Economics and Philosophy, **20** (2004) 89–115 Copyright © Cambridge University Press
DOI: 10.1017/S0266267104001269

BELIEF REVISION IN GAMES OF PERFECT INFORMATION

Thorsten Clausing
University of Magdeburg

A syntactic formalism for the modeling of belief revision in perfect information games is presented that allows to define the rationality of a player's choice of moves relative to the beliefs he holds as his respective decision nodes have been reached. In this setting, true common belief in the structure of the game and rationality held before the start of the game does not imply that backward induction will be played. To derive backward induction, a "forward belief" condition is formulated in terms of revised rather than initial beliefs. Alternative notions of rationality as well as the use of knowledge instead of belief are also studied within this framework.

1. INTRODUCTION

The defining property of a game of perfect information is that moves are made sequentially, and whenever a player is to move, he is informed about which moves have already been made. Thus, at a decision node other than the root, the player has some additional information as compared to the situation before the start of the game. And even if he had already decided what he would play at this node if it were reached, when it actually has been reached he may use this additional information to reconsider his original decision and check whether the intended move still appears rational. Doubts about this may arise in particular in case the player did not expect the node where he presently finds himself to be reached, as this indicates that his opponents did not behave the way he expected them

I would like to thank Wlodek Rabinowicz and three anonymous referees for very helpful comments.

to do in the game so far, and this in turn might indicate that they will not behave as expected in the remaining part of the game. If the original decision about which move to choose was based on the player's beliefs about what opponents would choose at subsequent nodes, as one would expect from a rational player, he might well revise his initial plan in such a situation.

Of course, his opponents should have in turn considered the possibility that their move might lead to such a revision of his beliefs and plans when determining their optimal choice at preceding nodes.

Thus the behavior of a rational player at one of his decision nodes depends on the beliefs he holds at this node, i.e. beliefs revised upon the information that this node is reached, which may often differ from his initial beliefs. For a formal analysis of rational play one therefore needs a formalism that can handle belief revision. The subject of how rational agents change their beliefs in the face of additional information turns out to be a difficult one, however, and only comparatively weak rules for belief revision are generally accepted (see, e.g. [10]). Thus it is assumed that a rational agent will not change his beliefs if he receives a piece of information which he already believes to be true, and if he receives information that does not contradict his present beliefs, this will not cause him to doubt anything which he presently believes to be true. Furthermore, it is often claimed that in case the additional information does contradict his current beliefs, a rational agent should nevertheless try to maintain as much of his present beliefs as he can without having contradictory beliefs.

This paper tries to develop a formalism for the study of belief revision in games on the basis of a modal logic with binary belief operators. I do not start with initial beliefs but rather treat revised beliefs as the basic concept in the language that I will employ. However, with the standard assumptions from belief revision theory given in the preceding paragraph and the assumption that an agent's beliefs are always logically closed, initial beliefs can easily be derived from revised beliefs: They are exactly what is believed after receiving the additional information that some tautology (like $\phi \vee \neg\phi$) is true.

The formalism also features two other modal elements, namely common initial belief and subjunctive conditionals. There is common initial belief in something if every player initially believes it, every player initially believes that every player initially believes it, and so on. A subjunctive conditional is a statement of the form "if ϕ were true, then ψ would be true," which is to be interpreted as saying that in a hypothetical world where ϕ is true, but that is otherwise as similar to the actual world as possible, ψ is also true. Common initial belief is introduced because it is often claimed in the literature that common belief before the start of the game in rationality and the structure of the game implies that the players

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

behave according to the Nash equilibrium or backward induction solution concepts. Subjunctive conditionals will be needed to describe the structure of the game.

The remainder of the paper is organized as follows. The next section presents the syntactic formalism that I will work with, namely belief revision logic, in detail. Section 3 shows how the structure of a generic game of perfect information can be described in the language of belief revision logic. I will restrict the analysis to this class of games throughout. Section 4 describes rational behavior in terms of revised beliefs and presents results on sufficient conditions for backward induction play. A further discussion of my approach and a comparison to similar work by other authors is given in section 5, proofs are presented in section 6.

2. BELIEF REVISION LOGIC

I will define a propositional language $\mathcal{L}_\Gamma$ with respect to a given generic game of perfect information Γ. Its supply of primitive propositions PP consists of node formulas v and payoff formulas $\pi_i = x$ such that $v \in PP$ if and only if there is a node v in the tree of Γ and $\pi_i = x \in PP$ if and only if player i gets a payoff of x at some terminal node in Γ. The intended interpretation of v is "the move leading to node v will be made," and the intended interpretation of $\pi_i = x$ is "player i will receive a payoff of x." (I will not notationally distinguish between nodes and node formulas.)

Well-formed formulas of $\mathcal{L}_\Gamma$ are defined recursively as follows. Any primitive proposition is a well-formed formula, and if ϕ and ψ are well-formed formulas, so are $\neg\phi$, $\phi \Rightarrow \psi$, $B_i(\phi \mid \psi)$, $CB\phi$, and $\phi \hookrightarrow \psi$.

The first two of these formulas have their usual interpretation as negation and material implication. $B_i(\phi \mid \psi)$ is to be interpreted as a statement about the revised beliefs of player i, namely as "upon receiving the information that ψ is true, player i would come to believe that ϕ is true". $CB\phi$ is intended to stand for "there is common initial belief in ϕ," and $\phi \hookrightarrow \psi$ is a subjunctive conditional.

I will freely use standard abbreviations like $\phi \wedge \psi$ or $\phi \vee \psi$, and take $\otimes$ to denote the exclusive or, i.e. $\phi \otimes \psi :\Leftrightarrow (\phi \vee \psi) \wedge \neg(\phi \wedge \psi)$. I will also use the following extended version of the exclusive or:

$$\bigotimes_{i \in J} \phi_i :\Leftrightarrow \bigvee_{i \in J} \left(\phi_i \wedge \bigwedge_{j \in J \setminus \{i\}} \neg\phi_j \right)$$

For the conditional part of the logic, consider the following set of axioms:

$C0$ *all propositional tautologies*
$C1$ $\phi \hookrightarrow \phi$
$C2$ $((\phi \hookrightarrow \psi_1) \wedge (\phi \hookrightarrow \psi_2)) \Rightarrow (\phi \hookrightarrow (\psi_1 \wedge \psi_2))$
$C3$ $((\phi_1 \hookrightarrow \psi) \wedge (\phi_2 \hookrightarrow \psi)) \Rightarrow ((\phi_1 \vee \phi_2) \hookrightarrow \psi)$
$C4$ $((\phi \hookrightarrow \psi) \wedge (\psi \hookrightarrow \phi)) \Rightarrow ((\phi \hookrightarrow \sigma) \Rightarrow (\psi \hookrightarrow \sigma))$
$C5$ $((\phi_1 \hookrightarrow \phi_2) \wedge (\phi_1 \hookrightarrow \psi)) \Rightarrow ((\phi_1 \wedge \phi_2) \hookrightarrow \psi)$
$C6$ $\phi \Rightarrow ((\phi \hookrightarrow \psi) \Leftrightarrow \psi)$
$C7$ $(\phi \hookrightarrow \psi) \vee (\phi \hookrightarrow \neg\psi)$

In words, $C1$ simply says that any proposition ϕ conditionally implies itself, and $C2$ says that if ϕ conditionally implies both ψ_1 and ψ_2, it also conditionally implies their conjunction. Similarily, $C3$ states that if both ϕ_1 and ϕ_2 conditionally imply ψ, so does their disjunction. $C4$ can be interpreted to mean that if two propositions are conditionally equivalent in the sense of conditionally implying each other, anything conditionally implied by one is also conditionally implied by the other. $C5$ says that if ϕ_1 conditionally implies both ϕ_2 and ψ, then ϕ_1 and ϕ_2 together also conditionally imply ψ. $C6$ states that if ϕ is true, it conditionally implies ψ exactly if ψ is also true. In other words, this means that if ϕ is true in the actual world, then the world most similar to the actual world where ϕ is true is just the actual world.

$C7$ excludes the possibility that the truth value of ψ conditional on ϕ may not be determined. This amounts to saying that the world most similar to the actual one where ϕ is true is uniquely determined. In the philosophical literature, $C7$ appears to be more controversial than axioms $C0$–$C6$, which may be seen as standard properties of subjunctive conditionals. For a detailed discussion, the reader is referred to e.g. [15] or [16].

For the belief revision part of the logic, consider the following set of axioms, where I take $\top$ to stand for some fixed propositional tautology like $\phi \vee \neg\phi$ and $\bot$ refers to some fixed propositional contradiction like $\phi \wedge \neg\phi$. m denotes the number of players.

$B1$ $B_i(\phi \mid \phi)$
$B2$ $(B_i(\psi \mid \phi) \wedge B_i(\sigma \mid \phi)) \Rightarrow B_i(\psi \wedge \sigma \mid \phi)$
$B3$ $\neg B_i(\bot \mid \top)$
$B4$ $(B_i(\phi_2 \mid \phi_1) \wedge B_i(\psi \mid \phi_1)) \Rightarrow B_i(\psi \mid \phi_1 \wedge \phi_2)$
$B5$ $(B_i(\psi \mid \phi) \wedge B_i(\phi \mid \psi)) \Rightarrow (B_i(\sigma \mid \phi) \Rightarrow B_i(\sigma \mid \psi))$
$B6$ $\neg B_i(\neg\psi \mid \phi) \Rightarrow (B_i(\sigma \mid \phi \wedge \psi) \Leftrightarrow B_i(\psi \Rightarrow \sigma \mid \phi))$
$B7$ $CB\phi \Rightarrow \bigwedge_{i=1}^{m} B_i(CB\phi \wedge \phi \mid \top)$

Axiom $B1$ says that if a player receives additional information, he believes that this information is true. One may take this to mean that attention is restricted to additional information of such a kind that a rational player cannot doubt its correctness. Note that the information that one of his decision nodes has been reached in a game of perfect information must certainly be of this kind. $B2$ says that if a player comes to believe both ψ and σ, he also comes to believe their conjunction. Together with inference rule BR below, this implies that revised beliefs are logically closed.

$B3$ captures the assumption that initial beliefs are consistent. Thus attention is restricted to the case where players are not confused at the outset, even though they may become so later when they revise their beliefs. A possible way of extending the consistency assumption to revised beliefs will be discussed in section 5.1.

The meaning of $B4$ is that if upon learning that ϕ_1 is true, the player comes to believe that both ϕ_2 and ψ are true, then upon learning that both ϕ_1 and ϕ_2 are true, he still comes to believe that ψ is true. $B5$ states that if the information that ψ is true makes the player believe that ϕ is true and vice versa, then his beliefs revised by ψ are the same as those revised by ϕ.

To interpret $B6$, first replace ϕ by $\top$. Then the axiom says that upon receiving the information that some formula ϕ is true which he initially considered possible, the player comes to believe that exactly those formulas are true of which he initially believed that their negation is incompatible with ϕ. This means that he revises his beliefs by adding ϕ to his initial set of beliefs and then believing in the logical closure of this enlarged set. Thus he makes maximal use of the additional information in terms of drawing conclusions, but he restricts himself to such conclusions that are indeed logically implied by this information. An analogous interpretation applies for the given version of $B6$ with ϕ instead of $\top$. Finally, $B7$ makes sure that common initial belief in ϕ implies that every player initially believes that . . . every player initially believes that ϕ is true, as demanded by the interpretation of common initial belief given above.

The axioms $C0$–$C7$ and $B1$–$B7$ are completely context independent in the sense that they can be used to analyse revised beliefs about any unspecified subject. Consequently, without further axioms, any game specific assumptions would have to be made explicitly. This explicitness might be seen as an advantage of such a formalism. Nevertheless, one might wonder whether the specific context of games of perfect information should not impose some additional restrictions on the player's belief revision. Therefore I will also consider the following game specific axioms, where $i(v)$ denotes the player who is to move at decision node v and $M(v)$ refers to the set of immediate successors of node v.

$G1 \quad B_i(\phi \mid v) \Rightarrow (v \hookrightarrow B_i(\phi \mid v))$

$G2 \quad (v \hookrightarrow w) \Rightarrow B_{i(v)}(w \mid v)$ *for* $w \in M(v)$

Axiom $G1$ says that if a player would come to believe in ϕ after receiving the information that node v is reached, then he would revise

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

his beliefs in this way if v were reached. The axiom thus makes sure that the beliefs a player would come to hold if he were informed that v is reached and the beliefs he would hold if v were reached are consistent with each other. As by the definition of a game of perfect information a player is necessarily informed that v is reached if it is, this axiom appears to be a straightforward consequence of the intended interpretation of the revised belief operator.

Axiom $G2$ means that if a player were to choose the move leading to node w at node v, he would come to believe that he will make this move as v is reached. Thus the axiom says that players are certain and not mistaken about their own choices at a decision node once this node is reached. As Rabinowicz points out, this can be interpreted to mean that the beliefs under analysis are those held by a player after he has determined his choice, as opposed to the beliefs a player holds before making a decision. From a decision theoretic perspective, however, it might appear that the relevant beliefs are those on the basis of which a decision is made, i.e. those held before the decision is made. For a detailed discussion, see [19]. Note also that nothing in the definition of a game of perfect information implies that a player cannot be mistaken about the move he will make. Nevertheless, an assumption corresponding to $G2$ is made in almost all doxastic or epistemic analyses of games.

The following rules of inference will be employed:

MP	*From* ϕ *and* $\phi \Rightarrow \psi$ *infer* ψ
CR	*From* $\phi_1 \Rightarrow \phi_2$ *infer* $(\psi \hookrightarrow \phi_1) \Rightarrow (\psi \hookrightarrow \phi_2)$
BR	*From* $\phi_1 \Rightarrow \phi_2$ *infer* $B_i(\phi_1 \mid \psi) \Rightarrow B_i(\phi_2 \mid \psi)$
CBR	*From* $\phi \Rightarrow \bigwedge_{i=1}^{m} B_i(\phi \wedge \psi \mid \top)$ *infer* $\phi \Rightarrow CB\psi$

To give a semantics to this logic, consider belief revision models $(\Omega, f_0, f_1, \ldots, f_m, p)$. Ω is a non-empty set of states, $f_i : \Omega \times \mathcal{L}_\Gamma \to 2^\Omega$ is a state selection function and $p : \Omega \times PP \to \{true,\ false\}$ a valuation function assigning truth values to primitive propositions. For an intuitive interpretation, $f_0(\omega, \phi)$ can be understood to be the set of states where ϕ is true which are most similar to state ω. For $i = 1, \ldots, m$, $f_i(\omega, \phi)$ can be understood to be the set of states which player i would consider possible at state ω if he received the information ϕ.

Let $[\phi]$ stand for $\{\omega \in \Omega \mid \omega \models \phi\}$, where $\omega \models \phi$ as usual denotes that ϕ is true at state ω. Consider the following restrictions on the state selection functions:

$R1$	$f_i(\omega, \phi) \subset [\phi]$	$i \in \{0, 1, \ldots, m\}$
$R2$	$f_i(\omega, \phi_1 \wedge \phi_2) \subset f_i(\omega, \phi_1)$	*if* $f_i(\omega, \phi_1) \subset [\phi_2], i \in \{0, 1, \ldots, m\}$
$R3$	$f_0(\omega, \phi_1 \vee \phi_2) \subset f_0(\omega, \phi_1) \cup f_0(\omega, \phi_2)$	

$R4$ $\quad f_i(\omega,\phi) = f_i(\omega,\psi)$ $\quad$ *if* $f_i(\omega,\phi) \subset [\psi]$ *and* $f_i(\omega,\psi) \subset [\phi], i \in \{0, 1, \ldots, m\}$

$R5$ $\quad f_0(\omega,\phi) = \{\omega\}$ $\quad$ *if* $\omega \in [\phi]$

$R6$ $\quad f_i(\omega,\top) \neq \emptyset$ $\quad$ $i \in \{1, \ldots, m\}$

$R7$ $\quad f_i(\omega,\phi \wedge \psi) = f_i(\omega,\phi) \cap [\psi]$ $\quad$ *if* $f_i(\omega,\phi) \cap [\psi] \neq \emptyset, i \in \{1, \ldots, m\}$

$R8$ $\quad f_0(\omega,\phi)$ *is empty or a singleton*

$R9$ $\quad f_i(\omega',v) \subset f_i(\omega,v)$ $\quad$ *if* $\omega' \in f_0(\omega,v), i \in \{1, \ldots, m\}$

$R10$ $\quad f_{i(v)}(\omega,v) \subset [w]$ $\quad$ *if* $f_0(\omega,v) \subset [w], w \in M(v)$

Let $f(\omega) := \{\omega' \in \Omega \mid \exists\omega_1, \ldots, \omega_n, \exists j_1, \ldots, j_{n-1} : \omega_1 = \omega, \omega_n = \omega', \omega_{i+1} \in f_{j_i}(\omega_i,\top), j_i \in \{1, \ldots, m\}\}$. Truth values for well-formed formulas are defined recursively by the following rules:

$\omega \models \phi$ $\quad$ *if* $\phi \in PP$ *and* $p(\omega)(\phi) = true$

$\omega \models \neg\phi$ $\quad$ *if not* $\omega \models \phi$

$\omega \models \phi \Rightarrow \psi$ $\quad$ *if* $\omega \models \psi$ *or not* $\omega \models \phi$

$\omega \models B_i(\psi \mid \phi)$ $\quad$ *if* $\omega' \models \psi \ \forall\omega' \in f_i(\omega,\phi)$

$\omega \models \phi \hookrightarrow \psi$ $\quad$ *if* $\omega' \models \psi \ \forall\omega' \in f_0(\omega,\phi)$

$\omega \models CB\phi$ $\quad$ *if* $\omega' \models \phi \ \forall\omega' \in f(\omega)$

In the following, I will use a small and a large version of belief revision logic. Let S be a set of axioms and inference rules consisting of $C0$–$C6$, $B1$–$B7$, MP, CR, BR and CBR. S is intended to yield a basic logic without any game specific assumptions, so that any result proven on the basis of this logic has a high degree of generality. Alternatively, let L include $C0$–$C7$, $B1$–$B7$, $G1$–$G2$, MP, CR, BR and CBR. L thus also comprises game specific axioms and assumptions like $C7$ and $G2$ which are frequently employed, but often criticized in the literature. I will use this stronger version of the logic for unprovability results to show that the unprovability does not result from the absence of these controversial assumptions.

Using techniques for conditional and epistemic logic from e.g. [11] and [14], one can show that S is a sound and complete axiomatization of the class of all belief revision models satisfying restrictions $R1$–$R7$. L yields a sound and complete axiomatization of the class of all belief revision models satsifying restrictions $R1$–$R10$.

To denote provability, I will write $X \vdash \phi$ to say that ϕ can be proven from the axioms and inference rules in X. Correspondingly, $X \not\vdash \phi$ denotes that ϕ is not provable from the axioms and inference rules in X, i.e. $\neg\phi$ is consistent in the logical system X.

3. STRUCTURE OF THE GAME

I will now define a formula that describes the structure of a given game of perfect information. Such a formula should contain the information that

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

one can derive from the game tree. Therefore it should say that every move to a terminal node conditionally implies that every player gets a certain payoff associated with this node, that every move to a decision node conditionally implies that exactly one of the direct successors of this node will be reached, that every move to a node conditionally implies that all predecessors of this node will be reached, and that the root of the game tree will be reached. The game from Figure 1 can thus be represented by the following formula:

$$\begin{aligned}&(e \hookrightarrow (\pi_1 = 1 \wedge \pi_2 = 0 \wedge a)) \wedge (f \hookrightarrow (\pi_1 = 0 \wedge \pi_2 = 2 \wedge a \wedge b))\\ &\wedge (g \hookrightarrow (\pi_1 = 3 \wedge \pi_2 = 1 \wedge a \wedge b \wedge c))\\ &\wedge (d \hookrightarrow (\pi_1 = 2 \wedge \pi_2 = 3 \wedge a \wedge b \wedge c))\\ &\wedge (a \hookrightarrow (e \otimes b)) \wedge (b \hookrightarrow ((c \otimes f) \wedge a))\\ &\wedge (c \hookrightarrow ((d \otimes g) \wedge a \wedge b)) \wedge a\end{aligned}$$

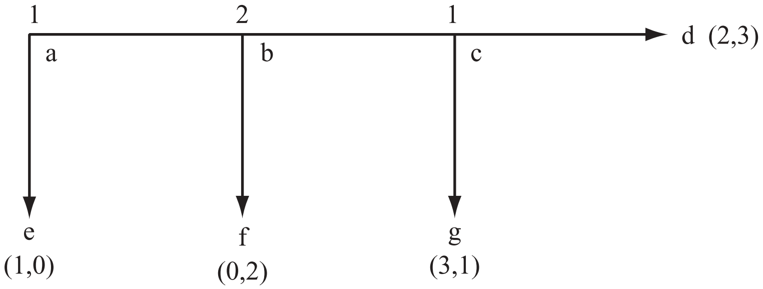

FIGURE 1. A short centipede.

For a general formulation, let T denote the set of terminal nodes of the given generic game of perfect information Γ, V the set of decision nodes of Γ, $P(v)$ the set of predecessors of v, $\pi_i(t)$ player i's payoff at the terminal node t and, as above, $M(v)$ the set of possible moves at v; r stands for the root of the game tree. Now the structure of Γ can be described by the following formula S_Γ:

$$\begin{aligned}S_\Gamma :\Leftrightarrow r \wedge \bigwedge_{t\in T}\left(t \hookrightarrow \left(\bigwedge_{i=1}^{m} \pi_i = \pi_i(t) \wedge \bigwedge_{v\in P(t)} v\right)\right.\\ \wedge \bigwedge_{v\in V}\left(v \hookrightarrow \left(\bigwedge_{w\in P(v)} w \wedge \bigotimes_{w\in M(v)} w\right)\right)\end{aligned}$$

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

For a formula describing backward induction play, define BIP to be the conjunction of all node formulas standing for nodes on the backward induction path of Γ. For a formula describing backward induction in the stronger sense of backward induction strategies, let v^* refer to the backward induction move at node v. A formula BI can then be defined as follows:

$$BI :\Leftrightarrow r \wedge \bigwedge_{v \in V} (v \hookrightarrow v^*)$$

In this manner, one could take a formula of the kind $\bigwedge_{v \in V_i} (v \hookrightarrow v')$, where $v' \in M(v)$ and V_i denotes the set of decision nodes owned by player i, to describe a strategy for player i. Note that $C7$ then makes sure that indeed each player has a strategy, which might serve as a justification for including this axiom in the system L.

4. RATIONALITY

I will refer to a player as rational if he always chooses moves that are optimal relative to his beliefs. As argued in the introduction, the relevant beliefs here are those held at the node where the given move has to be made, i.e. beliefs revised on the information that this node is reached. Thus assume that upon being informed that node v is reached, a rational player came to believe that for some given move at this node, there is an alternative move which would give him a higher payoff. Then if node v were indeed reached (and the player consequently informed of this), he would not play the given move there. This consequence of rationality can naturally be expressed by the following material implication with $x > y$ and $u, w \in M(v)$:

$$(1) \quad B_{i(v)}\big((u \hookrightarrow \pi_{i(v)} = x) \wedge (w \hookrightarrow \pi_{i(v)} = y)\big| v\big) \Rightarrow (v \hookrightarrow \neg w)$$

This formula can be thought of as a scheme. Replacing the payoff formulas by other payoff formulas (such that the first one refers to a higher payoff than the second one) and the node formulas by other node formulas from $M(v)$ also yields consequences of rationality. The same is true if one replaces the payoff formulas by disjunctions of different payoff formulas such that all payoffs referred to in the first disjunction are higher than those referred to in the second one. Due to the finiteness of PP, there are only finitely many possible replacements of this kind. Therefore defining R_v as the conjunction of all of these possible replacements yields a well-formed formula of $\mathcal{L}_\Gamma$ standing for "rationality at node v." A formula standing for rationality of player i can be defined as $R_i :\Leftrightarrow \bigwedge_{v \in V_i} R_v$, and a formula standing for rationality as $R :\Leftrightarrow \bigwedge_{v \in V} R_v$.

With these definitions, one can turn to a question that has received much attention in the literature, namely whether common initial belief in rationality and the structure of the game implies that rational players behave according to the backward induction solution concept. The answer is no.

Proposition 1. *True common initial belief in rationality and the structure of the game does not imply backward induction play.*

$$L \nvdash (CB(R \wedge S_\Gamma) \wedge R \wedge S_\Gamma) \Rightarrow BIP$$

Furthermore, for any perfect information game Γ, true initial common belief in rationality and the structure of the game is a consistent assumption.

$$L \nvdash \neg(CB(R \wedge S_\Gamma) \wedge R \wedge S_\Gamma)$$

Let me note that the second part of proposition 1 is a general statement about all games of perfect information and thus does not follow from the first part, which says that there is a game in which BIP is not implied by the stated condition.

For an intuitive explanation of the negative result on backward induction, consider the game from figure 1. It is compatible with common initial belief in rationality and the structure of the game that node f will be reached. Note that this is not even a Nash equilibrium outcome.

Let there be correct beliefs about the structure of the game throughout. Assume that before the game starts, player 1 thinks that player 2 initially believes that 1 is rational, will play down at node a and would play down at node c. However, let 1 expect that upon being informed that 1 actually played across at node a, 2 would come to believe that 1 will also play across at c, and therefore rationally decide to play across at node b. In this situation, the rational choice for 1 is, of course, to play across at node a and to plan to play down at node c. Assume furthermore that player 2 is not deceived by this manoeuver and believes at node b that 1 would play down at node c. He therefore rationally plays down at node b. Both players act rationally and initially believe their opponent to be rational; indeed, they maintain this belief throughout the game, but as player 1 mistakenly expects player 2 to abandon this belief at node b, a non-Nash outcome arises. Nevertheless, the players' initial beliefs may well be commonly held. Note also that the players hold correct beliefs about their own choices throughout. A belief revision model corresponding to this situation, which seems to capture a criticism of backward induction advanced e.g. in [17], is constructed in the proof of proposition 1.

Thus just considering initial beliefs is not enough for finding a sufficient condition for behavior consistent with standard solution concepts. One needs a concept that refers to beliefs held at later nodes in the game tree as well. I will therefore now develop an analogue of the concept of forward knowledge used e.g. in [3] and [19] on the basis of belief instead of knowledge. The idea of forward belief in ϕ is the following: After

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

having been informed that one of his decision nodes has been reached, each player believes ϕ, and he believes that at each subsequent decision node, the players who have to make a move there will also come to believe in ϕ, and they will furthermore come to believe that at all subsequent decision nodes the players who will have to move there will also come to believe in ϕ ... and so on. To formalize this idea in a recursive way, let forward belief from node v to node v in ϕ be equivalent to ϕ:

$$B_{v\to v}\phi :\Leftrightarrow \phi$$

Now let u be on the path leading to v and assume that forward belief from node w to node v has already been defined, where w is the immediate successor of u on the path to v. Then forward belief from node u to node v is defined as follows:

$$B_{u\to v}\phi :\Leftrightarrow B_{i(u)}(B_{w\to v}\phi \mid u) \wedge B_{w\to v}\phi$$

With this concept, the following result on sufficient conditions for backward induction can be established.

Proposition 2. *Forward belief from the root to all decision nodes v in rationality at v and in true revised belief about the structure of the game at v implies backward induction.*

$$S \vdash \left(S_\Gamma \wedge \bigwedge_{v\in V} B_{r\to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big)\right) \Rightarrow BI$$

Proposition 3. *Forward belief from the root to all decision nodes v in rationality at v and in true revised belief about the structure of the game at v is a consistent assumption.*

$$L \nvdash \neg \left(S_\Gamma \wedge \bigwedge_{v\in V} B_{r\to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big)\right)$$

5. DISCUSSION

5.1. The AGM theory of belief revision

The best-known theory of belief revision is the so-called AGM theory (see [10]). It treats revised beliefs of a single agent and is based formally on a set of states B'. Initial beliefs are represented by a non-empty subset $B \subset B'$, i.e. the agent initially believes that exactly those events are true which are supersets of B. Revised beliefs are captured by a function $B() : 2^{B'} \to 2^{B'}$ such that $B(\Phi)$ represents the agent's beliefs after having received the information that the event Φ is true. This function is assumed to have the following properties:

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

$$\begin{array}{llll} P1 & B(\Phi) \subset \Phi & & \\ P2 & \textit{If}\ \ B \cap \Phi \neq \emptyset & \textit{then} & B(\Phi) = B \cap \Phi \\ P3 & \textit{If}\ \ \Phi \neq \emptyset & \textit{then} & B(\Phi) \neq \emptyset \\ P4 & \textit{If}\ \ B(\Phi) \cap \Psi \neq \emptyset & \textit{then} & B(\Phi \cap \Psi) = B(\Phi) \cap \Psi \end{array}$$

Comparing this formalism to the semantics of belief revision logic as developed in section 2, one sees that B corresponds to $f_i(\omega, \top)$ and $B(\Phi)$ to $f_i(\omega, \phi)$ (where ϕ is a proposition corresponding to the event Φ and ω some fixed state). Therefore the non-emptiness of B directly corresponds to restriction $R6$ (and thus axiom $B3$), property $P1$ corresponds to $R1$ and $B1$, and $P2$ and $P4$ correspond to $R7$ and $B6$ (to see this for the case of $P2$, set $\phi = \top$ in $R7$ and $B6$). Also, one obviously finds both that $B(\Phi) \subset \Psi$ and $B(\Phi) \subset \Sigma$ implies $B(\Phi) \subset \Psi \cap \Sigma$ and that $f_i(\omega, \phi) \subset [\psi]$ and $f_i(\omega, \phi) \subset [\sigma]$ implies $f_i(\omega, \phi) \subset [\psi \wedge \sigma]$, which corresponds to axiom $B2$. Furthermore, it follows from $P3$ and $P4$ together that $B(\Phi) \subset \Psi$ and $B(\Psi) \subset \Phi$ imply $B(\Phi) = B(\Psi)$, which corresponds to $R4$ and $B5$. Finally, assume $B(\Phi) \subset \Psi$. One either has $B(\Phi) \cap \Psi \neq \emptyset$ or $B(\Phi) = \emptyset$. In the first case, $P4$ yields $B(\Phi \cap \Psi) = B(\Phi) \cap \Psi$; in the second case, $P3$ implies $\Phi = \emptyset$ and $P1$ $B(\Phi \cap \Psi) = \emptyset$. Therefore in either case, $B(\Phi \cap \Psi) \subset B(\Phi)$, which corresponds to $R2$ and $B4$.

Note that repeated belief revision can be captured in the AGM framework by defining functions $X() : 2^{B'} \rightarrow 2^{B'}$ for all $X \subset B'$ instead of just for some fixed set B. $P4$ then says that if the event Ψ is not excluded after receiving the information that the event Φ is true, revising first by Φ and then by Ψ gives the same final beliefs as a revision by $\Phi \cap \Psi$. Even though repeated belief revision cannot be directly represented in the formalism from section 2, the correspondence between $P4$ and $R7$ allows to give an alternative interpretation to axiom $B6$, namely as an assumption about when a repeated belief revision first by ϕ and then by ψ can be represented as a revision by $\phi \wedge \psi$.

In summary, all non-game-specific one-person belief revision axioms from section 2 have a counterpart in the AGM theory. On the other hand, the semantics in section 2 do not contain an analogue to $P3$, which is a consistency requirement for revised beliefs. However, such an analogue can easily be formulated as an extension of restriction $R6$:

$$R6' \quad f_i(\omega, \phi) \neq \emptyset \quad \textit{if} \quad [\phi] \neq \emptyset,\ i \in \{1, \ldots, m\}$$

An axiom system that is sound and complete for the class of all belief revision models satisfying $R6'$ can be obtained by adding the following axiom $B8$ to the axioms of belief revision logic, as shown in [11]. Here $\Box_i \phi$ stands for $B_i(\bot |\, \neg\phi)$ and $\Diamond_i \phi$ for $\neg\Box_i \neg\phi$ with $i \in \{1, \ldots, m\}$.

$$\begin{array}{lll} B8 & (a) & \Box_i \phi \Rightarrow \left(\phi \wedge \bigwedge_{j=1}^{m} B_j(\Box_i \phi \mid \psi_j) \wedge (\psi_0 \hookrightarrow \Box_i \phi)\right) \\ & (b) & \Diamond_i \phi \Rightarrow \left(\bigwedge_{j=1}^{m} B_j(\Diamond_i \phi \mid \psi_j) \wedge (\psi_0 \hookrightarrow \Diamond_i \phi)\right) \end{array}$$

I have not included this axiom in the system of section 2 as it does not seem to have an intuitive interpretation. Like $P3$, however, it does imply a consistency assumption for revised beliefs, as one can show the following: $S, B8 \vdash \neg CB(\neg\phi) \Rightarrow \neg B_i(\perp|\phi)$.

Note that even though $B8$ is not contained in S and L, all models constructed in the proofs in section 6 satisfy restriction $R6'$. Consequently, all unprovability and consistency results in this paper are valid for a notion of revised belief that corresponds to the AGM notion. Furthermore, the possibility of inconsistent revised beliefs does not play any role for these results.

5.2. Knowledge

Much of the literature on solution concepts for extensive form games is based on knowledge instead of belief (for an overview, see e.g. [9]). The conceptual difference between knowledge and belief is that while beliefs may be mistaken, what is known is necessarily true. Thus for the operator B_i to denote knowledge, it must fulfil the veridicality condition $B_i(\phi \mid \top) \Rightarrow \phi$. One may go on to argue that in the context of revised knowledge, veridicality should not only hold for what is initially known. One way of extending veridicality in this context is to demand that if a player comes to know ψ upon learning ϕ, then this knowledge revision should be correct in the sense that ψ would indeed be true if ϕ were. As I am about to state an unprovability result, I will use this strong notion of revised knowledge, which is captured by the following axiom:

$B9 \quad B_i(\psi \mid \phi) \Rightarrow (\phi \hookrightarrow \psi)$

The semantic counterpart to $B9$ is the restriction $R11$:

$R11 \quad f_0(\omega, \phi) \subset f_i(\omega, \phi) \quad i \in \{1, \ldots, m\}$

While revising one's knowledge on information consistent with what one already knows does not pose any conceptual problems, one may argue that it does not make sense to consider knowledge revised on information that contradicts what is already known. As what is known must be true, a situation in which a player actually receives 'knowledge contravening' information is not possible. Nevertheless, even if a player knows that a given piece of information cannot be received, one may still ask what he or his opponents would come to know in a situation where it is received, as Stalnaker argues convincingly in [22]. Indeed, the description of rational decision making in perfect information games given in the introduction demands that this question be asked, and this description would seem to be as valid if one considers knowledge as it is when one talks about beliefs.

I formulate the main line of argumentation of this paper in terms of belief rather than knowledge because I feel that this is the more appropriate

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

concept in the given context. Note that I use full beliefs, i.e. a player believes something if and only if he is completely convinced that it is true. If a player bases his decisions on deliberation, as is assumed here, then what counts is how he perceives the world to be. If he is completely convinced that some move would give him a higher payoff than an alternative one, he will decide not to play the alternative one independently of whether his conviction is right or not. In particular, there is no reason to assume that people can only be completely convinced of something if it is true.

Further reasons against the use of knowledge can be found in [6].

However, even if one were not to follow this argumentation and to replace true initial common belief by common initial knowledge in the formulation of proposition 1, backward induction play would still not be implied.

Remark 1. *Common initial knowledge of rationality and the structure of the game does not imply backward induction play.*

$$L, B9 \nvdash CB(R \wedge S_\Gamma) \Rightarrow BIP$$

Furthermore, for any perfect information game Γ*, common initial knowledge of rationality and the structure of the game is a consistent assumption.*

$$L, B9 \nvdash \neg CB(R \wedge S_\Gamma)$$

5.3. Ex ante rationality

In [1], Aumann uses a notion of ex ante rationality defined in terms of the players' initial information. Even though this notion is formally based on the choice of strategies, one can easily formulate a move-based version of ex ante rationality in $\mathcal{L}_\Gamma$ with the help of the following analogue of (1) (with $x > y$ and $u, w \in M(v)$):

(2) $B_{i(v)}\big((u \hookrightarrow \pi_{i(v)} = x) \wedge (w \hookrightarrow \pi_{i(v)} = y) \mid \top\big) \Rightarrow (v \hookrightarrow \neg w)$

In the same way as R was defined from (1) above, let the formula $R^{ex\,ante}$ be defined from (2). One can then proceed to derive the following result.

Remark 2. *True common initial belief in ex ante rationality and the structure of the game implies backward induction.*

$$S \vdash (CB(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante} \wedge S_\Gamma) \Rightarrow BI$$

Note that even though the analysis in [1] is set in terms of knowledge, axiom $B9$ is not needed to prove this result.

However, Aumann himself writes that a rationality concept based on the information the players have as their respective decision node has been reached is more natural than one based exclusively on their initial information. But he claims that as the information received at such a node

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

is additional, a player who is rational relative to what he knows at his decision node must also be rational relative to his initial information, because if the initial information allowed to deduce that there is a better move than the one he decides to play, he would a fortiori know this later on. This reasoning is certainly correct with respect to additional information that a player actually receives, but it would not necessarily be valid for additional information that he would receive if a node were reached that is not actually reached.

Nevertheless, in [2], Aumann presents a model in which rationality defined in terms of revised knowledge (which he calls ex post rationality) indeed implies ex ante rationality. This is, however, achieved by redefining the term "knowledge at node v." Aumann interprets this to mean "knowledge revised upon the information *whether* node v will be reached," i.e. if v is reached, the knowledge is revised on v, and if v is not reached, it is revised on $\neg v$. He maintains that this refers to the moment when a player decides what to play at v. One then has the following formula from which to define the notion of ex post rationality:

$$\begin{aligned}
&\big(\big(v \Rightarrow B_{i(v)}\big(\big(u \hookrightarrow \pi_{i(v)}=x\big) \wedge \big(w \hookrightarrow \pi_{i(v)}=y\big) \mid v\big)\big)\\
&\wedge \big(\neg v \Rightarrow B_{i(v)}\big(\big(u \hookrightarrow \pi_{i(v)}=x\big)\\
&\wedge \big(w \hookrightarrow \pi_{i(v)}=y\big) \mid \neg v\big)\big)\big) \Rightarrow (v \hookrightarrow \neg w)
\end{aligned} \tag{3}$$

Again, let the formula $R^{ex\ post}$ be defined from (3) in the same way as R was defined from (1).

Remark 3. *With knowledge instead of belief, ex post rationality implies ex ante rationality.*

$$S, B9 \vdash R^{ex\,post} \Rightarrow R^{ex\,ante}$$

Furthermore, common initial knowledge of ex post rationality and the structure of the game is a consistent assumption.

$$L, B9 \nvdash \neg C\,B(S_\Gamma \wedge R^{ex\,post})$$

Together with remark 2, this means that common initial knowledge of ex post rationality and the structure of the game implies backward induction. However, this result can only be derived by abandoning the view advocated in the introduction that the relevant knowledge for the choice of a rational player at an unreached node v is what would be known if v were reached and instead considering only knowledge that the players actually have.

5.4. BI terminating games

A natural question to ask if common initial knowledge of rationality and the structure of the game does not imply backward induction for the class of all generic games of perfect information is whether there is a subclass of games for which it does. To answer this question, consider the class of backward induction terminating games introduced by Rabinowicz. A generic game of perfect information belongs to this class if at all decision nodes the backward induction move terminates the game. Formally, one thus has $v^* \in T$ for all $v \in V$. Let me use Θ instead of Γ to denote such a game. In [19], Rabinowicz shows that for these games a comparatively weak version of forward knowledge that refers only to revised knowledge at nodes that are actually reached can be used to formulate a sufficient condition for backward induction play.

The centipede is the best-known example of a backward induction terminating game. It is probably also the game where backward induction reasoning has been most fervently attacked. As the example in section 4 shows, true common belief in rationality and the structure of the game does not suffice to bring about backward induction play even if players necessarily hold correct beliefs about their own choices. However, in contrast to the general case, for backward induction terminating games replacing belief by knowledge changes this result.

Remark 4. *With knowledge of own choices, common initial knowledge of rationality and the structure of the game implies backward induction play in backward induction terminating games.*

$$S, B9, G2 \vdash CB(R \wedge S_\Theta) \Rightarrow BIP$$

As can be seen from the proof of remark 4, this result continues to hold with the weaker version of knowledge where $B9$ is replaced by $B_i(\phi \mid \top) \Rightarrow \phi$.

5.5. Further related literature

In writing this paper, I was strongly influenced by Stalnaker's analysis in [22]. The main difference between his approach and mine is that his formalism is based on strategic form representations of extensive games. Stalnaker studies different belief revision policies of the players, i.e. restrictions on how they revise their beliefs if confronted with information contradicting some of their previously held beliefs. He presents such policies that, if commonly adopted, would make true common initial belief a sufficient condition for backward (and forward) induction, but argues that there are no reasons to assume that rational players adopt or should adopt these policies. A similar line of argumentation is also put forward in [21].

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

In [20], Samet presents a condition for backward induction that looks very similar to the forward belief condition in proposition 2. His analysis is, however, not based on revised beliefs but on the notion of hypothetical knowledge. Instead of what a player would come to know if he learned that some event is true, this refers to the initial knowledge (he thinks) he would have if he did not initially know that the event is not true. Halpern shows in [12] how this notion can be captured in a setting with conditionals and unary initial knowledge operators.

While Samet's notion treats moves as the object of choice as the notion defined in section 4, it is much weaker than the latter one because it only allows to say something about the player's actual moves, not the ones he would make if actually unreached nodes were reached. This corresponds to replacing $(v \hookrightarrow \neg w)$ by $\neg w$ in (1). Consequently the non-backward induction result presented in [20] is weaker than proposition 1.

In [13] Halpern compares the contradictory results on common knowledge of rationality and backward induction in [1] and [22] in a framework with state selection functions. He concludes that the contradiction stems from a different interpretation of counterfactual conditionals in the two papers. He notes that the difference can also be understood in terms of belief revision, the possibility of which is taken into account in [22], but not in [1]. However, Halpern's formalism does not treat revised beliefs explicitly.

A different strand of the literature considers probabilistic beliefs using type space models. The type of a player specifies his strategy, his probability distribution over the possible types of his opponents, and how this distribution changes/would change as he learns which nodes of the game tree have been reached. If the player is not surprised by this information, he updates his beliefs according to Bayes' rule. Note that axiom $B6$ can be seen as a qualitative version of this rule.

In [5], Ben-Porath uses such a setting to show that common certainty of rationality (which may be seen as the equivalent in his formalism of true common initial belief in rationality) does not imply Nash outcomes. Rather the set of possible outcomes in his model if common certainty of rationality obtains is characterized by the Dekel-Fudenberg procedure, i.e. one round of deletion of weakly dominated strategies and iterated deletion of strictly dominated strategies. Battigalli and Siniscalchi extend this analysis in [4] by introducing the notion of strong belief. Something is believed strongly if the belief is not given up after new information has been received as long as this information does not directly contradict this belief. On this basis a notion of higher order correct strong belief can be formulated such that correct strong belief of rationality of sufficiently high order implies backward (as well as forward) induction.

One way to interpret this correct strong belief condition is that it ensures that forward belief as in proposition 2 obtains. Thus the results

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

on backward induction derived in state space models and in the present paper largely coincide, which confirms the intuition expressed in [8] that "simply as a theoretical matter, probabilities play an inessential role in games" of perfect information.

All of the aforementioned papers use semantic formalisms and do not treat belief in or knowledge of the structure of the game. The first attempt to employ a formal (propositional) logic to describe the structure of the game was undertaken by Bonanno in [7]. This has been elaborated upon by Vilks in [23]. A recent addition to the syntactic literature with conditionals and time indexed epistemic and doxastic operators is presented by Priest in [18]. His focus is on the surprise test paradox and the centipede game, and for the latter he states a sufficient condition for backward induction in terms of knowledge of rationality at sucessive points of time that can be interpreted as an analogue of the forward belief condition from section 4 for this particular game. Priest shows that this condition is not implied by knowledge of rationality at the beginning of the game by any valid principle of persistence of knowledge.

6. PROOFS

I will make use of the following lemma.

Lemma 1. *Let Ω be finite. For all $\omega \in \Omega$ and $i = 0, 1, \ldots, m$, let there be injective functions $r_i^\omega : \Omega \to N_{|\Omega|}$ such that $r_0^\omega(\omega) = 1$. Define f_i as follows:*

$$f_i(\omega, \phi) = \left\{\omega' \mid r_i^\omega(\omega') = \min\left\{j \mid \exists\omega'' \in [\phi],\, j = r_i^\omega(\omega'')\right\}\right\}$$

(a) The state selection functions thus defined satisfy restrictions $R1$–$R8$ and $R6'$.

(b) If one has $r_0^\omega = r_i^\omega$ for all $\omega \in \Omega$ and $i = 1, \ldots, m$, the selection functions satisfy restrictions $R9$–$R11$.

Proof of lemma 1. (a) $R1$, $R6$ and $R6'$ are obviously satisfied. $R5$ follows from $r_0^\omega(\omega) = 1$ and $R8$ is implied by the injectiveness of r_i^ω. Furthermore, $R6'$ and $R8$ together imply $R4$.

For $R7$, assume $\omega' \in f_i(\omega, \phi) \cap [\psi]$, which means $\omega' \models \phi \wedge \psi$. Because of $\min\{j \mid \exists\omega'' \in [\phi], j = r_i^\omega(\omega'')\} \leq \min\{j \mid \exists\omega'' \in [\phi \wedge \psi], j = r_i^\omega(\omega'')\}$, this gives $\omega' \in f_i(\omega, \phi \wedge \psi)$.

Now assume $\omega' \in f_i(\omega, \phi \wedge \psi)$. Thus $\omega' \in [\psi]$. If $\omega' \notin f_i(\omega, \phi)$, then there is a state ω'' such that $r_i^\omega(\omega'') < r_i^\omega(\omega')$, $\omega'' \models \phi$, $\omega'' \not\models \psi$ and therefore $f_i(\omega, \phi) \cap [\psi] = \emptyset$.

For $R2$, let $f_i(\omega, \phi_1) \subset [\phi_2]$. Thus $\omega' \models \phi_2$ for ω' such that $r_i^\omega(\omega') = \min\{j \mid \exists\omega'' \in [\phi_1], j = r_i^\omega(\omega'')\}$ and thus $\min\{j \mid \exists\omega'' \in [\phi_1 \wedge \phi_2], j = r_i^\omega(\omega'')\} = \min\{j \mid \exists\omega'' \in [\phi_1], j = r_i^\omega(\omega'')\}$.

$R3$ follows from $\min\{j \mid \exists\omega'' \in [\phi_1 \vee \phi_2], j = r_i^\omega(\omega'')\} = \min\{\min\{j \mid \exists\omega'' \in [\phi_1], j = r_i^\omega(\omega'')\}, \min\{j \mid \exists\omega'' \in [\phi_2], j = r_i^\omega(\omega'')\}\}$.

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

(b) $R10$ and $R11$ are obvious. For $R9$, $\omega' \in f_0(\omega, v)$ implies $\omega' \models v$, and thus because of $r_i^{\omega'}(\omega') = 1$ and the injectiveness of $r_i^{\omega'}$ $f_i(\omega', v) = \{\omega'\} \subset f_0(\omega, v) = f_i(\omega, v)$. ■

Proof of proposition 1. To show that $(CB(R \wedge S_\Gamma) \wedge R \wedge S_\Gamma) \Rightarrow BIP$ is not provable, I will construct a belief revision model with a state where this formula is false. To this end, consider a model $(\Omega, f_0, f_1, f_2, p)$ based on the game in Figure 1. Let $\Omega = T$ and define $p(t)(v) =$ *true* exactly if v stands for a node on the path leading to t and $p(t)(\pi_i = x) =$*true* exactly if at t, player i gets a payoff of x. Furthermore, let $r_0^d(d) = 1, r_0^d(g) = 2, r_0^d(f) = 3, r_0^d(e) = 4; r_0^e(e) = 1, r_0^e(f) = 2, r_0^e(g) = 3, r_0^e(d) = 4; r_0^f(f) = 1, r_0^f(g) = 2, r_0^f(e) = 3, r_0^f(d) = 4$ and $r_0^g(g) = 1, r_0^g(f) = 2, r_0^g(e) = 3, r_0^g(d) = 4$.

For $\omega \in \{d, f, g\}$ let $r_1^\omega(g) = 1, r_1^\omega(f) = 2, r_1^\omega(e) = 3$ and $r_1^\omega(d) = 4$, and $r_1^e(e) = 1, r_1^e(g) = 2, r_1^e(f) = 3, r_1^e(d) = 4$.

For $\omega \in \{d, e, g\}$ let $r_2^\omega(e) = 1, r_2^\omega(f) = 2, r_2^\omega(d) = 3$ and $r_2^\omega(g) = 4$, and $r_2^f(f) = 1, r_2^f(d) = 2, r_2^f(g) = 3, r_2^f(e) = 4$. Now let the state selection functions be defined as in lemma 1.

One can check that these selection functions satisfy resriction $R9$. Obviously $R9$ cannot be violated in case $f(\omega, v) = \{\omega\}$. Furthermore, the restriction must be fullfilled for all terminal nodes t in the given model because of $f_i(\omega, t) = \{t\}$ for all $\omega \in \Omega$ and $i \in \{1, 2\}$. Thus it only remains to check the cases $f_0(f, c) = \{g\}$, $f_0(e, b) = \{f\}$ and $f_0(e, c) = \{g\}$. In the first case, one finds $f_1(g, c) = \{g\} = f_1(f, c)$ and $f_2(g, c) = \{d\} = f_2(f, c)$, in the second $f_1(f, b) = \{g\} = f_1(e, b)$ and $f_2(f, b) = \{f\} = f_2(e, b)$, and in the third $f_1(g, c) = \{g\} = f_1(e, c)$ and $f_2(g, c) = \{d\} = f_2(e, c)$. One can also easily see that the model respects restriction $R10$.

Obviously, $f \not\models BIP$. It remains to show $f \models CB(R \wedge S_\Gamma) \wedge R \wedge S_\Gamma$. Because of $f(\omega) = \{e, f, g\}$, this is the case if e, f, and g satisfy $R \wedge S_\Gamma$. It is easy to check that all states in this model satsify S_Γ.

To see $f \models R$, observe that one has $f \models (a \hookrightarrow \neg e) \wedge (b \hookrightarrow \neg c) \wedge (c \hookrightarrow \neg d)$ because of $f_0(f, a) = f_0(f, b) = \{f\}$ and $f_0(f, c) = \{g\}$. Thus the only subformulas of type (1) of R that can be false at f are material implications with the consequent $a \hookrightarrow \neg b$, $b \hookrightarrow \neg f$ or $c \hookrightarrow \neg g$.

Because of $f_1(f, a) = \{g\}$, player 1 believes at node a in state f exactly those formulas that are true in state g. Therefore $f_0(g, e) = \{e\}$ and $f_0(g, b) = \{g\}$ implies $f \models B_1((e \hookrightarrow \bigvee_{i \in I} \pi_1 = x_i) \wedge (b \hookrightarrow \bigvee_{j \in J} \pi_1 = y_j) \mid a)$ exactly if for some $i \in I$ $x_i = 1$ and for some $j \in J$ $y_j = 3$. Consequently it cannot be that for all $i \in I$ and all $j \in J$, $x_i > y_j$, and therefore for any material implication of type (1) with the consequent $a \hookrightarrow \neg b$, the antecedent must be false.

In the same way, $f_2(f, b) = \{f\}, f_0(f, f) = \{f\}$ and $f_0(f, c) = \{g\}$ imply $f \models B_2((c \hookrightarrow \bigvee_{i \in I} \pi_2 = x_i) \wedge (f \hookrightarrow \bigvee_{j \in J} \pi_2 = y_j) \mid b)$ exactly if $x_i = 1$ for some $i \in I$ and $y_j = 2$ for some $j \in J$. Thus any material implication of type (1) with the consequent $b \hookrightarrow \neg f$ must be true.

Furthermore $f_1(f, c) = \{g\}$ together with $f_0(g, g) = \{g\}$ and $f_0(g, d) = \{d\}$ yields $f \models B_1((d \hookrightarrow \bigvee_{i \in I} \pi_1 = x_i) \wedge (g \hookrightarrow \bigvee_{j \in J} \pi_1 = y_j) \mid c)$ only if $x_i = 2$ for some $i \in I$ and $y_j = 3$ for some $j \in J$, which means that no material implication of type (1) with the consequent $c \hookrightarrow \neg g$ can be false. This establishes $f \models R$.

It is left to the reader to verify $g \models R$ and $e \models R$ in a completely analogous manner.

For the second part of the proposition, see the proof of proposition 3. ■

For the formulation of the next lemma, I introduce some additional notation. For a decision node v, let $\pi_i(v)$ denote the payoff of player i if v were reached and afterwards only backward induction moves played. $S(u)$ stands for the set of decision nodes weakly succeeding u (i.e. $u \in S(u)$). Furthermore, define an empty conjunction to be true.

Lemma 2. *For all $u \in V$ the following is valid:*

$$S \vdash \left(S_\Gamma \wedge \bigwedge_{t \in S(u)} (t \hookrightarrow t^*) \right) \Rightarrow \left(u \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(u) \right)$$

Proof of lemma 2. The proof proceeds by induction on the game tree. For the base case, let u be a node at which only terminal moves are possible. Consider the following instance of $C4$:

$$S \vdash ((u \hookrightarrow u^*) \wedge (u^* \hookrightarrow u)) \Rightarrow \left(\left(u^* \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(u^*) \right) \right.$$
$$\left. \Rightarrow \left(u \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(u^*) \right) \right)$$

From this and $\pi_i(u^*) = \pi_i(u)$ follows:

$$S \vdash (S_\Gamma \wedge (u \hookrightarrow u^*)) \Rightarrow \left(u \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(u) \right)$$

For the induction step, assume that the following has been shown for all decision nodes $v \in M(u)$:

$$S \vdash \left(S_\Gamma \wedge \bigwedge_{t \in S(v)} (t \hookrightarrow t^*) \right) \Rightarrow \left(v \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(v) \right)$$

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

Then one finds as in the base case due to $C4$:

$$S \vdash (S_\Gamma \wedge (u \hookrightarrow u^*)) \Rightarrow \left(\left(u^* \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(u^*)\right) \Rightarrow \left(u \hookrightarrow \bigwedge_{i=1}^{m} \pi_i = \pi_i(u^*)\right)\right)$$

Together with the induction hypothesis and $\pi_i(u^*) = \pi_i(u)$, this yields the desired result. ■

Proof of proposition 2. For any $u \in V$, I will show the following by induction on the game tree:

$$S \vdash \left(S_\Gamma \wedge \bigwedge_{v \in S(u)} B_{u \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big)\right) \Rightarrow \bigwedge_{v \in S(u)} (v \hookrightarrow v^*)$$

With $u = r$, the assertion of the proposition then follows directly.

For the base case, let v be a decision node at which only terminal moves are possible. For any non backward induction move $v' \in M(v)$ one finds:

$$S \vdash B_{i(v)}(S_\Gamma \mid v) \Rightarrow B_{i(v)}\Big(\big(v^* \hookrightarrow \pi_{i(v)} = \pi_{i(v)}(v^*)\big) \wedge \big(v' \hookrightarrow \pi_{i(v)} = \pi_{i(v)}(v')\big)\Big| v\Big)$$

Because of $\pi_{i(v)}(v^*) > \pi_{i(v)}(v')$ this gives:

$$S \vdash \big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \Rightarrow (v \hookrightarrow \neg v')$$

As v' was arbitrary, this yields:

$$S \vdash \big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v \wedge S_\Gamma\big) \Rightarrow (v \hookrightarrow v^*)$$

For the induction step, assume the induction claim has been shown for all decision nodes $u \in M(s)$. By the definition of forward belief, one finds for any such u:

$$\begin{aligned} S \vdash & \bigwedge_{v \in S(s)} B_{s \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \\ & \Rightarrow \left(B_{i(s)}\left(S_\Gamma \wedge \bigwedge_{v \in S(u)} B_{u \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \mid s\right) \wedge R_s\right) \end{aligned}$$

BR together with the induction hypothesis and lemma 2 now yield:

$$S \vdash \bigwedge_{v \in S(s)} B_{s \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \Rightarrow \big(B_{i(s)}\big(u \hookrightarrow \pi_{i(s)} = \pi_{i(s)}(u) \mid s\big) \wedge R_s\big)$$

For any $u \neq s^*$, this means:

$$S \vdash \bigwedge_{v \in S(s)} B_{s \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \Rightarrow (s \hookrightarrow \neg u)$$

From this follows:

$$S \vdash \left(S_\Gamma \wedge \bigwedge_{v \in M(s)} B_{s \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big)\right) \Rightarrow (s \hookrightarrow s^*)$$

Together with the induction hypothesis, this completes the induction. ■

Proof of proposition 3. For a given game Γ, consider a belief revision model $(\Omega, f_0, f_1, \ldots, f_m, p)$ with $\Omega = T$ and p defined as in the model in the proof of proposition 1.

Let b denote the backward induction outcome and $i \in \{0, 1, \ldots, m\}$. For an inductive definition of r_i^b, let $r_i^b(b) = 1$. Now assume that exactly the numbers $1, \ldots, n$ have already been assigned to states and $r_i^b(t) = n$. Let v be the last node on the path to t that is also on the path to a terminal node that has not yet been numbered. From the paths to unnumbered terminal nodes trough v, consider only those where from the immediate successor of v onwards, only backward induction moves are played. From these paths, take the one where player $i(v)$ gets the highest payoff. Let t' denote the terminal node of this path and set $r_i^b(t') = n + 1$. The reader may check that with this procedure, r_i^b assigns a unique number to all states in Ω.

For all states $\omega \neq b$, define $r_i^\omega(\omega) = 1$, $r_i^\omega(\omega') = r_i^b(\omega') + 1$ if $r_i^b(\omega') < r_i^b(\omega)$ and $r_i^\omega(\omega') = r_i^b(\omega')$ if $r_i^b(\omega') > r_i^b(\omega)$.

Now for all $i \in \{0, 1, \ldots, m\}$, f_i can be defined as in lemma 1. Note that with this definition, $f_i(\omega, v) = \{t\}$ implies either $\omega = t$ or that on the path from v to t, only backward induction moves are played.

One can easily see that S_Γ is satisfied at all states, i.e. $\omega \models S_\Gamma \wedge B_{i(v)}(S_\Gamma \mid v)$ for all $v \in V, \omega \in \Omega$.

Let t be such that either t is the backward induction outcome of the subgame starting at v or $t \not\models v$. Then for $\omega \in f_0(t, v) = f_{i(v)}(t, v)$, one has $\omega \models v^*$ and thus $t \models (v \hookrightarrow \neg v')$ for $v' \in M(v) \setminus \{v^*\}$. For $\omega' \in f_0(\omega, v')$ one finds $\omega' \models \pi_{i(v)} = \pi_{i(v)}(v')$ and thus

$$t \models B_{i(v)}\left(\left(v' \hookrightarrow \bigvee_{k \in K} \pi_{i(v)} = x_k\right) \wedge \left(v^* \hookrightarrow \bigvee_{j \in J} \pi_{i(v)} = y_j\right) \mid v\right)$$

exactly if $x_k = \pi_{i(v)}(v')$ for some $k \in K$ and $y_j = \pi_{i(v)}(v)$ for some $j \in J$. Due to $\pi_{i(v)}(v) > \pi_{i(v)}(v')$, this implies $t \models R_v$.

One thus finds

$$t \models B_{v \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \tag{4}$$

I will use induction on the game tree to show the following for any $u \in V$ if t is either the backward induction outcome of the subgame starting at u or $t \not\models u$:

$$t \models \bigwedge_{v \in S(u)} B_{u \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big)$$

With $u = r$ and $t = b$, this establishes the proposition.

The base case for a decision node u where only terminal moves are possible follows directly from (4).

For the induction step, assume that the induction claim has already been shown for all decision nodes u' in $M(u)$. Let t be either the backward induction outcome of the subgame starting at u or $t \not\models u$. Then t also has this property with respect to any decision node $u' \in M(u)$ and the induction hypothesis yields:

$$t \models \bigwedge_{v \in S(u')} B_{u' \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big)$$

Furthermore, by the construction of $f_{i(u)}$, the same is true of $t' \in f_{i(u)}(t, u)$, which gives:

$$t \models \bigwedge_{v \in S(u')} B_{i(u)}\big(B_{u' \to v}\big(B_{i(v)}(S_\Gamma \mid v) \wedge R_v\big) \mid u\big)$$

Together with (4), this completes the induction.

Note that the induction implies in particular $b \models R$. Furthermore it is straightforward to show $b \models R^{ex\,post}$, and therefore also $b \models R^{ex\,ante}$ as the models satisfies $R9$. Because of $f(b) = \{b\}$, this yields $b \models CB(R \wedge S_\Gamma)$, $b \models CB(R^{ex\,post} \wedge S_\Gamma)$ and $b \models CB(R^{ex\,ante} \wedge S_\Gamma)$, which proves the second part of proposition 1 and of the remarks 1 and 3. ■

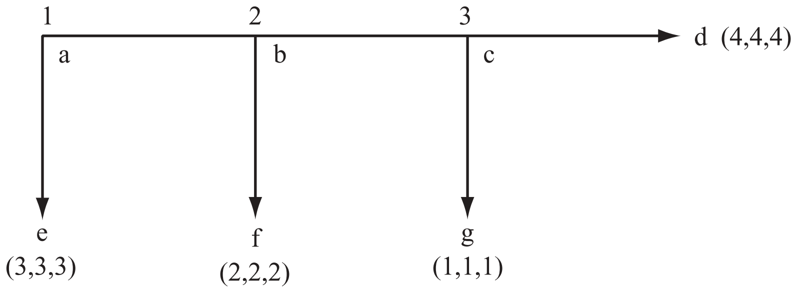

FIGURE 2. A three player game.

Proof of remark 1. Consider a belief revision model $(\Omega, f_0, f_1, f_2, f_3, p)$ based on the game in figure 2 with $\Omega = T$ and p defined as in the model in the proof of proposition 1. For $i \in \{0, 1, 2, 3\}$, let $r_i^e(e) = 1, r_i^e(f) = 2, r_i^e(d) = 3, r_i^e(g) = 4; r_i^f(f) = 1, r_i^f(g) = 2, r_i^f(d) = 3, r_i^f(e) = 4;$ $r_i^g(g) = 1, r_i^g(f) = 2, r_i^g(d) = 3, r_i^g(e) = 4$ and $r_i^d(d) = 1, r_i^d(f) = 2, r_i^d(e) = 3, r_i^d(g) = 4$. Define f_i as in lemma 1.

One finds $e \not\models BIP$. I will show $e \models CB(R \wedge S_\Gamma)$. Because of $f(e) = \{e\}$, it suffices to show $e \models R \wedge S_\Gamma$. Clearly all states in the given models satisfy S_Γ. Furthermore because of $e \models (a \hookrightarrow \neg b) \wedge (b \hookrightarrow \neg c) \wedge (c \hookrightarrow \neg g)$, the only subformulas of R of type (1) that could be false at e are material implications with the consequent $(a \hookrightarrow \neg e)$, $(b \hookrightarrow \neg f)$ or $(c \hookrightarrow \neg d)$. In the first case, $f_1(e,a) = \{e\}$ together with $f_0(e,b) = \{f\}$ and $f_0(e,e) = \{e\}$ yields $e \models B_1((b \hookrightarrow \bigvee_{i \in I} \pi_1 = x_i) \wedge (e \hookrightarrow \bigvee_{j \in J} \pi_1 = y_j) \mid a)$ exactly if for some $i \in I$ $x_i = 2$ and for some $j \in J$ $y_j = 3$. In the second case, $f_2(e,b) = \{f\}$ together with $f_0(f,c) = \{g\}$ and $f_0(f,f) = \{f\}$ yields $e \models B_2((c \hookrightarrow \bigvee_{i \in I} \pi_2 = x_i) \wedge (f \hookrightarrow \bigvee_{j \in J} \pi_2 = y_j) \mid b)$ exactly if for some $i \in I$ $x_i = 1$ and for some $j \in J$ $y_j = 2$. In the third case, $f_3(e,c) = \{d\}$ together with $f_0(d,g) = \{g\}$ and $f_0(d,d) = \{d\}$ yields $e \models B_3((g \hookrightarrow \bigvee_{i \in I} \pi_3 = x_i) \wedge (d \hookrightarrow \bigvee_{j \in J} \pi_3 = y_j) \mid a)$ exactly if for some $i \in I$ $x_i = 1$ and for some $j \in J$ $y_j = 4$. Thus in all cases, the antecedents of the material implication subformulas of R are false, which gives $e \models R$.

For the second part of the remark, see the proof of proposition 3. ■

Proof of remark 2. For this proof, I will use the following additional notation. Let $l(v)$ denote the maximum number of decision nodes on any path of the subgame starting at node $v \in V$. Furthermore, let $B^n\phi$ be defined as follows:

$$\begin{array}{lll} B^1\phi & :\Leftrightarrow & \bigwedge_{i=1}^{n} B_i(\phi \mid \top) \\ B^n\phi & :\Leftrightarrow & \bigwedge_{i=1}^{n} B_i(B^{n-1}\phi \mid \top) \wedge B^{n-1}\phi \quad \textit{for } n > 1 \end{array}$$

Let $V^n := \{v \in V \mid l(v) \leq n\}$. I will use induction to show the following for all $n \in N$:

$$(5) \quad S \vdash (B^n(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante} \wedge S_\Gamma) \Rightarrow \bigwedge_{v \in V^n} (v \hookrightarrow v^*)$$

For the base case, assume that v is a decision node at which only terminal moves are possible, i.e. $l(v) = 1$. One finds for $t \neq v^*$, $t \in M(v)$

$$S \vdash B^1(R^{ex\,ante} \wedge S_\Gamma) \Rightarrow B_{i(v)}\big((v^* \hookrightarrow \pi_{i(v)}(v)) \wedge \big(t \hookrightarrow \pi_{i(v)}(t)\big) \mid \top\big)$$

Because of $\pi_{i(v)}(v) > \pi_{i(v)}(t)$ one has

$$S \vdash (B^1(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante}) \Rightarrow (v \hookrightarrow \neg t)$$

As this is true for any $t \in M(v)$ with $t \neq v^*$, one gets

$$S \vdash (B^1(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante} \wedge S_\Gamma) \Rightarrow (v \hookrightarrow v^*)$$

Now assume that (5) has been shown for all $n < m$ and let $l(v) = m$. One finds

$$S \vdash B^m(R^{ex\,ante} \wedge S_\Gamma) \Rightarrow B_{i(v)}(B^{m-1}(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante} \wedge S_\Gamma \mid \top)$$

The induction hypothesis and the definition of S_Γ then imply for $u \neq v^*$, $u \in M(v)$

$$S \vdash B^m(R^{ex\,ante} \wedge S_\Gamma) \Rightarrow B_{i(v)}((v^* \hookrightarrow \pi_{i(v)}(v)) \wedge (t \hookrightarrow \pi_{i(v)}(t)) \mid \top)$$

This in turn implies

$$S \vdash (B^m(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante}) \Rightarrow (v \hookrightarrow \neg u)$$

and thus

$$S \vdash (B^m(R^{ex\,ante} \wedge S_\Gamma) \wedge R^{ex\,ante}) \Rightarrow (v \hookrightarrow v^*)$$

Together with the induction hypothesis, this completes the induction. Furthermore, as $B7$ implies $S \vdash CB(R^{ex\,ante} \wedge S_\Gamma) \Rightarrow B^{n_r}(R^{ex\,ante} \wedge S_\Gamma)$ with $n_r = l(r)$ and thus $V^{n_r} = V$, this also establishes the claim of the remark. ■

Proof of remark 3. I will show $S, B9 \vdash \neg R^{ex\,ante} \Rightarrow \neg R^{ex\,post}$. Therefore assume $R^{ex\,ante}$ is false. This means that some formula of the form

$$B_{i(v)}\left(\left(u \hookrightarrow \bigvee_{k \in K} \pi_{i(v)} = x_k\right) \wedge \left(w \hookrightarrow \bigvee_{j \in J} \pi_{i(v)} = y_j\right) \mid \top\right) \wedge \neg(v \hookrightarrow \neg w)$$

with $x_k > y_j$ for all $k \in K$ and all $j \in J$ is true. I will show that then the formula

$$\left(\left(v \Rightarrow B_{i(v)}\left(\left(u \hookrightarrow \bigvee_{k \in K} \pi_{i(v)} = x_k\right) \wedge \left(w \hookrightarrow \bigvee_{j \in J} \pi_{i(v)} = y_j\right) \mid v\right)\right)\right.$$

$$\wedge \left(\neg v \Rightarrow B_{i(v)}\left(\left(u \hookrightarrow \bigvee_{k \in K} \pi_{i(v)} = x_k\right)\right.\right.$$

$$\left.\left.\left.\wedge \left(w \hookrightarrow \bigvee_{j \in J} \pi_{i(v)} = y_j\right) \mid \neg v\right)\right)\right)$$

$$\wedge \neg(v \hookrightarrow \neg w)$$

must also be true, which yields that $R^{ex\ post}$ must be false.

For this it suffices to show

$$S, B9 \vdash B_i(\phi \mid \top) \Rightarrow ((v \Rightarrow B_i(\phi \mid v) \wedge (\neg v \Rightarrow B_i(\phi \mid \neg v)))$$

This is in turn implied by the following:

(6) $S, B9 \vdash (\psi \wedge B_i(\phi \mid \top)) \Rightarrow B_i(\phi \mid \psi)$

To see that (6) is true, note that applying $B9$ yields $S, B9 \vdash \psi \Rightarrow \neg B_i(\neg\psi \mid \top)$. This together with $B6$ gives $S, B9 \vdash \psi \Rightarrow (B_i(\phi \mid \psi) \Leftrightarrow B_i(\psi \Rightarrow \phi \mid \top))$. As BR and propositional reasoning yield $S, B9 \vdash B_i(\phi \mid \top) \Rightarrow B_i(\psi \Rightarrow \phi \mid \top)$, this establishes (6).

For the second part of the remark, see the proof of proposition 3. ■

Proof of remark 4. I will show $S, B9, G2 \vdash CB(R \wedge S_\Theta) \Rightarrow (v \Rightarrow v^*)$ for all $v \in V$ by induction on the game tree. With $v = r$, this yields the first part of the assertion of the remark.

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

For the base case, let v stand for a decision node where only terminal moves are possible. Because of (6), one finds for all $t \in M(v)$:

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(v)}\big(t \hookrightarrow \pi_{i(v)} = \pi_{i(v)}(t) \mid v\big)$$

From this follows $S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow (v \hookrightarrow \neg t)$ for all $t \in M(v) \setminus \{v^*\}$. Propositional reasoning, $C6$ and the definition of S_Θ then give $S, B9, G2 \vdash CB(R \wedge S_\Theta) \Rightarrow (v \Rightarrow v^*)$.

For the induction step, assume the induction claim has been shown for all decision nodes $v \in M(w)$. By $B7$ and BR one has

$$S, B9, G2, \vdash CB(R \wedge S_\Theta) \Rightarrow B_{i(w)}(v \Rightarrow v^* \mid \top)$$

Because of (6), this implies

$$S, B9, G2 \vdash CB(v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}(v \Rightarrow v^* \mid v)$$

Via $B1$, $C6$ and BR this in turn gives

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}(v \hookrightarrow v^* \mid v)$$

By $C6$ and $G2$ one has $S, B9, G2 \vdash (w \wedge v) \Rightarrow B_{i(w)}(v \mid w)$. Furthermore, (6) gives

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}(v \hookrightarrow w \mid v)$$

and thus via $C1$, $C6$ and BR

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}(w \mid v)$$

As one has $S, B9, G2 \vdash S_\Theta \Rightarrow (v \Rightarrow w)$, one can apply $B5$ to derive

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}(v \hookrightarrow v^* \mid w)$$

Due to

$$S, B9, G2 \vdash S_\Theta \Rightarrow \big((v \hookrightarrow v^*) \Rightarrow \big(v \hookrightarrow \pi_{i(w)} = \pi_{i(w)}(v)\big)\big)$$

one can apply BR to derive

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}\big(v \hookrightarrow \pi_{i(w)} = \pi_{i(w)}(v) \mid w\big)$$

for all decision nodes $v \in M(w)$. As one furthermore has

$$S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}\big(w^* \hookrightarrow \pi_{i(w)} = \pi_{i(w)}(w) \mid w\big),$$

one finds $S, B9, G2 \vdash (v \wedge CB(R \wedge S_\Theta)) \Rightarrow (w \hookrightarrow \neg v)$ and thus $S, B9, G2 \vdash CB(R \wedge S_\Theta) \Rightarrow \neg v$ for all decision nodes $v \in M(w)$.

Similarly$S, B9, G2 \vdash (w \wedge CB(R \wedge S_\Theta)) \Rightarrow B_{i(w)}(t \hookrightarrow \pi_{i(w)} = \pi_{i(w)}(t) \mid w)$ for all terminal nodes $t \in M(v)$ gives $S, B9, G2 \vdash (w \wedge CB(R \wedge S_\Theta)) \Rightarrow (w \hookrightarrow \neg t)$ for all terminal nodes $t \in M(w) \setminus \{w^*\}$. One can now derive $S, B9, G2 \vdash CB(R \wedge S_\Theta) \Rightarrow (w \Rightarrow w^*)$, which completes the induction. ■

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press

REFERENCES

[1] Aumann, R. 1995. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8:6–19

[2] Aumann, R. 1998. On the centipede game. *Games and Economic Behavior*, 23:97–105

[3] Balkenborg, D. and E. Winter. 1997. A necessary and sufficient epistemic condition for playing backward induction. *Journal of Mathematical Economics*, 27:325–45

[4] Battigalli, P. and M. Siniscalchi. 2000. Interactive beliefs and forward induction. Mimeo

[5] Ben-Porath, E. 1997. Rationality, Nash-equilibrium and backward induction in perfect information games. *Review of Economic Studies*, 64:23–46

[6] Binmore, K. and H. Shin. 1992. Algorithmic knowledge and game theory, in Bicchieri and Dalla Chiara (eds.), *Knowledge, Belief, and Strategic Interaction*. Cambridge, MA:141–54

[7] Bonanno, G. 1991. The logic of rational play in games of perfect information. *Economics and Philosophy*, 7:37–65

[8] Brandenburger, A. 1998. On the existence of a "Complete" belief model. *HBS Working Paper* 99–056

[9] Dekel, E. and F. Gul. 1996. Rationality and common knowledge in game theory, in Kreps and Wallis (eds.), *Advances in Economics and Econometrics: Theory and Application*, vol. I, Cambridge: 87–172

[10] Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Cambridge, MA

[11] Halpern, J. 1999. Set-theoretic completeness for epistemic and conditional logic. *Annals of Mathematics and Artificial Intelligence*, 26: 1–27

[12] Halpern, J. 1999. Hypothetical knowledge and counterfactual reasoning. *International Journal of Game Theory*, 28:315–30

[13] Halpern, J. 1998. Substantive rationality and backward induction. Mimeo

[14] Halpern, J. and Y. Moses. 1992. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:311–79

[15] Lewis, D. 1973. *Counterfactuals*. Oxford

[16] Nute, D. 1984. Conditional logic, in Gabbay and Guenthner (eds.), *Handbook of Philosophical Logic*, vol. II, Dordrecht:387–439

[17] Pettit, P. and R. Sugden. 1989. The backward induction paradox. *The Journal of Philosophy*, 86:169–82

[18] Priest, G. 2000. The logic of backwards inductions. *Economics and Philosophy*, 16:267–85

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

[20] Samet, D. 1996. Hypothetical knowledge and games with perfect information. *Games and Economic Behavior*, 17:230–51

[21] Stalnaker, R. 1996. Knowledge, belief and counterfactual reasoning in Games. *Economics and Philosophy*, 12:133–63

[22] Stalnaker, R. 1998. Belief revision in games: forward and backward induction. *Mathematical Social Sciences*, 36:31–56

[23] Vilks, A. 1999. Knowledge of the game, relative rationality, and backwards induction without counterfactuals. *Working Paper* No. 25, Leipzig Graduate School of Management

https://doi.org/10.1017/S0266267104001269 Published online by Cambridge University Press