Games and Economic Behavior 110 (2018) 120–138

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Why forward induction leads to the backward induction outcome: A new proof for Battigalli's theorem ☆

Andrés Perea

EpiCenter and Dept. of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, the Netherlands

ARTICLE INFO

Article history:
Received 29 March 2017
Available online 10 April 2018

JEL classification:
C72
C73

Keywords:
Backward induction
Forward induction
Extensive-form rationalizability
Battigalli's theorem
Order independence
Monotonicity

ABSTRACT

Battigalli (1997) has shown that in dynamic games with perfect information and without relevant ties, the forward induction concept of extensive-form rationalizability yields the backward induction outcome. In this paper we provide a new proof for this remarkable result, based on four steps. We first show that extensive-form rationalizability can be characterized by the iterated application of a special reduction operator, the strong belief reduction operator. We next prove that this operator satisfies a mild version of monotonicity, which we call monotonicity on reachable histories. This property is used to show that for this operator, every possible order of elimination leads to the same set of outcomes. We finally show that backward induction yields a possible order of elimination for the strong belief reduction operator. These four properties together imply Battigalli's theorem.

1. Introduction

Extensive-form rationalizability (Pearce, 1984; Battigalli, 1997) is a natural forward induction concept, based on the idea that a player, whenever possible, must rationalize the opponents' past choices. *Backward induction* reasoning is different in that a player need no longer reason about past choices. Instead, a player must believe that his opponents will choose rationally in the future, regardless of what these opponents have done in the past.

Despite this difference, Battigalli (1997) shows in his Theorem 4 that both lines of reasoning lead to exactly the same outcome in dynamic games with perfect information and without relevant ties. This remarkable and surprising result is important for the foundations of game theory, as backward induction and forward induction both play a prominent role in the theory of dynamic games. It therefore seems relevant to not only know *that* Battigalli's theorem holds, but also *why* it holds. The purpose of this paper is to make a step forward in that direction, by delivering a new proof for Battigalli's theorem which we hope leads to an even better understanding of *why* it holds.

Our proof is based on the following four steps. We first introduce a special reduction operator, the *strong belief reduction operator,* which eliminates strategies from any given set of strategy profiles in the game, and show that the extensive-form rationalizable strategies can be characterized by the iterated application of this strong belief reduction operator to the full set of strategy profiles. For any given set D of strategy profiles, the strong belief reduction operator only keeps those

☆ I wish to thank Jing Chen, János Flesch, Aviad Heifetz, Philip Reny, an associate editor, two referees, seminar audiences at Maastricht and Liverpool, and the audience at TARK 2017, for their valuable feedback.

E-mail address: a.perea@maastrichtuniversity.nl.
URL: http://www.epicenter.name/Perea/.

https://doi.org/10.1016/j.geb.2018.04.001
0899-8256/

strategies from D that are optimal, at every history reachable under D, for a conditional belief that assigns full probability to opponents' strategies in D.

In the next step we show that this reduction operator satisfies a mild version of monotonicity that we call *monotonicity on reachable histories.* To explain what this monotonicity condition entails, consider sets of strategy profiles D and E where E can be reached by some order of elimination, and where D is equivalent, in terms of behavior on histories reachable under D, to some partial reduction of E. Here, by a partial reduction of E we mean the result of eliminating some, but not necessarily all, strategies that can be eliminated from E. Monotonicity on reachable histories then states that for any two such sets D and E, the full reduction of D, when restricted to histories reachable under D, must be contained in the full reduction of E, when restricted to these same histories.

In the third step we show that every reduction operator that is monotone on reachable histories will be *order independent with respect to outcomes*. That is, every order of elimination that is possible for this reduction operator eventually yields the same set of induced outcomes. Together with the second step, this implies that the strong belief reduction operator is order independent with respect to outcomes.

In the final step, we prove that backward induction yields a possible order of elimination for the strong belief reduction operator. This result, together with the other steps, implies Battigalli's theorem.

The outline of this paper is as follows. In Section 2 we introduce dynamic games with observable past choices, define the concept of extensive-form rationalizability, present the strong belief reduction operator, and show that the extensive-form rationalizable strategies are obtained by the iterated application of this operator. In Section 3 we introduce the notion of monotonicity on reachable histories, show that the strong belief reduction operator satisfies this mild form of monotonicity, and prove that every reduction operator that is monotone on reachable histories will also be order independent with respect to outcomes. Together with the previous result it then follows that the strong belief reduction operator is order independent with respect to outcomes. In Section 4 we prove that backward induction yields a possible order of elimination for the strong belief reduction operator, which finally enables us to prove Battigalli's theorem. The main body of the paper ends in Section 5 with some concluding remarks. Section 6, finally, contains all the proofs, except the very last one where we prove Battigalli's theorem, relying on all the results outlined above. This short proof is presented in the main body, in Section 4.

Although Battigalli's theorem only applies to dynamic games with perfect information, our Sections 2 and 3 apply to the more general class of games with observable past choices which allow for simultaneous moves. Only Section 4 restricts to games with perfect information.

2. Strong belief reduction operator

In this section we start by formally introducing the class of dynamic games with observable past choices, and present the extensive-form rationalizability procedure for this class. Subsequently, we define the strong belief reduction operator and show that its repeated application yields the set of extensive-form rationalizable strategies.

2.1. Dynamic games with observable past choices

In Sections 2 and 3 of this paper we will focus on finite dynamic games with observable past choices. Such games allow for simultaneous moves, but at every stage of the game every active player knows exactly which choices have been made by the opponents in the past. Formally, a *finite dynamic game with observable past choices* is a tuple

$$G = (I, H, Z, (H_i)_{i\in I}, (C_i(h))_{i\in I, h\in H_i}, (u_i)_{i\in I})$$

where

(a) $I = \{1, 2, ..., n\}$ is the finite set of *players*;

(b) H is the finite set of *histories*, consisting of *non-terminal* and *terminal* histories. At every non-terminal history, one or more players must make a choice, whereas at every terminal history the game ends. By $\varnothing$ we denote the *root* of the game, which is the non-terminal history where the game starts;

(c) $Z \subseteq H$ is the set of terminal histories;

(d) $H_i \subseteq H$ is the set of non-terminal histories where player i must make a choice. For a given non-terminal history h, we denote by $I(h) := \{i \in I \mid h \in H_i\}$ the set of *active* players at h. We allow $I(h)$ to contain more than one player, that is, we allow for *simultaneous moves*. At the same time, we require $I(h)$ to be non-empty for every non-terminal history h;

(e) $C_i(h)$ is the finite set of choices available to player i at a history $h \in H_i$; and

(f) $u_i : Z \to \mathbb{R}$ is player i's utility function, assigning to every terminal history $z \in Z$ some utility $u_i(z)$.

For every non-terminal history h and choice combination $(c_i)_{i\in I(h)}$ in $\times_{i\in I(h)} C_i(h)$, we denote by $h' = (h, (c_i)_{i\in I(h)})$ the (terminal or non-terminal) history that immediately follows this choice combination at h. In this case, we say that h' immediately follows h. We say that a history h *follows* a non-terminal history h' if there is a sequence of histories $h^1, ..., h^K$ such that $h^1 = h'$, $h^K = h$, and h^{k+1} immediately follows h^k for all $k \in \{1, ..., K-1\}$. A history h is said to *weakly follow* h' if either h follows h' or $h = h'$. In the obvious way, we can then also define what it means for h to (*weakly*) *precede* another history h'.

We view a strategy for player i as a *plan of action* (Rubinstein, 1991), assigning choices only to those histories $h \in H_i$ that are not precluded by previous choices. Formally, consider a set of non-terminal histories $\hat{H}_i \subseteq H_i$, and a mapping $s_i : \hat{H}_i \to \cup_{h \in \hat{H}_i} C_i(h)$ assigning to every history $h \in \hat{H}_i$ some available choice $s_i(h) \in C_i(h)$. We say that a history $h \in H$ is *reachable* under s_i if at every history $h' \in \hat{H}_i$ preceding h, the choice $s_i(h')$ is the unique choice that leads to h. The mapping $s_i : \hat{H}_i \to \cup_{h \in \hat{H}_i} C_i(h)$ is called a *strategy* if $\hat{H}_i$ contains exactly those histories in H_i that are reachable under s_i.

By S_i we denote the set of strategies for player i. For every history $h \in H$ and player i, we denote by $S_i(h)$ the set of strategies for player i under which h is reachable. Similarly, for a given strategy s_i we denote by $H_i(s_i)$ the set of histories in H_i that are reachable under s_i.

Finally, we say that the game is with *perfect information* if at every non-terminal history there is only one active player. This is the class of games we will focus on in Section 4.

2.2. Extensive-form rationalizability

We now introduce the extensive-form rationalizability procedure (Pearce, 1984; Battigalli, 1997) which recursively eliminates, at every round, some strategies and conditional belief vectors for the players. Our definition is closest to Battigalli's (1997) Definition 2, which is an equivalent adjustment of Pearce's (1984) original definition of extensive-form rationalizability. To formally state extensive-form rationalizability, we need some additional definitions.

For a finite set X, we denote by $\Delta(X)$ the set of probability distributions on X. For a player i and history $h \in H_i$, let $S_{-i}(h) := \times_{j \neq i} S_j(h)$ be the set of opponents' strategy combinations under which h is reachable.

A *conditional belief vector* for player i is tuple $b_i = (b_i(h))_{h \in H_i}$ where $b_i(h) \in \Delta(S_{-i}(h))$ for every $h \in H_i$. Here, $b_i(h)$ represents the conditional probabilistic belief that i holds at h about the opponents' strategy choices. We say that the conditional belief vector b_i satisfies *Bayesian updating* if for every $h, h' \in H_i$ where h' follows h and $b_i(h)(S_{-i}(h')) > 0$, it holds that

$$b_i(h')(s_{-i}) = \frac{b_i(h)(s_{-i})}{b_i(h)(S_{-i}(h'))} \text{ for all } s_{-i} \in S_{-i}(h').$$

By B_i we denote the set of conditional belief vectors for player i that satisfy Bayesian updating.

For a given conditional belief vector b_i, a set $E \subseteq S_{-i}$ of opponents' strategy combinations, and a history $h \in H_i$, we say that $b_i(h)$ *strongly believes* E if $b_i(h)(E) = 1$ whenever $S_{-i}(h) \cap E \neq \emptyset$. That is, $b_i(h)$ assigns full probability to E whenever E is logically consistent with the event that h has been reached. We say that b_i *strongly believes* E if $b_i(h)$ strongly believes E at every $h \in H_i$.

For a strategy combination $s = (s_i)_{i \in I}$ we denote by $z(s)$ the induced terminal history. For a history $h \in H_i$, a strategy $s_i \in S_i(h)$, and a conditional belief $b_i(h) \in \Delta(S_{-i}(h))$, we denote by

$$u_i(s_i, b_i(h)) := \sum_{s_{-i} \in S_{-i}(h)} b_i(h)(s_{-i}) \cdot u_i(z(s_i, s_{-i}))$$

the induced expected utility at h. We say that strategy s_i is *rational at* h for the conditional belief vector b_i if $u_i(s_i, b_i(h)) \geq u_i(s_i', b_i(h))$ for all $s_i' \in S_i(h)$. That is, strategy s_i yields the highest possible expected utility at h under the belief $b_i(h)$.

For a given strategy s_i and a collection $\hat{H} \subseteq H$ of histories, we say that strategy s_i is *rational at* $\hat{H}$ for b_i if s_i is rational at every $h \in \hat{H} \cap H_i(s_i)$ for b_i. Finally, we say that strategy s_i is *rational* for the conditional belief vector b_i if s_i is rational at H for b_i.

The *extensive-form rationalizability procedure* iteratively eliminates strategies and conditional belief vectors, as follows.

Definition 2.1 *(Extensive-Form Rationalizability).* Consider a finite dynamic game G with observable past choices.

(Induction start) Set $S_i^0 := S_i$ and $B_i^0 := B_i$ for all players i.

(Induction step) Let $k \geq 1$, and assume that S_i^{k-1} and B_i^{k-1} have already been defined for all players i. Then, define for all players i

$$S_i^k := \{s_i \in S_i^{k-1} \mid s_i \text{ rational for some } b_i \in B_i^{k-1}\},$$
$$B_i^k := \{b_i \in B_i^{k-1} \mid b_i \text{ strongly believes } S_{-i}^k\}.$$

A strategy $s_i \in S_i$ is called extensive-form rationalizable if $s_i \in S_i^k$ for all $k \geq 0$.

Here, by S_{-i}^k we denote the set $\times_{j \neq i} S_j^k$. Since there are only finitely many strategies in the game, there must be some $K \geq 0$ such that $S_i^{K+1} = S_i^K$ for every player i. That is, the procedure will terminate after K steps. By $S_i^{efr} := S_i^K$ we denote the set of extensive-form rationalizable strategies for player i.

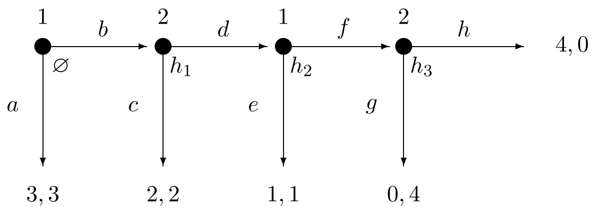

Fig. 1. Reny's game.

As an illustration, consider the game G in Fig. 1, which is based on Fig. 3 in Reny (1992). It may be verified that

$$S_1^1 = \{a, (b, f)\} \text{ and } S_2^1 = \{c, (d, g)\}.$$

Note that the strategies (b, e) and (d, h) can never be rational for any conditional belief vector. By construction, we then have that

$$B_1^1 = \{b_1 \in B_1 \mid b_1(\varnothing)(\{c, (d, g)\}) = 1 \text{ and } b_1(h_2)(\{(d, g)\}) = 1\}$$

and

$$B_2^1 = \{b_2 \in B_2 \mid b_2(h_1)(\{(b, f)\}) = b_2(h_3)(\{(b, f)\}) = 1\}.$$

Note that a is the only strategy for player 1 that is rational for a conditional belief vector in B_1^1. Similarly, (d, g) is the only strategy for player 2 that is rational for the unique conditional belief vector in B_2^1. Hence,

$$S_1^2 = \{a\} \text{ and } S_2^2 = \{(d, g)\},$$

which implies that

$$B_1^2 = \{b_1 \in B_1 \mid b_1(\varnothing)(\{(d, g)\}) = b_1(h_2)(\{(d, g)\}) = 1\} \text{ and } B_2^2 = B_2^1.$$

After this round the procedure terminates, as $S_1^3 = S_1^2$ and $S_2^3 = S_2^2$. Hence, the extensive-form rationalizable strategies are a for player 1 and (d, g) for player 2, which implies that the unique extensive-form rationalizable outcome is the terminal history a. We thus conclude that the unique extensive-form rationalizable outcome is the same as the backward induction outcome in this game G. Note, however, that the extensive-form rationalizable strategy (d, g) for player 2 is *different* from his backward induction strategy c.

2.3. *Strong belief reduction operator*

We next show that the extensive-form rationalizable strategies can be obtained by the iterated application of a certain reduction operator, which we call the *strong belief reduction operator*. Before doing so, we first define what we mean by a reduction operator in general.

A *product of strategy sets* is a Cartesian product $D = \times_{i \in I} D_i$, where $D_i \subseteq S_i$ is a subset of strategies for every player i. A *reduction operator* is a mapping r that assigns to every product of strategy sets D a product of strategy sets $r(D) \subseteq D$ that is contained in it. Hence, whenever $r(D) \neq D$ then $r(D)$ is obtained from D by eliminating some strategies. For two products of strategy sets D and E we say that D is a *partial reduction* of E if $r(E) \subseteq D \subseteq E$. That is, D is obtained from E by eliminating some, but not necessarily all, strategies that can be eliminated according to r. Hence, the notion of partial reduction is always defined relative to a specific reduction operator r. The set $D = r(E)$ is called the *full reduction* of E. For every $k \geq 1$, we denote by

$$r^k(D) := \underbrace{(r \circ \ldots \circ r)}_{k \text{ times}}(D)$$

the k-fold application of r to the product of strategy sets D, and we set $r^0(D) := D$.

For a given product of strategy sets D, let $H(D) \subseteq H$ be the set of histories that are reached by strategy combinations in D.

Definition 2.2 *(Strong belief reduction operator).* The strong belief reduction operator sb assigns to every product of strategy sets $D = \times_{i \in I} D_i$ the set $\times_{i \in I} sb_i(D)$, where for every i

$$sb_i(D) := \{s_i \in D_i \mid s_i \text{ is rational at } H(D) \text{ for some } b_i \in B_i \text{ that strongly believes } D_{-i}\}.$$

Note that $sb(D) \subseteq D$ by definition, and that the additional restrictions imposed by $sb_i(D)$ are rationality conditions at histories reachable under D. In that sense, it is similar to Chen and Micali's (2013) notion of *distinguishable dominance*, where dominance is only required at histories that are reachable under D. In the following subsection we will show that the extensive-form rationalizable strategies are obtained by the iterated application of the strong belief reduction operator to the full set of strategies.

2.4. Characterization of extensive-form rationalizable strategies

Remember from Definition 2.1 that S_i^k denotes the set of strategies for player i that survives round k of the extensive-form rationalizability procedure. In the following theorem we show that S_i^k is obtained by the k-fold application of the strong belief reduction operator to the product of full strategy sets. In particular, the extensive-form rationalizable strategies are exactly those that survive the iterated application of this reduction operator.

Theorem 2.1 *(Characterization of EFR strategies). For every $k \geq 0$, let S_i^k be the set of strategies for player i that survive round k of the extensive-form rationalizability procedure, and let $S^k := \times_{i \in I} S_i^k$ be the induced product of strategy sets. Then, for every $k \geq 0$ we have that*

$$S^k = (sb)^k(S)$$

where $S := \times_{i \in I} S_i$.

As we already mentioned, the procedure $(S^k)_{k \geq 0}$ as we present it in this paper is close to Battigalli's (1997) formulation of extensive-form rationalizability in his Definition 2. In turn, the procedure $((sb)^k(S))_{k \geq 0}$ is very similar to Pearce's original definition of extensive-form rationalizability, which also appears – with a slight but inessential change – as the procedure $(P_c(k))_{k \geq 0}$ in Battigalli (1997)'s Definition 3. Like $((sb)^k(S))_{k \geq 0}$, also Pearce's procedure $(P_c(k))_{k \geq 0}$ imposes at every round k only optimality restrictions at histories that are reachable by $P_c(k-1)$. This is in contrast to $(S^k)_{k \geq 0}$, where at every round k optimality restrictions are imposed at *all* histories.

Battigalli (1997) has shown in Theorem 1 that his formulation of extensive-form rationalizability is equivalent to Pearce's original definition $(P_c(k))_{k \geq 0}$. In that sense, our Theorem 2.1 is very similar to Battigalli's Theorem 1. This similarity also applies to the proof techniques being used. Indeed, in both Battigalli's proof and our proof, the key observation is the following: The restrictions that $(S^k)_{k \geq 0}$ imposes on the conditional belief vectors at histories that are *not reachable* under the strategies from the previous round, are already captured by the restrictions of the preceding rounds. Hence, at every round k it suffices to impose new restrictions only at histories that are reachable under the strategies from the previous round, which is exactly what $((sb)^k(S))_{k \geq 0}$ and $(P_c(k))_{k \geq 0}$ do.

3. Monotonicity on reachable histories

In this section we present a mild version of monotonicity for reduction operators, that we call *monotonicity on reachable histories*. We then show that the strong belief reduction operator, characterizing extensive-form rationalizability, satisfies this type of monotonicity. We finally prove that every reduction operator satisfying monotonicity on reachable histories is guaranteed to be *order independent with respect to outcomes*. That is, every possible order of elimination will yield the same set of induced *outcomes*. In particular, we conclude that the strong belief reduction operator is order independent with respect to outcomes.

3.1. Definition of monotonicity on reachable histories

To formally state monotonicity on reachable histories, we first define the *restriction* of strategies and strategy sets to subcollections of histories. For a given strategy $s_i \in S_i$ and a collection of histories $\hat{H} \subseteq H$, let

$$s_i|_{\hat{H}} := (s_i(h))_{h \in H_i(s_i) \cap \hat{H}}$$

be its restriction to histories in $\hat{H}$. For a set of strategies $D_i \subseteq S_i$, we denote by $D_i|_{\hat{H}} := \{s_i|_{\hat{H}} \mid s_i \in D_i\}$ the restriction of the set D_i to histories in $\hat{H}$. Moreover, for a product of strategy sets $D = \times_{i \in I} D_i$, we define $D|_{\hat{H}} := \times_{i \in I} D_i|_{\hat{H}}$.

For a given reduction operator r, an elimination order for r is a finite sequence of successive partial reductions, and can be formalized as follows.

Definition 3.1 *(Elimination order for r).* An elimination order for a reduction operator r is a finite sequence $(D^0, D^1, ..., D^K)$ of products of strategy sets where (a) $D^0 = S$, (b) $r(D^k) \subseteq D^{k+1} \subseteq D^k$ for every $k \in \{0, ..., K-1\}$, and (c) $r(D^K) = D^K$.

Condition (b) thus states that D^{k+1} is a partial reduction of D^k, whereas condition (c) guarantees that r allows no further eliminations after round K. We say that a product of strategy sets E is *possible in an elimination order* for r if there is an elimination order $(D^0, D^1, ..., D^K)$ for r such that $E = D^k$ for some $k \in \{0, ..., K\}$. We are now fully equipped to define monotonicity on reachable histories.

Definition 3.2 *(Monotonicity on reachable histories).* A reduction operator r is monotone on reachable histories if for every two products of strategy sets D and E where E is possible in an elimination order for r and

$$r(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)},$$

it holds that

$$r(D)|_{H(D)} \subseteq r(E)|_{H(D)}.$$

If in the above definition we would replace $H(D)$ by H, then we obtain exactly the condition of *1-monotonicity** in Luo et al. (2016), which is shown to imply *order independence with respect to strategies.* The latter is more restrictive than order independence with respect to outcomes, as it states that all possible orders of elimination yield the same sets of *strategies*, and not only the same sets of induced *outcomes*.

Note, however, that 1-monotonicity* does not automatically imply monotonicity on reachable histories. The reason is that 1-monotonicity* restricts to sets D and E with $r(E) \subseteq D \subseteq E$, whereas our restrictions on the sets D and E are milder.

It can be shown that $r(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}$ if and only if there is a partial reduction D' of E with $D'|_{H(D)} = D|_{H(D)}$. To see this, suppose first that there is a partial reduction D' of E with $D'|_{H(D)} = D|_{H(D)}$. Since $r(E) \subseteq D' \subseteq E$ and $D'|_{H(D)} = D|_{H(D)}$, it immediately follows that $r(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}$. Assume next that $r(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}$. Since $D|_{H(D)} \subseteq E|_{H(D)}$ there is a mapping $f : D \to E$ with $f(s)|_{H(D)} = s|_{H(D)}$ for every $s \in D$. Then, it may be verified that $D' := f(D) \cup r(E)$ is a partial reduction of E with $D'|_{H(D)} = D|_{H(D)}$.

Hence, monotonicity on reachable histories states that, whenever E is possible in an elimination order for r, and D is equivalent, in terms of behavior on $H(D)$, to a partial reduction of E, then the full reduction of D, when restricted to behavior on $H(D)$, is contained in the full reduction of E, when restricted to behavior on $H(D)$.

3.2. Strong belief reduction operator is monotone on reachable histories

The main step, but also the most difficult step, in our proof of Battigalli's theorem is to show that the strong belief reduction operator satisfies monotonicity on reachable histories.

Theorem 3.1 *(Monotonicity theorem). The strong belief reduction operator sb is monotone on reachable histories.*

Suppose we would remove the restriction in Definition 3.2 that E must be possible in an elimination order for r. Then, the strong belief reduction operator *sb* would no longer satisfy this stronger version of monotonicity. To see this, consider the game in Fig. 1 and take the sets $D = \{a\} \times \{c\}$ and $E = \{a, (b, f)\} \times \{c\}$. Then, it may be verified that $sb(D) = D$ and $sb(E) = \emptyset$. As a consequence, $sb(D)|_{H(D)} \nsubseteq sb(E)|_{H(D)}$ despite the fact that $sb(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}$.

The reason for this failure is that E is not possible in any elimination order for *sb*. Indeed, take any elimination order $(D^0, ..., D^K)$ for *sb* and suppose that $E = D^k$ for some $k \in \{1, ..., K\}$. Then, $sb(D^{k-1}) \subseteq E \subseteq D^{k-1}$. Since $E = D^k$ does not contain strategy (d, g) for player 2, there must be some $m \leq k-1$ such that $(d, g) \in D_2^m$ but $(d, g) \notin D_2^{m+1}$. On the other hand, since $E \subseteq D^{k-1} \subseteq D^m$, it must be that D_1^m contains strategy (b, f) for player 1. As $(d, g) \in D_2^m$, and (d, g) is rational at $H(D^m)$ for the conditional belief vector b_2 with $b_2(h_1) = b_2(h_3) = (b, f)$, which strongly believes D_1^m, it follows that $(d, g) \in D_2^{m+1}$. This, however, is a contradiction. Therefore, we conclude that E is not possible in any elimination order for *sb*. We thus see that in the definition of monotonicity on reachable histories we need to restrict to sets E that are possible in an elimination order for r, otherwise Theorem 3.1 would no longer hold.

3.3. Order independence with respect to outcomes

We next show that every reduction operator r that is monotone on reachable histories, will automatically be *order independent with respect to outcomes.* That is, every order of elimination allowed by r yields the same set of induced outcomes at the end. To define this formally, we denote by $Z(D) := Z \cap H(D)$ the set of terminal histories that are reachable under a product of strategy sets D.

Definition 3.3 *(Order independence with respect to outcomes).* A reduction operator r is order independent with respect to outcomes if for every two elimination orders $(D^0, ..., D^K)$ and $(E^0, ..., E^L)$ for r we have that $Z(D^K) = Z(E^L)$.

We show that monotonicity on reachable histories implies order independence with respect to outcomes.

Theorem 3.2 *(Sufficient condition for order independence with respect to outcomes). Every reduction operator r that is monotone on reachable histories is order independent with respect to outcomes.*

This result is analogous to Luo et al. (2016) who show that their notion of *1-monotonicity** implies order independence with respect to *strategies*. That is, every elimination order for r not only induces the same set of induced outcomes, but also the same set of *strategies*.

Since we have seen in Theorem 3.1 that the strong belief reduction operator is indeed monotone on reachable histories, we immediately obtain the following result.

Corollary 3.1 *(Order independence theorem). The strong belief reduction operator is order independent with respect to outcomes.*

As we will see, this corollary plays a crucial role in establishing the proof of Battigalli's theorem in the following section.

4. Proof of Battigalli's theorem

With Theorem 2.1 and Corollary 3.1 at hand we are finally able to prove Battigalli's theorem. Note that so far we have considered general dynamic games with observable past choices, and all results obtained up to this point hold for that general class. In this section we turn to the more special class of games with *perfect information* and *without relevant ties* – the class of games to which Battigalli's theorem applies.

In this section we proceed as follows. We first define this more special class of games, and give a formal statement of Battigalli's theorem. Next, we show that in every perfect information game without relevant ties, backward induction yields an elimination order for the strong belief reduction operator. We finally use this result, together with Theorem 2.1 and Corollary 3.1, to prove Battigalli's theorem.

4.1. Statement of Battigalli's theorem

Consider a finite dynamic game G with perfect information. That is, at every non-terminal history there is exactly one active player. Following Battigalli (1997), we say that G is *without relevant ties* if for every player i, every $h \in H_i$, every two different choices $c_i, c_i' \in C_i(h)$, every terminal history z weakly following (h, c_i), and every terminal history z' weakly following (h, c_i'), we have that $u_i(z) \neq u_i(z')$. It is easily verified that every such game has a unique backward induction outcome $z^{bi} \in Z$.

Theorem 4.1 *(Battigalli's theorem). Let G be a finite dynamic game with perfect information and without relevant ties. Let z^{bi} be the unique backward induction outcome, let S_i^{efr} be the set of extensive-form rationalizable strategies for every player i, and let $S^{efr} := \times_{i\in I} S_i^{efr}$. Then, $Z(S^{efr}) = \{z^{bi}\}$.*

That is, the backward induction outcome is the unique outcome induced by extensive-form rationalizability.

4.2. Backward induction yields elimination order for sb

We define the *backward induction sequence* $(D^{bi,0}, D^{bi,1}, ..., D^{bi,K})$ as follows. Let K be the maximal number of consecutive choices between the root and a terminal history in the game. For every $k \in \{1, ..., K\}$, let H^k be the collection of non-terminal histories h such that for every terminal history z following h there are at most k consecutive choices between h and z.

We define the products of strategy sets $D^{bi,0}, ..., D^{bi,K}$ inductively by setting $D_i^{bi,0} := S_i$ for every player i, and

$$D_i^{bi,k} := \{s_i \in S_i \mid s_i(h) \text{ is the backward induction choice at } h \text{ for all } h \in H_i(s_i) \cap H^k\}$$

for every player i and every $k \in \{1, ..., K\}$.

Hence, $D_i^{bi,K}$ contains only one strategy for player i, which is his unique backward induction strategy. In particular, it follows that $Z(D^{bi,K}) = \{z^{bi}\}$.

In order to show Battigalli's theorem it is therefore sufficient, in view of Theorem 2.1 and Corollary 3.1, to prove that the backward induction sequence above is an elimination order for *sb*.

Lemma 4.1 *(Backward induction yields elimination order for sb). Let G be a finite dynamic game with perfect information and without relevant ties. Then, the backward induction sequence $(D^{bi,0}, ..., D^{bi,K})$ defined above is an elimination order for sb.*

With Lemma 4.1, Theorem 2.1 and Corollary 3.1 at hand, we are finally able to prove Battigalli's theorem.

4.3. Proof of Battigalli's theorem

Take the backward induction sequence $(D^{bi,0}, D^{bi,1}, ..., D^{bi,K})$ defined above. Then we know, by Lemma 4.1, that this is an elimination order for *sb*. Moreover, the elimination order $(E^0, E^1, ..., E^L)$ obtained by the iterated application of *sb* "at full speed" clearly yields another elimination order for *sb*. But then, by Corollary 3.1 we conclude that $Z(E^L) = Z(D^{bi,K})$. As $Z(D^{bi,K}) = \{z^{bi}\}$ and, by Theorem 2.1, $Z(E^L) = Z(S^{efr})$, it follows that $Z(S^{efr}) = \{z^{bi}\}$, which completes the proof of Battigalli's theorem. □

5. Concluding remarks

5.1. Literature on backward and forward induction reasoning

Backward induction and forward induction are two fundamentally different lines of reasoning in dynamic games. In backward induction, a player believes throughout the game that his opponents will choose rationally in the future, regardless of what these opponents have done in the past. This principle is the basis for the well-known backward induction procedure in dynamic games with perfect information, and for the concept of common belief in future rationality (Perea (2014), see also Penta (2015) and Baltag et al. (2009) for related lines of reasoning) for general dynamic games. The backward induction principle is also implicitly present in equilibrium concepts like subgame perfect equilibrium (Selten, 1965) and sequential equilibrium (Kreps and Wilson, 1982). A common feature of all these backward induction concepts is thus that players are not required to reason about the opponents' past choices, but instead are required to believe that the opponents will act rationally in the future independent of what these opponents have done in the past.

Forward induction, on the other hand, *does* require the players to actively reason about the opponents' past choices. Although there is no unique definition of forward induction in the literature, the main idea is that a player, whenever possible, tries to interpret the opponent's past moves as being part of a rational strategy, and that he bases his belief about the opponent's *future* moves on this hypothesis. *Extensive-form rationalizability* (Pearce, 1984; Battigalli, 1997) is a very basic and natural forward induction concept, based on the idea that a player, whenever possible, must believe that his opponents are implementing rational strategies. This idea can be formalized by the epistemic condition of *strong belief in the opponents' rationality* (Battigalli and Siniscalchi, 2002), which provides the basis for *common strong belief in rationality* – a concept that characterizes extensive-form rationalizability on an epistemic level.

5.2. Other proofs of Battigalli's theorem

This paper is not the first to prove Battigalli's theorem. Much credit should of course go to Battigalli (1997), who was the first to prove this result by relying on certain properties of fully stable sets (Kohlberg and Mertens, 1986). Battigalli's proof, in turn, was inspired by Reny (1992)[1] who used a similar proof technique to show that a different forward induction concept – explicable equilibrium – also leads to the backward induction outcome in the class of games we consider. Battigalli's theorem also follows from Chen and Micali (2013), who show that the iterated elimination of distinguishably dominated strategies is order independent with respect to outcomes, and that performing this procedure "at full speed" is equivalent to the iterated conditional dominance procedure (Shimoji and Watson, 1998). Since Shimoji and Watson (1998) show that the iterated conditional dominance procedure characterizes the extensive-form rationalizable strategies, and the backward induction outcome can be obtained by a specific order of elimination of distinguishably dominated strategies, Battigalli's theorem follows. Luo et al. (2016) provide an alternative proof for the fact that the iterated elimination of distinguishably dominated strategies is order independent with respect to outcomes. Heifetz and Perea (2015) prove Battigalli's theorem via a different route. The main step in their proof is to show that the extensive-form rationalizable outcomes of a game do not change if we *truncate* the game, by eliminating the suboptimal choices at every last non-terminal history. Catonini (2017) studies the concept of *strong* Δ*-rationalizability,* which combines the logic of extensive-form rationalizability with a set Δ of restrictions on *first-order beliefs.* He shows that, if Δ corresponds to "strong belief in a certain path of play", then the set of outcomes induced by strong Δ-rationalizability is smaller than under extensive-form rationalizability. Catonini (2017) then uses this result to prove Battigalli's theorem. Arieli and Aumann (2015) prove Battigalli's theorem for the special case where every player is only active at one history in the game. The key step in their proof is to show that the extensive-form rationalizable outcomes in such games can be characterized by their *pruning process* – a procedure that iteratively eliminates histories from the game. Features that distinguish our approach from the papers above are our use of the *strong belief reduction operator,* and our focus on *monotonicity on reachable histories* as a tool to prove order independence with respect to outcomes.

5.3. Monotonicity on reachable histories

The new notion of *monotonicity on reachable histories* plays a crucial role in our proof of Battigalli's theorem. This condition enters the proof at two different stages: We first show, in Theorem 3.1, that the strong belief reduction operator is

[1] See Battigalli (1997), footnote 13.

monotone on reachable histories, whereas Theorem 3.2 guarantees that monotonicity on reachable histories implies order independence with respect to outcomes. These two steps are our key to proving Battigalli's theorem.

We believe that Theorem 3.2 may also be of interest outside the specific setting of this paper, since it provides an easy to verify sufficient condition for order independence with respect to outcomes. Indeed, suppose we consider a game-theoretic concept for dynamic games that can be characterized by the iterated application of a certain reduction operator r. If we wish to prove that this concept is order independent with respect to outcomes, then, by Theorem 3.2, it would be sufficient to show that the reduction operator r is monotone on reachable histories.

5.4. Reny's theorem

Proposition 3 in Reny (1992) is, in terms of content and proof, very similar to Battigalli's theorem. It shows that in every dynamic game with perfect information and without relevant ties, the forward induction concept of *explicable equilibrium* yields a unique outcome: the backward induction outcome. Like Battigalli (1997), also Reny (1992) proves this result by using properties of fully stable sets (Kohlberg and Mertens, 1986). It would be interesting to see whether the proof techniques in this paper can be used to develop an alternative proof for Reny's theorem.

5.5. Games with imperfect information

Common belief in future rationality (Perea, 2014) represents a backward induction concept that is also applicable to dynamic games with *imperfect information.* We believe that a similar proof as the one in this paper can be used to show that in such games, the set of outcomes induced by extensive-form rationalizability is always smaller than (or equal to) the set of outcomes induced by common belief in future rationality.

6. Proofs

6.1. Proof of Theorem 2.1

We prove the statement by induction on k. For $k = 0$ the statement trivially holds as $S^0 = (sb)^0(S) = S$.

Consider now some $k \geq 1$, and assume that $S^{k-1} = (sb)^{k-1}(S)$. In order to show that $S^k = (sb)^k(S)$, we first prove that (a) $S^k \subseteq (sb)^k(S)$, and then show that (b) $(sb)^k(S) \subseteq S^k$.

(a) We first show that $S^k \subseteq (sb)^k(S)$. Take some player i and some $s_i \in S_i^k$. We must show that $s_i \in sb_i((sb)^{k-1}(S))$. As, by the induction assumption, $(sb)^{k-1}(S) = S^{k-1}$, it suffices to show that $s_i \in sb_i(S^{k-1})$.

Since $s_i \in S_i^k$ we know, by definition of S_i^k, that $s_i \in S_i^{k-1}$, and that s_i is rational for some conditional belief vector $b_i \in B_i^{k-1}$. Here, B_i^{k-1} is the set of conditional belief vectors that survive round $k-1$ of the extensive-form rationalizability procedure. By definition of B_i^{k-1}, it follows that $b_i \in B_i$ and that b_i strongly believes S_{-i}^{k-1}. Hence, $s_i \in S_i^{k-1}$ and s_i is rational for some $b_i \in B_i$ that strongly believes S_{-i}^{k-1}. In particular, s_i is rational at $H(S^{k-1})$ for b_i. As such, $s_i \in sb_i(S^{k-1})$. Together with the induction assumption that $S^{k-1} = (sb)^{k-1}(S)$, we conclude that $s_i \in sb_i((sb)^{k-1}(S))$. This holds for every player i and every $s_i \in S_i^k$, and hence $S^k \subseteq (sb)^k(S)$.

(b) We next show that $(sb)^k(S) \subseteq S^k$, which amounts to proving that $sb_i((sb)^{k-1}(S)) \subseteq S_i^k$ for every player i. Consider some player i and some $s_i \in sb_i((sb)^{k-1}(S))$. By the induction assumption we know that $(sb)^{k-1}(S) = S^{k-1}$, from which we conclude that $s_i \in sb_i(S^{k-1})$. Hence, $s_i \in S_i^{k-1}$ and s_i is rational at $H(S^{k-1})$ for a conditional belief vector $b_i \in B_i$ that strongly believes S_{-i}^{k-1}. As $s_i \in S_i^{k-1}$ we know that s_i is rational (at H) for a conditional belief vector $b_i^{k-2} \in B_i^{k-2}$ which, by definition, strongly believes each of the sets $S_{-i}^0, S_{-i}^1, ..., S_{-i}^{k-2}$.

We now construct a new conditional belief vector b_i^{k-1}, from b_i and b_i^{k-2}, as follows. For every $h \in H_i$, let

$$b_i^{k-1}(h) := \begin{cases} b_i(h), & \text{if } S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset \\ b_i^{k-2}(h), & \text{otherwise} \end{cases}.$$

We will show that $b_i^{k-1} \in B_i^{k-1}$, and that s_i is rational for b_i^{k-1}.

In order to prove that $b_i^{k-1} \in B_i^{k-1}$ we must show that b_i^{k-1} satisfies Bayesian updating, and that b_i^{k-1} strongly believes each of the sets $S_{-i}^0, S_{-i}^1, ..., S_{-i}^{k-1}$.

We start by proving Bayesian updating. Consider some $h, h' \in H_i$ where h' follows h and $b_i^{k-1}(h)(S_{-i}(h')) > 0$. We distinguish two cases: (i) that $S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset$, and (ii) that $S_{-i}^{k-1} \cap S_{-i}(h) = \emptyset$.

(i) Suppose first that $S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset$. Then, $b_i^{k-1}(h) = b_i(h)$. We know, by assumption, that b_i strongly believes S_{-i}^{k-1}, and hence $b_i(h)(S_{-i}^{k-1}) = 1$. We are also assuming that $b_i^{k-1}(h)(S_{-i}(h')) > 0$, which implies that $b_i(h)(S_{-i}(h')) > 0$. By combining the insights that $b_i(h)(S_{-i}^{k-1}) = 1$ and $b_i(h)(S_{-i}(h')) > 0$, we obtain that $S_{-i}^{k-1} \cap S_{-i}(h') \neq \emptyset$. This means, in turn, that

$b_i^{k-1}(h') = b_i(h')$. We thus see that $b_i^{k-1}(h) = b_i(h)$ and $b_i^{k-1}(h') = b_i(h')$. As we assume that b_i satisfies Bayesian updating, we conclude that b_i^{k-1} will satisfy Bayesian updating if the game moves from h to h'.

(ii) Suppose next that $S_{-i}^{k-1} \cap S_{-i}(h) = \emptyset$. Since h' follows h, we know that $S_{-i}^{k-1} \cap S_{-i}(h') = \emptyset$ as well. Therefore, by definition, $b_i^{k-1}(h) = b_i^{k-2}(h)$ and $b_i^{k-1}(h') = b_i^{k-2}(h')$. As $b_i^{k-2} \in B_i^{k-2}$, we know that b_i^{k-2} satisfies Bayesian updating, and therefore b_i^{k-1} will satisfy Bayesian updating as well if the game moves from h to h'. By combining the cases (i) and (ii) we conclude that b_i^{k-1} satisfies Bayesian updating.

We next show that b_i^{k-1} strongly believes each of the sets $S_{-i}^0, S_{-i}^1, ..., S_{-i}^{k-1}$. Consider some arbitrary history $h \in H_i$. We again consider two cases: (i) that $S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset$, and (ii) that $S_{-i}^{k-1} \cap S_{-i}(h) = \emptyset$.

(i) If $S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset$, then $b_i^{k-1}(h) = b_i(h)$. Since b_i strongly believes S_{-i}^{k-1}, we conclude that $b_i^{k-1}(h)(S_{-i}^{k-1}) = b_i(h)(S_{-i}^{k-1}) = 1$. As $S_{-i}^0, ..., S_{-i}^{k-2}$ are supersets of S_{-i}^{k-1}, it follows that $b_i^{k-1}(h)(S_{-i}^0) = ... = b_i^{k-1}(h)(S_{-i}^{k-2}) = 1$ as well. Therefore, $b_i^{k-1}(h)$ strongly believes each of the sets $S_{-i}^0, ..., S_{-i}^{k-1}$.

(ii) If $S_{-i}^{k-1} \cap S_{-i}(h) = \emptyset$, then $b_i^{k-1}(h)$ automatically strongly believes S_{-i}^{k-1}. By definition, we have that $b_i^{k-1}(h) = b_i^{k-2}(h)$. As, by assumption, $b_i^{k-2}(h)$ strongly believes the sets $S_{-i}^0, ..., S_{-i}^{k-2}$, we conclude that $b_i^{k-1}(h)$ strongly believes each of the sets $S_{-i}^0, ..., S_{-i}^{k-1}$. By combining the cases (i) and (ii) we obtain that b_i^{k-1} strongly believes the sets $S_{-i}^0, ..., S_{-i}^{k-1}$. Together with the insight above that b_i^{k-1} satisfies Bayesian updating, we conclude that $b_i^{k-1} \in B_i^{k-1}$.

We finally show that s_i is rational for b_i^{k-1}. Consider some arbitrary history $h \in H_i(s_i)$. We again consider the same two cases: (i) that $S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset$, and (ii) that $S_{-i}^{k-1} \cap S_{-i}(h) = \emptyset$.

(i) If $S_{-i}^{k-1} \cap S_{-i}(h) \neq \emptyset$, then $b_i^{k-1}(h) = b_i(h)$. Moreover, as h is reachable under s_i and $s_i \in S_i^{k-1}$, it follows that $h \in H(S^{k-1})$. Since, by assumption, s_i is rational at $H(S^{k-1})$ for b_i, we conclude that s_i is rational for b_i^{k-1} at h.

(ii) If $S_{-i}^{k-1} \cap S_{-i}(h) = \emptyset$, then $b_i^{k-1}(h) = b_i^{k-2}(h)$. By assumption, s_i is rational for b_i^{k-2}, and hence we see that s_i is rational for b_i^{k-1} at h. By combining the cases (i) and (ii) we may conclude that s_i is rational for b_i^{k-1}.

Altogether, we see that s_i is rational for a conditional belief vector $b_i^{k-1} \in B_i^{k-1}$, and hence $s_i \in S_i^k$. As this holds for every $s_i \in sb_i((sb)^{k-1}(S))$, we conclude that $sb_i((sb)^{k-1}(S)) \subseteq S_i^k$. This applies to every player i, and hence we see that $(sb)^k(S) \subseteq S^k$.

By combining parts (a) and (b) we conclude that $S^k = (sb)^k(S)$. By induction, this holds for every k, and hence the proof is complete. □

6.2. Proof of Theorem 3.1

We prove the theorem by a series of eight preparatory results, of which only the last three concern the strong belief reduction operator. The first result compares two products of strategy sets D and E. The lemma states that, if the behavior in D is more restrictive than the behavior in E, when restricted to histories that are reachable under D, then all histories that are reachable under D are also reachable under E. The same holds if we restrict to histories that are reachable under E. Both the result, and the proof, are rather intuitive.

Lemma 6.1 *(From choice monotonicity to outcome monotonicity). Consider two products of strategy sets D and E such that $D|_{H(D)} \subseteq E|_{H(D)}$ or $D|_{H(E)} \subseteq E|_{H(E)}$. Then, $H(D) \subseteq H(E)$.*

Proof. Assume first that $D|_{H(D)} \subseteq E|_{H(D)}$. Take some $h \in H(D)$. Then, there is some strategy combination s in D that reaches h. As $D|_{H(D)} \subseteq E|_{H(D)}$ there is some strategy combination s' in E with $s|_{H(D)} = s'|_{H(D)}$. Since every history preceding h is also in $H(D)$, it follows that s' and s coincide at all histories preceding h. But then, also s' reaches h. Since $s' \in E$, it follows that $h \in H(E)$. We thus conclude that $H(D) \subseteq H(E)$.

Suppose next that $D|_{H(E)} \subseteq E|_{H(E)}$. For every $k \geq 0$, let H^k be the set of histories that are preceded by k other histories. We show, by induction on k, that $H(D) \cap H^k \subseteq H(E)$ for every $k \geq 0$.

For $k = 0$, the statement is trivial as H^0 only contains the beginning of the game $\emptyset$, which clearly is in $H(E)$. Now, consider some $k \geq 1$, and suppose that $H(D) \cap H^{k-1} \subseteq H(E)$. Take some $h \in H(D) \cap H^k$, and let h' be the history immediately preceding h. Then, $h' \in H(D) \cap H^{k-1}$, and hence by the induction assumption we know that $h' \in H(E)$. This implies that all histories preceding h are in $H(E)$. Since $h \in H(D)$, there is some strategy combination s in D that reaches h. As $D|_{H(E)} \subseteq E|_{H(E)}$, there is some strategy combination s' in E with $s|_{H(E)} = s'|_{H(E)}$. In particular, s and s' coincide at all histories preceding h, as we have seen that all these histories are in $H(E)$. But then, also s' reaches h, which implies that $h \in H(E)$. It thus follows that $H(D) \cap H^k \subseteq H(E)$. By induction on k we conclude that $H(D) \subseteq H(E)$. This completes the proof. □

For the next result we consider a product of strategy sets D, a conditional belief $b_i(h)$ for player i at a history $h \in H(D)$ that strongly believes D_{-i}, and a strategy s_i for player i in D_i. Now suppose we replace $b_i(h)$ by a new belief $b_i'(h)$ that preserves the probabilities on the induced opponents' behavior at histories in $H(D)$, and replace s_i by a new strategy s_i' that preserves the induced behavior for player i at histories in $H(D)$. Then, we show that the expected utility of s_i under the belief $b_i(h)$ is the same as the expected utility of s_i' under $b_i'(h)$.

Lemma 6.2 *(Only behavior on reachable histories matters). Consider a product of strategy sets $D = \times_{i \in I} D_i$, a player i, and a mapping $f_{-i} : D_{-i} \to S_{-i}$ with $f_{-i}(s_{-i})|_{H(D)} = s_{-i}|_{H(D)}$ for every $s_{-i} \in D_{-i}$. Consider for player i a history $h \in H_i \cap H(D)$, a conditional belief $b_i(h) \in \Delta(S_{-i}(h) \cap D_{-i})$, and a conditional belief $b_i'(h) \in \Delta(S_{-i}(h))$ such that*

$$b_i'(h)(s_{-i}) = b_i(h)(f_{-i}^{-1}(s_{-i})) \text{ for every } s_{-i} \in S_{-i}(h).$$

Then, for every $s_i \in D_i \cap S_i(h)$ and every $s_i' \in S_i$ with $s_i|_{H(D)} = s_i'|_{H(D)}$, we have

$$u_i(s_i, b_i(h)) = u_i(s_i', b_i'(h)).$$

Proof. By definition,

$$\begin{aligned} u_i(s_i', b_i'(h)) &= \sum_{s_{-i}' \in S_{-i}} b_i'(h)(s_{-i}') \cdot u_i(z(s_i', s_{-i}')) = \sum_{s_{-i}' \in S_{-i}} b_i(h)(f_{-i}^{-1}(s_{-i}')) \cdot u_i(z(s_i', s_{-i}')) \\ &= \sum_{s_{-i} \in D_{-i}} b_i(h)(s_{-i}) \cdot u_i(z(s_i', f_{-i}(s_{-i}))), \end{aligned} \tag{6.1}$$

where the second equality follows from the definition of $b_i'(h)$, and the third equality follows from the fact that $f_{-i}^{-1}(S_{-i}) = D_{-i}$. Consider now some $s_{-i} \in D_{-i}$ and the induced terminal history $z(s_i', f_{-i}(s_{-i}))$. By assumption, $s_i'|_{H(D)} = s_i|_{H(D)}$ with $s_i \in D_i$, and $f_{-i}(s_{-i})|_{H(D)} = s_{-i}|_{H(D)}$ with $s_{-i} \in D_{-i}$. As $z(s_i, s_{-i})$ is only preceded by non-terminal histories in $H(D)$, and $(s_i', f_{-i}(s_{-i}))$ coincides with (s_i, s_{-i}) at all these histories, it follows that $z(s_i', f_{-i}(s_{-i})) = z(s_i, s_{-i})$.

Since this holds for every $s_{-i} \in D_{-i}$, it follows with (6.1) that

$$\begin{aligned} u_i(s_i', b_i'(h)) &= \sum_{s_{-i} \in D_{-i}} b_i(h)(s_{-i}) \cdot u_i(z(s_i', f_{-i}(s_{-i}))) = \sum_{s_{-i} \in D_{-i}} b_i(h)(s_{-i}) \cdot u_i(z(s_i, s_{-i})) \\ &= u_i(s_i, b_i(h)), \end{aligned}$$

where the last equality follows from the fact that $b_i(h) \in \Delta(S_{-i}(h) \cap D_{-i})$. This completes the proof. □

The following result is well-known in the literature on dynamic games. It states that a strategy which is rational for a conditional belief vector at h will remain rational at a later history h' if the conditional belief at h' is obtained from the conditional belief at h through Bayesian updating. A formal proof for this result can be found, for instance, in Perea (2012, Proof of Lemma 8.14.9).

Lemma 6.3 *(Bayesian updating preserves optimality). Consider a player i, a strategy s_i, a conditional belief vector $b_i \in B_i$, and two histories $h, h' \in H_i(s_i)$ such that h' follows h and $b_i(h)(S_{-i}(h')) > 0$. If s_i is rational for b_i at h, then s_i is also rational for b_i at h'.*

We next show that under Bayesian updating we can always construct a strategy that is rational at *all* histories. We even show a little more than this: for every history h^* we can always construct a strategy that makes h^* reachable and that is rational at all histories weakly following h^*. The reader will notice in the proof that we construct the strategy by a *forward procedure*, in which we first define it at early stages of the game, after which we extend it to later stages. In the literature, the construction of such strategies typically proceeds by a *backward procedure,* in which the strategy is first defined at the final stages of the game, after which it is inductively defined at earlier stages.

Lemma 6.4 *(Existence of rational strategies). Consider a player i, a conditional belief vector $b_i \in B_i$ and a non-terminal history $h^* \in H$. Then, there is a strategy $s_i \in S_i(h^*)$ that is rational for b_i at all $h \in H_i(s_i)$ that weakly follow h^*.*

Proof. We inductively define collections of histories $H_i^1, H_i^{1+}, H_i^2, H_i^{2+} ...$ as follows. Let

$$\begin{aligned} H_i^1 :&= \{h \in H_i \mid h \text{ weakly follows } h^*, \text{ and there is no } h' \in H_i \\ &\qquad \text{that weakly follows } h^* \text{ and precedes } h\}, \text{ and} \\ H_i^{1+} :&= \{h \in H_i \mid \text{there is some } h' \in H_i^1 \text{ preceding } h \text{ with } b_i(h')(S_{-i}(h)) > 0\}. \end{aligned}$$

For a given $k \geq 2$, assume that H_i^{k-1+} has already been defined. Then, let

$$H_i^k := \{h \in H_i \mid h \text{ follows some } h' \in H_i^{k-1} \cup H_i^{k-1+}, \text{ and there is no } h'' \in H_i \text{ that follows } h' \text{ and precedes } h\}, \text{ and}$$

$$H_i^{k+} := \{h \in H_i \mid \text{there is some } h' \in H_i^k \text{ preceding } h \text{ with } b_i(h')(S_{-i}(h)) > 0\}.$$

For every $k \geq 1$ and every $h \in H_i^k$, let s_i^h be a strategy in $S_i(h)$ that is rational for b_i at h. For every $h \in H_i^k \cup H_i^{k+}$, let $h_i^k[h]$ be the unique history in H_i^k that weakly precedes h. Finally, let s_i be a strategy in $S_i(h^*)$ such that for every $k \geq 1$ and every $h \in H_i(s_i) \cap (H_i^k \cup H_i^{k+})$,

$$s_i(h) := s_i^{h_i^k[h]}(h). \tag{6.2}$$

We now show that s_i is rational for b_i at all $h \in H_i(s_i)$ that weakly follow h^*. Take an arbitrary $h \in H_i(s_i)$ that weakly follows h^*, and let $k \geq 1$ be such that $h \in H_i^k \cup H_i^{k+}$. We distinguish two cases: (i) $h \in H_i^k$, and (ii) $h \in H_i^{k+}$.

(i) Consider some $h \in H_i^k$. By construction of H_i^{k+}, every s_{-i} with $b_i(h)(s_{-i}) > 0$ is such that (s_i, s_{-i}) only reaches player i histories weakly following h which are in $H_i^k \cup H_i^{k+}$. Note that $h_i^k[h] = h$, because $h \in H_i^k$. Therefore, by (6.2), s_i and s_i^h coincide on all these histories in $H_i^k \cup H_i^{k+}$ weakly following h, and hence $u_i(s_i, b_i(h)) = u_i(s_i^h, b_i(h))$. Since s_i^h is rational for b_i at h, we conclude that s_i is rational for b_i at h as well.

(ii) Assume next that $h \in H_i^{k+}$. Then, there is some $h' \in H_i^k$ preceding h with $b_i(h')(S_{-i}(h)) > 0$. Since we know from (i) that s_i is rational for b_i at h', it follows from Lemma 6.3 that s_i is rational for b_i at h as well.

From (i) and (ii) we conclude that s_i is rational for b_i at all $h \in H_i(s_i) \cap (H_i^k \cup H_i^{k+})$. As this holds for every $k \geq 1$, we obtain that s_i is rational for b_i at all $h \in H_i(s_i)$ weakly following h^*. This completes the proof. □

The next result shows that for checking the optimality of a strategy with respect to a conditional belief vector, it is sufficient to compare the strategy to alternative strategies that are *rational* for that belief vector *at all histories.*

Lemma 6.5 *(Comparison to optimal strategies suffices). Consider a player i, a strategy s_i, a conditional belief vector $b_i \in B_i$ and a history $h^* \in H_i(s_i)$ such that s_i is not rational for b_i at h^*. Then, there is a history $h^{**} \in H_i$ weakly preceding h^* and a strategy $\tilde{s}_i \in S_i(h^{**})$ that is rational for b_i such that $u_i(s_i, b_i(h^{**})) < u_i(\tilde{s}_i, b_i(h^{**}))$.*

Proof. Let h^{**} be the first history in H_i weakly preceding h^* at which s_i is not rational for b_i. Note that h^{**} can be equal to h^* itself. Let H_i^{pre} be the set of histories in H_i preceding h^{**}, and let

$$H_i^+ := \{h \in H_i \backslash H_i^{pre} \mid \text{there is } h' \in H_i^{pre} \text{ preceding } h \text{ with } b_i(h')(S_{-i}(h)) > 0\}.$$

Note that H_i^{pre} and H_i^+ can be empty if h^{**} is not preceded by any history in H_i. Finally, let H_i^0 be the collection of histories h in $H_i \backslash (H_i^{pre} \cup H_i^+)$ such that h is not preceded by any other $h' \in H_i \backslash (H_i^{pre} \cup H_i^+)$. For every $h \in H_i \backslash (H_i^{pre} \cup H_i^+)$, let $h_i^0[h]$ be the unique history in H_i^0 that weakly precedes h.

We know by Lemma 6.4 that for every $h \in H_i^0$ there is a strategy $s_i^h \in S_i(h)$ that is rational for b_i at all histories in $H_i(s_i^h)$ weakly following h. We define the strategy $\tilde{s}_i$ by

$$\tilde{s}_i(h) := \begin{cases} s_i(h), & \text{if } h \in H_i^{pre} \cup H_i^+ \\ s_i^{h_i^0[h]}(h), & \text{if } h \in H_i \backslash (H_i^{pre} \cup H_i^+) \end{cases}$$

for every $h \in H_i(\tilde{s}_i)$.

We first show that $\tilde{s}_i$ is rational for b_i. That is, we must show that $\tilde{s}_i$ is rational for b_i at every $h \in H_i(\tilde{s}_i)$. We distinguish three cases: (i) $h \in H_i^{pre}$, (ii) $h \in H_i^+$, and (iii) $h \in H_i \backslash (H_i^{pre} \cup H_i^+)$.

(i) Take first some $h \in H_i^{pre}$. Then, h precedes h^{**} and hence, by the choice of h^{**}, strategy s_i is rational for b_i at h. By construction, every opponents' strategy combination $s_{-i} \in S_{-i}(h)$ with $b_i(h)(s_{-i}) > 0$ has the property that $(\tilde{s}_i, s_{-i})$ only reaches player i histories in $H_i^{pre} \cup H_i^+$. As $\tilde{s}_i$ and s_i coincide on $H_i^{pre} \cup H_i^+$, it follows that $u_i(\tilde{s}_i, b_i(h)) = u_i(s_i, b_i(h))$. Since s_i is rational for b_i at h, strategy $\tilde{s}_i$ is rational for b_i at h as well.

(ii) Consider next some $h \in H_i^+$. Then, there is some $h' \in H_i^{pre}$ preceding h with $b_i(h')(S_{-i}(h)) > 0$. Since we have seen in (i) that $\tilde{s}_i$ is rational for b_i at h', we know from Lemma 6.3 that $\tilde{s}_i$ is rational for b_i at h.

(iii) Suppose finally that $h \in H_i \backslash (H_i^{pre} \cup H_i^+)$. Let $h_i^0[h]$ be the unique history in H_i^0 that weakly precedes h. Since $\tilde{s}_i$ coincides with $s_i^{h_i^0[h]}$ at all player i histories weakly following $h_i^0[h]$, and $s_i^{h_i^0[h]}$ is rational for b_i at all histories in $H_i(s_i^{h_i^0[h]})$ weakly following $h_i^0[h]$, it follows that $\tilde{s}_i$ is rational for b_i at h.

By (i), (ii) and (iii) it follows that $\tilde{s}_i$ is rational for b_i.

We next show that $\tilde{s}_i \in S_i(h^{**})$. Since $s_i \in S_i(h^*)$ and h^{**} weakly precedes h^* it follows immediately that $s_i \in S_i(h^{**})$. By definition, all player i histories preceding h^{**} are in H_i^{pre}. As $\tilde{s}_i$ and s_i coincide on H_i^{pre}, they coincide in particular on the player i histories preceding h^{**}. From the fact that $s_i \in S_i(h^{**})$ it then follows that $\tilde{s}_i \in S_i(h^{**})$ as well.

Summarizing, we see that $\tilde{s}_i$ is in $S_i(h^{**})$, and that $\tilde{s}_i$ is rational for b_i. In particular, $\tilde{s}_i$ is rational for b_i at h^{**}. Since s_i is not rational for b_i at h^{**} we conclude that $u_i(s_i, b_i(h^{**})) < u_i(\tilde{s}_i, b_i(h^{**}))$, which completes the proof. □

In order to state the last three results we need to introduce a new operator sb^*, as follows. For a product of strategy sets $D = \times_{i \in I} D_i$, we define for every player i the set

$$sb_i^*(D) := \{s_i \in S_i \mid s_i \text{ is rational at } H(D) \text{ for some } b_i \in B_i \text{ that strongly believes } D_{-i}\}$$

and set

$$sb^*(D) := \times_{i \in I} sb_i^*(D).$$

The difference with the operator sb is thus that $sb_i(D)$ only considers strategies s_i inside D_i, whereas $sb_i^*(D)$ also considers strategies outside D_i. As a consequence, $sb_i^*(D)$ is not necessarily a subset of D_i, in contrast to $sb_i(D)$.

The objective of the last three results is to show that for every product of strategy sets D that is possible in an elimination order for sb, we have that $sb(D)|_{H(D)} = sb^*(D)|_{H(D)}$. That is, for the induced behavior on $H(D)$ it does not matter whether we apply the operator sb or the weaker operator sb^* above. We prove this result in two steps. The first lemma below shows, for any product of strategy sets D, that $sb(D)|_{H(D)} = sb^*(D)|_{H(D)}$ whenever $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$. In the second lemma below we prove that every set D that is possible in an elimination order for sb satisfies the latter property that $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$. In combination with the first lemma, it thus follows that $sb(D)|_{H(D)} = sb^*(D)|_{H(D)}$ for every set D that is possible in an elimination order for sb, which is what we want to show.

Lemma 6.6 *(Consequence of being closed under rational behavior). For every product of strategy sets D with $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$, it holds that $sb^*(D)|_{H(D)} = sb(D)|_{H(D)}$.*

Here, the sufficient condition $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$ reduces to Basu and Weibull's (1991) notion of being *closed under rational behavior* if the game G is a static game, with $\varnothing$ as the only non-terminal history. For that reason, we will say that D is closed under rational behavior whenever $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$.

Proof. By definition we have that $sb(D)|_{H(D)} \subseteq sb^*(D)|_{H(D)}$. It therefore only remains to show that $sb^*(D)|_{H(D)} \subseteq sb(D)|_{H(D)}$. To that purpose, we show that for every player i, and every strategy $s_i \in sb_i^*(D)$, there is some $s_i' \in sb_i(D)$ with $s_i|_{H(D)} = s_i'|_{H(D)}$.

Take some player i and some $s_i \in sb_i^*(D)$. Then, s_i is rational at $H(D)$ for some $b_i \in B_i$ that strongly believes D_{-i}. As $sb_i^*(D)|_{H(D)} \subseteq D_i|_{H(D)}$, there is some $s_i' \in D_i$ with $s_i|_{H(D)} = s_i'|_{H(D)}$.

We show that s_i' is rational at $H(D)$ for b_i. Take some history $h \in H_i(s_i') \cap H(D)$. As $h \in H(D)$ and $s_i|_{H(D)} = s_i'|_{H(D)}$, we conclude that $h \in H_i(s_i) \cap H(D)$ as well. By assumption, s_i is rational at $H(D)$ for b_i, which implies in particular that s_i is rational at h for b_i. That is,

$$u_i(s_i, b_i(h)) \geq u_i(s_i'', b_i(h)) \text{ for all } s_i'' \in S_i(h). \tag{6.3}$$

Since $h \in H(D)$ and b_i strongly believes D_{-i}, we conclude that

$$b_i(h)(D_{-i}) = 1. \tag{6.4}$$

Let $Z(D)$ be the set of terminal histories that are reachable by strategy combinations in D. As $s_i' \in D_i$, we have that $z(s_i', s_{-i}) \in Z(D)$ for all $s_{-i} \in D_{-i}$. Moreover, as $s_i|_{H(D)} = s_i'|_{H(D)}$, it follows that

$$z(s_i, s_{-i}) = z(s_i', s_{-i}) \text{ for all } s_{-i} \in D_{-i}. \tag{6.5}$$

By combining (6.3), (6.4) and (6.5), we conclude that

$$\begin{aligned} u_i(s_i', b_i(h)) &= \sum_{s_{-i} \in D_{-i}} b_i(h)(s_{-i}) \cdot u_i(z(s_i', s_{-i})) \\ &= \sum_{s_{-i} \in D_{-i}} b_i(h)(s_{-i}) \cdot u_i(z(s_i, s_{-i})) \\ &= u_i(s_i, b_i(h)) \geq u_i(s_i'', b_i(h)) \text{ for all } s_i'' \in S_i(h). \end{aligned}$$

Here, the first and third equality follow from (6.4), the second equality follows from (6.5), and the inequality follows from (6.3). We thus see that s_i' is rational at h for b_i. Since this holds for every $h \in H_i(s_i') \cap H(D)$, it follows that s_i' is rational at $H(D)$ for b_i. Together with the facts that $s_i' \in D_i$ and that b_i strongly believes D_{-i}, this implies that $s_i' \in sb_i(D)$.

Remember that $s_i|_{H(D)} = s_i'|_{H(D)}$. We thus have shown that for every $s_i \in sb_i^*(D)$ there is some $s_i' \in sb_i(D)$ with $s_i|_{H(D)} = s_i'|_{H(D)}$. As this holds for every player i, we conclude that $sb^*(D)|_{H(D)} \subseteq sb(D)|_{H(D)}$, which was to show. $\square$

Using the result above, we can show that every product of strategy sets D that is possible in an elimination order for sb is closed under rational behavior.

Lemma 6.7 *(sb leads to sets closed under rational behavior). Every product of strategy sets D that is possible in an elimination order for sb satisfies $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$.*

Proof. Take an arbitrary elimination order $(D^0, ..., D^K)$ for sb. We prove, by induction on k, that $sb^*(D^k)|_{H(D^k)} \subseteq D^k|_{H(D^k)}$ for every $k \in \{0, ..., K\}$. For $k = 0$ this statement is trivial since $D^0 = S$. Consider now some $k \geq 1$ and assume that $sb^*(D^{k-1})|_{H(D^{k-1})} \subseteq D^{k-1}|_{H(D^{k-1})}$. We show that $sb^*(D^k)|_{H(D^k)} \subseteq D^k|_{H(D^k)}$.

Define $D := D^k$ and $E := D^{k-1}$. Then, $sb(E) \subseteq D \subseteq E$, and $sb^*(E)|_{H(E)} \subseteq E|_{H(E)}$. We will show that

$$sb^*(D)|_{H(D)} \subseteq D|_{H(D)}. \tag{6.6}$$

As $sb^*(E)|_{H(E)} \subseteq E|_{H(E)}$, it follows from Lemma 6.6 that $sb^*(E)|_{H(E)} = sb(E)|_{H(E)}$. Hence, we conclude that $sb^*(E)|_{H(D)} = sb(E)|_{H(D)}$ since $H(D) \subseteq H(E)$. Since $sb^*(E)|_{H(D)} = sb(E)|_{H(D)}$ and $sb(E) \subseteq D$, it thus suffices to prove that

$$sb^*(D)|_{H(D)} \subseteq sb^*(E)|_{H(D)} \tag{6.7}$$

in order to show (6.6).

To prove (6.7), take some player i and some $s_i^D \in sb_i^*(D)$. We show that there is some $s_i^E \in sb_i^*(E)$ with $s_i^D|_{H(D)} = s_i^E|_{H(D)}$. Since $s_i^D \in sb_i^*(D)$, there is some conditional belief vector $b_i^D \in B_i$ that strongly believes D_{-i} such that s_i^D is rational for b_i^D at $H(D)$. We proceed in two steps: In step 1 we transform b_i^D into a conditional belief vector b_i^E that strongly believes E_{-i}. In step 2 we finally construct a strategy s_i^E that is rational for b_i^E and coincides with s_i^D on $H(D)$. Consequently, s_i^E will be in $sb_i^*(E)$ and $s_i^D|_{H(D)} = s_i^E|_{H(D)}$, as was to show.

Step 1. We first transform b_i^D into a new conditional belief vector $b_i^E \in B_i$ that strongly believes E_{-i}, as follows.

(i) For all histories $h \in H_i(D) := H_i \cap H(D)$, let

$$b_i^E(h) := b_i^D(h). \tag{6.8}$$

(ii) Define $H_i^+ := \{h \in H_i \backslash H_i(D) \mid b_i^E(h')(S_{-i}(h)) > 0 \text{ for some } h' \in H_i(D) \text{ that precedes } h\}$. For all histories $h \in H_i^+$, let

$$b_i^E(h)(s_{-i}) := \frac{b_i^E(h')(s_{-i})}{b_i^E(h')(S_{-i}(h))} \text{ for all } s_{-i} \in S_{-i}(h), \tag{6.9}$$

where h' is the last history in $H_i(D)$ that precedes h.

(iii) Define $H_i^0 := H_i \backslash (H_i(D) \cup H_i^+)$. For every history $h \in H_i^0$, define

$$b_i^E(h) := b_i(h) \tag{6.10}$$

where b_i is an arbitrary conditional belief vector in B_i that strongly believes E_{-i}.

The reader may easily verify that b_i^E satisfies Bayesian updating.

We next show that b_i^E strongly believes E_{-i}. That is, we must show that for every $h \in H_i$ with $S_{-i}(h) \cap E_{-i} \neq \emptyset$, it holds that $b_i^E(h)(E_{-i}) = 1$. We distinguish three cases: (i) $h \in H_i(D)$, (ii) $h \in H_i^+$, and (iii) $h \in H_i^0$.

(i) Consider first some $h \in H_i(D)$. Then, by (6.8), $b_i^E(h)(E_{-i}) = b_i^D(h)(E_{-i})$. Since b_i^D strongly believes D_{-i} and $h \in H_i(D)$, we know that $b_i^D(h)(D_{-i}) = 1$. This implies that $b_i^D(h)(E_{-i}) = 1$, as $D_{-i} \subseteq E_{-i}$. We thus conclude that $b_i^E(h)(E_{-i}) = b_i^D(h)(E_{-i}) = 1$.

(ii) Consider next some $h \in H_i^+$, and let h' be the last history in $H_i(D)$ that precedes h. Suppose that $b_i^E(h)(s_{-i}) > 0$. Then, by (6.9), $b_i^E(h')(s_{-i}) > 0$. Since we have shown in (i) that $b_i^E(h')(E_{-i}) = 1$, it must hold that $s_{-i} \in E_{-i}$. We thus see that $b_i^E(h)(s_{-i}) > 0$ only if $s_{-i} \in E_{-i}$, which guarantees that $b_i^E(h)(E_{-i}) = 1$.

(iii) Consider finally some $h \in H_i^0$ with $S_{-i}(h) \cap E_{-i} \neq \emptyset$. Then, by (6.10), $b_i^E(h)(E_{-i}) = b_i(h)(E_{-i}) = 1$ since b_i strongly believes E_{-i}.

By combining the cases (i), (ii) and (iii) we conclude that b_i^E strongly believes E_{-i}.

Step 2. We now construct a strategy s_i^E that is rational for b_i^E and coincides with s_i^D on $H(D)$. For every $h \in H_i^0$, let $h^0[h]$ be the first history in H_i^0 that weakly precedes h. As b_i^E satisfies Bayesian updating, we know by Lemma 6.4 that for every first history $h \in H_i^0$ there is some strategy $s_i^h \in S_i(h)$ that is rational for b_i^E at every $h' \in H_i(s_i^h)$ that weakly follows h.

Let the strategy s_i^E be such that

$$s_i^E(h) := \begin{cases} s_i^D(h), & \text{if } h \in H_i(D) \cup H_i^+ \\ s_i^{h^0[h]}(h), & \text{if } h \in H_i^0 \end{cases} \tag{6.11}$$

for all $h \in H_i(s_i^E)$. Then, it immediately follows that $s_i^D|_{H(D)} = s_i^E|_{H(D)}$.

We now show that s_i^E is rational for b_i^E. That is, we must show that s_i^E is rational for b_i^E at every $h \in H_i(s_i^E)$. We distinguish three cases: (i) $h \in H_i(D)$, (ii) $h \in H_i^+$, and (iii) $h \in H_i^0$.

(i) Assume first that $h \in H_i(D)$. Since $b_i^E(h) = b_i^D(h)$, it follows by definition of H_i^+ that every $s_{-i} \in S_{-i}(h)$ with $b_i^E(h)(s_{-i}) > 0$ is such that (s_i^E, s_{-i}) only reaches player i histories weakly following h that are in $H_i(D) \cup H_i^+$. Since, by (6.11), s_i^E and s_i^D coincide on $H_i(D) \cup H_i^+$, it follows that $u_i(s_i^E, b_i^E(h)) = u_i(s_i^D, b_i^E(h))$. As, by (6.8), $b_i^E(h) = b_i^D(h)$, and s_i^D is rational for b_i^D at $H(D)$, it follows that s_i^E is rational for b_i^E at h.

(ii) Assume next that $h \in H_i^+$. Let h' be the last history in $H_i(D)$ that precedes h. Then, by (6.9), $b_i^E(h)$ is obtained through Bayesian updating from $b_i^E(h')$. Since we have seen in (i) that s_i^E is rational for b_i^E at h', it follows from Lemma 6.3 that s_i^E is rational for b_i^E at h as well.

(iii) Assume finally that $h \in H_i^0$. As, by assumption, $s_i^{h^0[h]}$ is rational for b_i^E at all histories in $H_i(s_i^{h^0[h]})$ weakly following $h^0[h]$, it follows in particular that $s_i^{h^0[h]}$ is rational for b_i^E at h. But then, by (6.11), s_i^E is rational at h for b_i^E.

Altogether, we see that for all $h \in H_i(s_i^E)$, strategy s_i^E is rational at h for b_i^E. That is, s_i^E is rational for b_i^E.

Since $b_i^E \in B_i$ and b_i^E strongly believes E_{-i}, we conclude that $s_i^E \in sb_i^*(E)$. We know from above that $s_i^D|_{H(D)} = s_i^E|_{H(D)}$. Hence, there is some $s_i^E \in sb_i^*(E)$ with $s_i^D|_{H(D)} = s_i^E|_{H(D)}$. As this holds for every player i and every $s_i^D \in sb_i^*(D)$, we conclude that $sb^*(D)|_{H(D)} \subseteq sb^*(E)|_{H(D)}$, which establishes (6.7). As we saw above, this implies that $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$, that is, $sb^*(D^k)|_{H(D^k)} \subseteq D^k|_{H(D^k)}$. By induction, this holds for every $k \in \{0, ..., K\}$. As this applies to every elimination order $(D^0, ..., D^K)$, we conclude that every product of strategy sets D that is possible in an elimination order for sb satisfies $sb^*(D)|_{H(D)} \subseteq D|_{H(D)}$. This completes the proof. □

An immediate consequence of the two lemmas above is that every set D that is possible in an elimination order for sb satisfies $sb^*(D)|_{H(D)} = sb(D)|_{H(D)}$.

Corollary 6.1 *(Property of sets in elimination order). For every product of strategy sets D that is possible in an elimination order for sb it holds that $sb^*(D)|_{H(D)} = sb(D)|_{H(D)}$.*

With these preparatory results at hand we are now fully equipped to prove Theorem 3.1.

Proof of Theorem 3.1. Consider some products of strategy sets D and E where E is possible in an elimination order for sb and $sb(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}$. We must show, for every player i, that $sb_i(D)|_{H(D)} \subseteq sb_i(E)|_{H(D)}$.

Consider some player i. As $D|_{H(D)} \subseteq E|_{H(D)}$ we have, in particular, that $D_{-i}|_{H(D)} \subseteq E_{-i}|_{H(D)}$. Hence, there is some function $f_{-i} : D_{-i} \to S_{-i}$ such that $f_{-i}(D_{-i}) \subseteq E_{-i}$ and

$$s_{-i}|_{H(D)} = f_{-i}(s_{-i})|_{H(D)} \text{ for every } s_{-i} \in D_{-i}. \tag{6.12}$$

Take some strategy $s_i^D \in sb_i(D)$. We will prove that there is some $s_i^E \in sb_i(E)$ with $s_i^D|_{H(D)} = s_i^E|_{H(D)}$. By definition, $s_i^D \in D_i$ and s_i^D is rational at $H(D)$ for some conditional belief vector $b_i^D \in B_i$ that strongly believes D_{-i}. We proceed in three steps: In step 1 we transform b_i^D into a conditional belief vector b_i^E in B_i that strongly believes E_{-i}. In step 2 we construct a strategy $\tilde{s}_i^E$ that is rational for b_i^E and for which $\tilde{s}_i^E|_{H(D)} = s_i^D|_{H(D)}$. In step 3 we transform $\tilde{s}_i^E$ into a strategy $s_i^E \in sb_i(E)$ with $s_i^E|_{H(D)} = s_i^D|_{H(D)}$.

Step 1. We transform b_i^D into a conditional belief vector b_i^E in B_i that strongly believes E_{-i}, as follows.

(i) For all histories $h \in H_i(D) := H_i \cap H(D)$, let

$$b_i^E(h)(s_{-i}) := b_i^D(h)(f_{-i}^{-1}(s_{-i})) \text{ for all } s_{-i} \in S_{-i}. \tag{6.13}$$

(ii) Define $H_i^+ := \{h \in H_i \backslash H_i(D) \mid b_i^E(h')(S_{-i}(h)) > 0 \text{ for some } h' \in H_i(D) \text{ preceding } h\}$. For all histories $h \in H_i^+$, let

$$b_i^E(h)(s_{-i}) := \frac{b_i^E(h')(s_{-i})}{b_i^E(h')(S_{-i}(h))} \text{ for all } s_{-i} \in S_{-i}(h), \tag{6.14}$$

where h' is the last history in $H_i(D)$ that precedes h.

(iii) Define $H_i^0 := H_i \backslash (H_i(D) \cup H_i^+)$. For every history $h \in H_i^0$, define

$$b_i^E(h) := \hat{b}_i^E(h) \tag{6.15}$$

where $\hat{b}_i^E$ is an arbitrary conditional belief vector in B_i that strongly believes E_{-i}.

We first show that b_i^E is a *well-defined* conditional belief vector. That is, for every $h \in H_i$ we must show that $b_i^E(h)(s_{-i}) > 0$ only if $s_{-i} \in S_{-i}(h)$, and that $\sum_{s_{-i} \in S_{-i}} b_i^E(h)(s_{-i}) = 1$. We consider three cases: (i) $h \in H_i(D)$, (ii) $h \in H_i^+$, and (iii) $h \in H_i^0$.

(i) Consider first some $h \in H_i(D)$. Suppose that $b_i^E(h)(s_{-i}) > 0$. Then, by (6.13), there is some $s'_{-i} \in D_{-i}$ with $f_{-i}(s'_{-i}) = s_{-i}$ and $b_i^D(h)(s'_{-i}) > 0$. Since b_i^D is a well-defined conditional belief vector, we must have that $s'_{-i} \in S_{-i}(h)$. By (6.12) we know that $s'_{-i}|_{H(D)} = f_{-i}(s'_{-i})|_{H(D)} = s_{-i}|_{H(D)}$. Since $h \in H_i(D)$, all histories preceding h will also be in $H(D)$. Hence, s'_{-i} and s_{-i} coincide at all histories preceding h. As $s'_{-i} \in S_{-i}(h)$, it follows that $s_{-i} \in S_{-i}(h)$ as well. We thus see that $b_i^E(h)(s_{-i}) > 0$ only if $s_{-i} \in S_{-i}(h)$.

Moreover, by (6.13),

$$\sum_{s_{-i} \in S_{-i}} b_i^E(h)(s_{-i}) = \sum_{s_{-i} \in S_{-i}} b_i^D(h)(f_{-i}^{-1}(s_{-i})) = \sum_{s'_{-i} \in D_{-i}} b_i^D(h)(s'_{-i}) = 1.$$

The latter equality follows from the facts that $h \in H_i(D)$ and that b_i^D strongly believes D_{-i}.

For cases (ii) and (iii), these properties follow automatically from (6.14) and (6.15).

We next show that b_i^E satisfies Bayesian updating. Consider two histories $h, h' \in H_i$ such that h' follows h, and $b_i^E(h)(S_{-i}(h')) > 0$. We must show that

$$b_i^E(h')(s_{-i}) = \frac{b_i^E(h)(s_{-i})}{b_i^E(h)(S_{-i}(h'))} \text{ for all } s_{-i} \in S_{-i}(h'). \tag{6.16}$$

The only problematic case is where $h, h' \in H_i(D)$. For the cases where at least one of these two histories is in H_i^+ or H_i^0, (6.16) follows rather immediately from (6.14) or (6.15), and we leave these cases to the reader.

Let us therefore assume that $h, h' \in H_i(D)$. For every $s_{-i} \in D_{-i}$ we have by (6.12) that $f_{-i}(s_{-i})|_{H(D)} = s_{-i}|_{H(D)}$. As $h' \in H(D)$, all histories preceding h' are also in $H(D)$. It thus follows that $s_{-i} \in D_{-i} \cap S_{-i}(h')$ if and only if $f_{-i}(s_{-i}) \in S_{-i}(h')$. Consequently,

$$f_{-i}^{-1}(S_{-i}(h')) = D_{-i} \cap S_{-i}(h'). \tag{6.17}$$

By (6.13) we then have for every $s_{-i} \in S_{-i}(h')$ that

$$\begin{aligned}\frac{b_i^E(h)(s_{-i})}{b_i^E(h)(S_{-i}(h'))} &= \frac{b_i^D(h)(f_{-i}^{-1}(s_{-i}))}{b_i^D(h)(f_{-i}^{-1}(S_{-i}(h')))} = \frac{b_i^D(h)(f_{-i}^{-1}(s_{-i}))}{b_i^D(h)(D_{-i} \cap S_{-i}(h'))} \\ &= \frac{b_i^D(h)(f_{-i}^{-1}(s_{-i}))}{b_i^D(h)(S_{-i}(h'))} = b_i^D(h')(f_{-i}^{-1}(s_{-i})) = b_i^E(h')(s_{-i}),\end{aligned}$$

where the first equality follows from (6.13), the second equality from (6.17), the third equality from the facts that $h \in H_i(D)$ and that b_i^D strongly believes D_{-i}, the fourth equality from the fact that b_i^D satisfies Bayesian updating, and the last equality from (6.13). Hence, (6.16) holds, which was to show.

We finally show that b_i^E strongly believes E_{-i}. That is, we must show that $b_i^E(h)(E_{-i}) = 1$ whenever $S_{-i}(h) \cap E_{-i} \neq \emptyset$. Consider now an arbitrary $h \in H_i$ with $S_{-i}(h) \cap E_{-i} \neq \emptyset$. We distinguish three cases: (i) $h \in H_i(D)$, (ii) $h \in H_i^+$, and (iii) $h \in H_i^0$.

(i) Suppose first that $h \in H_i(D)$. Consider some $s_{-i} \in S_{-i}(h)$ with $b_i^E(h)(s_{-i}) > 0$. By (6.13), it then follows that there is some $s'_{-i} \in D_{-i}$ with $f_{-i}(s'_{-i}) = s_{-i}$. Hence, $s_{-i} \in E_{-i}$. We thus see that $b_i^E(h)(s_{-i}) > 0$ only if $s_{-i} \in E_{-i}$, that is, $b_i^E(h)(E_{-i}) = 1$.

(ii) Suppose next that $h \in H_i^+$. Consider some $s_{-i} \in S_{-i}(h)$ with $b_i^E(h)(s_{-i}) > 0$. By (6.14) it then follows that $b_i^E(h')(s_{-i}) > 0$, where h' is the last history in $H_i(D)$ that precedes h. Since $S_{-i}(h) \cap E_{-i} \neq \emptyset$ and h' precedes h, we know that $S_{-i}(h') \cap E_{-i} \neq \emptyset$ also. As $b_i^E(h')(s_{-i}) > 0$, we know by (i) above that $s_{-i} \in E_{-i}$. We thus see that $b_i^E(h)(s_{-i}) > 0$ only if $s_{-i} \in E_{-i}$, that is, $b_i^E(h)(E_{-i}) = 1$.

(iii) Suppose finally that $h \in H_i^0$. Then, by (6.15), $b_i^E(h)(E_{-i}) = \hat{b}_i^E(h)(E_{-i}) = 1$, since $\hat{b}_i^E$ strongly believes E_{-i}.

Overall, we conclude that b_i^E strongly believes E_{-i}.

Summarizing, we have shown that b_i^E is a well-defined conditional belief vector that satisfies Bayesian updating and that strongly believes E_{-i}. That is, $b_i^E \in B_i$ and b_i^E strongly believes E_{-i}.

Step 2. We next construct a strategy $\tilde{s}_i^E$ that is rational for b_i^E and coincides with s_i^D on $H(D)$. For every $h \in H_i^0$, let $h^0[h]$ be the first history in H_i^0 that weakly precedes h. Since b_i^E satisfies Bayesian updating, we know by Lemma 6.4 that for every first history h in H_i^0 there is some strategy $s_i^h \in S_i(h)$ that is rational for b_i^E at all histories in $H_i(s_i^h)$ that weakly follow h. Let $\tilde{s}_i^E$ be the strategy given by

$$\tilde{s}_i^E(h) := \begin{cases} s_i^D(h), & \text{if } h \in H_i(D) \cup H_i^+ \\ s_i^{h^0[h]}(h), & \text{if } h \in H_i^0 \end{cases} \tag{6.18}$$

for all $h \in H_i(\tilde{s}_i^E)$. Then, it immediately follows that $s_i^D|_{H(D)} = \tilde{s}_i^E|_{H(D)}$.

We will now show that strategy $\tilde{s}_i^E$ is rational for b_i^E. That is, we must show that, for all $h \in H_i(\tilde{s}_i^E)$, strategy $\tilde{s}_i^E$ is rational at h for b_i^E. We again consider three cases: (i) $h \in H_i(D)$, (ii) $h \in H_i^+$, and (iii) $h \in H_i^0$.

(i) Assume first that $h \in H_i(D)$. Suppose, contrary to what we want to show, that $\tilde{s}_i^E$ is not rational at h for b_i^E. Since b_i^E satisfies Bayesian updating there is, by Lemma 6.5, a history $h' \in H_i$ weakly preceding h and a strategy $s_i'' \in S_i(h')$ such that s_i'' is rational for b_i^E and

$$u_i(\tilde{s}_i^E, b_i^E(h')) < u_i(s_i'', b_i^E(h')). \tag{6.19}$$

As $h \in H_i(D)$ and h' precedes h we know that $h' \in H_i(D)$ as well. Hence, $b_i^D(h') \in \Delta(S_{-i}(h') \cap D_{-i})$ since b_i^D strongly believes D_{-i}. Moreover, $b_i^E(h')$ is given by (6.13) above where, by (6.12), $s_{-i}|_{H(D)} = f_{-i}(s_{-i})|_{H(D)}$ for every $s_{-i} \in D_{-i}$. Since $s_i^D|_{H(D)} = \tilde{s}_i^E|_{H(D)}$ and $s_i^D \in D_i$, it follows by Lemma 6.2 that

$$u_i(\tilde{s}_i^E, b_i^E(h')) = u_i(s_i^D, b_i^D(h')). \tag{6.20}$$

Recall from (6.19) that $u_i(\tilde{s}_i^E, b_i^E(h')) < u_i(s_i'', b_i^E(h'))$ for some $s_i'' \in S_i(h')$ that is rational for b_i^E. As b_i^E strongly believes E_{-i}, it follows that $s_i'' \in sb_i^*(E)$. Since E is possible in an elimination order for sb, we know from Corollary 6.1 that $sb_i^*(E)|_{H(E)} = sb_i(E)|_{H(E)}$. As, by the assumptions in the theorem, $D|_{H(D)} \subseteq E|_{H(D)}$, we know from Lemma 6.1 that $H(D) \subseteq H(E)$, and hence $sb_i^*(E)|_{H(D)} = sb_i(E)|_{H(D)}$. Moreover, from the other assumption in the theorem, $sb_i(E)|_{H(D)} \subseteq D_i|_{H(D)}$. By combining these two insights we obtain that $sb_i^*(E)|_{H(D)} \subseteq D_i|_{H(D)}$. As $s_i'' \in sb_i^*(E)$, we conclude that there is some $\hat{s}_i^D \in D_i$ with $s_i''|_{H(D)} = \hat{s}_i^D|_{H(D)}$. But then it follows, in the same way as above, from Lemma 6.2 that

$$u_i(s_i'', b_i^E(h')) = u_i(\hat{s}_i^D, b_i^D(h')). \tag{6.21}$$

By combining (6.19), (6.20) and (6.21) it then follows that $u_i(s_i^D, b_i^D(h')) < u_i(\hat{s}_i^D, b_i^D(h'))$, which contradicts our assumption that s_i^D is rational for b_i^D at $H(D)$. We therefore conclude that $\tilde{s}_i^E$ is rational at h for b_i^E.

(ii) Assume next that $h \in H_i^+$. Let h' be the last history in $H_i(D)$ that precedes h. Since we have shown in (i) that $\tilde{s}_i^E$ is rational at h' for b_i^E, it follows from (6.14) and Lemma 6.3 that $\tilde{s}_i^E$ is rational at h for b_i^E as well.

(iii) Assume finally that $h \in H_i^0$. Then, by (6.18) we know that

$$\tilde{s}_i^E(h') = s_i^{h^0[h]}(h') \text{ for all } h' \in H_i(\tilde{s}_i^E) \text{ weakly following } h. \tag{6.22}$$

As, by assumption, $s_i^{h^0[h]}$ is rational for b_i^E at all histories in $H_i(s_i^{h^0[h]})$ weakly following $h^0[h]$, it follows in particular that $s_i^{h^0[h]}$ is rational for b_i^E at h. But then, by (6.22), also $\tilde{s}_i^E$ is rational at h for b_i^E, which was to show.

Altogether, we see that for all $h \in H_i(\tilde{s}_i^E)$, strategy $\tilde{s}_i^E$ is rational at h for b_i^E. That is, $\tilde{s}_i^E$ is rational for b_i^E.

Step 3. We finally transform $\tilde{s}_i^E$ into a strategy $s_i^E \in sb_i(E)$ that coincides with s_i^D on $H(D)$. Since we know from above that $\tilde{s}_i^E$ is rational for b_i^E, that $b_i^E \in B_i$ and that b_i^E strongly believes E_{-i}, we conclude that $\tilde{s}_i^E \in sb_i^*(E)$. Since we have seen above that $sb_i^*(E)|_{H(D)} = sb_i(E)|_{H(D)}$, there is some $s_i^E \in sb_i(E)$ with $s_i^E|_{H(D)} = \tilde{s}_i^E|_{H(D)}$. As, by Step 2, $\tilde{s}_i^E|_{H(D)} = s_i^D|_{H(D)}$, it follows that $s_i^E|_{H(D)} = s_i^D|_{H(D)}$. Since $s_i^D \in sb_i(D)$ was chosen arbitrarily, we see that for every $s_i^D \in sb_i(D)$ there is some $s_i^E \in sb_i(E)$ with $s_i^D|_{H(D)} = s_i^E|_{H(D)}$. That is, $sb_i(D)|_{H(D)} \subseteq sb_i(E)|_{H(D)}$, which completes the proof. □

6.3. Proof of Theorem 3.2

Before we can prove this theorem, we will first discuss a preparatory result that is needed. Consider a reduction operator r that is monotone on reachable histories, an elimination order $(D^0, ..., D^K)$ for r, and two subsequent sets F and G in this elimination order. The lemma shows that if we iteratively apply the reduction operator r "at full speed" to F and G respectively, then the induced elimination orders will be nested at every round in terms of behavior on reachable histories. As a consequence, both elimination orders will eventually yield the same set of outcomes.

Lemma 6.8 *(Sandwich lemma). Consider a reduction operator r that is monotone on reachable histories, and let $(D^0, ..., D^K)$ be an elimination order for r. For some $m \in \{0, ..., K-1\}$, let $F := D^{m+1}$ and $G := D^m$. Then, for every $k \geq 0$,*

$$r^{k+1}(G)|_{H(r^k(F))} \subseteq r^k(F)|_{H(r^k(F))} \subseteq r^k(G)|_{H(r^k(F))},$$

and

$$H(r^{k+1}(G)) \subseteq H(r^k(F)) \subseteq H(r^k(G)).$$

Proof. We prove the statement by induction on k. Consider first $k = 0$. As $r(G) \subseteq F \subseteq G$, it immediately follows that $r(G)|_{H(F)} \subseteq F|_{H(F)} \subseteq G|_{H(F)}$ and $H(r(G)) \subseteq H(F) \subseteq H(G)$, which was to show.

Consider now some $k \geq 1$, and suppose that

$$r^k(G)|_{H(r^{k-1}(F))} \subseteq r^{k-1}(F)|_{H(r^{k-1}(F))} \tag{6.23}$$

and

$$r^{k-1}(F)|_{H(r^{k-1}(F))} \subseteq r^{k-1}(G)|_{H(r^{k-1}(F))}. \tag{6.24}$$

We first show that

$$r^k(F)|_{H(r^k(F))} \subseteq r^k(G)|_{H(r^k(F))}. \tag{6.25}$$

If we set $D := r^{k-1}(F)$ and $E := r^{k-1}(G)$, then (6.23) and (6.24) state that

$$r(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}. \tag{6.26}$$

Clearly, E is possible in an elimination order for r, as G is possible in an elimination order for r and $E = r^{k-1}(G)$. But then, together with (6.26) and the fact that r is monotone on reachable histories, we conclude that $r(D)|_{H(D)} \subseteq r(E)|_{H(D)}$, which can be restated as

$$r^k(F)|_{H(r^{k-1}(F))} \subseteq r^k(G)|_{H(r^{k-1}(F))}. \tag{6.27}$$

This automatically implies (6.25), since $H(r^k(F)) \subseteq H(r^{k-1}(F))$.

We next show that

$$r^{k+1}(G)|_{H(r^k(F))} \subseteq r^k(F)|_{H(r^k(F))}. \tag{6.28}$$

Set $D := r^k(G)$ and $E := r^{k-1}(F)$. Hence, (6.28) can be restated as

$$r(D)|_{H(r(E))} \subseteq r(E)|_{H(r(E))}. \tag{6.29}$$

By (6.27) and (6.23) we know that $r(E)|_{H(E)} \subseteq D|_{H(E)} \subseteq E|_{H(E)}$, which by Lemma 6.1 implies that $H(D) \subseteq H(E)$. We can thus conclude that

$$r(E)|_{H(D)} \subseteq D|_{H(D)} \subseteq E|_{H(D)}. \tag{6.30}$$

As F is possible in an elimination order for r and $E = r^{k-1}(F)$, it follows that E is possible in an elimination order for r as well. But then, by (6.30) and the fact that r is monotone on reachable histories, we can conclude that $r(D)|_{H(D)} \subseteq r(E)|_{H(D)}$. Since we have seen above that $r(E)|_{H(D)} \subseteq D|_{H(D)}$, we conclude by Lemma 6.1 that $H(r(E)) \subseteq H(D)$. As $r(D)|_{H(D)} \subseteq r(E)|_{H(D)}$, this implies (6.29), which is equivalent to (6.28) that had to be shown.

Finally, the set inclusions

$$H(r^{k+1}(G)) \subseteq H(r^k(F)) \subseteq H(r^k(G))$$

follow directly from (6.28), (6.25) and Lemma 6.1. By induction on k, the proof is therefore complete. □

We are now ready to prove Theorem 3.2. As we will see, it follows rather directly from Lemma 6.8.

Proof of Theorem 3.2. Consider a reduction operator r that is monotone on reachable histories. We must show that r is order independent with respect to outcomes.

Let $M := \sum_{i \in I} |S_i|$ be the total number of strategies in the game. Then, $r^{M+1}(D) = r^M(D)$ for every product of strategy sets D. Consider an arbitrary elimination order $(D^0, ..., D^K)$ for r and some $k \in \{0, ..., K-1\}$. Then, we know from Lemma 6.8 that

$$H(r^{M+1}(D^k)) \subseteq H(r^M(D^{k+1})) \subseteq H(r^M(D^k)).$$

As $r^{M+1}(D^k) = r^M(D^k)$, it follows that $H(r^M(D^{k+1})) = H(r^M(D^k))$, and hence, in particular, $Z(r^M(D^{k+1})) = Z(r^M(D^k))$.

Since this holds for every $k \in \{0, ..., K-1\}$, we conclude that $Z(r^M(D^0)) = Z(r^M(D^K))$. As $r(D^K) = D^K$, it follows that $r^M(D^K) = D^K$. We thus conclude that

$$Z(D^K) = Z(r^M(D^K)) = Z(r^M(D^0)) = Z(r^M(S)).$$

As this holds for every elimination order $(D^0, ..., D^K)$ for r, we conclude that r is order independent with respect to outcomes. □

6.4. *Proof of Lemma 4.1*

In order to show that $(D^{bi,0}, ..., D^{bi,K})$ is an elimination order for sb, we must show properties (a), (b) and (c) in Definition 3.1. As properties (a) and (c) hold by construction, we need only concentrate on (b). The inclusion $D^{bi,k+1} \subseteq D^{bi,k}$ in (b) again holds by construction. Hence, it only remains to show that $sb_i(D^{bi,k}) \subseteq D_i^{bi,k+1}$ for every player i.

Take some $s_i \in sb_i(D^{bi,k})$. Then, $s_i \in D_i^{bi,k}$ and s_i is rational at $H(D^{bi,k})$ for some b_i that strongly believes $D_{-i}^{bi,k}$. Since $D^{bi,k}$ only puts restrictions on choices at histories in H^k, we have that $H^{k+1} \backslash H^k \subseteq H(D^{bi,k})$, and hence it follows that s_i is rational at $H^{k+1} \backslash H^k$ for b_i. Take some $h \in H_i(s_i) \cap (H^{k+1} \backslash H^k)$. Since b_i strongly believes $D_{-i}^{bi,k}$ and $h \in (H^{k+1} \backslash H^k) \subseteq H(D^{bi,k})$, the conditional belief $b_i(h)$ only assigns positive probability to opponents' strategies that prescribe the backward induction choice at every history that follows. As s_i is rational at h for b_i, the prescribed choice $s_i(h)$ at h must be the backward induction choice.

We thus conclude that $s_i(h)$ is the backward induction choice at every $h \in H_i(s_i) \cap (H^{k+1} \backslash H^k)$. Since s_i is in $D_i^{bi,k}$, we also know that $s_i(h)$ is the backward induction choice for every $h \in H_i(s_i) \cap H^k$. Therefore, $s_i(h)$ is the backward induction choice at every $h \in H_i(s_i) \cap H^{k+1}$. But then, by definition, $s_i \in D_i^{bi,k+1}$. As this holds for every $s_i \in sb_i(D^{bi,k})$, we conclude that $sb_i(D^{bi,k}) \subseteq D_i^{bi,k+1}$, which was to show.

We thus conclude that the backward induction sequence is an elimination order for sb. This completes the proof. □

References

Arieli, I., Aumann, R.J., 2015. The logic of backward induction. J. Econ. Theory 159, 443–464.
Baltag, A., Smets, S., Zvesper, J.A., 2009. Keep 'hoping' for rationality: a solution to the backward induction paradox. Synthese 169, 301–333 (Knowledge, Rationality and Action 705–737).
Basu, K., Weibull, J.W., 1991. Strategy subsets closed under rational behavior. Econ. Letters 36, 141–146.
Battigalli, P., 1997. On rationalizability in extensive games. J. Econ. Theory 74, 40–61.
Battigalli, P., Siniscalchi, M., 2002. Strong belief and forward induction reasoning. J. Econ. Theory 106, 356–391.
Catonini, E., 2017. On Non-monotonic Strategic Reasoning. Working paper.
Chen, J., Micali, S., 2013. The order independence of iterated dominance in extensive games. Theoretical Econ. 8, 125–163.
Heifetz, A., Perea, A., 2015. On the outcome equivalence of backward induction and extensive form rationalizability. Int. J. Game Theory 44, 37–59.
Kohlberg, E., Mertens, J.-F., 1986. On the strategic stability of equilibria. Econometrica 54, 1003–1037.
Kreps, D.M., Wilson, R., 1982. Sequential equilibria. Econometrica 50, 863–894.
Luo, X., Qian, X., Qu, C., 2016. Unified Approach to Iterated Elimination Procedures in Strategic Games. Working paper.
Pearce, D.G., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica 52, 1029–1050.
Penta, A., 2015. Robust dynamic implementation. J. Econ. Theory 160, 280–316.
Perea, A., 2012. Epistemic Game Theory: Reasoning and Choice. Cambridge University Press.
Perea, A., 2014. Belief in the opponents' future rationality. Games Econ. Behav. 83, 231–254.
Reny, P.J., 1992. Backward induction, normal form perfection and explicable equilibria. Econometrica 60, 627–649.
Rubinstein, A., 1991. Comments on the interpretation of game theory. Econometrica 59, 909–924.
Selten, R., 1965. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragezeit. Z. ges. Staatswissen. (JITE) 121, 301–324, 667–689.
Shimoji, M., Watson, J., 1998. Conditional dominance, rationalizability, and game forms. J. Econ. Theory 83, 161–195.