
ELSEVIER
EDITORS
Simon Grant
Jean-François Laslier
EDITORIAL BOARD
Fuad Aleskerov
Carmen Bevia
Francis Bloch
Anna Bogomolnaia
Walter Bossert
Raouf Boucekkine
Steven J. Brams
Alain Chateauneuf
Alessandro Citanna
Hervé Crès
Gabrielle Demange
Jean Paul Doignon
John Duggan
Bhaskar Dutta
Aviad Heifetz
Ron Holzman
Mamoru Kaneko
Edi Karni
Bettina Klaus
Luc Lauwers
Ngo Van Long
Mark Machina
François Maniquet
Bernard Monjardet
Hervé Moulin
Patrick Moyes
Klaus Nehring
Wojciech Olszewski
Hans Peters
Martine Quinzii
Klaus Ritzberger
John Roemer
Zvi Safra
Remzi Sanver
Karl Schlag
Arunava Sen
Roberto Serrano
John Tomas Sjostrom
Rann Smorodinsky
Yves Sprumont
Jean-Marc Tallon
William Thomson
Timothy Van Zandt
William Zwicker
FOUNDING EDITOR
Ki Hang Kim
Available online at www.sciencedirect.com
ScienceDirect
Volume 56, Number 1, July 2008
ISSN 0165-4896
mathematical social sciences
CONTENTS
A. PEREA, Minimal belief revision leads to backward induction 1
Y. VIOSSAT, Evolutionary dynamics may eliminate all strategies used in correlated equilibrium 27
J. RENAULT, S. SCARLATTI, M. SCARSINI, Discounted and finitely repeated minority games with public signals 44
A. ATTAR, D. MAJUMDAR, G. PIASER, N. PORTEIRO, Common agency games: Indifference and separable preferences 75
L.A. GUARDIOLA, A. MECA, J. PUERTO, Production-inventory games and PMAS-games: Characterizations of the Owen point 96
H.-J. CHEN, M.-C. LI, Productive public expenditures, expectation formations and nonlinear dynamics 109
V. TILSON, Monotonicity properties of wholesale price contracts 127
A. AL-NOWAIHI, S. DHAMI, Corrigendum to "A note on the Loewenstein–Prelec theory of intertemporal choice" [Math. Social Sci. 52 (1) (2006) 99–108] 144
www.elsevier.com/locate/econbase
Full text articles (PDF); abstracts of 12,800 articles
ELSEVIER
AMSTERDAM – BOSTON – JENA – LONDON – NEW YORK – OXFORD – PARIS – PHILADELPHIA – SAN DIEGO – ST LOUIS

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Mathematical Social Sciences 56 (2008) 1–26
mathematical social sciences
www.elsevier.com/locate/econbase

Minimal belief revision leads to backward induction☆

Andrés Perea*

Maastricht University, Department of Quantitative Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Received 22 February 2007; received in revised form 1 October 2007; accepted 9 December 2007
Available online 20 January 2008

Abstract

We present an epistemic model for games with perfect information in which players, upon observing an unexpected move, may revise their belief about the opponents' preferences over outcomes. For a given profile P of preference relations over outcomes, we impose the following conditions: (1) players initially believe that opponents have preference relations as specified by P; (2) players believe at every instance of the game that each opponent is carrying out a sequentially rational strategy; (3) if a player revises his belief about an opponent's type, he must search for a "new" type that disagrees with the "old" type on a minimal number of statements about this opponent; (4) if a player revises his belief about an opponent's preference relation over outcomes, he must search for a "new" preference relation that disagrees with the "old" preference relation on a minimal number of pairwise rankings. It is shown that every player whose preference relation is given by P, and who throughout the game respects common belief in the events (1)–(4), has a unique sequentially rational strategy, namely his backward induction strategy in the game induced by P.

JEL classification: C72

Keywords: Belief revision; Minimal belief change; Backward induction; Dynamic games

1. Introduction

In this paper we are concerned with the problem of how to model rationality in dynamic games. In a purely static setting, rational choice can be formalized by the requirement that players

☆ I wish to thank Geir Asheim, Giacomo Bonanno, Dov Samet, Robert Sugden, Shmuel Zamir, an associate editor and some referees for their helpful comments.

* Fax: +31 3884874.
E-mail address: a.perea@ke.unimaas.nl.

0165-4896/$ - see front matter
doi:10.1016/j.mathsocsci.2007.12.003

hold beliefs about the opponents' strategy choices, and choose strategies that are optimal against these beliefs. In a dynamic game, however, it may happen that a player's initial belief about the opponents' strategy choices will be contradicted by the opponents' real behavior later on in the game. The player must then revise his belief about the opponents so as to explain the observed behavior. The two basic questions that we shall focus on are: How should the player revise his beliefs? and What consequences does this have for the player's own behavior?

Throughout this paper, we shall restrict attention to games with perfect information. In order to formalize how players reason about their opponents, and in particular how players may revise their beliefs about the opponents' characteristics, we present a formal epistemic model in which the relevant characteristics of each player are represented by a *type*. Within our setup, a type specifies a strict preference relation over the terminal nodes in the game, and defines at every decision node a conditional belief about the opponents' strategy choices *and types*. Since different opponents' types may have different preference relations over terminal nodes, a type may, in particular, revise his belief about the opponents' preference relations over terminal nodes during the game. Intuitively, a type for player i may be interpreted as a list of expressions of the kind "player i has preference relation P_i over terminal nodes", "player i believes at decision node h_i that player j chooses strategy s_j", "player i believes at decision node h_i that player j has preference relation P_j over terminal nodes", "player i believes at h_i that player j believes at h_j that player k chooses strategy s_k", and so forth. We refer to such expressions as *statements about player i*.

We consider a scenario in which players, throughout the game, believe that their opponents carry out sequentially rational strategies, that is, strategies that are optimal at every decision node. We refer to this condition as *structural belief in sequential rationality* (SBSR). Hence, if a player currently believes that an opponent chooses a certain strategy, but finds out later that his opponent has chosen otherwise, then the player must seek for a new belief about the opponent that rationalizes this unexpected event, while maintaining his belief that the opponent carries out a sequentially rational strategy. This may be done, for instance, by changing the belief about the opponent's preferences over terminal nodes, or by changing the belief about the opponent's beliefs, or both.

A generally accepted principle in belief revision theory is that belief changes should be as small as possible, while being able to explain the newly observed information (see Schulte (2002) for an excellent discussion of the idea of minimal belief revision, and an overview of the various formalizations thereof in belief revision theory). The intuition behind this principle is that the current beliefs of a decision maker reflect, in some sense, the "best possible theory" that he can produce about the state of affairs given his current information. If these beliefs are contradicted by new observations, the decision maker therefore attempts to explain these new events by disturbing his previous beliefs as little as possible.

We attempt to incorporate this idea of minimal belief revision into our epistemic model. To illustrate the basic idea, consider a player i who believes at his decision node h_i that player j has type t_j. Suppose there is some future decision node h_i', also controlled by player i, that cannot be reached if type t_j chooses a sequentially rational strategy. If h_i' is nevertheless reached, player i should clearly change his belief about player j's type. The requirement of SBSR, namely, imposes that player i should still believe at h_i' that player j chooses a sequentially rational strategy. Since player i may revise his belief about j's preference relation over terminal nodes, there are many player j's types which have a sequentially rational strategy leading to h_i', and which could be chosen by player i as a revised belief about player j's type. Our version of minimal belief revision (MBR) states that, among these many different new beliefs about player j's type that he could choose to rationalize the event of reaching h_i', he should choose a player

j type t'_j that is "most similar" to his previous belief t_j. By most similar, we mean that t'_j should disagree with t_j on as few statements about player j as possible, and, given the previous condition, the preference relations over terminal nodes of t'_j and t_j should disagree on a minimal number of pairwise rankings.

A last condition we impose is that there should be some profile P of preference relations over terminal nodes, specifying one preference relation for each player, such that every player initially believes that the opponents' preference relations are given by P. We refer to this event as initial belief in P (IBP).

The first main result in this paper, Theorem 5.1, shows that common structural belief in the events IBP, SBSR and MBR is possible. That is, requiring that every player always believes that every player always believes that... that every player always believes that every player satisfies IBP, SBSR and MBR does not lead to logical contradictions.

In the second main result, Theorem 5.2, we prove that common structural belief in IBP, SBSR and MBR leads to backward induction. More precisely, if a type t_i for player i has preference relation P_i as specified by P, and if t_i respects common structural belief in IBP, SBSR and MBR, then this type has a unique sequentially rational strategy, namely his backward induction strategy in the game induced by the profile P of preference relations.

The concept of common structural belief in IBP, SBSR and MBR may thus be viewed as a possible foundation for backward induction, which constitutes one of the oldest ideas in game theory. Other epistemic foundations for the backward induction *strategy profile* include Asheim (2002), Asheim and Perea (2005), Aumann (1995), Balkenborg and Winter (1997), Clausing (2003, 2004), Feinberg (2005), Quesada (2002, 2003), Samet (1996) and Stalnaker (1998). A detailed description and comparison of these foundations can be found in Perea (2007b). Battigalli and Siniscalchi (2002) and Brandenburger et al. (2007) provide epistemic models that lead to the backward induction *outcome* but not necessarily to the backward induction strategy profile. The main difference between the above foundations and our model is that in our setting, players are assumed to interpret every unexpected move by an opponent as a rational move, whereas this is not the case in the other foundations. Moreover, in our model players are allowed to revise their beliefs about the opponents' preference relations over terminal nodes in order to rationalize such unexpected moves, while the aforementioned foundations do not model this possibility, at least not explicitly. In fact, Reny (1992a, 1993) has already illustrated that there are only very few games with perfect information in which common structural belief in sequential rationality is possible *without* allowing players to revise their beliefs about the opponents' preferences over terminal nodes. Other foundations for backward induction that *do* allow players to revise their beliefs about the opponents' utilities during the game can be found in Perea (2006, 2007a). The main difference with our approach here is that the latter two foundations use *proper belief revision,* rather than MBR, as a criterion to restrict the possible belief revision procedures. Proper belief revision states that whenever player i at decision node h_i revises his belief about player j, then he must not change his belief about player j's relative ranking of two strategies s_j and s'_j, if both s_j and s'_j could have led to h_i. The intuition is that such belief changes would be "unnecessary " in order to explain the event that h_i has been reached. In Perea (2006), Theorem 7.1, it is shown that common structural belief in SBSR and proper belief revision leads to the backward induction strategy profile in every generic game with perfect information. In Theorem 6.8 we establish a formal relationship between MBR and proper belief revision, which proves to be important for deriving Theorem 5.2.

The outline of this paper is as follows. In Section 2 we provide an example that illustrates why common structural belief in IBP, SBSR and MBR leads to backward induction. In Section 3 we

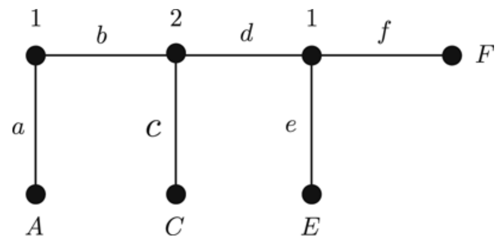

Fig. 1. Common structural belief in IBP, SBSR and MBR leads to backward induction.

develop an epistemic model for games with perfect information. Formal definitions of SBSR and MBR are given in Section 4. In Section 5 we present the main results. Section 6 discusses some important implications of (common structural belief in) SBSR and MBR, which will be used in Section 7 to prove the two main results. Some concluding remarks are given in Section 8. Proofs of technical lemmas and most preparatory results can be found in the Appendix.

2. Why minimal belief revision leads to backward induction

2.1. An example

As to illustrate why common structural belief in IBP, SBSR and MBR leads to backward induction, consider the game tree depicted in Fig. 1. The symbols A, C, E and F denote the different terminal nodes that can be reached at the end. Consider the profile $P = (P_1, P_2)$ of preference relations over terminal nodes, where $P_1 = EAFC$ and $P_2 = FCEA$. Here, $EAFC$ means that player 1 strictly prefers E to A, strictly prefers A to F and strictly prefers F to C. Similarly for $FCEA$. Let us denote the three decision nodes by h_1, h_2 and h_3, respectively. Consider a type t_2 for player 2 that has preference relation P_2 over terminal nodes and respects common structural belief in IBP, SBSR and MBR. We show that t_2 has a unique sequentially rational strategy, namely c, which is his backward induction strategy induced by P. We proceed by the following four steps.

Step 1.We show that type t_2, at h_2, does not revise his belief about player 1's conditional beliefs. That is, if type t_2 revises a belief at h_2, it must be about player 1's preference relation over the terminal nodes.

Proof of Step 1. Suppose that t_2 initially believes that player 1 has type t_1^{ini}, but believes at h_2 that player 1 has type t_1^{rev}. If player 2 is at h_2, meaning that he has observed move b by player 1, then he can always rationalize this move by believing that player 1's preference relation is $EFCA$, while maintaining his previous belief about player 1's beliefs. Therefore, player 2 can rationalize the move b by a revised belief about player 1's type that only differs from t_1^{ini} on at most one statement, namely player 1's preference relation over terminal nodes.

Since type t_2 is assumed to satisfy MBR, the revised belief t_1^{rev} about player 1's type should differ from the initial belief t_1^{ini} on at most one statement. Suppose, contrary to what we want to prove, that type t_2 changes his belief about player 1's beliefs. That is, suppose that t_1^{rev} and t_1^{ini} would have different beliefs about player 2's strategy choice and/or beliefs. To keep our argument simple, assume that t_1^{rev} and t_1^{ini} would have different first-order beliefs at h_1 about player 2's choice. Hence, t_1^{ini} initially believes that player 2 makes choice v at h_2, whereas t_1^{rev} initially believes that player 2 makes the other choice $w \neq v$. Suppose also that t_1^{ini} initially believes that player 2 believes, at h_2, that player 1 makes the choice x at h_3, and that t_1^{rev} initially believes that player 2 believes, at h_2, that player 1 makes the choice y at h_3. Since t_2 respects common structural belief in SBSR, both t_1^{ini} and t_1^{rev} must satisfy SBSR. Since t_1^{ini} initially believes that player 2 chooses v at h_2, choice v must optimal for player 2 at h_2, given t_1^{ini}'s initial belief that player 2's preference relation over terminal nodes is $FCEA$, and given t_1^{ini}'s initial belief that player 2 believes at h_2 that player 1 chooses x at h_3. Similarly, as t_1^{rev} initially believes that player 2 chooses w at h_2, choice w must optimal for player 2 at h_2, given t_1^{rev}'s initial belief that player 2's preference relation over terminal nodes is $FCEA$, and given t_1^{rev}'s initial belief that player 2 believes at h_2 that player 1 chooses y at h_3. However, since $v \neq w$, this can only be the case if $x \neq y$ as well. That is, t_1^{ini} and t_1^{rev} must not only differ on the initial belief they have about player 2's choice, but also on the initial belief they have about the belief that player 2 has at h_2 about player 1's choice at h_3. So, t_1^{ini} and t_1^{rev} must differ on at least two statements, which would contradict our assumption that t_2 satisfies MBR. We may thus conclude that t_1^{rev} and t_1^{ini} must have the same belief, at h_1, about player 2's choice.

By a similar argument we could also show that t_1^{rev} and t_1^{ini} must have the same belief, at h_1 and at h_3, about the belief that player 2 has, initially and at h_2, about player 1's strategy choice. Also, t_1^{rev} and t_1^{ini} must have the same belief, at h_1 and at h_3, about the belief that player 2 has, initially and at h_2, about the belief that player 1 has, at h_1 and h_3, about player 2's choice, and so on. That is, t_1^{rev} and t_1^{ini} must have exactly the same conditional beliefs about player 2. So, type t_2 does not change his belief about player 1's conditional beliefs, which completes the proof of step 1.

Step 2. We show that type t_2, at h_2, does not revise his belief about player 1's ranking of the terminal nodes that follow h_2.

Proof of Step 2. Recall that t_2 initially believes that player 1 is of type t_1^{ini} with preference relation $P_1 = EAFC$ over the terminal nodes. So, type t_2 initially believes that player 1 ranks outcome E over outcome F. Suppose, contrary to what we want to prove, that type t_2, at h_2, revises his belief about player 1's ranking of the outcomes E and F. That is, type t_2 believes, at h_2, that player 1 is of type t_1^{rev} with a preference relation P_1^{rev} over terminal nodes that ranks F over E. Since t_2 satisfies SBSR, type t_2 must believe at h_2 that player 1 is carrying out a sequentially rational strategy. Hence, choice b must be optimal for player 1, given the preference relation P_1^{rev} and the initial belief that t_1^{rev} has about player 2's choice.

Now, consider the preference relation $\tilde{P}_1^{\text{rev}}$ over terminal nodes which is obtained from P_1^{rev} by switching the roles of outcomes E and F, but leaves the roles of the other outcomes unchanged. In particular, $\tilde{P}_1^{\text{rev}}$ ranks E over F, just as P_1 does. Recall that P_1^{rev} and P_1 disagree on the ranking of E and F. It can be verified (see Lemma 6.7 for a formal proof) that P_1 and $\tilde{P}_1^{\text{rev}}$ disagree on less pairwise rankings than P_1 and P_1^{rev}.

Consider now the type $\tilde{t}_1^{\text{rev}}$ which has preference relation $\tilde{P}_1^{\text{rev}}$, but has the same conditional beliefs about player 2 as t_1^{rev}. Then, for type $\tilde{t}_1^{\text{rev}}$ is would still be optimal to choose b at h_1, since $\tilde{t}_1^{\text{rev}}$ holds the same belief at h_1 about player 2's choice as t_1^{rev}, and $\tilde{P}_1^{\text{rev}}$ has only switched the roles of E and F. This means, however, that there is another possible revised belief about player 1's type, $\tilde{t}_1^{\text{rev}}$, which rationalizes the event that player 1 has chosen b, disagrees with t_1^{ini} on the same number of statements as t_1^{rev} does, but which disagrees with t_1^{ini} on less pairwise rankings of terminal nodes than t_1^{rev} does. This would contradict our assumption that t_2 satisfies MBR. We may thus conclude that t_1^{rev} must rank E over F. So, type t_2, at h_2, does not revise his belief about player 1's ranking of the terminal nodes that follow h_2, as was to show.

Step 3. We show that type t_2, at h_2, does not change his belief about player 1's ranking of his strategies (b, e) and (b, f) that lead to h_2.

Proof of Step 3. We know from steps 1 and 2 that type t_2, at h_2, does not change his belief about player 1's conditional beliefs, nor about player 1's ranking of the outcomes E and F. This implies, however, that type t_2 does not change his belief about player 1's ranking of his strategies (b, e) and (b, f).

Step 4. We show that type t_2 has only one sequentially rational strategy, namely c, which is his backward induction strategy induced by P.

Proof of Step 4. Type t_2 initially believes that player 1 has preference relation $P_1 = EAFC$. So, type t_2 initially believes that player 1, at h_3, ranks strategy (b, e) over strategy (b, f). By step 3, it then follows that type t_2, at h_2, still believes that player 1, at h_3, ranks strategy (b, e) over strategy (b, f). As t_2 satisfies SBSR, type t_2 must believe at h_2 that player 1 is carrying out the strategy (b, e). Since type t_2 has preference relation $P_2 = FCEA$, type t_2 must choose c at h_2. This completes the proof.

2.2. *Outline of the proof*

The proof of our Theorem 5.2, which states that, in general, common structural belief in IBP, SBSR and MBR leads to backward induction, follows the same steps as the argument above. Here is a short outline: Take a type t_i for player i that respects common structural belief in IBP, SBSR and MBR.

We first show in Lemma 6.5 that t_i never revises his beliefs about the opponents' conditional beliefs. This is step 1 in our example above.

We then show, by means of Lemma 6.7 and Theorem 6.8, that t_i, at information set h_i, does not revise his belief about opponent j's ranking of strategies s_j and s'_j if both s_j and s'_j lead to h_i. Formally, we say that type t_i satisfies *proper belief revision* (see Perea (2006, 2007a)). This corresponds to steps 2 and 3 in our example.

We finally show, in Section 7.2, that t_i only has one sequentially rational strategy, namely his backward induction strategy induced by P. This final conclusion is based on the insight that t_i satisfies proper belief revision. This corresponds to step 4 in the example.

3. The epistemic model

3.1. *Games with perfect information*

A dynamic game is said to be with *perfect information* if every player, at each instance of the game, observes the opponents' moves that have been made until then. Formally, an *extensive form structure $\mathcal{S}$ with perfect information* consists of a finite game tree, a finite set I of players, for every player i a finite set H_i of decision nodes, for every decision node $h_i \in H_i$ a finite set $A(h_i)$ of available actions, and a finite set Z of terminal nodes. Perfect information is modeled by the assumption that each decision node by itself constitutes an information set. By A we denote the set of all actions, whereas H denotes the collection of all decision nodes. We assume that no chance moves occur. The definition of a strategy we shall employ coincides with the concept of a *plan of action,* as discussed in Rubinstein (1991). The difference with the usual definition is that we require a strategy only to prescribe an action at those decision nodes that the same strategy does not avoid. Formally, let $\tilde{H}_i \subseteq H_i$ be a collection of player i decision nodes, not necessarily containing all decision nodes, and let $s_i : \tilde{H}_i \to A$ be a mapping prescribing at every $h_i \in \tilde{H}_i$ some available action $s_i(h_i) \in A(h_i)$. For a given decision node $h \in H$, not necessarily belonging to player i, we say that s_i *avoids* h if there is some $h_i \in \tilde{H}_i$ on the path to

h at which the prescribed action $s_i(h_i)$ deviates from the path to h. Such a mapping $s_i : \tilde{H}_i \to A$ is called a *strategy* for player i if $\tilde{H}_i$ is exactly the collection of player i decision nodes not avoided by s_i. Obviously, every strategy s_i can be obtained by first prescribing an action at all player i decision nodes, that is, constructing a strategy in the classical sense, and then deleting those player i decision nodes that are avoided by it. For a given strategy $s_i \in S_i$, we denote by $H_i(s_i)$ the collection of player i decision nodes that are not avoided by s_i. Let S_i be the set of player i strategies. For a given decision node $h \in H$ and player i, we denote by $S_i(h)$ the set of player i strategies that do not avoid h. Then, it is clear that a profile $(s_i)_{i \in I}$ of strategies reaches a decision node h if and only if $s_i \in S_i(h)$ for all players i.

3.2. Types

We shall now formally model the players in the extensive form structure $\mathcal{S}$ as decision makers under uncertainty. Our primary assumption is that every player i holds a strict, complete and transitive preference relation P_i over the set of terminal nodes, and holds at the beginning of the game, as well as at every decision node $h_i \in H_i$, a conditional belief about the opponents' strategy choices and the opponents' preference relations over terminal nodes. Throughout this paper, whenever we write "preference relation over terminal nodes", we always assume that it is strict, complete and transitive. On top of this we assume that every player, throughout the game, holds a conditional belief about the opponents' conditional beliefs about the other players' strategy choices and a conditional belief about the opponents' conditional beliefs about the other players' preference relations over terminal nodes. Moreover, each player also holds, at every instance, a conditional belief about the opponents' conditional beliefs about the other players' conditional beliefs about their opponents' strategy choices and preference relations over terminal nodes, and so on. Repeating this argument leads to infinite hierarchies of conditional beliefs.

Similarly to Ben-Porath (1997), Battigalli and Siniscalchi (1999) and Perea (2006), we model such hierarchies of conditional beliefs by means of *epistemic types*. We differ, however, in our notion of belief. Whereas the aforementioned papers work with probabilistic beliefs, we use the simpler notion of single-valued possibility sets to express the players' beliefs. By the latter we mean that a player, at each of his decision nodes, only deems possible one strategy choice and one preference relation for every opponent. That is, the first-order belief of every player is single-valued. Not only this, we also require that a player, at each of his decision nodes, only deems possible one single-valued first-order belief for every opponent. So, also the second-order belief of a player should be single-valued. Similarly, we also require that a player, for every k, only deems possible one single-valued kth-order belief for every opponent at each of his decision nodes.

We make this assumption in order to keep our model and definitions as simple as possible. In Section 8 we show how the model could be extended to multi-valued possibility sets without affecting the main results. For the formal representation of our epistemic model we need some terminology. Let h_0 be the decision node that marks the beginning of the game, and let $H_i^* = H_i \cup \{h_0\}$. By $\mathcal{P}$ we denote the set of strict, complete and transitive preference relations over the terminal nodes.

Definition 3.1 (*Epistemic Model*). An epistemic model in our setting is a tuple

$$(T_i, P_i, (s_{ij})_{j \neq i}, (t_{ij})_{j \neq i})_{i \in I}$$

where, for every player i, (1) T_i is a nonempty set, (2) P_i is a function from T_i to $\mathcal{P}$, and for every opponent j, (3) s_{ij} is a function from T_i to $\times_{h_i \in H_i^*} S_j(h_i)$, and (4) t_{ij} is a function from T_i to $\times_{h_i \in H_i^*} T_j$.

The interpretation is that T_i represents the set of types for player i, $P_i(t_i)$ is t_i's preference relation over terminal nodes, $s_{ij}(t_i)$ is t_i's conditional belief vector about player j's strategy choice, and $t_{ij}(t_i)$ is t_i's conditional belief vector about player j's type. Let $s_{ij}(t_i, h_i) \in S_j(h_i)$ denote t_i's conditional belief at h_i about j's strategy choice, and let $t_{ij}(t_i, h_i) \in T_j$ be t_i's conditional belief at h_i about j's type. In order to reduce notation, we write $s_j(t_i, h_i)$ instead of $s_{ij}(t_i, h_i)$, and $t_j(t_i, h_i)$ instead of $t_{ij}(t_i, h_i)$. This cannot cause any confusion, since the index i in $s_j(t_i, h_i)$ and $t_j(t_i, h_i)$ already indicates that these beliefs belong to player i.

We say that the epistemic model is *complete* if for every player i, every possible preference relation over terminal nodes, and every possible conditional belief vector about the opponents' strategy-type pairs there is a type for player i with these characteristics. More formally, we have the following definition:

Definition 3.2 (*Complete Epistemic Model*). An epistemic model

$$(T_i, P_i, (s_{ij})_{j \neq i}, (t_{ij})_{j \neq i})_{i \in I}$$

is *complete* if for every player i, every $\tilde{P}_i \in \mathcal{P}$, every opponent j, every $\tilde{s}_{ij} \in \times_{h_i \in H_i^*} S_j(h_i)$, and every $\tilde{t}_{ij} \in \times_{h_i \in H_i^*} T_j$ there is some $t_i \in T_i$ with $P_i(t_i) = \tilde{P}_i$, $s_{ij}(t_i) = \tilde{s}_{ij}$ for all opponents j, and $t_{ij}(t_i) = \tilde{t}_{ij}$ for all opponents j.

As is well-known, the existence of complete epistemic models is a nontrivial problem. In the Appendix, however, we show how to construct such a complete epistemic model. In the remainder of this paper, whenever we speak about an epistemic model, we assume that it is complete.

The reason we insist on a *complete* epistemic model is because we need it later for our definition of MBR. Intuitively, this condition states that, whenever player i revises his belief about player j's type upon observing a new move a by player j, then he should look for the player j type that (1) rationalizes the newly observed move a, and (2) which is "as similar as possible" to the previous belief that player i had about player j's type. So, when player i searches for the most similar type for player j that rationalizes the newly observed move a, player i should consider *all possible preference relations over terminal nodes and all possible belief hierarchies* that player j could possibly have. That is, we need a complete epistemic model.[1]

Notice that from our epistemic model we can derive for every type a belief hierarchy. Since every type t_i holds conditional beliefs about player j's type, and since player j's type specifies player j's preference relation over terminal nodes, one can derive t_i's conditional belief vector about player j's preference relation over terminal nodes. In particular, t_i may change his belief about j's preference relation over terminal nodes as the game proceeds. Since player j's type also specifies j's conditional belief vector about the opponents' strategy choices, one can derive t_i's conditional belief vector about j's conditional beliefs about the opponents' strategy choices, and so on.

3.3. Belief and common belief

Fix an epistemic model. By $T = \cup_{i \in I} T_i$ we denote the collection of all types for all players. Let $E \subseteq T$ be some subset of types, and let t_i be a specific type for player i. We say that t_i

[1] We do not need a *universal* type space, though. (Universality means that every epistemic model can be mapped into the model by a beliefs-preserving morphism.)

initially believes E if $t_j(t_i, h_0) \in E$ for every opponent j. That is, t_i believes at the beginning of the game that the opponents' types belong to E. We say that t_i *structurally believes* E if $t_j(t_i, h_i) \in E$ for every opponent j and every $h_i \in H_i^*$. In words, t_i believes at every instance of the game that the opponents' types belong to E. We recursively define

$$B^1(E) = \{t \in E | t \text{ structurally believes } E\}$$

and

$$B^k(E) = \{t \in B^{k-1}(E) | t \text{ structurally believes } B^{k-1}(E)\}$$

for every $k \geq 2$. We say that t_i respects *common structural belief* in E if $t \in B^k(E)$ for every k. Hence, t_i belongs to E, believes at every instance that all opponents' types belong to E, believes at every instance that all opponents' types believe at every instance that all other players' types belong to E, and so on.

4. Restrictions on belief revision policies

If a player currently believes that an opponent chooses a certain strategy, but finds out later that his opponent has not, he must revise his belief about the opponent's strategy choice. The first assumption we make is that players believe, at each of their decision nodes, that opponents choose sequentially rational strategies. Consequently, if a player changes his belief about an opponent's strategy choice, he must in general also change his belief about the opponent's type, since the newly believed opponent's strategy should be sequentially rational for the newly believed opponent's type. It may be necessary, for instance, to change the belief about the opponent's preference relation over terminal nodes as to rationalize the newly observed move. The player may also decide to change his belief about the opponent's conditional beliefs. The second assumption we make is that players should revise their belief about an opponent's strategy choice and type in some minimal way. That is, if player i currently believes that player j has type t_j, but finds out later (perhaps surprisingly) that decision node h_i has been reached, he should look for a player j's type t_j' that (1) has a sequentially rational strategy that leads to h_i, and (2) is as "similar" to t_j as possible. By the latter, we mean that t_j' should disagree with t_j on as few "statements" as possible, and, moreover, the preference relations of t_j' and t_j over terminal nodes should disagree on as few pairwise rankings as possible.

We shall now formalize these two assumptions within our epistemic model. We start by defining sequentially rational strategies. Choose a strategy s_i and a type t_i for player i. Recall that $H_i(s_i)$ is the set of player i decision nodes that are not avoided by s_i. At a given decision node $h_i \in H_i(s_i)$, let $z(s_i, t_i, h_i)$ denote the terminal node that is reached if the game would start at h_i, player i would choose according to s_i, and every opponent j would choose according to the conditional belief $s_j(t_i, h_i)$ that t_i holds at h_i about j's strategy choice.

Definition 4.1 (*Sequentially Rational Strategy*). Strategy s_i is *sequentially rational* for type t_i if for every decision node $h_i \in H_i(s_i)$ there is no strategy $s_i' \in S_i(h_i)$ such that $P_i(t_i)$ strictly prefers the terminal node $z(s_i', t_i, h_i)$ to the terminal node $z(s_i, t_i, h_i)$.

If we would use the "classical" definition of a strategy, prescribing an action at *every* decision node, then the definition above would coincide with that of a *weakly sequentially rational* strategy (see Reny (1992b)). Weak sequential rationality only requires the player's behavior to be optimal at those information sets that can actually be reached by the strategy at hand, but not necessarily at information sets which the strategy itself avoids.

We may now formalize what it means that a type always believes that his opponents choose sequentially rational strategies.

Definition 4.2 (*Structural Belief in Sequential Rationality*). Type t_i *structurally believes in sequential rationality* if for every $h_i \in H_i^*$ and every opponent j it holds that $s_j(t_i, h_i)$ is sequentially rational for $t_j(t_i, h_i)$.

In order to introduce our notion of minimal belief revision, we must define a similarity relation between types. This similarity relation is based on two components: (1) comparing the "statements about a player" between types, and (2) comparing the preference relations over terminal nodes between types.

Definition 4.3 (*Statement About a Player*). A *first-order statement* about player i is either a statement of the type "player i has preference relation $\tilde{P}_i$", or a statement of the type "player i believes at h_i that player j chooses $s_j \in S_j(h_i)$". Assuming that $(k-1)$th-order statements about player j have been defined for every player j, a *kth-order statement* about player i is either a $(k-1)$th-order statement about player i, or a statement of the type "player i believes at h_i that φ", where φ is a $(k-1)$th-order statement about some player $j \neq i$. A *statement* about player i is a kth-order statement about player i for some k.

We say that two types $t_i, t_i' \in T_i$ *disagree* on an statement φ about player i if φ is true at t_i but not at t_i', or vice versa. Now, consider two preference relations P^1 and P^2 over terminal nodes. For every pair of terminal nodes $\{z, z'\}$, we say that P^1 and P^2 *disagree on the pairwise ranking of* $\{z, z'\}$ if P^1 ranks z over z' and P^2 ranks z' over z, or vice versa.

Definition 4.4 (*Similarity Between Types*). Consider three types t_i, t_i' and t_i'' for player i. We say that t_i' *is more similar to* t_i *than* t_i'' if and only if (1) t_i' disagrees with t_i on fewer statements about player i than t_i'' does, or (2) t_i' disagrees with t_i on as many statements about player i as t_i'' does, but $P_i(t_i')$ disagrees with $P_i(t_i)$ on fewer pairwise rankings than $P_i(t_i'')$ does.

In this definition, every statement about player i carries equal weight. Hence, an important implicit assumption we make in this notion of similarity is that all beliefs of *any order* are viewed as "equally important". That is, the belief that player i has about player j's strategy choice is considered "as important" as player i's belief about player j's belief about the other players' strategy choices. This assumption seems natural once we impose common structural belief in SBSR, since in this case player i's belief about player j's belief about his opponents' strategies serves as a justification for player i's belief about player j 's strategy choice. Common structural belief in SBSR implies, namely, that player i should believe that player j's strategy choice is optimal given player i's belief about player j's preference relation over terminal nodes, and given player i's belief about player j's conditional beliefs about the opponents' strategy choices. Hence, player i's second-order beliefs *justify* player i's first-order beliefs about the opponents' strategy choices, and therefore both beliefs may be viewed as "equally important". Similarly, common structural belief in SBSR implies that player i's kth-order beliefs justify his $(k-1)$th-order beliefs about the opponents' strategy choices for any k. For this reason, we assume that beliefs of all possible orders are viewed as "equally important" in our model, thereby justifying the notion of similarity as it is stated.

In order to define minimal belief revision, we need some more terminology. For every decision node h and every player i, let $T_i^{sr}(h)$ be the set of types for player i for which there is a sequentially rational strategy in $S_i(h)$. That is, if player i structurally believes in sequential rationality, and finds out that some information set h_i has been reached, he should believe that every opponent j has a type in $T_j^{sr}(h_i)$. Say that decision node $h_i^2 \in H_i^*$ immediately follows $h_i^1 \in H_i^*$ if (1) h_i^2 follows h_i^1, and (2) there is no player i's decision node between h_i^1 and h_i^2.

Definition 4.5 (*Minimal Belief Revision*). Type t_i is said to satisfy *minimal belief revision* if for every two decision nodes $h_i^1, h_i^2 \in H_i^*$ such that h_i^2 immediately follows h_i^1, and for every opponent j, the type $t_j(t_i, h_i^2)$ is in $T_j^{sr}(h_i^2)$ and is most similar to $t_j(t_i, h_i^1)$ among all types in $T_j^{sr}(h_i^2)$.

5. The main results

Let $\mathcal{S}$ be an extensive form structure with perfect information, and $\tilde{P} = (\tilde{P}_i)_{i\in I}$ a profile of preference relations on the set of terminal nodes. Let t_i be a type for player i. We say that t_i *initially believes in* $\tilde{P}$ if $P_j(t_j(t_i, h_0)) = \tilde{P}_j$ for every opponent j. That is, at the beginning t_i believes that the opponents' preference relations over terminal nodes are as specified by $\tilde{P}$. Let IBP denote the event that a type initially believes in $\tilde{P}$. Moreover, let SBSR denote the event that a type structurally believes in sequential rationality, and let MBR be the event that a type satisfies minimal belief revision. The first main result states that common structural belief in the events IBP, SBSR and MBR is possible. Hence, common structural belief in these events does not lead to logical contradictions.

Theorem 5.1. *Let $\mathcal{S}$ be an extensive form structure with perfect information, and $\tilde{P} = (\tilde{P}_i)_{i\in I}$ a profile of strict, complete and transitive preference relations on the set of terminal nodes. Then, there is an epistemic model such that for every player i there is a type t_i that respects common structural belief in the events IBP, SBSR and MBR.*

If one would use the "classical" definition of a strategy, prescribing an action at *every* information set, then the theorem above would still hold.

The second main result states that common structural belief in the events IBP, SBSR and MBR leads to backward induction. In order to state this result formally, we need the following definitions. Let $\mathcal{S}$ be an extensive form structure with perfect information, and $\tilde{P} = (\tilde{P}_i)_{i\in I}$ a profile of strict, complete and transitive preference relations on the set of terminal nodes. Then, the pair $(\mathcal{S}, \tilde{P})$ may be interpreted as a *game,* and the backward induction procedure in the game $(\mathcal{S}, \tilde{P})$ leads to a unique backward induction action $a^*(h_i)$ at every decision node h_i. For every player i, let s_i^* be the unique strategy that chooses the backward induction action $a^*(h_i)$ at every $h_i \in H_i(s_i^*)$. We refer to s_i^* as the *backward induction strategy* for player i in $(\mathcal{S}, \tilde{P})$.

Theorem 5.2. *Let $\mathcal{S}$ be an extensive form structure with perfect information, and $\tilde{P} = (\tilde{P}_i)_{i\in I}$ a profile of strict, complete and transitive preference relations on the set of terminal nodes. Let t_i be a type with preference relation $\tilde{P}_i$, respecting common structural belief in the events IBP, SBSR and MBR. Then, there is a unique sequentially rational strategy for t_i, namely player i's backward induction strategy in $(\mathcal{S}, \tilde{P})$.*

If we would use the "classical" definition of a strategy, then the theorem above would still hold.

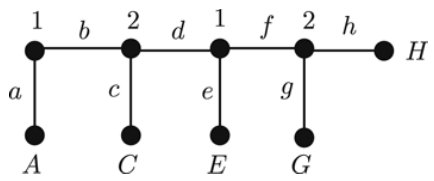

Fig. 2. Not every type has a sequentially rational strategy.

6. Implications of SBSR and MBR

Before turning to the proofs of the main results, we discuss some important implications of (common structural belief in) SBSR and MBR. Proofs of these results, except for Lemma 6.2 and Theorem 6.8, may be found in the Appendix. In the following section, we shall use these implications for proving Theorems 5.1 and 5.2.

6.1. Existence of sequentially rational strategies

It is important to note that not every type has a sequentially rational strategy. Consider, for instance, the extensive form structure in Fig. 2. Take a type t_1 for player 1 with the preference relation $HEGCA$ over the terminal nodes. Let player 1's decision nodes be denoted by h_1^1 and h_1^2, respectively. Suppose that t_1 believes at h_1^1 that player 2 chooses the strategy (d, g), but believes at h_1^2 that player 2 chooses (d, h). The unique strategy that is optimal for t_1 at h_1^1 is (b, e). However, (b, e) is not optimal for t_1 at h_1^2, which implies that t_1 has no sequentially rational strategy. The reason for this is that t_1's conditional beliefs at h_1^2 contradict Bayesian updating: t_1's belief at h_1^1 about player 2's behavior is compatible with the event of reaching h_1^2, and therefore Bayesian updating implies that t_1's belief at h_1^2 should coincide with his belief at h_1^1.

We shall now provide a formalization of the above mentioned Bayesian updating requirement and show in Lemma 6.2 that it guarantees the existence of a sequentially rational strategy. Let h_i^1 and h_i^2 be two decision nodes in H_i^* such that h_i^2 immediately follows h_i^1.

Definition 6.1. We say that t_i satisfies *Bayesian updating* at h_i^2 if for every opponent j for which $s_j(t_i, h_i^1) \in S_j(h_i^2)$, it holds that $s_j(t_i, h_i^2) = s_j(t_i, h_i^1)$.

In other words, if t_i's belief at h_i^1 about player j's strategy choice does not contradict the event of reaching h_i^2, then t_i should maintain at h_i^2 his previous belief about player j's strategy choice. We say that t_i satisfies Bayesian updating if it does so at every decision node.

Lemma 6.2. *Every type that satisfies Bayesian updating has a sequentially rational strategy.*

The above result follows from the claim in Battigalli (1997) on p. 54, and hence we do not provide a proof here.

Lemma 6.3. *Let t_i be a type that satisfies SBSR and MBR. Then, t_i satisfies Bayesian updating.*

By combining Lemmas 6.2 and 6.3, we obtain the following corollary.

Corollary 6.4. *Let t_i be a type that satisfies SBSR and MBR. Then, t_i has a sequentially rational strategy.*

6.2. Maintaining "beliefs about beliefs"

We next show that common structural belief in SBSR and MBR implies that a player never changes his belief about an opponent's beliefs about other players. In order to state the result formally, we need the following definition. For a given type $t_i \in T_i$ and preference relation $\tilde{P}_i$ over the terminal nodes, let $(t_i, \tilde{P}_i)$ denote the type that has preference relation $\tilde{P}_i$ and holds the same conditional beliefs about the opponents' strategies and types as t_i. That is, $s_j((t_i, \tilde{P}_i), h_i) = s_j(t_i, h_i)$ and $t_j((t_i, \tilde{P}_i), h_i) = t_j(t_i, h_i)$ for every $j \neq i$ and every $h_i \in H_i^*$.

Lemma 6.5. *Suppose that $t_i \in T_i$ respects common structural belief in SBSR and MBR. Let $h_i^1, h_i^2 \in H_i^*$ be such that h_i^2 immediately follows h_i^1, let j be an opponent, let $t_j^1 := t_j(t_i, h_i^1)$ and $t_j^2 := t_j(t_i, h_i^2)$. Then, $t_j^2 = (t_j^1, \tilde{P}_j)$ for some preference relation $\tilde{P}_j$ over terminal nodes.*

6.3. Relation with proper belief revision

We next prove that common structural belief in SBSR and MBR leads to *proper belief revision*: a concept that has been put forward in Perea (2006, 2007a). This result will prove to be crucial for proving Theorem 5.2. Informally, proper belief revision states that a player who wishes to revise his belief at decision node h about opponent j, should not change his belief about the opponent's relative ranking of two strategies s_j and s_j' if both s_j and s_j' could have led to h. The intuition is that the player, upon arriving at h, cannot exclude any of the opponent's strategies s_j and s_j', and therefore there is no reason for him to change his belief about the opponent's relative ranking of s_j and s_j'. In order to introduce proper belief revision formally, we need some more notation and definitions. Let t_i be a type for player i, and $h_i \in H_i^*$ some decision node. For a given strategy $s_i \in S_i(h_i)$, recall that $z(s_i, t_i, h_i)$ denotes the terminal node that would be reached if the game would start at h_i, player i would choose according to s_i, and every opponent j would choose according to $s_j(t_i, h_i)$. For two strategies $s_i, s_i' \in S_i(h_i)$, we say that t_i strictly prefers strategy s_i to strategy s_i' at decision node h_i if t_i strictly prefers the terminal node $z(s_i, t_i, h_i)$ to the terminal node $z(s_i', t_i, h_i)$. Now, let t_i be a type for player i, let $j \neq i$ be an opponent, let h_i and h_j be decision nodes for players i and j, respectively, and let s_j, s_j' be two strategies for player j in $S_j(h_j)$. We say that t_i believes at h_i that player j at h_j strictly prefers strategy s_j to strategy s_j' if type $t_j(t_i, h_i)$ strictly prefers s_j to s_j' at h_j.

Now, let t_i be a type for player i, and let h_i^1, h_i^2 be two decision nodes in H_i^* such that h_i^2 immediately follows h_i^1.

Definition 6.6 (*Proper Belief Revision*). We say that t_i satisfies *proper belief revision* at h_i^2 if for every opponent j, every decision node $h_j \in H_j$ and every two strategies s_j, s_j' that belong to both $S_j(h_j)$ and $S_j(h_i^2)$ the following holds: t_i believes at h_i^2 that player j at h_j strictly prefers s_j to s_j' if and only if t_i believes so at h_i^1.

Note that $s_j, s_j' \in S_j(h_i^2)$ implies that both s_j and s_j' could have led to h_i^2. We say that type t_i satisfies proper belief revision if t_i does so at each of his decision nodes.

Before showing that common structural belief in SBSR and MBR implies proper belief revision, we prove the following lemma. It states that the number of pairwise rankings on which two preference relations P^1 and P^2 over terminal nodes disagree can be reduced strictly by applying the following procedure: First, take a pair $\{a, b\}$ of terminal nodes on which P^1 and P^2

disagree, and then interchange the roles of a and b in P^2 without changing the roles of the other nodes.

Lemma 6.7. *Let P^1 and P^2 be two strict, complete and transitive preference relations on the set of terminal nodes, and let $\{a, b\}$ be a pair of terminal nodes on which P^1 and P^2 disagree. Let u^2 be an arbitrary utility representation of P^2, and let the utility function $\tilde{u}^2$ be given by*

$$\tilde{u}^2(z) = \begin{cases} u^2(b), & \text{if } z = a, \\ u^2(a), & \text{if } z = b, \\ u^2(z), & \text{otherwise.} \end{cases}$$

Let $\tilde{P}^2$ be the preference relation induced by $\tilde{u}_2$. Then, P^1 and $\tilde{P}^2$ disagree on less pairwise rankings than P^1 and P^2.

We are now able to prove the following result.

Theorem 6.8. *Let t_i be a type that respects common structural belief in SBSR and MBR. Then, t_i satisfies proper belief revision.*

Proof. For a given type $t_i \in T_i$ and decision node $h_i \in H_i^*$, let

$$Z(t_i, h_i) = \{z(s_i, t_i, h_i) | s_i \in S_i(h_i)\}$$

be the set of terminal nodes that can be reached if the game would start at h_i and every opponent j of player i would act according to $s_j(t_i, h_i)$.

Let t_i be a type for player i that respects common structural belief in SBSR and MBR. We prove that t_i satisfies proper belief revision. Suppose, contrary to what we want to prove, that t_i does not satisfy proper belief revision. Then, there must be two decision nodes $h_i^1, h_i^2 \in H_i^*$ such that h_i^2 immediately follows h_i^1, an opponent j, a decision node $h_j^* \in H_j$ and two strategies $s_j, s_j' \in S_j(h_j^*) \cap S_j(h_i^2)$ such that: t_i believes at h_i^1 that player j strictly prefers s_j to s_j' at h_j^*, but does not believe so at h_i^2.[2] Let $t_j^1 = t_j(t_i, h_i^1)$ and $t_j^2 = t_j(t_i, h_i^2)$, and let P_j^1 and P_j^2 denote the preference relations of t_j^1 and t_j^2 over terminal nodes. Since t_i respects common structural belief in SBSR and MBR, Lemma 6.5 guarantees that t_j^1 and t_j^2 hold the same conditional beliefs. In particular, $s_k(t_j^1, h_j^*) = s_k(t_j^2, h_j^*)$ for every $k \neq j$.

Since t_i believes at h_i^1 that player j strictly prefers s_j to s_j' at h_j^*, but does not believe so at h_i^2, we may conclude that P_j^1 strictly prefers $z(s_j, t_j^1, h_j^*)$ to $z(s_j', t_j^1, h_j^*)$, but P_j^2 strictly prefers $z(s_j', t_j^1, h_j^*)$ to $z(s_j, t_j^1, h_j^*)$. Let u_j^2 be some arbitrary utility representation of P_j^2, and let the utility function $\tilde{u}_j^2$ be given by

$$\tilde{u}_j^2(z) = \begin{cases} u_j^2(z(s_j', t_j^1, h_j^*)), & \text{if } z = z(s_j, t_j^1, h_j^*), \\ u_j^2(z(s_j, t_j^1, h_j^*)), & \text{if } z = z(s_j', t_j^1, h_j^*), \\ u_j^2(z), & \text{otherwise.} \end{cases} \tag{6.1}$$

[2] Note that if t_i believes at h_i^1 that player j is indifferent at h_j^* between s_j and s_j', then necessarily $z(s_j, t_j(t_i, h_i^1), h_j^*) = z(s_j', t_j(t_i, h_i^1), h_j^*)$. By Lemma 6.5, we have that $t_j(t_i, h_i^1)$ and $t_j(t_i, h_i^2)$ hold the same conditional beliefs, and hence $z(s_j, t_j(t_i, h_i^2), h_j^*) = z(s_j', t_j(t_i, h_i^2), h_j^*)$, which implies that t_i believes at h_i^2 that player j is indifferent between s_j and s_j'.

Let $\tilde{P}_j^2$ be the preference relation induced by $\tilde{u}_j^2$. Since P_j^1 and P_j^2 disagree on $\{z(s_j, t_j^1, h_j^*), z(s_j', t_j^1, h_j^*)\}$, we know by Lemma 6.7 that P_j^1 and $\tilde{P}_j^2$ disagree on less pairwise rankings than P_j^1 and P_j^2.

Let $\tilde{t}_j^2 := (t_j^1, \tilde{P}_j^2)$. We now prove that $\tilde{t}_j^2$ has a sequentially rational strategy $\tilde{s}_j^2 \in S_j(h_i^2)$. Since t_i respects common structural belief in SBSR and MBR, Lemma 6.3 guarantees that t_i respects common structural belief in the event that types satisfy Bayesian updating. Since $t_j^1 = t_j(t_i, h_i^1)$, and t_i believes that player j satisfies Bayesian updating, it follows that t_j^1 satisfies Bayesian updating. Since t_j^2 and $\tilde{t}_j^2$ have the same conditional beliefs as t_j^1, we have that also t_j^2 and $\tilde{t}_j^2$ satisfy Bayesian updating. By Lemma 6.2 we know that t_j^2 and $\tilde{t}_j^2$ have a sequentially rational strategy, which must then be unique. Let s_j^2 and $\tilde{s}_j^2$ be the unique sequentially rational strategies for types t_j^2 and $\tilde{t}_j^2$, respectively. Recall that, by definition, $t_j^2 = t_j(t_i, h_i^2)$. By SBSR, $s_j^2 \in S_j(h_i^2)$. We now show that $\tilde{s}_j^2 \in S_j(h_i^2)$.

For every $h_j \in H_j$ preceding h_i^2, let $a(h_j, h_i^2)$ be the unique action at h_j leading to h_i^2. In order to show that $\tilde{s}_j^2 \in S_j(h_i^2)$, we prove that $\tilde{s}_j^2(h_j) = a(h_j, h_i^2)$ for all $h_j \in H_j(\tilde{s}_j^2)$ preceding h_i^2. Choose some $h_j \in H_j(\tilde{s}_j^2)$ preceding h_i^2. As $s_j^2 \in S_j(h_i^2)$, we have that $h_j \in H_j(s_j^2)$ and $s_j^2(h_j) = a(h_j, h_i^2)$. By assumption, s_j^2 is sequentially rational for $t_j^2 = (t_j^1, P_j^2)$, which means in particular that s_j^2 is optimal for t_j^2 at h_j. Hence, P_j^2 strictly prefers $z(s_j^2, t_j^1, h_j)$ to all other nodes in $Z(t_j^1, h_j)$. We distinguish two cases.

Case 1. Suppose that $z(s_j^2, t_j^1, h_j) \neq z(s_j', t_j^1, h_j^*)$. Recall that P_j^1 strictly prefers $z(s_j, t_j^1, h_j^*)$ to $z(s_j', t_j^1, h_j^*)$, but that P_j^2 strictly prefers $z(s_j', t_j^1, h_j^*)$ to $z(s_j, t_j^1, h_j^*)$. Since P_j^2 strictly prefers $z(s_j^2, t_j^1, h_j)$ to all other nodes in $Z(t_j^1, h_j)$, and $z(s_j^2, t_j^1, h_j) \neq z(s_j', t_j^1, h_j^*)$, we have by (6.1) that $\tilde{P}_j^2$ also strictly prefers $z(s_j^2, t_j^1, h_j)$ to all other nodes in $Z(t_j^1, h_j)$. This implies that s_j^2 is optimal for $\tilde{t}_j^2$ at h_j. Since we know that $\tilde{s}_j^2$ is optimal for $\tilde{t}_j^2$ at h_j, it follows that $\tilde{s}_j^2(h_j) = s_j^2(h_j) = a(h_j, h_i^2)$, which was to show.

Case 2. Suppose that $z(s_j^2, t_j^1, h_j) = z(s_j', t_j^1, h_j^*)$. In this case, the terminal node $z(s_j^2, t_j^1, h_j)$ follows both h_j and h_j^*. Hence, it must be the case that h_j precedes or follows h_j^*. We distinguish two subcases.

Case 2.1. Suppose that h_j precedes h_j^*. Since $z(s_j^2, t_j^1, h_j)$ follows h_j^*, it must be the case that $s_k(t_j^1, h_j) \in S_k(h_j^*)$ for every $k \neq j$. We have seen above that t_j^1 satisfies Bayesian updating, which then implies that $s_k(t_j^1, h_j^*) = s_k(t_j^1, h_j)$ for every $k \neq j$. As $s_j \in S_j(h_j^*)$, it follows that $s_j \in S_j(h_j)$ and that $z(s_j, t_j^1, h_j) = z(s_j, t_j^1, h_j^*)$. Since P_j^2 strictly prefers $z(s_j^2, t_j^1, h_j) = z(s_j', t_j^1, h_j^*)$ to all other nodes in $Z(t_j^1, h_j)$, it follows by (6.1) that $\tilde{P}_j^2$ strictly prefers $z(s_j, t_j^1, h_j) = z(s_j, t_j^1, h_j^*)$ to all other nodes in $Z(t_j^1, h_j)$. Hence, s_j is optimal for $\tilde{t}_j^2$ at h_j. Since, by assumption, $\tilde{s}_j^2$ is optimal for $\tilde{t}_j^2$ at h_j, it follows that $\tilde{s}_j^2(h_j) = s_j(h_j)$. Since $s_j \in S_j(h_i^2)$, we have that $s_j(h_j) = a(h_j, h_i^2)$. Hence, $\tilde{s}_j^2(h_j) = a(h_j, h_i^2)$, which was to show.

Case 2.2. Suppose that h_j^* precedes h_j. As $z(s_j', t_j^1, h_j^*) = z(s_j^2, t_j^1, h_j)$ follows h_j, we must have that $s_k(t_j^1, h_j^*) \in S_k(h_j)$ for every $k \neq j$. By Bayesian updating of t_j^1, we may then conclude that $s_k(t_j^1, h_j) = s_k(t_j^1, h_j^*)$ for every $k \neq j$. Since $s_j \in S_j(h_i^2)$ and h_j precedes h_i^2, we have that $s_j \in S_j(h_j)$ as well. Combined with the fact that $s_k(t_j^1, h_j) = s_k(t_j^1, h_j^*)$

for every $k \neq j$, this implies that $z(s_j, t_j^1, h_j) = z(s_j, t_j^1, h_j^*)$. Since P_j^2 strictly prefers $z(s_j^2, t_j^1, h_j) = z(s_j', t_j^1, h_j^*)$ to all other nodes in $Z(t_j^1, h_j)$, it follows by (6.1) that $\tilde{P}_j^2$ strictly prefers $z(s_j, t_j^1, h_j) = z(s_j, t_j^1, h_j^*)$ to all other nodes in $Z(t_j^1, h_j)$. We may thus conclude that s_j is optimal for $\tilde{t}_j^2$ at h_j. As $\tilde{s}_j^2$ is optimal for $\tilde{t}_j^2$ at h_j as well, it follows that $\tilde{s}_j^2(h_j) = s_j(h_j)$. By assumption, $s_j \in S_j(h_i^2)$, implying that $s_j(h_j) = a(h_j, h_i^2)$. Hence, we may conclude that $\tilde{s}_j^2(h_j) = a(h_j, h_i^2)$, which was to show.

From Case 1 and 2 we may therefore conclude that $\tilde{s}_j^2(h_j) = a(h_j, h_i^2)$ for all decision nodes $h_j \in H_j(\tilde{s}_j^2)$ preceding h_i^2. This, in turn, implies that $\tilde{s}_j^2 \in S_j(h_i^2)$. Summarizing, we have found a strategy-type pair $(\tilde{s}_j^2, \tilde{t}_j^2)$ with $\tilde{s}_j^2 \in S_j(h_i^2)$ such that (1) $\tilde{s}_j^2$ is sequentially rational for $\tilde{t}_j^2$, (2) $\tilde{t}_j^2$ has the same conditional beliefs as t_j^2, and (3) $P_j(t_j^1)$ and $P_j(\tilde{t}_j^2)$ disagree on less pairwise rankings than $P_j(t_j^1)$ and $P_j(t_j^2)$. This, however, is a contradiction to the assumption that t_i satisfies MBR. Therefore, the assumption that t_i does not satisfy proper belief revision cannot be true. Hence, t_i must satisfy proper belief revision. This completes the proof of our theorem. ■

7. Proof of the main results

7.1. Proof of Theorem 5.1

For every player i, decision node $h_i \in H_i^*$ and opponent $j \neq i$, let $s_j^*(h_i)$ be the unique strategy for player j with the following properties: (1) at every decision node $h_j \in H_j(s_j^*(h_i))$ preceding h_i, the strategy $s_j^*(h_i)$ prescribes the unique action that leads to h_i, and (2) at every decision node $h_j \in H_j(s_j^*(h_i))$ not preceding h_i, it prescribes the backward induction action $a^*(h_j)$ in the game $(\mathcal{S}, \tilde{P})$. Then, by construction, $s_j^*(h_i)$ is a strategy in $S_j(h_i)$. Moreover, $s_j^*(h_0)$ coincides with the backward induction strategy s_j^* in $(\mathcal{S}, \tilde{P})$.

For every player i, denote by β_i the conditional belief vector about the opponents' strategy choices in which player i, at every decision node $h_i \in H_i^*$, believes that each opponent j chooses the strategy $s_j^*(h_i) \in S_j(h_i)$. By construction, the unique strategy that is sequentially rational for player i with respect to the conditional belief vector β_i and the preference relation $\tilde{P}_i$ is his backward induction strategy s_i^* in $(\mathcal{S}, \tilde{P})$.

Fix a player i and an opponent $j \neq i$. For every decision node $h_i \in H_i^*$ we shall define a conditional belief $P_j(h_i)$ for player i about player j's preference relation over the terminal nodes. We proceed recursively, starting from h_0. At h_0, let $P_j(h_0) = \tilde{P}_j$. Now, take a decision node $h_i^2 \in H_i^*$ and suppose that $P_j(h_i^1)$ has already been defined for all $h_i^1 \in H_i^*$ that precede h_i^2. Let h_i^1 be the unique decision node in H_i^* that immediately precedes h_i^2. By assumption, $P_j(h_i^1)$ has already been defined. We can now choose a preference relation $P_j(h_i^2)$ with the following properties: (1) there is a strategy $s_j \in S_j(h_i^2)$ that is sequentially rational with respect to the conditional belief vector β_j and the preference relation $P_j(h_i^2)$ over the terminal nodes, and (2) there is no preference relation $\hat{P}_j(h_i^2)$ satisfying (1) that disagrees with $P_j(h_i^1)$ on less pairwise rankings than $P_j(h_i^2)$ does. In this way, a conditional belief $P_j(h_i)$ about player j's preference relation over terminal nodes can be defined for every player i, every opponent j, and every decision node $h_i \in H_i^*$.

We may now construct an epistemic model, and subsequently select a set of types

$$T^* = \{t_j(h_i) | i, j \in I, i \neq j \text{ and } h_i \in H_i^*\}$$

with the following properties:

(1) the preference relation over terminal nodes for $t_j(h_i)$ is equal to $P_j(h_i)$;
(2) the conditional belief vector of $t_j(h_i)$ about the opponents' strategy choices is given by β_j;
(3) the conditional belief of $t_j(h_i)$ at decision node $h_j \in H_j^*$ about opponent k's type is equal to $t_k(h_j)$.

We now prove that every type $t_j(h_i) \in T^*$ respects common structural belief in IBP, SBSR and MBR. By construction, every type $t \in T^*$ believes, at each of his decision nodes, that each of his opponents' types belongs to T^*. It is therefore sufficient to show that every type $t_j(h_i) \in T^*$ satisfies IBP, SBSR and MBR

IBP. Choose an arbitrary type $t_j(h_i) \in T^*$. By definition, $t_j(h_i)$ believes at h_0 that every opponent k is of type $t_k(h_0)$. Since $t_k(h_0)$ has preference relation $P_k(h_0)$ over terminal nodes and since, by construction, $P_k(h_0) = \tilde{P}_k$, we have that $t_j(h_i)$ believes at h_0 that every opponent k has preference relation $\tilde{P}_k$ over terminal nodes. Hence, $t_j(h_i)$ satisfies IBP.

SBSR. Let $t_j(h_i) \in T^*$ and let $k \neq j$ be some opponent. We first show that $t_j(h_i)$ *initially* believes in sequential rationality. By definition, $t_j(h_i)$ believes at h_0 that player k chooses strategy $s_k^*(h_0)$ and has type $t_k(h_0)$. Since type $t_k(h_0)$'s conditional belief vector about the opponents' strategies is β_k, type $t_k(h_0)$'s preference relation over terminal nodes is $\tilde{P}_k$ and $s_k^*(h_0)$ is player k's backward induction strategy in $(\mathcal{S}, \tilde{P})$, it follows that $s_k^*(h_0)$ is sequentially rational for $t_k(h_0)$. Hence, $t_j(h_i)$ initially believes in sequential rationality.

In order to prove that $t_j(h_i)$ *structurally* believes in sequential rationality, we need the following claim.

Claim. *Type $t_j(h_i)$ satisfies proper belief revision.*

Proof of the claim. Suppose, contrary to what we want to prove, that $t_j(h_i)$ does not satisfy proper belief revision. Then, there must be two decision nodes $h_j^1, h_j^2 \in H_j^*$ such that h_j^2 immediately follows h_j^1, an opponent k, a decision node $h_k^* \in H_k$ and two strategies $s_k, s_k' \in S_k(h_k^*) \cap S_k(h_j^2)$ such that: $t_j(h_i)$ believes at h_j^1 that player k strictly prefers s_k to s_k' at h_k^*, but does not believe so at h_j^2. By definition, $t_j(h_i)$ believes at h_j^1 that player k is of type $t_k(h_j^1)$, while he believes at h_j^2 that player k is of type $t_k(h_j^2)$. Moreover, $t_k(h_j^1)$ and $t_k(h_j^2)$ only differ by their preference relations over terminal nodes, $P_k(h_j^1)$ and $P_k(h_j^2)$. Hence, similarly to the proof of Theorem 6.8, we may conclude that $P_k(h_j^1)$ strictly prefers $z(s_k, t_k(h_j^1), h_k^*)$ to $z(s_k', t_k(h_j^1), h_k^*)$, but $P_k(h_j^2)$ strictly prefers $z(s_k', t_k(h_j^1), h_k^*)$ to $z(s_k, t_k(h_j^1), h_k^*)$. Let u_k^2 be some arbitrary utility representation of $P_k(h_j^2)$, and let the utility function $\tilde{u}_k^2$ be given by

$$\tilde{u}_k^2(z) = \begin{cases} u_k^2(z(s_k', t_k(h_j^1), h_k^*)), & \text{if } z = z(s_k, t_k(h_j^1), h_k^*), \\ u_k^2(z(s_k, t_k(h_j^1), h_k^*)), & \text{if } z = z(s_k', t_k(h_j^1), h_k^*), \\ u_k^2(z), & \text{otherwise.} \end{cases}$$

Let $\tilde{P}_k^2$ be the preference relation induced by $\tilde{u}_k^2$. Since $P_k(h_j^1)$ and $P_k(h_j^2)$ disagree on the pairwise ranking of $\{z(s_k, t_k(h_j^1), h_k^*), z(s_k', t_k(h_j^1), h_k^*)\}$, we know by Lemma 6.7 that $P_k(h_j^1)$ and $\tilde{P}_k^2$ disagree on less pairwise rankings than $P_k(h_j^1)$ and $P_k(h_j^2)$.

By construction of $P_k(h_j^2)$, there is some strategy $s_k \in S_k(h_j^2)$ that is sequentially rational for type $(t_k(h_j^1), P_k(h_j^2))$. Similarly to the proof of Theorem 6.8, it can be shown that there is also some sequentially rational strategy $s_k' \in S_k(h_j^2)$ for type $(t_k(h_j^1), \tilde{P}_k^2)$. Hence, we have found a preference relation $\tilde{P}_k^2$ such that (1) there is some strategy $s_k' \in S_k(h_j^2)$ that is sequentially rational with respect to $\tilde{P}_k^2$ and the conditional belief vector β_j, and (2) $P_k(h_j^1)$ and $\tilde{P}_k^2$ disagree on less pairwise rankings than $P_k(h_j^1)$ and $P_k(h_j^2)$. However, this contradicts the choice of $P_k(h_j^2)$, and hence we must conclude that $t_j(h_i)$ satisfies proper belief revision. This completes the proof of the claim. □

We now show that type $t_j(h_i)$ *structurally* believes in sequential rationality. Choose a decision node $h_j \in H_j^*$ and some opponent k. By definition, $t_j(h_i)$ believes at h_j that opponent k has type $t_k(h_j)$ and chooses strategy $s_k^*(h_j)$. We prove that $s_k^*(h_j)$ is sequentially rational for $t_k(h_j)$. We do so by induction on the number of decision nodes in H_j^* that precede h_j.

Assume first that h_j is not preceded by any decision node in H_j^*, that is, $h_j = h_0$. We have seen above that $s_k^*(h_0)$ is sequentially rational for $t_k(h_0)$, and hence there is nothing left to prove here.

Now, take some decision node $h_j^2 \in H_j^* \setminus \{h_0\}$ and assume that for every $h_j^1 \in H_j^*$ preceding h_j^2 it holds that $s_k^*(h_j^1)$ is sequentially rational for $t_k(h_j^1)$. We prove that $s_k^*(h_j^2)$ is sequentially rational for $t_k(h_j^2)$. Hence, we must prove for every $h_k \in H_k(s_k^*(h_j^2))$ that $s_k^*(h_j^2)$ is optimal for $t_k(h_j^2)$ at h_k. We distinguish two cases.

Case 1. Assume that $h_k \in H_k(s_k^*(h_j^2))$ and that h_k does not precede h_j^2. Then, by definition of $s_k^*(h_j^2)$, we have that $s_k^*(h_j^2)$ prescribes the backward induction action $a^*(h_k')$ at every player k's decision node h_k' equal to or following h_k. Suppose, contrary to what we want to prove, that $s_k^*(h_j^2)$ is not optimal for $t_k(h_j^2)$ at h_k. Hence, there is some $s_k(h_j^2) \in S_k(h_k)$ such that $t_k(h_j^2)$ strictly prefers $s_k(h_j^2)$ to $s_k^*(h_j^2)$ at h_k. Now, let the strategy $\tilde{s}_k(h_j^2)$ be such that (1) $\tilde{s}_k(h_j^2)$ coincides with $s_k(h_j^2)$ at h_k and at all decision nodes in $H_k(\tilde{s}_k(h_j^2))$ following h_k, and (2) $\tilde{s}_k(h_j^2)$ coincides with $s_k^*(h_j^2)$ at all other decision nodes in $H_k(\tilde{s}_k(h_j^2))$. Since $s_k^*(h_j^2) \in S_k(h_k) \cap S_k(h_j^2)$, and h_k does not precede h_j^2, it follows that $\tilde{s}_k(h_j^2) \in S_k(h_k) \cap S_k(h_j^2)$ as well. Moreover, as $\tilde{s}_k(h_j^2)$ coincides with $s_k(h_j^2)$ in the subgame starting at h_k, we may conclude that $t_k(h_j^2)$ strictly prefers $\tilde{s}_k(h_j^2)$ to $s_k^*(h_j^2)$ at h_k. Since $t_j(h_i)$ believes at h_j^2 that player k is of type $t_k(h_j^2)$, the following holds:

$$t_j(h_i) \text{ believes at } h_j^2 \text{ that player } k,\ \text{at } h_k,\ \text{strictly prefers } \tilde{s}_k(h_j^2) \text{ to } s_k^*(h_j^2), \tag{7.1}$$

where both $\tilde{s}_k(h_j^2)$ and $s_k^*(h_j^2)$ are in $S_k(h_k) \cap S_k(h_j^2)$.

We have seen in our claim that $t_j(h_i)$ satisfies proper belief revision. Now, let h_j^1 be the unique decision node in H_j^* that immediately precedes h_j^2. Since both $\tilde{s}_k(h_j^2)$ and $s_k^*(h_j^2)$ are in $S_k(h_k) \cap S_k(h_j^2)$, proper belief revision of $t_j(h_i)$, together with (7.1), imply the following:

$$t_j(h_i) \text{ believes at } h_j^1 \text{ that player } k,\ \text{at } h_k,\ \text{strictly prefers } \tilde{s}_k(h_j^2) \text{ to } s_k^*(h_j^2). \tag{7.2}$$

As h_j^1 precedes h_j^2, and h_k does not precede h_j^2, we must have that h_k does not precede h_j^1. Hence, $s_k^*(h_j^1)$ prescribes at every player k's decision node h_k' equal to or following h_k the backward induction $a^*(h'_k)$, just as $s^*_k(h^2_j)$ does. Together with (7.2), this yields:

$t_j(h_i)$ believes at h^1_j that player k, at h_k, strictly prefers $\tilde{s}_k(h^2_j)$ to $s^*_k(h^1_j)$.

Since $t_j(h_i)$ believes at h^1_j that player k has type $t_k(h^1_j)$, it follows that $s^*_k(h^1_j)$ is not sequentially rational for $t_k(h^1_j)$, which contradicts our induction assumption that $s^*_k(h^1_j)$ is sequentially rational for $t_k(h^1_j)$. Hence, we may conclude that $s^*_k(h^2_j)$ is optimal for $t_k(h^2_j)$ at every $h_k \in H_k(s^*_k(h^2_j))$ not preceding h^2_j. This completes Case 1.

Case 2. Assume that $h_k \in H_k(s^*_k(h^2_j))$ precedes h^2_j. By our construction, $t_k(h^2_j)$ has a sequentially rational strategy $s_k(h^2_j)$ in $S_k(h^2_j)$. Suppose, contrary to what we want to prove, that $s^*_k(h^2_j)$ is not optimal for $t_k(h^2_j)$ at h_k. Then, necessarily,

$$t_k(h^2_j) \text{ strictly prefers } z(s_k(h^2_j), t_k(h^2_j), h_k) \text{ to } z(s^*_k(h^2_j), t_k(h^2_j), h_k). \tag{7.3}$$

Since $s_k(h^2_j)$ and $s^*_k(h^2_j)$ are both in $S_k(h^2_j)$, they coincide on all player k's decision nodes preceding h^2_j. Hence, by (7.3), there must be some player k 's decision node h'_k not preceding h^2_j such that (1) $s_l(t_k(h^2_j), h_k) \in S_l(h'_k)$ for every $l \neq k$, and (2) $(s_k(h^2_j), (s_l(t_k(h^2_j)))_{l\neq k}, h_k))$ and $(s^*_k(h^2_j), (s_l(t_k(h^2_j)))_{l\neq k}, h_k))$ both reach h'_k. By construction, $t_k(h^2_j)$ satisfies Bayesian updating, and hence we have that $s_l(t_k(h^2_j), h'_k) = s_l(t_k(h^2_j), h_k)$ for every $l \neq k$. This implies that

$$z(s_k(h^2_j), t_k(h^2_j), h_k) = z(s_k(h^2_j), t_k(h^2_j), h'_k) \quad \text{and}$$
$$z(s^*_k(h^2_j), t_k(h^2_j), h_k) = z(s^*_k(h^2_j), t_k(h^2_j), h'_k).$$

Together with (7.3), we may conclude that

$$t_k(h^2_j) \text{ strictly prefers } z(s_k(h^2_j), t_k(h^2_j), h'_k) \text{ to } z(s^*_k(h^2_j), t_k(h^2_j), h'_k),$$

which means that $s^*_k(h^2_j)$ is not optimal for $t_k(h^2_j)$ at h'_k. However, this contradicts our findings in Case 1, as h'_k does not precede h^2_j. Therefore, $s^*_k(h^2_j)$ must be optimal for $t_k(h^2_j)$ at h_k. This completes Case 2.

By combining the cases 1 and 2, we have shown for every $h_k \in H_k(s^*_k(h^2_j))$ that $s^*_k(h^2_j)$ is optimal for $t_k(h^2_j)$ at h_k. As such, $s^*_k(h^2_j)$ is sequentially rational for $t_k(h^2_j)$. Since $t_j(h_i)$ believes at h^2_j that player k is of type $t_k(h^2_j)$ and chooses strategy $s^*_k(h^2_j)$, and since this holds for every h^2_j and every opponent k, it follows that $t_j(h_i)$ structurally believes in sequential rationality, which was to show.

MBR. Take some decision nodes $h^1_j, h^2_j \in H^*_j$ such that h^2_j immediately follows h^1_j. By definition, $t_j(h_i)$ believes at h^1_j that player k has type $t_k(h^1_j)$ and chooses strategy $s^*_k(h^1_j)$, and believes at h^2_j that player k has type $t_k(h^2_j)$ and chooses strategy $s^*_k(h^2_j)$. We have already seen above that $s^*_k(h^2_j)$ is sequentially rational for $t_k(h^2_j)$. By construction of $t_k(h^1_j)$ and $t_k(h^2_j)$ we know that $t_k(h^1_j)$ has preference relation $P_k(h^1_j)$ over terminal nodes, that $t_k(h^2_j)$ has preference relation $P_k(h^2_j)$ over terminal nodes, and that $t_k(h^1_j)$ and $t_k(h^2_j)$ have identical conditional beliefs about the opponents' strategies and types. As such,

$$t_k(h^2_j) = (t_k(h^1_j), P_k(h^2_j)).$$

In particular, it follows that $t_k(h_j^1)$ and $t_k(h_j^2)$ disagree on at most one statement about player k, namely k's preference relation over terminal nodes.

Moreover, by construction of the preference relation $P_k(h_j^2)$, we know that there is no preference relation P_k' such that (1) P_k' and $P_k(h_j^1)$ disagree on less pairwise rankings than $P_k(h_j^2)$ and $P_k(h_j^1)$ do, and (2) the type $(t_k(h_j^1), P_k')$ has a sequentially rational strategy in $S_k(h_j^2)$. Hence, $t_j(h_i)$ satisfies MBR.

We may thus conclude that every type $t \in T^*$ satisfies IBP, SBSR and MBR. As every type $t \in T^*$ structurally believes that all opponents' types are in T^*, it holds that every type $t \in T^*$ respects common structural belief in IBP, SBSR and MBR. This completes the proof of this theorem.

7.2. Proof of Theorem 5.2

For a given player i, decision node $h_i \in H_i^*$ and opponent j, let $S_j^*(h_i)$ be the set of player j strategies s_j such that (1) $s_j \in S_j(h_i)$, and (2) at every $h_j \in H_j(s_j)$ following h_i, the strategy s_j prescribes the backward induction action $a^*(h_j)$ in $(\mathcal{S}, \tilde{P})$. We prove the following property.

Claim. *Let t_i be a type for player i that respects common structural belief in IBP, SBSR and MBR. Then,*

$$s_j(t_i, h_i) \in S_j^*(h_i)$$

for every $h_i \in H_i^*$ and every opponent j.

Proof of the claim. We prove the claim by induction on the number of decision nodes following h_i. If h_i is not followed by any decision node, the statement is trivial since $S_j^*(h_i) = S_j(h_i)$. Suppose now that the claim holds for all pairs (i', j') of players and every decision node $h_{i'}$ followed by at most $K - 1$ decision nodes. Choose h_i with the property that h_i is followed by exactly K decision nodes. We prove that $s_j(t_i, h_i) \in S_j^*(h_i)$ for every opponent j. Hence, we must show that for every decision node $h_j \in H_j(s_j(t_i, h_i))$ following h_i, the strategy $s_j(t_i, h_i)$ prescribes the backward induction action $a^*(h_j)$.

Let $t_j^* = t_j(t_i, h_i)$ and $s_j^* = s_j(t_i, h_i)$. Choose a decision node $h_j \in H_j(s_j^*)$ following h_i. We shall prove that $s_j^*(h_j) = a^*(h_j)$. As t_i respects common structural belief in IBP, SBSR and MBR and since t_i believes at h_i that player j is of type t_j^*, it follows that t_j^* respects common structural belief in IBP, SBSR and MBR. Since h_j is followed by at most $K - 1$ decision nodes, we thus know by the induction assumption that

$$s_k(t_j^*, h_j) \in S_k^*(h_j)$$

for every opponent $k \neq j$. Consequently, t_j^* believes at h_j that all opponents choose their backward induction actions in $(\mathcal{S}, \tilde{P})$ at the decision nodes following h_j.

As t_i satisfies IBP, it follows that $t_j(t_i, h_0)$ has preference relation $\tilde{P}_j$. Moreover, since t_i respects common structural belief in SBSR and MBR, we know by Lemma 6.5 that t_i does not change his belief about player j's beliefs. Hence, it must be the case that $t_j(t_i, h_0)$ has the same conditional belief vector as $t_j(t_i, h_i) = t_j^*$. We may thus conclude that $t_j(t_i, h_0)$ believes at h_j that all opponents choose their backward induction actions in $(\mathcal{S}, \tilde{P})$ at the decision nodes following h_j. Together with the fact that $t_j(t_i, h_0)$ has preference relation $\tilde{P}_j$, it follows that $t_j(t_i, h_0)$'s optimal strategies at h_j all prescribe the backward induction action $a^*(h_j)$ at h_j.

More precisely, for every $s_j \in S_j(h_j)$ not prescribing $a^*(h_j)$ at h_j there is some $s'_j \in S_j(h_j)$ prescribing $a^*(h_j)$ at h_j such that $t_j(t_i, h_0)$ strictly prefers s'_j to s_j at h_j. This, in turn, means that t_i believes at h_0 that for every $s_j \in S_j(h_j)$ not prescribing $a^*(h_j)$ at h_j there is some $s'_j \in S_j(h_j)$ prescribing $a^*(h_j)$ at h_j such that player j strictly prefers s'_j to s_j at h_j.

Since t_i respects common structural belief in SBSR and MBR, we know by Theorem 6.8 that t_i satisfies proper belief revision. Therefore, t_i's belief at h_i about player j's preference relation at h_j over strategies in $S_j(h_j) \cap S_j(h_i)$ should coincide with t_i's belief at h_0 about player j's preference relation at h_j over strategies in $S_j(h_j) \cap S_j(h_i)$. Since, by assumption, h_j follows h_i we have that $S_j(h_j) \subseteq S_j(h_i)$. Hence, t_i's belief at h_i about player j's preference relation over strategies in $S_j(h_j)$ should coincide with t_i's belief at the beginning about player j's preference relation at h_j over strategies in $S_j(h_j)$. Since we have seen that t_i believes at h_0 that for every $s_j \in S_j(h_j)$ not prescribing $a^*(h_j)$ at h_j there is some $s'_j \in S_j(h_j)$ prescribing $a^*(h_j)$ at h_j such that player j strictly prefers s'_j to s_j at h_j, it follows that t_i believes so at h_i. This implies, however, that t_i believes at h_i that player j's optimal strategies at h_j all prescribe the backward induction action $a^*(h_j)$ at h_j.

Since t_i structurally believes in sequential rationality and $s^*_j = s_j(t_i, h_i)$, we must have that s^*_j is optimal for $t_j(t_i, h_i)$ at h_j. By the above, it follows that s^*_j must prescribe the backward induction $a^*(h_j)$ at h_j, which was to show. This completes the proof of the claim.

Now, let t_i be a type that has preference relation $\tilde{P}_i$, and that respects common structural belief in IBP, SBSR and MBR. By the claim, we know that t_i believes at every decision node h_i that his opponents will choose the backward induction actions in $(\mathcal{S}, \tilde{P})$ at every decision node following h_i. Since t_i has preference relation $\tilde{P}_i$, the unique sequentially rational strategy for t_i is his backward induction strategy in $(\mathcal{S}, \tilde{P})$. This completes the proof. ■

8. Concluding remarks

We conclude by discussing some explicit and implicit assumptions in our model. In this paper, we have decided to model the players' beliefs by single-valued possibility sets so as to make our definitions and proofs as transparent as possible. However, with some additional effort all of our definitions and results can be extended to an epistemic model with multi-valued possibility sets. A first additional difficulty that would arise here is that complete type spaces would no longer exist if one wishes to allow for all possible multi-valued belief sets (see Brandenburger (2003)). This problem can be solved by constructing a type space that is not complete, but that incorporates all beliefs in the belief hierarchy up to a specific order, where this order could be chosen equal to the length of the game tree. In fact, for our purposes here we only need beliefs up to this order. A second difficulty would be that the definition of minimal belief revision would become more elaborate. A type t_i in the new, set-valued model would hold, at every decision node h_i and for every opponent j, a set $B_j(t_i, h_i)$ of strategy-type pairs which t_i deems possible at h_i. Now, consider two decision nodes $h_i^1, h_i^2 \in H_i^*$ such that h_i^2 immediately follows h_i^1, and let $T_j(t_i, h_i^1)$ and $T_j(t_i, h_i^2)$ be the sets of j's types that t_i deems possible at h_i^1 and h_i^2, respectively. In order to define MBR, one should formalize what it means that the set $T_j(t_i, h_i^2)$ is "as similar as possible" to the set $T_j(t_i, h_i^1)$. A possible way to do so would be to say that for every $t_j^2 \in T_j(t_i, h_i^2)$ there should be some $t_j^1 \in T_j(t_i, h_i^1)$ such that t_j^2 is as similar as possible to t_j^1 among all types in $T_j^{sr}(h_i^2)$. Here, similarity between types could be defined as in Definition 4.4. The definition of sequential rationality could easily be adapted as follows: For a given decision node h_i, let $S_j(t_i, h_i)$ be the set of j's strategies that t_i deems possible at h_i. A strategy s_i can then be called sequentially rational for type t_i if at every decision node $h_i \in H_i(s_i)$ there is no alternative strategy $s_i' \in S_i(h_i)$ such that *for every* $(s_j)_{j \neq i} \in \times_{j \neq i} S_j(t_i, h_i)$ the terminal node $z(s_i', (s_j)_{j \neq i}, h_i)$ is preferred to $z(s_i, (s_j)_{j \neq i}, h_i)$. The definitions of SBSR and IBP can then be stated in the obvious way. By choosing these new definitions of IBP, SBSR and MBR, one could then still prove Theorems 5.1 and 5.2. The proof of Theorem 5.1 could in fact be copied as it is, since the types with single-valued beliefs constructed in this proof are simply special cases of types in the new set-valued model. The proof of Theorem 5.2 could be adapted easily to the new model, replacing the original claim by the following:

Claim. *Let t_i be a type for player i that respects common structural belief in IBP, SBSR and MBR. Then,*

$$S_j(t_i, h_i) \subseteq S_j^*(h_i)$$

for all $h_i \in H_i^*$ and all opponents j.

The proof for this claim would go along the same lines as the original proof. Summarizing, our model and results can be extended to multi-valued possibility sets, at the cost of additional complications and more elaborate definitions.

Appendix

A.1. Construction of a complete epistemic model

We show that, within our model, it is always possible to construct a *complete* epistemic model. Recall that we model the players' beliefs by *single-valued possibility sets*. That is, a player, at each of his decision nodes, only deems possible one strategy choice and one preference relation over terminal nodes for every opponent. Recall that $\mathcal{P}$ denotes the set of strict, complete and transitive preference relations over terminal nodes. Then, the set of possible *first-order* beliefs for player i is given by

$$B_i^1 := \times_{h_i \in H_i^*} \times_{j \neq i} (S_j(h_i) \times \mathcal{P}).$$

Namely, a first-order belief $b_i^1 \in B_i^1$ should specify at every decision node h_i and for every opponent j a conditional belief $s_{ij}(h_i)$ about player j's strategy choice (so, a member of $S_j(h_i)$), and a conditional belief $P_{ij}(h_i)$ about player j's preference relation over terminal nodes (so, a member of $\mathcal{P}$).

A second-order belief for player i should specify at every decision node h_i and for every opponent j a (point-) belief about player j's first-order belief. Hence, the set of possible *second-order* beliefs for player i is given by

$$B_i^2 := \times_{h_i \in H_i^*} \times_{j \neq i} B_j^1.$$

Similarly, the sets of third-order, fourth-order and higher-order beliefs for player i are given recursively by

$$B_i^k := \times_{h_i \in H_i^*} \times_{j \neq i} B_j^{k-1}$$

for $k \geq 3$. A *type* for player i describes a preference relation over terminal nodes, and a belief hierarchy consisting of a first-order belief, a second-order belief, a third-order belief, and so on. So, the set of all possible types for player i is given by

$$T_i := \mathcal{P} \times (\times_{k=1}^{\infty} B_i^k).$$

By construction, the set T_i is homeomorphic to the set

$$\mathcal{P} \times (\times_{h_i \in H_i^*} \times_{j \neq i} (S_j(h_i) \times \mathcal{P} \times (\times_{k=1}^{\infty} B_j^k)))$$

which is equal to the set

$$\mathcal{P} \times (\times_{h_i \in H_i^*} \times_{j \neq i} (S_j(h_i) \times T_j)).$$

The latter set, in turn, is homeomorphic to the set

$$\mathcal{P} \times (\times_{j \neq i} (\times_{h_i \in H_i^*} (S_j(h_i))) \times (\times_{h_i \in H_i^*} T_j)).$$

Hence, for every player i there is a homeomorphism

$$b_i : T_i \to \mathcal{P} \times (\times_{j \neq i} (\times_{h_i \in H_i^*} (S_j(h_i))) \times (\times_{h_i \in H_i^*} T_j)).$$

In particular, the function b_i is onto. Let the function P_i be the projection of b_i on $\mathcal{P}$, and let, for every opponent j, the function s_{ij} be the projection of b_i on $\times_{h_i \in H_i^*} S_j(h_i)$ and the function t_{ij} be the projection of b_i on $\times_{h_i \in H_i^*} T_j$. Since b_i is onto, we can find, for every preference relation $\tilde{P}_i \in \mathcal{P}$, every opponent j, every $\tilde{s}_{ij} \in \times_{h_i \in H_i^*} S_j(h_i)$, and every $\tilde{t}_{ij} \in \times_{h_i \in H_i^*} T_j$ some type $t_i \in T_i$ with $P_i(t_i) = \tilde{P}_i$, $s_{ij}(t_i) = \tilde{s}_{ij}$ for all opponents j, and $t_{ij}(t_i) = \tilde{t}_{ij}$ for all opponents j. However, this means that the epistemic model constructed above is complete. So, it is always possible to construct a complete epistemic model.

A.2. *Proofs*

Proof of Lemma 6.3. Choose a type t_i that satisfies SBSR and MBR. Let h_i^1, h_i^2 be two decision nodes in H_i^* such that h_i^2 immediately follows h_i^1. Let j be an opponent for which $s_j(t_i, h_i^1)$ belongs to $S_j(h_i^2)$. By SBSR it must be the case that $s_j(t_i, h_i^1)$ is sequentially rational for $t_j(t_i, h_i^1)$. Since $s_j(t_i, h_i^1) \in S_j(h_i^2)$ it holds that $t_j(t_i, h_i^1) \in T_j^{sr}(h_i^2)$, and MBR implies that $t_j(t_i, h_i^2) = t_j(t_i, h_i^1)$. Since $s_j(t_i, h_i^1)$ is the unique sequentially rational strategy for $t_j(t_i, h_i^1)$, it follows that $s_j(t_i, h_i^2) = s_j(t_i, h_i^1)$, which implies that t_i satisfies Bayesian updating. This completes the proof. ■

Proof of Lemma 6.5. Let h_i^2 immediately follow h_i^1, and let j be an opponent. Define $s_j^1 := s_j(t_i, h_i^1)$, $t_j^1 := t_j(t_i, h_i^1)$, $s_j^2 := s_j(t_i, h_i^2)$ and $t_j^2 := t_j(t_i, h_i^2)$. We distinguish two cases.

Case 1. If $s_j^1 \in S_j(h_i^2)$. Then, we know by the proof of Lemma 6.3 that $t_j^2 = t_j^1$ and $s_j^2 = s_j^1$ and hence the statement in the lemma holds.

Case 2. If $s_j^1 \notin S_j(h_i^2)$. Then, necessarily, $s_j^2 \neq s_j^1$. Since s_j^1 is the unique sequentially rational strategy for t_j^1, it follows that s_j^2 is not sequentially rational for t_j^1.

We now construct a type t_j' such that (1) s_j^2 is sequentially rational for t_j', and (2) $t_j' = (t_j^1, P_j')$ for some P_j'. We may construct a preference relation $P_j' \neq P_j(t_j^1)$ over terminal nodes such that s_j^2 is the unique sequentially rational strategy, regardless of the conditional beliefs about the opponents' strategies. Define $t_j' := (t_j^1, P_j')$. Then, s_j^2 is sequentially rational for t_j'. Since $P_j(t_j') \neq P_j(t_j^1)$, it follows that t_j^1 and t_j' disagree on exactly one statement about player j, namely player j's preference relation.

Assume, contrary to what we want to prove, that t_j^1 and t_j^2 have different conditional beliefs. Hence, there is some $h_j \in H_j^*$ and $k \neq j$ with $s_k(t_j^1, h_j) \neq s_k(t_j^2, h_j)$ or $t_k(t_j^1, h_j) \neq t_k(t_j^2, h_j)$.

Suppose first that $s_k(t_j^1, h_j) \neq s_k(t_j^2, h_j)$. By common structural belief in SBSR, t_i believes at h_i^1 that player j structurally believes in sequential rationality. As $t_j(t_i, h_i^1) = t_j^1$, it follows that t_j^1 structurally believes in sequential rationality. Consequently, $s_k(t_j^1, h_j)$ must be sequentially rational for $t_k(t_j^1, h_j)$. Similarly, $s_k(t_j^2, h_j)$ must be sequentially rational for $t_k(t_j^2, h_j)$. Since $s_k(t_j^1, h_j)$ is the unique sequentially rational strategy for $t_k(t_j^1, h_j)$, and $s_k(t_j^1, h_j) \neq s_k(t_j^2, h_j)$, it follows that $t_k(t_j^1, h_j) \neq t_k(t_j^2, h_j)$. Hence, we have that $s_k(t_j^1, h_j) \neq s_k(t_j^2, h_j)$ and $t_k(t_j^1, h_j) \neq t_k(t_j^2, h_j)$, which implies that t_j^1 and t_j^2 differ at least on two statements. Since t_j^1 and t_j' disagree on exactly one statement, this contradicts the assumption that t_i satisfies MBR. Hence, we may conclude that $s_k(t_j^1, h_j) = s_k(t_j^2, h_j)$.

Suppose, next, that $t_k(t_j^1, h_j) \neq t_k(t_j^2, h_j)$. We prove that this is impossible. To that purpose, we show that either $P_j(t_j^1) \neq P_j(t_j^2)$ or $s_l(t_j^1, h_j') \neq s_l(t_j^2, h_j')$ for some $h_j' \in H_j^*$ and $l \neq j$. Assume, namely, that $P_j(t_j^1) = P_j(t_j^2)$ and $s_l(t_j^1, h_j') = s_l(t_j^2, h_j')$ for all $h_j' \in H_j^*$ and $l \neq j$. Then, t_j^1 and t_j^2 must have the same sequentially rational strategy. By SBSR, t_i structurally believes in sequential rationality, and hence s_j^1 is sequentially rational for t_j^1 and s_j^2 is sequentially rational for t_j^2. Since t_j^1 and t_j^2 have the same sequentially rational strategy, it follows that $s_j^1 = s_j^2$, which is a contradiction to the assumption that $s_j^1 \notin S_j(h_i^2)$. Hence, $P_j(t_j^1) \neq P_j(t_j^2)$ or $s_l(t_j^1, h_j') \neq s_l(t_j^2, h_j')$ for some $h_j' \in H_j^*$ and $l \neq j$. Together with the assumption that $t_k(t_j^1, h_j) \neq t_k(t_j^2, h_j)$, we may conclude that t_j^1 and t_j^2 differ at least on two statements. By the same reasoning as above, this leads to a contradiction.

We may thus conclude that $s_k(t_j^1, h_j) = s_k(t_j^2, h_j)$ and $t_k(t_j^1, h_j) = t_k(t_j^2, h_j)$ for all $h_j \in H_j^*$ and all $k \neq j$, which completes the proof. ■

Proof of Lemma 6.7. Let u^1 be an arbitrary utility representation of P^1, and let the utility functions u^2 and $\tilde{u}^2$ be as stated in the lemma. Let $D(P^1, P^2)$ be the set of pairs of terminal nodes on which P^1 and P^2 disagree. Similarly, we define $D(P^1, \tilde{P}^2)$. Without loss of generality, let a and b in the lemma be chosen such that $u^1(a) > u^1(b)$. Then, by construction, $u^2(a) < u^2(b)$ and $\tilde{u}^2(a) > \tilde{u}^2(b)$. We prove our result through a series of smaller facts. The proof for each of these facts is given in the lines immediately following the statement of the fact.

Fact 1. It holds that $\{a, b\} \notin D(P^1, \tilde{P}^2)$, but $\{a, b\} \in D(P^1, P^2)$. This follows directly from the observation that $u^1(a) > u^1(b)$, $\tilde{u}^2(a) > \tilde{u}^2(b)$ but $u^2(a) < u^2(b)$.

Fact 2. Let $\{x, y\} \in D(P^1, \tilde{P}^2)$, and $x, y \notin \{a, b\}$. Then, $\{x, y\} \in D(P^1, P^2)$. This follows directly from the observation that $\tilde{u}^2(x) = u^2(x)$ and $\tilde{u}^2(y) = u^2(y)$.

Fact 3. Let $\{a, y\} \in D(P^1, \tilde{P}^2)$ such that $\tilde{u}^2(y) > \tilde{u}^2(a)$. Then, $\{a, y\} \in D(P^1, P^2)$. Since $\{a, y\} \in D(P^1, \tilde{P}^2)$ and $\tilde{u}^2(a) < \tilde{u}^2(y)$, we must have that $u^1(a) > u^1(y)$. On the other hand, by construction of $\tilde{u}^2$, we know that $u^2(a) = \tilde{u}^2(b)$ and $u^2(y) = \tilde{u}^2(y)$. Since $\tilde{u}^2(y) > \tilde{u}^2(a)$ and $\tilde{u}^2(a) > \tilde{u}^2(b)$, it follows that $u^2(a) = \tilde{u}^2(b) < \tilde{u}^2(y) = u^2(y)$, which implies that $\{a, y\} \in D(P^1, P^2)$.

Fact 4. Let $\{a, y\} \in D(P^1, \tilde{P}^2)$ such that $\tilde{u}^2(a) > \tilde{u}^2(y) > \tilde{u}^2(b)$. Then, $\{b, y\} \in D(P^1, P^2)$. Since $\{a, y\} \in D(P^1, \tilde{P}^2)$ and $\tilde{u}^2(a) > \tilde{u}^2(y)$, we must have that $u^1(a) < u^1(y)$. By assumption, $u^1(a) > u^1(b)$, and hence $u^1(b) < u^1(y)$. By definition of $\tilde{u}^2$, we have that $u^2(b) = \tilde{u}^2(a)$ and $u^2(y) = \tilde{u}^2(y)$. Since $\tilde{u}^2(a) > \tilde{u}^2(y)$, we have that $u^2(b) > u^2(y)$, which implies that $\{b, y\} \in D(P^1, P^2)$.

Fact 5. Let $\{a, y\} \in D(P^1, \tilde{P}^2)$ such that $\tilde{u}^2(y) < \tilde{u}^2(b)$. Then, $\{a, y\} \in D(P^1, P^2)$. As $\tilde{u}^2(y) < \tilde{u}^2(b)$ and $\tilde{u}^2(a) > \tilde{u}^2(b)$, we may conclude that $\tilde{u}^2(a) > \tilde{u}^2(y)$. Since $\{a, y\} \in D(P^1, \tilde{P}^2)$ we must have that $u^1(a) < u^1(y)$. By definition of $\tilde{u}^2$, it is seen that $u^2(y) = \tilde{u}^2(y)$ and $u^2(a) = \tilde{u}^2(b)$. As $\tilde{u}^2(b) > \tilde{u}^2(y)$, it follows that $u^2(a) > u^2(y)$, and hence $\{a, y\} \in D(P^1, P^2)$.

Fact 6. Let $\{b, y\} \in D(P^1, \tilde{P}^2)$ such that $\tilde{u}^2(y) > \tilde{u}^2(a)$. Then, $\{b, y\} \in D(P^1, P^2)$. As $\tilde{u}^2(b) < \tilde{u}^2(a) < \tilde{u}^2(y)$, and $\{b, y\} \in D(P^1, \tilde{P}^2)$, we must have that $u^1(b) > u^1(y)$. By definition of $\tilde{u}^2$, it holds that $u^2(b) = \tilde{u}^2(a)$ and $u^2(y) = \tilde{u}^2(y)$. Since $\tilde{u}^2(a) < \tilde{u}^2(y)$, we know that $u^2(b) < u^2(y)$, and hence $\{b, y\} \in D(P^1, P^2)$.

Fact 7. Let $\{b, y\} \in D(P^1, \tilde{P}^2)$ such that $\tilde{u}^2(a) > \tilde{u}^2(y) > \tilde{u}^2(b)$. Then, $\{a, y\} \in D(P^1, P^2)$. As $\tilde{u}^2(b) < \tilde{u}^2(y)$ and $\{b, y\} \in D(P^1, \tilde{P}^2)$, we may conclude that $u^1(b) > u^1(y)$. Since $u^1(a) > u^1(b)$, it follows that $u^1(a) > u^1(y)$. On the other hand, we know by definition of $\tilde{u}^2$ that $u^2(a) = \tilde{u}^2(b)$ and $u^2(y) = \tilde{u}^2(y)$. As $\tilde{u}^2(b) < \tilde{u}^2(y)$, it follows that $u^2(a) < u^2(y)$, and hence $\{a, y\} \in D(P^1, P^2)$.

Fact 8. Let $\{b, y\} \in D(P^1, \tilde{P}^2)$ such that $\tilde{u}^2(y) < \tilde{u}^2(b)$. Then, $\{b, y\} \in D(P^1, P^2)$. Since $\tilde{u}^2(b) > \tilde{u}^2(y)$ and $\{b, y\} \in D(P^1, \tilde{P}^2)$, it must be the case that $u^1(b) < u^1(y)$. By construction of $\tilde{u}^2$, it holds that $u^2(b) = \tilde{u}^2(a)$ and $u^2(y) = \tilde{u}^2(y)$. As $\tilde{u}^2(a) > \tilde{u}^2(b) > \tilde{u}^2(y)$, we have that $u^2(b) > u^2(y)$, and hence $\{b, y\} \in D(P^1, P^2)$.

From Facts 1–8, it follows that $D(P^1, \tilde{P}^2)$ contains strictly less pairs than $D(P^1, P^2)$, which completes the proof. ■

References

Asheim, G.B., 2002. On the epistemic foundation for backward induction. Mathematical Social Sciences 44, 121–144.

Asheim, G.B., Perea, A., 2005. Sequential and quasi-perfect rationalizability in extensive games. Games and Economic Behavior 53, 15–42.

Aumann, R., 1995. Backward induction and common knowledge of rationality. Games and Economic Behavior 8, 6–19.

Balkenborg, D., Winter, E., 1997. A necessary and sufficient epistemic condition for playing backward induction. Journal of Mathematical Economics 27, 325–345.

Battigalli, P., 1997. On rationalizability in extensive games. Journal of Economic Theory 74, 40–61.

Battigalli, P., Siniscalchi, M., 1999. Hierarchies of conditional beliefs, and interactive epistemology in dynamic games. Journal of Economic Theory 88, 188–230.

Battigalli, P., Siniscalchi, M., 2002. Strong belief and forward induction reasoning. Journal of Economic Theory 106, 356–391.

Ben-Porath, E., 1997. Rationality, Nash equilibrium and backwards induction in perfect-information games. Review of Economic Studies 64, 23–46.

Brandenburger, A., 2003. On the existence of a "complete possibility structure". In: Dimitri, N., Basili, M., Gilboa, I. (Eds.), Cognitive Processes and Economic Behavior. pp. 30–34.

Brandenburger, A., Friedenberg, A., Keisler, H.J., 2007. Admissibility in games, Available at: http://pages.stern.nyu.edu/~abranden/.

Clausing, T., 2003. Doxastic conditions for backward induction. Theory and Decision 54, 315–336.

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

Feinberg, Y., 2005. Subjective reasoning — dynamic games. Games and Economic Behavior 52, 54–93.

Perea, A., 2006. Proper belief revision and rationalizability in dynamic games. International Journal of Game Theory 34, 529–559.

Perea, A., 2007a. Proper belief revision and equilibrium in dynamic games. Journal of Economic Theory 136, 572–586.

Perea, A., 2007b. Epistemic foundations for backward induction: An overview. In: van Benthem, Johan, Gabbay, Dov, Loewe, Benedikt (Eds.), Texts in Logic and Games, Volume 1: Interactive Logic. Amsterdam University Press, Amsterdam, pp. 159–193.

Quesada, A., 2002. Belief system foundations of backward induction. Theory and Decision 53, 393–403.

Quesada, A., 2003. From common knowledge of rationality to backward induction. In: International Game Theory Review.

Reny, P.J., 1992a. Rationality in extensive-form games. Journal of Economic Perspectives 6, 103–118.
Reny, P.J., 1992b. Backward induction, normal form perfection and explicable equilibria. Econometrica 60, 627–649.
Reny, P.J., 1993. Common belief and the theory of games with perfect information. Journal of Economic Theory 59, 257–274.
Rubinstein, A., 1991. Comments on the interpretation of game theory. Econometrica 59, 909–924.
Samet, D., 1996. Hypothetical knowledge and games with perfect information. Games and Economic Behavior 17, 230–251.
Schulte, O., 2002. Minimal belief change, Pareto-optimality and logical consequence. Economic Theory 19, 105–144.
Stalnaker, R., 1998. Belief revision in games: Forward and backward induction. Mathematical Social Sciences 36, 31–56.