respect to $\mathcal{A}_i$) and adapting the utility functions correspondingly. We write

$$nsd_i^{\Gamma}(X_1, \ldots, X_N)$$

for the pure strategies that are not strictly dominated for player i in the subgame of Γ spanned by $\prod_i X_i$, and nwd_i^{Γ} analogously for weak dominance. With some abuse of notation, this applies to functions with different domains. Assuming some enumeration, player i's strategies are written i_1, i_2, and so on. A multi-matrix $(r_{i,k_1,\ldots,k_N})_{i,k_1,\ldots,k_N}$ containing reals $r_{i,k_1,\ldots,k_N}$ is used to build constructs of the form

$$nsd_i^{\Gamma}(X_1, \ldots, X_N, (r_{i,k_1,\ldots,k_N})_{i,k_1,\ldots,k_N})$$

denoting the set of pure strategies that are not strictly dominated in the subgame of Γ spanned by $\prod_i X_i$ in which the utility functions u_i are replaced by utility functions u_i'

defined by

$$u_i'(1_{k_1}, \ldots, N_{k_N}) = r_{i,k_1,\ldots,k_N}.$$

That is, take Γ, remove all strategies in the complement of X_i, substitute u_i by u_i', and collect all strategies that are not strictly dominated in the resulting game. A multi-matrix containing sets of reals $(D_{i,k_1,\ldots,k_N})_{i,k_1,\ldots,k_N}$ is used to build constructs of the form

$$nsd_i^{\Gamma}(X_1, \ldots, X_N, (D_{i,k_1,\ldots,k_N})_{i,k_1,\ldots,k_N})$$

denoting the set of pure strategies of player i that are not strictly dominated in any subgame spanned by $\prod_i X_i$ in which the utility functions u_i are replaced by utility functions u_i' satisfying

$$u_i'(1_{k_1}, \ldots, N_{k_N}) \in D_{i,k_1,\ldots,k_N}.$$

That is, take as many copies of Γ as there are u_i' satisfying this condition, remove all strategies in the complement of X_i, substitute u_i by u_i' in the corresponding copy of Γ, and collect all strategies that are not strictly dominated in any resulting game.

Using the nsd_i^{Γ} and the nwd_i^{Γ}, the Dekel-Fudenberg procedure is defined recursively by

$$\begin{aligned} \mathsf{DF}_i^0 &= \mathcal{A}_i, \\ \mathsf{DF}_i^1 &= nwd_i^{\Gamma}(\mathsf{DF}_1^0, \ldots, \mathsf{DF}_N^0), \\ \mathsf{DF}_i^{n+1} &= nsd_i^{\Gamma}(\mathsf{DF}_1^n, \ldots, \mathsf{DF}_N^n) \quad (n \geq 1); \end{aligned}$$

iterated strict dominance, by

$$\begin{aligned} \mathsf{S}_i^0 &= \mathcal{A}_i, \\ \mathsf{S}_i^{n+1} &= nsd_i^{\Gamma}(\mathsf{S}_1^n, \ldots, \mathsf{S}_N^n) \quad (n \geq 0); \end{aligned}$$

and the concept we characterise, which we call 'mixed iterated strict weak dominance,' by

$$\begin{aligned} \mathsf{M}_i^0 &= \mathcal{A}_i, \\ \mathsf{M}_i^{n+1} &= nwd_i^{\Gamma}(\mathsf{S}_1^n, \ldots, \mathsf{M}_i^n, \ldots, \mathsf{S}_N^n) \quad (n \geq 0), \end{aligned}$$

showing mixed recursion. Informally put, the idea is that player i considers a sequence of games spanned by, for opponent strategies, the relevant stages from the sequence of iterated strict dominance, and for himself, the strategies that are not weakly dominated for him in the previous stage. Stage zero is $\mathcal{A}_i$. To obtain stage one, he removes from the entire game those strategies of his that are weakly dominated. To obtain stage two, he considers the game spanned by, for opponent strategies, the sets S_j^1, $j \neq i$, and for

himself, the set obtained at stage one (that is, $nwd_i^{\Gamma}(\mathcal{A}_1, \ldots, \mathcal{A}_i, \ldots, \mathcal{A}_N)$), and he removes from this game those strategies of his that are weakly dominated. And so he continues.

Players are 'rational' if they conform to Dekel and Fudenberg's (1990) maxim not to play *weakly* dominated strategies.

2.2 Logic

We use logical symbols $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (implication), and $\leftrightarrow$ (equivalence). No quantifiers are needed. The conjunction (disjunction) of all sentences from a finite set Σ is abbreviated by $\bigwedge \Sigma (\bigvee \Sigma)$. If the φ_i enumerate Σ we often write $\bigwedge_i \varphi_i$ ($\bigvee_i \varphi_i$).

The $\Box_i$ operator has an epistemic reading ('i knows that…'). In fact, we shall formally cast the epistemic characterisation theorem in a proof system without veridicality, that is, in terms of 'true belief' ($\varphi \wedge \Box_i \varphi$). But we follow standard convention informally to phrase epistemic characterisation theorems in terms of knowledge instead of true belief. The $\mathbf{E}_{\mathcal{I}}$ operator stands for 'every player $i \in \mathcal{I}$ knows that…' The $\mathbf{C}_{\mathcal{I}}$ is used to speak about 'common' knowledge: 'all players know that…, and all players know that all players know that…, and all players know that all players know that all players know that…, and so on *ad inf.*' An abbreviation for $\mathbf{E}_{\mathcal{I}} \cdots \mathbf{E}_{\mathcal{I}} \varphi$ with n occurrences of $\mathbf{E}_{\mathcal{I}}$ is $\mathbf{E}_{\mathcal{I}}^n \varphi$. Furthermore, $\mathbf{E}_{\mathcal{I}} \varphi \wedge \mathbf{E}_{\mathcal{I}}^2 \varphi \wedge \ldots \wedge \mathbf{E}_{\mathcal{I}}^n \varphi$ is written $\mathbf{E}_{\mathcal{I}}^{\leq n} \varphi$. This is referred to as common knowledge 'up to level n.' We write $\Box_i \Sigma$ for player i's knowledge that at least one of the propositions from Σ holds; that is, for $\Box_i \bigvee \Sigma$.

Proposition letters $\mathbf{i_m}$ are used for the statement 'i plays his mth strategy i_m.' The formal analogue of the statement that $u_i(1_{k_1}, \ldots, N_{k_N}) = r$ for some real number r is $\mathbf{u}_i(1_{k_1}, \ldots, N_{k_N}) = \mathbf{r}$. At most countably many symbols for real numbers are needed. Proposition letters $\mathbf{rat}_i$ are used for rationality in the sense of Dekel and Fudenberg's (1990) maxim to exclude playing weakly dominated strategies. The set of proposition letters for player i's pure strategies that survive at least n rounds of the Dekel-Fudenberg procedure is denoted by $\mathcal{DF}_i^n$, and its limit (the set of player i's pure strategies that survive the entire Dekel-Fudenberg procedure), by $\mathcal{DF}_i^{\infty}$. Sets $\mathcal{S}_i^n$ (for iterated strict dominance), and $\mathcal{M}_i^n$ (for the solution concept we characterise, mixed iterated strict weak dominance), and their limits, are defined similarly.

The following axioms from epistemic logic are standard:

Prop	All classical propositional tautologies.
K	$\Box_i(\varphi \rightarrow \psi) \rightarrow (\Box_i \varphi \rightarrow \Box_i \psi)$.
D	$\Box_i \varphi \rightarrow \neg \Box_i \neg \varphi$.
4	$\Box_i \varphi \rightarrow \Box_i \Box_i \varphi$.
E	$\mathbf{E}_{\mathcal{I}} \varphi \leftrightarrow \bigwedge_{i \in \mathcal{I}} \Box_i \varphi$.
C	$\mathbf{C}_{\mathcal{I}} \varphi \leftrightarrow \mathbf{E}_{\mathcal{I}}(\varphi \wedge \mathbf{C}_{\mathcal{I}} \varphi)$.

The following axioms describe situations of normal form game playing:

$\text{Strat}_{\geq 1}$	$\bigwedge_{i \in \mathcal{I}} \bigvee_m \mathbf{i_m}$.
$\text{Strat}_{\leq 1}$	$\bigwedge_{i \in \mathcal{I}} \bigwedge_{m \neq n} \neg(\mathbf{i_m} \wedge \mathbf{i_n})$.
KnStrat	$\bigwedge_{i \in \mathcal{I}} \bigwedge_m (\Box_i \mathbf{i_m} \leftrightarrow \mathbf{i_m})$.

Every player plays at least one strategy ($\text{Strat}_{\geq 1}$) and at most one strategy ($\text{Strat}_{\leq 1}$), and knows which strategy he chooses (KnStrat).

The following proof rules of modus ponens, necessitation, and induction are standard:

MP If $\vdash \varphi \rightarrow \psi$ and $\vdash \varphi$, then $\vdash \psi$.
Nec If $\vdash \varphi$, then $\vdash \Box_i \varphi$.
Ind If $\vdash \varphi \rightarrow \mathbf{E}_{\mathcal{I}}(\varphi \wedge \psi)$, then $\vdash \varphi \rightarrow \mathbf{C}_{\mathcal{I}}\psi$.

3 Formalisation

3.1 Payoff uncertainty

Without loss of generality we assume that we are dealing with a two-person normal form game, and that we wish to formalise the statement that player i has approximate knowledge about player j's utility function. What makes the formalisation task a non-trivial one is that we have to accomplish this in a context in which player i knows, too, that player j has exact knowledge about player j's utility function. We will first formalise knowledge about exact knowledge, and then turn to approximate knowledge.

Knowledge about exact knowledge about a utility function The most straightforward way to formalise the statement that player j has exact knowledge about the utility $r_{j,k,l}$ player j assigns to strategy profile (i_k, j_l) is

$$\Box_j \mathbf{u}_j(i_k, j_l) = \mathbf{r}_{j,k,l},$$

and as a result the most straightforward way to formalise the statement that player i knows that player j has exact knowledge about the utility player j assigns to strategy profile (i_k, j_l) is to add a $\Box_i$ to the above sentence, resulting in

$$\Box_i \Box_j \mathbf{u}_j(i_k, j_l) = \mathbf{r}_{j,k,l}. \tag{1}$$

Yet this cannot be coherent in a context where player i has only approximate knowledge about player j's utility function, for in such a context sentence (1) implies that player i has in mind the specific utility value $r_{j,k,l}$, illegitimately suggesting a kind of exact knowledge about player j's utility function.

What does work, though, is to cast player i's knowledge about player j's exact knowledge about the utility player j assigns to strategy profile (i_k, j_l) in knowledge about a conjunction of conditionals,

$$\Box_i \bigwedge_{r \in D_{j,k,l}} (\mathbf{u}_j(i_k, j_l) = \mathbf{r} \rightarrow \Box_j \mathbf{u}_j(i_k, j_l) = \mathbf{r}),$$

for a finite set $D_{j,k,l}$ containing $r_{j,k,l}$ and contained in a small environment of $r_{j,k,l}$. In words this expresses player i's knowledge that if the utility player j assigns to (i_k, j_l) is such and such, then player j knows that it is such and such, and this does *not* imply that player i has particular utility values in mind.

Generalising this,

$$\Box_i \bigwedge_{k,l} \bigwedge_{r \in D_{j,k,l}} (\mathbf{u}_j(i_k, j_l) = \mathbf{r} \rightarrow \Box_j \mathbf{u}_j(i_k, j_l) = \mathbf{r})$$

expresses the fact that player i knows that player j has exact knowledge about player j's utility function. We abbreviate this by $\Box_i \Box_j \upsilon_j$, and we abbreviate

$$\bigwedge_{k,l} \bigwedge_{r \in D_{j,k,l}} (\mathbf{u}_j(i_k, j_l) = \mathbf{r} \rightarrow \Box_j \mathbf{u}_j(i_k, j_l) = \mathbf{r})$$

by $\Box_j \upsilon_j$, where the fact that this notation does not fully capture the logical form of the statements is unproblematic.

Approximate knowledge about a utility function The careful formalisation of knowledge about exact knowledge about a utility function makes it straightforward now to formalise approximate knowledge about a utility function. Because of the above problem about illegitimate knowledge about specific utility values, it is no option to represent player i's approximate knowledge about the utility player j assigns to strategy profile (i_k, j_l) by means of

$$\Box_i \mathbf{u}_j(i_k, j_l) = \mathbf{r}'$$

for some r' sufficiently close to the real $r_{j,k,l} = u(i_k, j_l)$. What does work, by contrast, is to write

$$\Box_i \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r}$$

for a finite set of reals $D_{j,k,l}$ containing $r_{j,k,l}$ and contained in a small environment of $r_{j,k,l}$, and to define the degree of approximation by putting specific conditions on $D_{j,k,l}$. A natural such condition we adopt in this paper is the following, suggested by Dekel and Fudenberg (1990). Given a particular two-person normal form game, one can find $\varepsilon_{j,k,l} > 0$ such that if $|r - r_{j,k,l}| < \varepsilon_{j,k,l}$ for all $r \in D_{j,k,l}$, player i's knowledge is such that player i gets relations of strict dominance among the strategies of player j right, but not (necessarily) relations of weak dominance. We shall say that the $\varepsilon_{j,k,l}$ 'ensure knowledge about strict dominance' whenever these conditions hold with respect to the utility player j assigns to strategy profile (i_k, j_l).

3.2 Rationality

The rationality principle we wish to formalise is defined by its exclusion of *weakly* dominated strategies, and was suggested by Dekel and Fudenberg (1990). It is not our aim here to defend the conceptual or empirical plausibility of the principle, the purpose of this paper being to construct a logical formalism to capture structural aspects of

common knowledge of payoff uncertainty and rationality. However, we are well aware of the fact that weak dominance is generally considered to be more problematic than strict dominance, witness elaborate treatments of this issue in the literature referred to in Sect. 5 and references therein. But while a structural clarification of common knowledge of payoff uncertainty and rationality is the main objective of this paper, we do believe that our formalisation can help shed light on issues about plausibility by showing the precise logical consequences of such assumptions.

The following two axioms capture the rationality principle of excluding weakly dominated strategies:

$$\begin{aligned}
\mathrm{Rat}_{bas} \quad & (\mathbf{rat}_i \wedge \Box_i \textstyle\bigwedge_{i_k \in \mathcal{A}_i, j_l \in \mathcal{A}_j} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l}) \\
& \rightarrow nwd_i^{\Gamma}(\mathcal{A}_i, \mathcal{A}_j, (\mathbf{r}_{i,k,l})_{i,k,l}). \\
\mathrm{Rat}_{ind} \quad & (\mathbf{rat}_i \wedge \Box_i \textstyle\bigwedge_{i_k \in X_i, j_l \in X_j} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l} \wedge \Box_i X_i \wedge \Box_i X_j) \\
& \rightarrow nwd_i^{\Gamma}(X_i, X_j, (\mathbf{r}_{i,k,l})_{i,k,l}).
\end{aligned}$$

Without loss of generality, the rationality axioms are phrased for a two-person normal form game, for player i. With the convention that $j = 3 - i$ is player i's opponent, the rationality axioms for player j are analogous.

The axioms fix the meaning of the term $\mathbf{rat}_i$ in an inductive and implicit manner. The Rat_{bas} axiom states that if player i is rational, that then he chooses a strategy that is not weakly dominated in the entire game. Often, however, player i will be able to exclude more than only the strategies that are weakly dominated in the entire game because he will know that certain strategies of his opponent and of himself will not be chosen. Such knowledge is represented by means of the clauses $\Box_i X_i$ (player i knows that player i will choose a strategy from X_i) and $\Box_i X_j$ (player i knows that player j will choose a strategy from X_j). Given such knowledge, the rationality principle holds that player i chooses a strategy that is not weakly dominated in the subgame of the original game spanned by $X_i \times X_j$.

All in all, the two axioms take care of a situation without additional knowledge as well as of a situation with additional knowledge.

3.3 Knowledge about payoff uncertainty and rationality

To sum up, we have formalised payoff uncertainty (knowledge about exact knowledge about utility functions, and approximate knowledge about utility functions), and we have formalised rationality. To investigate the behavioural consequences of common knowledge of payoff uncertainty and common knowledge of rationality, we have to take one more step and add two axioms to ensure that common knowledge of payoff uncertainty and common knowledge of rationality do the job they are supposed to do.

It may come as a surprise that we need extra axioms to that end, since given our formalisation of rationality one could at first sight expect that, technically put, applying the rule of necessitation to the rationality axioms would suffice. But as we shall see, such a route is blocked for reasons that are quite similar to the reasons why we had to develop a special formalisation of knowledge about exact knowledge about utility functions. The assumptions of common knowledge of payoff uncertainty and

rationality remain powerless in a proof of an epistemic characterisation theorem if necessitation on the rationality axioms is the only way to proceed.

This is why. Necessitation for $\Box_j$ and the K axiom yield, applied to the Rat_{bas} axiom for player i,

$$(\Box_j \mathbf{rat}_i \wedge \Box_j \Box_i \bigwedge_{i_k \in \mathcal{A}_i, j_l \in \mathcal{A}_j} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l}) \rightarrow \Box_j nwd_i^{\Gamma}(\mathcal{A}_i, \mathcal{A}_j, (\mathbf{r}_{i,k,l})_{i,k,l}). \quad (2)$$

The antecedent of this sentence will, however, not be made true under the assumption of common knowledge of payoff uncertainty and rationality. This is because the second conjunct of the antecedent involves knowledge possessed by player j about player i's knowledge about specific utility values player i assigns to certain outcomes of the game, and as we saw, this is incoherent in a context in which player j has only approximate knowledge about player i's utility function.

In fact, it is a good thing that the antecedent of sentence (2) cannot be made true under the assumption of common knowledge of payoff uncertainty and rationality, for if it were true, player j would have knowledge about weak dominance relations among player i's strategies. Such knowledge was excluded because approximate knowledge about utility functions was defined in terms of getting strict dominance relations right, but not (necessarily) weak dominance relations.

Necessitation, then, does not get the right procedure off the ground without extra axioms. Our solution is to use clauses of the form

$$(\Box_j \mathbf{rat}_i \wedge \Box_j \Box_i \upsilon_i \wedge \Box_j \bigwedge_{k,l} \bigvee_{r \in D_{i,k,l}} \mathbf{u}_i(i_k, j_l) = \mathbf{r}) \rightarrow \Box_j nsd_i^{\Gamma}(\mathcal{A}_i, \mathcal{A}_j, (D_{i,k,l})_{i,k,l}),$$

to express the consequences of player j's knowledge about player i's rationality in a situation in which player j only has approximate knowledge about player i's utility functions. Clearly, the antecedent conditions are fulfilled once common knowledge of payoff uncertainty and rationality is assumed: player j knows that player i is rational (first conjunct), player j knows that player i has exact knowledge of player i's utility function (second conjunct), and player j has approximate knowledge about player i's utility function (third conjunct). Equally clearly, the consequent provides the appropriate knowledge: player j knows that player i will choose a strategy that is not *strictly* dominated.

More intricate, but structurally similar, reasoning occurs in the proof of the epistemic characterisation result, for which more intricate, but structurally similar, axioms are needed. In fact, again a distinction is made between two cases, depending on whether player i has additional knowledge in the form of $\Box_i X_i$ and $\Box_i X_j$, or not:

Knw_{bas} $(\Box_j \Box_i^n \mathbf{rat}_i \wedge \Box_j \Box_i^n \Box_i \upsilon_i \wedge \Box_j \bigwedge_{k,l} \bigvee_{r \in D_{i,k,l}} \mathbf{u}_i(i_k, j_l) = \mathbf{r})$
$\rightarrow \Box_j \Box_i^n nsd_i^{\Gamma}(\mathcal{A}_i, \mathcal{A}_j, (D_{i,k,l})_{i,k,l}).$

Knw_{ind} $(\Box_j \Box_i^n \mathbf{rat}_i \wedge \Box_j \Box_i^n \Box_i \upsilon_i \wedge \Box_j \Box_i^n X_i \wedge \Box_j \Box_i^n X_j$

$$\wedge\Box_j \bigwedge_{i_k, j_l} \bigvee_{r \in D_{i,k,l}} \mathbf{u}_i(i_k, j_l) = \mathbf{r}) \rightarrow \Box_j \Box_i^n nsd_i^{\Gamma}(X_i, X_j, (D_{i,k,l})_{i,k,l}).$$

Without loss of generality, the knowledge axioms are phrased for a two-person normal form game, to grasp the knowledge player j has about player i's rationality and knowledge ($n = 0$), or his knowledge about player i's knowledge about player i's rationality and knowledge ($n = 1$), or his knowledge about player i's knowledge about player i's knowledge about player i's rationality and knowledge… ($n > 1$). With the convention that $j = 3 - i$ is player i's opponent, the knowledge axioms for player j are analogous.

4 Epistemic characterisation theorem

We turn to statement and proof of the epistemic characterisation theorem of mixed iterated strict weak dominance in terms of common knowledge of payoff uncertainty and rationality.

The proof system consists of the following axioms: Prop, K, D, 4, E, C, the proof rules modus ponens, necessitation, and induction, the three axioms $\text{Strat}_{\geq 1}$, $\text{Strat}_{\leq 1}$, and KnStrat, plus the four axioms Rat_{bas}, Rat_{ind}, Knw_{bas}, and Knw_{ind}.

Theorem 1 *Let $\Gamma = (\mathcal{I}, (\mathcal{A}_i)_i, (u_i)_i)$ be a two-person normal form game, let $i = 1, 2$, $j = 3 - i$, and let $D_{i,k,l}$ be finite sets of reals such that $|r - u_{i,k,l}| < \varepsilon_{i,k,l}$ for all $r \in D_{i,k,l}$ and $\varepsilon_{i,k,l}$ ensuring knowledge about strict dominance. Then*

$$\begin{aligned}
\vdash \Bigg(& \bigwedge_{i,k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l} \wedge \bigwedge_i \mathbf{rat}_i \wedge \\
& \bigwedge_i \Box_i \bigwedge_{k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l} \wedge \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r} \\
& \wedge \mathbf{C} \bigwedge_i \Box_i \upsilon_i \wedge \mathbf{C} \bigwedge_i \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r} \wedge \mathbf{C} \bigwedge_i \mathbf{rat}_i \Bigg) \\
\rightarrow & \bigwedge_i \mathcal{M}_i^{\infty}
\end{aligned}$$

Proof We write φ^n for

$$\begin{aligned}
& \bigwedge_i \mathbf{rat}_i \wedge \bigwedge_{i,k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l} \wedge \\
& \bigwedge_i \Box_i \bigwedge_{k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l} \wedge \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r} \\
& \wedge \mathbf{E}^{\leq n} \Box_i \upsilon_i \wedge \mathbf{E}^{\leq n} \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r} \wedge \mathbf{E}^{\leq n} \bigwedge_i \mathbf{rat}_i
\end{aligned}$$

and φ_i^n for the part of φ^n starting with $\Box_i$ conjoined with the statement $\mathbf{rat}_i$, that is,

$$\mathbf{rat}_i \wedge \Box_i \bigwedge_{k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l} \wedge \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r}$$
$$\wedge \Box_i \mathbf{E}^{\leq n-1} \Box_i \upsilon_i \wedge \Box_i \mathbf{E}^{\leq n-1} \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r} \wedge \Box_i \mathbf{E}^{\leq n-1} \bigwedge_i \mathbf{rat}_i,$$

with the convention that if $n = 0$ the $\Box_i \mathbf{E}^{\leq n-1}$ vanish completely.

We prove

Claim 1 $\forall n \vdash \varphi_i^n \rightarrow \mathcal{M}_i^{n+1}$.

First, however, we prove

Claim 2 *All you need to prove the statement* $\forall n \vdash \varphi_i^n \rightarrow \mathcal{M}_i^{n+1}$ *from Claim 1 is:* $\forall n \vdash \varphi_i^n \rightarrow (\Box_i \mathcal{M}_i^n \wedge \Box_i \mathcal{S}_j^n)$.

Proof of Claim 2 Assume $\forall n \vdash \varphi_i^n \rightarrow \Box_i \mathcal{M}_i^n \wedge \Box_i \mathcal{S}_j^n$ is proved. By definition of φ_i^n we have, too,

$$\forall n \vdash \varphi_i^n \rightarrow (\mathbf{rat}_i \wedge \Box_i \bigwedge_{k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l}).$$

Apply Rat_{ind} for i to get

$$\forall n \vdash \varphi_i^n \rightarrow nwd_i^{\Gamma}(\mathcal{M}_i^n, \mathcal{S}_j^n),$$

which, observing that $\mathcal{M}_i^{n+1} = nwd_i^{\Gamma}(\mathcal{M}_i^n, \mathcal{S}_j^n)$, concludes the proof of Claim 2. □

One level deeper we prove

Claim 3 *All you need to prove the statement* $\forall n \vdash \varphi_i^n \rightarrow (\Box_i \mathcal{M}_i^n \wedge \Box_i \mathcal{S}_j^n)$ *from Claim 2 is*

$$\forall n \vdash \varphi_i^n \rightarrow (\Box_i \Box_i \mathcal{M}_i^{n-1} \wedge \Box_i \Box_i \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_i^{n-1}).$$

Proof of Claim 3 Assume $\forall n \vdash \varphi_i^n \rightarrow (\Box_i \Box_i \mathcal{M}_i^{n-1} \wedge \Box_i \Box_i \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_i^{n-1})$ is proved. To show

$$\forall n \vdash \varphi_i^n \rightarrow \Box_i \mathcal{M}_i^n,$$

we observe that by definition of φ_i^n and the assumption we have

$$\forall n \vdash \varphi_i^n \rightarrow (\Box_i \mathbf{rat}_i \wedge \Box_i \Box_i \bigwedge_{k,l} \mathbf{u}_i(i_k, j_l) = \mathbf{r}_{i,k,l}$$
$$\wedge \Box_i \Box_i \mathcal{M}_i^{n-1} \wedge \Box_i \Box_i \mathcal{S}_j^{n-1}).$$

Apply the rule of necessitation for i to the appropriate instance of Rat_{ind} for i, and observe that its consequent is what we wish to show and that its antecedent is what we have just shown. To show

$$\forall n \vdash \varphi_i^n \to \Box_i \mathcal{S}_j^n,$$

we observe that by definition of φ_i^n and the assumption we have

$$\forall n \vdash (\varphi_i^n \to \Box_i \mathbf{rat}_j \wedge \Box_i \Box_j \upsilon_j \wedge \Box_i \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_i^{n-1} \wedge \Box_i \bigwedge_{k,l} \bigvee_{r \in D_{j,k,l}} \mathbf{u}_j(i_k, j_l) = \mathbf{r}).$$

Take the appropriate instance of Knw_{ind} for i (the instance of the axiom that speaks about i's beliefs about j's iterated beliefs), and observe that its consequent is what we wish to show and that its antecedent is what we have just shown. This concludes the proof of Claim 3. □

Now we are ready for

Claim 4 *The statement* $\forall n \vdash \varphi_i^n \to (\Box_i \Box_i \mathcal{M}_i^{n-1} \wedge \Box_i \Box_i \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_i^{n-1})$ *from Claim 3 is true.*

Proof of Claim 4 Induction on n. The basis is left to the reader. Assume that the statement holds for some n. We prove the four implications for $n+1$ separately.

(i) To prove

$$\varphi_i^{n+1} \to \Box_i \Box_i \mathcal{M}_i^n,$$

it is sufficient to prove $\varphi_i^{n+1} \to (\Box_i \Box_i \Box_i \mathcal{M}_i^{n-1} \wedge \Box_i \Box_i \Box_i \mathcal{S}_j^{n-1})$. Assume the latter to be true. Apply the rule of necessitation for i to the appropriate instance of Rat_{ind} for i twice, and observe that its consequent is what we wish to show and that its antecedent follows from the definition of φ_i^{n+1} and the assumption.

To prove the truth of the assumption apply necessitation for i to the inductive hypothesis and observe that $\forall n \vdash \varphi_i^{n+1} \to \Box_i \varphi_i^n$.

(ii) To prove

$$\varphi_i^{n+1} \to \Box_i \Box_i \mathcal{S}_j^n,$$

it is sufficient to prove $\varphi_i^{n+1} \to (\Box_i \Box_i \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_i \Box_j \mathcal{S}_i^{n-1})$. Assume the latter to be true. Apply the rule of necessitation for i once to the appropriate instance of Knw_{ind} for i and for $n = 0$, and observe that its antecedent follows from the definition of φ_i^{n+1} and the assumption, and that the consequent is what we wish to prove.

To prove the truth of the assumption apply necessitation for i to the inductive hypothesis and observe that $\forall n \vdash \varphi^{n+1} \to \Box_i \varphi_i^n$.

(iii) To prove

$$\varphi_i^{n+1} \rightarrow \Box_i \Box_j \mathcal{S}_j^n,$$

it is sufficient to prove $\varphi_i^{n+1} \rightarrow (\Box_i \Box_j \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \Box_j \mathcal{S}_i^{n-1})$. Assume the latter to be true. Take the appropriate instance of Knw_{ind} for i and for $n = 1$, and observe that its consequent follows from the definition of φ_i^{n+1} and the assumption.

To prove the truth of the assumption observe that, first, $\forall n \vdash \varphi_i^{n+1} \rightarrow \varphi_i^n$, and second, $\forall n \vdash (\Box_i \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \mathcal{S}_i^{n-1}) \rightarrow (\Box_i \Box_j \Box_j \mathcal{S}_j^{n-1} \wedge \Box_i \Box_j \Box_j \mathcal{S}_i^{n-1})$. Together with the inductive hypothesis this shows the truth of the assumption.

(iv) To prove

$$\varphi_i^{n+1} \rightarrow \Box_i \Box_j \mathcal{S}_i^n,$$

it is sufficient to prove $\varphi_i^{n+1} \rightarrow (\Box_i \Box_j \Box_i \mathcal{S}_i^{n-1} \wedge \Box_i \Box_j \Box_i \mathcal{S}_j^{n-1})$. Assume the latter to be true. Apply necessitation for i to the appropriate instance of Knw_{ind} for j and for $n = 0$, and observe that its antecedent follows from the definition of φ_i^{n+1} and the assumption and that its consequent is what we wish to prove.

To prove the truth of the assumption we observe that, indeed, the inductive hypothesis may be assumed to hold for φ_j^n as well. Applying the rule of necessitation for i to it yields $\Box_i \varphi_j^n \rightarrow (\Box_i \Box_j \Box_i \mathcal{S}_i^{n-1} \wedge \Box_i \Box_j \Box_i \mathcal{S}_j^{n-1})$. What we now need is $\varphi_i^{n+1} \rightarrow \Box_i \varphi_j^n$. And that is easy to see. □

5 Discussion

5.1 Comparison with the literature

The motivation to prove epistemic characterisation theorems typically comes from one of two sides. One may start with a familiar game theoretic solution concept in mind, and ask under what epistemic conditions it will capture game play adequately. And one may start with certain epistemic conditions in mind, and investigate what solution concept follows, that is, in the game theoretic jargon, investigate the 'behavioural consequences' of certain epistemic conditions. This may lead, in the first case, to the discovery of new and surprising epistemic conditions, and in the latter case, to new and surprising solution concepts.

This paper investigates the behavioural consequences of epistemic conditions involving common knowledge of payoff uncertainty and rationality. Dekel and Fudenberg (1990) presented a formalisation of conditions involving payoff uncertainty, and linked them to the iterated elimination of strictly dominated strategies preceded by one round of elimination of weakly dominated strategies, the Dekel-Fudenberg procedure. Subsequent game theoretic and logical research zoomed in on the solution concept of the Dekel-Fudenberg procedure and provided epistemic characterisations in terms of approximate common knowledge of rationality (Börgers 1994), a lexicographic variant

thereof called 'common first-order knowledge' (Brandenburger 1992), the weakest perfect τ-theory (Gul 1996), in terms of players believing that opponents make errors with small (and correlated) probability (Herings and Vannetelbosch 2000), and in terms of common knowledge of perfect rationality (Stalnaker 1996).

By contrast to this literature, rather than starting with the solution concept in mind, we zoom in on the epistemic conditions. We provide an alternative formalisation of common knowledge of payoff uncertainty and rationality, and investigate its behavioural consequences. These consequences are quite different from the Dekel-Fudenberg procedure, and this is, of course, due to the fact that our formalisation of common knowledge of payoff uncertainty and rationality is different. Precisely to locate these differences, we discuss the conceptions of common knowledge of payoff uncertainty and rationality developed by Dekel and Fudenberg and Börgers (as they are fairly representative for the literature), and show by means of an example to what extent they are different from the model we propose.

The difference lies in the formalisation of payoff uncertainty, not in that of rationality, for as we noted, we explicitly adopt Dekel and Fudenberg's principle that a rational player does not choose weakly dominated strategies (Dekel and Fudenberg 1990). Payoff uncertainty, however, we model differently. Without going into too much technical detail, what Dekel and Fudenberg do is to model payoff uncertainty by means of elaborations of games as they were developed by Harsanyi (1967–1968). Roughly, a sequence of games is considered in which the utility functions of the players are slightly different than in the original game. A notion of convergence is defined, both on sequences of games and on sequences of strategies, and the main result is then that a strategy survives the Dekel-Fudenberg procedure just in case it is the limit of a sequence of strategies that survive the iterated elimination of weakly dominated strategies in elaborations converging to the original game.

Börgers (1994), in turn, defines a notion of approximate common knowledge, and shows that approximate common knowledge of rationality characterises the Dekel-Fudenberg procedure. A proposition φ is approximate common knowledge whenever everyone believes with high probability that φ, everybody believes with high probability that everyone believes with high probability that φ, and so on.

The difference between these two models and the model we propose becomes transparent once we compare the respective formalisations of statements such as 'player i has approximate knowledge about the utility player j assigns to strategy profile (i_k, j_l).' Dekel and Fudenberg and Börgers model this by having player i assign high probability to the correct, actual utility value, and low probability to alternative utility values. Under this reading, approximate knowledge is modelled in probabilistic terms. In our model, by contrast, player i has knowledge about a disjunction of statements one of which expresses the correct utility, and all of which fall in a small environment of the correct utility. Our non-probabilistic model has, one could say, player i assign equal probability to finitely many alternatives close to the actual value.

To show that this makes a difference, consider the games shown in Fig. 1. The game on the left correctly represents the utility functions of both players. It is the game that they actually play. The game on the right is the game as player 2 perceives it. It represents player 2's utility correctly, but it gives an approximation of the utility player 1 assigns to strategy profile $(1_1, 2_2)$. The intended interpretation is that player 2

	2_1	2_2	2_3
1_1	(1, 4)	(1, 8)	(1, 0)
1_2	(5, 8)	(1, 4)	(3, 0)
1_3	(3, 4)	(0, 8)	(2, 0)

	2_1	2_2	2_3
1_1	(1, 4)	(D,8)	(1, 0)
1_2	(5, 8)	(1, 4)	(3, 0)
1_3	(3, 4)	(0, 8)	(2, 0)

Fig. 1 The real game and the believed game

knows that $u_1(1_1, 2_2) \in D$ for some finite set D containing the actual utility value, and contained in a small environment of it, such that the conditions of the epistemic characterisation theorem are satisfied. That is, that player 2 has approximate knowledge about the utility player 1 assigns to $(1_1, 2_2)$ is formalised in our model by knowledge about a disjunction over D. Strictly speaking, the game on the right represents not one but $|D|$ games. For completeness, we should mention that player 1 perceives the game as it is.

The only Dekel-Fudenberg outcome is $(1_2, 2_1)$. If player 2 assigns high probability p to the correct utility player 1 assigns to $(1_1, 2_2)$, any other outcome disappears when p tends to infinity. Similar observations hold for Börgers' model. If we assume common knowledge of payoff uncertainty and rationality as we have formalised it, however, there is a second outcome possible, namely $(1_2, 2_2)$. An informal way to demonstrate this is the following. First note that under common knowledge of payoff uncertainty and rationality as we have formalised it player 1 will indeed play his second strategy: both his first one and his third one are weakly dominated. To see that player 2 can play his second strategy under these assumptions, observe that, since he is rational, he will not play the weakly dominated 2_3. Further, player 2 knows that player 1 is rational. From the game which player 2 thinks is being played (the matrix on the right), player 2 removes player 1's weakly dominated strategies, but the only such strategy is 1_3. In particular, since player 2 only has approximate knowledge about the utility function of player 1, he cannot justifiably remove 1_1: player 2 holds it possible that 1_1 will score more for player 1 against 2_2 than any other strategy of player 1. From player 2's point of view, player 1 will either play his first or his second strategy, and this makes, for player 2, playing his first as well as his second strategy rational. Rationality does not exclude 2_2, and hence $(1_2, 2_2)$ is a possible outcome.

5.2 Motivation of the axioms

A syntactic approach as proferred in this paper makes it possible to analyse the role of various levels of common knowledge about rationality and a particular form of payoff uncertainty, but it has the admitted drawback to make it more difficult to motivate the axioms in terms of the lexicographic beliefs characteristic of the semantic approach. While to some extent this is distinctive of the syntactic approach to interactive epistemology (see, e.g., Clausing 2004; van Benthem 2003), such a motivation may justifiably be demanded, and we shall attempt to give one here.

Let us start with the Rat_{bas} axiom. In our formalisation the purpose of this axiom is to capture the rationality principle according to which weakly dominated strategies are to be excluded, and it does so by stating that it can be concluded that player i will not play a strategy that is weakly dominated in the entire game on the basis of no more information than that player i is rational and that he knows his own utility function. A high-yielding place to look for a basis of a motivation for this axiom is a result due to Pearce (1984). It states that a strategy is not weakly dominated for player i in the entire game iff it is a best response for him to some probability distribution with full support (that is, assigning zero probability to no alternative) over player j's available strategies. Rat_{bas} indeed formalises such a situation in which no specific information is available about which alternatives player i's belief exclude.

In the process of iteratively eliminating strategies as discussed in this paper, player i's beliefs focus in on certain sets of strategies of player j and himself. By doing so his beliefs start excluding alternatives and as a result cease to have full support (the temporal phrasing of this process is rather metaphoric). At the first stage of the elimination process player i assigns zero probability to no strategy of his opponent. At the next stage of the elimination process, however, he will remove strategies that, given his own payoff uncertainty and knowledge about his opponent j's rationality, he believes player j will not play. If such strategies exist, player i's beliefs at this stage will exclude them and consequently not have full support. In our formalism this case is dealt with by Rat_{ind}.

How to motivate this axiom? First, consider the following attempt to motivate Rat_{ind}. If player i's beliefs are captured by $\Box_i X_i \wedge \Box_i X_j$ for $X_i \subseteq \mathcal{A}_i$ and a real subset $X_j \subset \mathcal{A}_j$, they have indeed no full support with respect to the entire game, but they do have full support with respect to the subgame spanned by $X_i \times X_j$. This shows, the attempt would go, that the result by Pearce still may be used to motivate why Rat_{ind} has player i choose a strategy that is not weakly dominated in that very subgame: simply reduce Pearce's result to the subgame.

This may sound like a plausible defense of Rat_{ind}, but it has an important defect. Going from the first to the second stage, player i has given up his full support beliefs from the first stage. While these beliefs have full support with respect to the subgame of the second stage, the fact that they no longer have full support with respect to the entire game makes it impossible simultaneously to adopt Rat_{bas} (which seems to presuppose full support beliefs with respect to the entire game), and Rat_{ind} (which seems to presuppose non-full support beliefs).

This is very much related to a puzzle presented by Larry Samuelson (1992). A solution to that puzzle by means of lexicographic probability systems by Brandenburger et al. (2007) proves very useful for our motivational task. Consider Brandenburger et al.'s version of Samuelson's game shown in Fig. 2. If player 2 is rational in the sense of not choosing weakly dominated strategies, he will not play 2_2. This means that if player 1 expects player 2 to be rational in that sense, he will not form a full support belief. But how is that possible if at the same time player 1 is rational, too, in the sense of excluding weakly dominated strategies. There seems to be a friction, then, between assuming someone to be rational in that sense, and assuming someone to believe others to be rational in that same sense.

	2_1	2_2
1_1	(1,1)	(0,1)
1_2	(0,2)	(1,0)

Fig. 2 The puzzle

Brandenburger et al.'s solution represents the players' beliefs by means of lexicographic probability systems. Player 1's primary measure assigns probability 1 to player 2 playing 2_1, thus grasping the consequences of his expectations about player 2's rationality. But player 1 has a secondary measure assigning probability 1 to the event that player 2 plays 2_2, and the interpretation of this is that player 1 considers it infinitely more likely that player 2 is rational and plays 2_1 than that he plays 2_2. The idea now is that player 1's beliefs at the end of their iterative elimination process (which is, of course, different from ours) can be represented by means of a lexicographic probability system the primary measure of which assigns positive probability to the surviving strategies only, while the remaining measures cover the subsequently eliminated strategies. In doing so, Brandenburger et al. rely on the fact that a strategy is not weakly dominated for player i in the entire game nor in the subgame spanned by X_i and X_j whenever it is a best response to a lexicographic probability system with full support (all alternatives receive non-zero probability in some measure in the system) according to which X_i is infinitely more likely than its complement. The situation described by this result is exactly the one covered by Rat_{ind} without any resulting incoherence with Rat_{bas}.

This motivates Rat_{bas} and Rat_{ind} from the perspective of lexicographic beliefs. What about Knw_{bas} and Knw_{ind}? These two axioms are intended to capture a specific conception of payoff uncertainty, namely, that if player j is fully informed about player i's rationality as well as about the fact that player i knows his own utility function, but player j is less than fully informed about player i's utility function, that then player j will form certain beliefs about player i's prospective strategies as well as about player i's beliefs about his own prospective strategies.

At first sight there may seem to be an incoherence, if not an inconsistency, as the antecedent rationality condition of Knw_{bas} and Knw_{ind} refers to a rationality concept that, by Rat_{bas} and Rat_{ind}, was cast in terms of excluding weakly dominated strategies, while the consequent refers to the exclusion of strong domination. It is important, however, to realise that in Knw_{bas} as well as in Knw_{ind} strong domination does not occur in the sense of an alternative to weak domination, and that it does not entail that player j would attribute to player i something like the possession of lexicographic (or non-lexicographic) beliefs. Rather its occurrence is motivated by the fact that player j is under the sway of a particular form of payoff uncertainty. He is so cautious as to be unwilling epistemically to exclude those strategies that are weakly dominated but not strongly dominated. An additional sign of his cautious epistemic attitude is the requirement, in the consequent of Knw_{bas} and Knw_{ind}, that these strategies are strongly dominated in *all* games that player i holds epistemically open. Player j may reason like this: I do not have any information about the likelihood of the

games captured by the sets $D_{i,k,l}$ (in crucial contrast, mentioned before, to probabilistic approaches to payoff uncertainty). Not being willing to take any epistemic risk, I therefore zoom in on strategies that are not strongly dominated in any of the games that I cannot epistemically exclude. This epistemic policy is motivated by combining my epistemic risk-aversity with my knowledge about player i's rationality of excluding weakly dominated strategies. If, for instance, I knew him to be irrational, and to play completely arbitrary, I would not adopt any specific such policy and stay with full support equiprobable beliefs.

Summarising, the assumptions of the epistemic characterisation theorem proved in this paper involve players who adopt, on the level of the practical rationality of strategy choice, a principle that is 'broad' in the sense that it allows for more than its competitor, based on strong dominance. On the level of the theoretical rationality of belief formation, by contrast, the players adopt a more 'narrow' and cautious policy.

While this motivates the axioms and makes them intelligible in terms of lexicographic beliefs and epistemic policies, it does not yet directly show them to be plausible. It is of course an empirical question whether actual players of games will adequately be described in these terms, and alternative assumptions can be studied which empirical research may reveal to be more realistic. Conceptually, however, we believe that nothing speaks against the assumptions the theorem makes. In particular, we believe that there is room for an analysis of the consequences of interpreting payoff uncertainty in equiprobable terms in a context of cautious belief formation policies, rather than as involving players whose uncertainty entails assigning different degrees of likelihood to alternatives. And while our approach does not formalise all ingredients of game playing situations that Brandenburger (2007) lists when he discusses open questions about logic and the epistemic program in game theory, we do exploit a uniquely logical tool to study the players' reasoning processes in game playing situations with payoff uncertainty, namely, implicit and inductive definitions in a recursive context to lay bare levels of knowledge and belief.

6 Conclusion

We have provided a way to formalise statements such as 'player i has approximate knowledge about the utility functions of player j,' and we have shown that on the basis of this formalisation, common knowledge of payoff uncertainty and rationality (in the sense of excluding weakly dominated strategies, due to Dekel and Fudenberg (1990)) characterises a new solution concept we have called 'mixed iterated strict weak dominance.'

It would be interesting to investigate the possibilities of extending the present framework to other iterated solution concepts. While this may be rather straightforward in many cases, such a logical analysis may not always lead to an increase in theoretical insight. In the present case, epistemic logic was fruitfully used to uncover asymmetries among the knowledge of players in settings of common knowledge of payoff uncertainty and rationality. A set of intricate axioms was applied in an epistemic characterisation theorem with a complex and informative proof. In the case of several other iterated concepts, though, we expect proofs in which conclusions follow from

assumptions without adding much insight to alternative proofs. Extending the formalism to include ingredients of game playing situations listed by Brandenburger (2007) seems a fruitful next step to take.

Acknowledgements Warmest thanks are due to Johan van Benthem, Giacomo Bonanno, Adam Brandenburger, Peter van Emde Boas, Paul Harrenstein, Wiebe van der Hoek, Robert Stalnaker, Martin Stokhof and two anonymous referees of this Journal for detailed comments on earlier versions of this paper.

References

Aumann, R. (1999). Interactive epistemology I: Knowledge. *International Journal of Game Theory, 28*, 263–300.

Baltag, A. (2002). A logic for suspicious players: Epistemic actions and belief-updates in games. *Bulletin of Economic Research, 54*, 1–45.

Board, O. (2004). Dynamic interactive epistemology. *Games and Economic Behavior, 49*, 49–80.

Bonanno, G. (2002). Modal logic and game theory: Two alternative approaches. *Risk, Decision, and Policy, 7*, 309–324.

Bonanno, G. (2003). A syntactic characterization of perfect recall in extensive games. *Research in Economics, 57*, 201–217.

Börgers, T. (1994). Weak dominance and approximate common knowledge. *Journal of Economic Theory, 64*, 265–276.

Brandenburger, A. (1992). Lexicographic probabilities and iterated admissibility. In P. Dasgupta, D. Gale, O. Hart, & E. Maskin (Eds.), *Economic analysis of markets and games* (pp. 282–276). Cambridge: The MIT Press.

Brandenburger, A. (2007). The power of paradox: Some recent developments in interactive epistemology. *International Journal of Game Theory, 35*, 465–492.

Brandenburger, A., Friedenberg, A., & Jerome Keisler, H. (2007). Admissibility in games. *Econometrica*, forthcoming.

Clausing, T. (2004). Belief revision in games of perfect information. *Economics and Philosophy, 20*, 89–115.

Dekel, E., & Fudenberg, D. (1990). Rational behavior with payoff uncertainty. *Journal of Economic Theory, 52*, 243–267.

Gul, F. (1996). Rationality and coherent theories of strategic behavior. *Journal of Economic Theory, 70*, 1–31.

Harsanyi, J. (1967–1968). Games with incomplete information played by Bayesian players, I–III. *Management Science, 14*, 159–182, 320–334, 486–502.

Heifetz, A., & Mongin, P. (2001). Probability logic for type spaces. *Games and Economic Behavior, 35*, 31–35.

Herings, P., & Vannetelbosch, V. (2000). The equivalence of the Dekel-Fudenberg iterative procedure and weakly perfect rationalizability. *Economic Theory, 15*(3), 677–687.

Kaneko, M. (2002). Epistemic logics and their game theoretic applications: Introduction. *Economic Theory, 19*, 7–62.

Pauly, M. (2002). A modal logic for coalitional power in games. *Journal of Logic and Computation, 12*, 149–166.

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

Rabinowicz, W. (1998). Grappling with the centipede: Defence of backward induction for BI-terminating games. *Economics and Philosophy, 14*, 95–126.

Samuelson, L. (1992). Dominated strategies and common knowledge. *Games and Economic Behavior, 4*, 284–313.

Stalnaker, R. (1996). Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy, 12*, 133–163. (Repr., with proofs, in C. Bicchieri, R. Jeffrey, & B. Skyrms (Eds.), *The logic of strategy*, New York: Oxford University Press, 1999).
van Benthem, J. (2001). Games in dynamic-epistemic logic. *Bulletin of Economic Research, 53*, 219–248.
van Benthem, J. (2003). Rational dynamics and epistemic logic in games. *International Journal of Game Theory*, forthcoming.