Int J Game Theory (2015) 44:37–59
DOI 10.1007/s00182-014-0418-x

On the outcome equivalence of backward induction and extensive form rationalizability

Aviad Heifetz · Andrés Perea

Accepted: 2 March 2014 / Published online: 19 March 2014

Abstract Pearce's (Econometrica 52:1029–1050, 1984) extensive-form rationalizablity (EFR) is a solution concept embodying a best-rationalization principle (Battigalli, Games Econ Behav 13:178–200, 1996; Battigalli and Siniscalchi, J Econ Theory 106:356–391, 2002) for forward-induction reasoning. EFR strategies may hence be distinct from backward-induction (BI) strategies. We provide a direct and transparent proof that, in perfect-information games with no relevant ties, the unique BI outcome is nevertheless identical to the unique EFR outcome, even when the EFR strategy profile and the BI strategy profile are distinct.

Keywords Backward induction · Extensive-form rationalizability · Forward induction

JEL Classification C72 · C73

1 Introduction

Subgame perfect equilibrium is one of the most fundamental solution concepts for extensive-form games with perfect information. In many such games, subgame perfection rules out implausible Nash equilibria which are based on incredible threats.

A. Heifetz (✉)
Department of Management and Economics, The Open University of Israel,
1 University Rd., Raanana, Israel
e-mail: aviadhe@openu.ac.il

A. Perea
EpiCenter and Department of Quantitative Economics, Maastricht University,
P.O. Box 616, 6200 MD, Maastricht, The Netherlands
e-mail: a.perea@maastrichtuniversity.nl

Moreover, there is a unique subgame perfect equilibrium in games with generic payoffs, or games that only have irrelevant payoff ties.[1] This unique subgame perfect equilibrium can be singled out via the backward induction (BI) procedure.

By definition, subgame-perfection (recursively) analyzes each subgame on its own, abstracting from the question what on earth could have led the players to reach this subgame in the first place. This question is particularly wanting for subgames which can only be reached if, on the path leading to the subgame, some player makes a dominated choice, or a choice which can only be rationalized if this player believes that one of her opponents will subsequently make a dominated choice, etc. In such subgames, it is therefore questionable why the other players should assume that this player will behave rationally in the future, or will believe her opponents will behave rationally in the future, and so forth. However, BI hinges on these assumptions of rationality and common belief in future rationality (see Perea 2014), unattending to past behavior. Thus, a hidden anchor underwriting BI is that all past instances of irrational behavior (or past beliefs about others' forthcoming irrationality, etc.) are to be interpreted as transient, unintended mistakes, which can be safely ignored when a given subgame is analyzed—no matter how often such instances of irrationality have already figured in the course of play leading to that subgame.

If, in contrast, one is reluctant to ignore past instances of irrationality as mere transient fluctuations in the player's healthy reasoning, a completely different set of considerations is called for, namely considerations of forward induction. The player's past behavior is then to be interpreted as bearing on her future choices and indicative for it. In particular, if a player has already manifested irrational behavior, there need not be any gurantee that she will behave optimally henceforth. Similarly, if a player's past behavior was optimal only under a belief of hers that her rivals will choose suboptimally in the future, there is no guarantee that she will attribute rationality to her rivals in any subsequent subgame; and so on.

Such forward induction considerations are very different from those underlying BI, and hence call for a different solution concept. Extensive-form rationalizability (EFR) is a solution concept formulated by Pearce (1984) which embodies such forward induction considerations. In particular, it relies on a best-rationalization principle (Battigalli 1996): when a player's future behavior is to be divined, she is ascribed the highest degree of rationality, belief in her rivals' rationality, belief in their belief of their rivals' rationality etc. compatible with the player's past observed behavior. In epistemic terms this best-rationalization principle is akin to the notion of 'strong belief in rationality'[2], which is indeed part and parcel of a characterization of EFR (Battigalli and Siniscalchi 2002).

Given the conceptual chasm between backward and forward induction, no wonder that there exist games in which some player's BI strategies are different from her EFR strategies. In particular, Reny (1992) gave an example of a generic game, in which

[1] That is, games where each player's payoffs are distinct from one another in the leaves which follow each of this player's decision nodes.

[2] Somewhat of a misnomer—if the player at the root of the game has a dominated move and she chooses it, at subsequent nodes the other players 'strongly believe she is rational' by attributing to her up-front irrationality.

a player has a unique BI strategy which is distinct from her unique EFR strategy. A related example appears in Perea (2012). Nevertheless, in these examples the outcome induced by the players' unique profile of BI strategies is identical to the outcome reached by their unique though distinct profile of EFR strategies. Even more surprisingly, Battigalli (1997) proved that in all games with no relevant ties, the unique BI outcome is identical to the unique EFR outcome.

This result is of fundamental importance, because it records why, in generic perfect-information games, BI never clashes—outcome-wise—with forward induction considerations. Had there existed such clashes, proponents of forward induction would have called to abandon subgame perfection as implausible at least in some games. But in view of the outcome equivalence theorem, any conceptual argument in favor of forward induction over backward induction simply remains mute for generic perfect-information games.

But what is the intuition behind the outcome equivalence of these very different modes of solving the game? Battigalli (1997)'s original proof was not transparent in this regard, because it relied on properties of fully stable sets (Kohlberg and Mertens 1986)—a complex notion which does not lend a direct intuition for confronting the two modes of reasoning. In fact, Battigalli's proof technique is based on Reny (1992), who also used properties of fully stable sets to show that in generic perfect-information games, the unique outcome induced by explicable equilibrium[3] is the BI outome. More recent proofs are not direct either. Perea (2012) suggests a method of proof which relies on the epistemic characterizations of backward and forward induction reasoning in games with potentially imperfect information. Chen and Micali (2011) provide a proof (see their Theorem 3) which relies on another result of theirs (Chen and Micali 2013) on the outcome equivalence under different orders of elimination of strategies in games with possibly imperfect information. The latter result relies in turn on the 'diamond property' of Church and Rosser (1936). Arieli and Aumann (2012) provide a proof designed for generic games with perfect information, but here again relying on an argument, devised by Gretlein (1983) for showing the outcome equivalence of various elimination orders of strategies in such games. In a similar vein, one can also prove the result by applying the notion of "nice weak dominance" (Marx and Swinkels 1997, 2000) to generic perfect-information games, and using Theorem 2 in Marx and Swinkels (1997).

In view of the complexity of all these arguments, our aim in this paper is to provide a direct proof of the outcome equivalence of BI and EFR, and thus shed more light on this important result.

We proceed, inductively, by trimming the leaves of the tree by BI. Obviously, the truncated tree has less room for forward induction considerations: wherever a player originally had to choose among several leaves, that decision node of hers is now replaced, in the truncated tree, by a leaf with payoffs corresponding to her original optimal choice at that decision node. Viewed from an earlier stage of the game, after the trimming the other players have now less decision nodes left to speculate about and wonder (on the basis of past behavior) how the player is going to choose in them.

[3] Like EFR, also explicable equilibrium is based on some kind of best-rationalization principle. However, the two concepts are different.

In Lemma 4—the 'truncation lemma'—we show, in particular, that this intuition is correct and has an important implication. If some outcome can, originally, be rationalized up to level $k + 1$ (i.e. with a profile of strategies each of which was rational, supported by a belief that others are rational when choosing strategies which are themselves rationalized by beliefs that the others' rivals are rational ... [$k + 1$ steps]), then in the truncated game the same outcome can be rationalized up to level k—though with, possibly, a different profile of strategies. The crux of the argument is to construct, explicitly, the possibly different strategies in the folded game embodying this possibly lower level of rationality. Conversely, if an outcome can be rationalized up to level k in the truncated game, it can also be rationalized up to level k in the original game—again, possibly with a different strategy profile, that we construct explicitly.

Hence, an EFR outcome of the original game, which was supported by a profile of strategies rationalized to level n for every $n \in \mathbb{N}$, is supported by a (possibly different) profile of strategies that, in the game truncated once by BI, are rationalized to level $n - 1$ for every $n \in \mathbb{N}$. But this is the same as saying that the strategies in this profile are rationalized for every $k \in \mathbb{N}$, and hence that the outcome induced by this profile is EFR also in the truncated game. And conversely, every outcome rationalized to level k in the trimmed game for every $k \in \mathbb{N}$ is rationalized to level k for every $k \in \mathbb{N}$ also in the original game for every $k \in \mathbb{N}$—implying that the EFR outcomes in the trimmed game are also EFR outcomes of the original game. This is the content of Lemma 5.

Continuing to fold the game by BI, the above argument shows that each EFR outcome is maintained after each folding step. As the game has no relevant ties and the BI procedure therefore halts after finitely many folding steps with a unique outcome, we conclude that this outcome is also the unique EFR outcome of all the folded games, and in particular also the unique EFR outcome of the full, original game—as we headed to show.

Strictly speaking, to prove Battigalli's theorem we only need one direction in Lemma 5, namely that every EFR outcome in the original game is also an EFR outcome in the truncated game. We do not really need the other inclusion. But since the other inclusion actually comes for free in our proof, we have decided to state Lemma 5 as described above. Moreover, we believe that the current content of Lemma 5, which states that the set of EFR outcomes actually remains identical when truncating the game, is interesting in its own right. More details on this issue can be found in Sect. 3.3.

The paper is organized as follows. Section 2 is dedicated to definitions. In Sect. 3 we state Battigalli's theorem, give a sketch of our proof, and illustrate the main steps in the proof by means of an example. A complete and formal proof is given in Sect. 4. In Sect. 5 we derive two auxiliary results that follow from our proof of Battigalli's theorem. The first result states that in perfect-information games with no relevant ties, truncating a game once by BI can decrease the rationality level of each decision node (i.e. the highest k for which the node is reached by a profile of strategies rationalized up to level k) by at most 1. This formalizes the idea described above that trimming the game once by BI leaves less room for forward induction reasoning. The second result states that whenever a decision node is followed by at most k actions, and this decision node can be reached by a profile of strategies that is rationalized to level k, then the player at that node will always make the BI choice in any such strategy profile that is rationalized to level k.

2 Definitions

2.1 Extensive-form games with perfect information

In this paper we focus on extensive-form games with perfect information. For every such game G we use the following notation: I denotes the set of players, N_i denotes the set of decision nodes for player i, at every decision node $n \in N_i$ we denote by $C_i(n)$ the set of available choices for player i at n, Z denotes the set of terminal nodes, and at every terminal node $z \in Z$ we denote by $u_i(z)$ the utility for player i at z.

Throughout the paper we assume that the sets I, N_i, $C_i(n)$ and Z are all finite—that is, we restrict attention to finite games—and that the game G is without relevant ties (Battigalli 1997) which means that for every player i, every decision node $n \in N_i$, and every two distinct terminal nodes $z, z' \in Z$ following n, we have that $u_i(z) \neq u_i(z')$.

2.2 Truncation and backward induction

Consider a finite extensive-form game G with perfect information and without relevant ties. Let N^{last} be the set of decision nodes that are not followed by any other decision node. Consider some last decision node $n \in N^{last}$ at which player i must make a choice. Since G is without relevant ties, there is a unique optimal choice $c^*(n) \in C_i(n)$ for player i at n. Let $tr(G)$ be the truncated game obtained from G if we replace every last decision node $n \in N^{last}$ by a terminal node z at which the utility for every player i coincides with the utility he obtains in G if the player moving at n chooses $c^*(n)$. Clearly, $tr(G)$ will again be a game without relevant ties.

If we repeatedly apply the truncation operator tr to the game G, we eventually end up with a trivial game G^* in which there is only one terminal node z^*. As G is without relevant ties, there is a unique terminal node z in the original game G for which the utilities of all players match with the utilities in G^* at z^*. This terminal node z in G is called the BI outcome of G.

2.3 Extensive-form rationalizability

For every player i we denote by S_i the set of strategies in the game G, whereas $S := \times_{i \in I} S_i$ denotes the set of strategy profiles. By $S_{-i} := \times_{j \neq i} S_j$ we denote the set of strategy profiles for i's opponents. For a given decision node n, let $S(n)$ be the set of strategy profiles $s \in S$ that reach n. Accordingly, let $S_i(n)$ be the set of strategies $s_i \in S_i$ for which there is some $s_{-i} \in S_{-i}$ such that $(s_i, s_{-i}) \in S(n)$. Similarly, let $S_{-i}(n)$ be the set of strategy profiles $s_{-i} \in S_{-i}$ for which there is some $s_i \in S_i$ such that $(s_i, s_{-i}) \in S(n)$. We say that $S_i(n)$ contains those strategies for player i that reach n, and that $S_{-i}(n)$ contains those strategy profiles for i's opponents that reach n.

A belief vector for player i is a vector $b_i = (b_i(n))_{n \in N_i}$ where $b_i(n) \in \Delta(S_{-i}(n))$ for every $n \in N_i$. Here, $\Delta(S_{-i}(n))$ denotes the set of probability distributions on $S_{-i}(n)$. That is, a belief vector b_i associates with every decision node $n \in N_i$ for player i some probability distribution $b_i(n)$ over the opponents' strategy profiles that reach n. We denote by B_i the set of all belief vectors for player i in G.

If s_i is a strategy of player i and $b_i(n) \in \Delta(S_{-i}(n))$ is a belief of player i at the node n, we say that $(s_i, b_i(n))$ reaches a node $\tilde{n} \in N$ if for some $s_{-i} \in S_{-i}(n)$ for which $b_i(n)(s_{-i}) > 0$, the strategy profile (s_i, s_{-i}) reaches $\tilde{n}$. Here, we denote by N the set of all decision nodes in the game. Accordingly, for a subset of nodes $\tilde{N} \subseteq N$, we say that $(s_i, b_i(n))$ only reaches $\tilde{N}$ if for every $s_{-i} \in S_{-i}(n)$ for which $b_i(n)(s_{-i}) > 0$, the strategy profile (s_i, s_{-i}) only reaches nodes in $\tilde{N}$, and no nodes outside $\tilde{N}$.

For a given strategy s_i, belief vector b_i, and decision node $n \in N_i$, let $u_i(s_i, b_i(n)|n)$ be the expected utility that results for player i if the game reaches n, player i holds the belief $b_i(n)$ at n over the opponents' strategy profiles, and chooses according to s_i in the subgame that starts at n. Strategy s_i is said to be optimal for player i at n for the belief $b_i(n)$ if $u_i(s_i, b_i(n)|n) \geq u_i(s_i', b_i(n)|n)$ for all $s_i' \in S_i$. Strategy s_i is said to be optimal for the belief vector b_i if, for every $n \in N_i$, the strategy s_i is optimal at n for $b_i(n)$. Note that we require optimality at n even when s_i does not reach n! This is different from Pearce (1984) and Battigalli (1997) notion of optimality, who require s_i only to be optimal at those decision nodes $n \in N_i$ that are actually reached by s_i. This difference will have consequences for the eventual strategies selected, but not for the plans of action (Rubinstein 1991) and hence neither for the outcomes selected by EFR. Since we are eventually interested in the outcomes selected by extensive-form rationalizability, this difference is not crucial for what we do in this paper.

The concept of EFR (Pearce 1984; Battigalli 1997) recursively defines, for every $k \geq 0$, sets of strategies S_i^k and sets of belief vectors B_i^k, as follows.

Definition 1 (*Extensive-form rationalizability*) Consider a finite extensive-form game G with perfect information and without relevant ties.

Induction start. Define $S_i^0 := S_i$ and $B_i^0 := B_i$ for all players i.

Induction step. Let $k \geq 1$, and suppose that S_i^{k-1} and B_i^{k-1} have already been defined for all players i. Then, S_i^k is defined as the set of strategies for player i that are optimal for some belief vector in B_i^{k-1}. Moreover, B_i^k is defined as the set of belief vectors $b_i \in B_i^{k-1}$ for which $b_i(n) \in \Delta(S_{-i}^k \cap S_{-i}(n))$ at every decision node $n \in N_i$ where $S_{-i}^k \cap S_{-i}(n)$ is nonempty.

Here, $S_{-i}^k := \times_{j \neq i} S_j^k$, and by $\Delta(S_{-i}^k \cap S_{-i}(n))$ we denote the set of probability distributions on $S_{-i}(n)$ that only assign positive probability to strategy profiles in S_{-i}^k. So, the condition that characterizes B_i^k states that, whenever there is an opponents' strategy profile in S_{-i}^k that reaches n, the belief $b_i(n)$ should *only* assign positive probability to strategy profiles in S_{-i}^k. This is actually the forward-induction element in the definition, reflecting the best-rationalization principle (Battigalli 1996).

For every player i, we denote by $S_i^\infty := \cap_{k \geq 0} S_i^k$ the set of extensive-form rationalizable strategies, whereas $S^\infty := \times_{i \in I} S_i^\infty$ is the set of extensive-form rationalizable strategy profiles. Every terminal node z that is reached by some profile $s \in S^\infty$ is called an extensive-form rationalizable outcome.

Our definition of EFR differs in two ways from the original definitions by Pearce (1984) and Battigalli (1997). Namely, in our definition we do not require the belief vectors in B_i^0 to satisfy Bayesian updating where possible, whereas the definitions by Pearce and Battigalli do[4]. However, Shimoji and Watson (1998) have shown that this difference is of no relevance[5]. We have chosen not to impose Bayesian updating merely for the sake of simplicity, as it makes the procedure above easier. By doing so, we actually view the EFR procedure above as a purely algorithmic procedure that yields us the extensive-form rationalizable strategies. One should bear in mind, however, that the underlying concept of EFR does assume Bayesian updating. This conceptually makes sense, as Bayesian updating guarantees that the players in the dynamic game will be dynamically consistent. As such, every player will have at least one strategy that is optimal at each of his information sets, and such strategies can be found through a folding back procedure.

The second difference, as we already mentioned above, lies in the way we define optimality of a strategy s_i for some belief vector b_i. We require that s_i be optimal for $b_i(n)$ at every decision node $n \in N_i$—no matter whether n is reached by s_i or not. In turn, Pearce and Battigalli only require s_i to be optimal for $b_i(n)$ at all $n \in N_i$ that are reached by s_i. This difference will have consequences for the set of strategies S_i^∞ obtained at the end, but not for the plans of action[6] (Rubinstein 1991) induced by the strategies in S_i^∞. In particular, this difference will not matter for the outcome(s) reached by S^∞, which is the primary focus of this paper.

In what follows, we often write $S_i^k(G)$ and $S_i^\infty(G)$ instead of S_i^k and S_i^∞, to indicate that it corresponds to a specific game G. The same holds for all the other objects we have defined in this section. So, we often write $S_i(G)$, $N_i(G)$, and so on, to make clear that these objects are defined for a specific game G.

3 Sketch of proof

In this section we state Battigalli's theorem, highlight the main steps in our proof, and illustrate these steps by means of an example. We hope that this section will help the reader to understand the formal proof of the theorem in the next section.

3.1 Main steps in proof

The purpose of this paper is to provide an elementary proof for the following well-known result.

[4] To be more precise, the condition in Pearce and Battigalli is equivalent to Bayesian updating, in the sense that it gives the same probability ratios for strategies that are not ruled out by information sets, but it does not require "normalization".

[5] It can be shown, namely, that whenever a strategy is optimal for a belief vector that strongly believes a set of opponents' strategy profiles, then it is also optimal for an equivalent belief vector that strongly believes this same set of strategy profiles, and which satisfies Bayesian updating.

[6] Here, a plan of action represents a class of behaviorally equivalent strategies, that is, a class of strategies that reach the same decision nodes for player i, and make the same choices at these decision nodes. Actually, it can be shown that not only the plans of action induced by S_i^∞ are the same in our setting as in Pearce's and Battigalli's, but also the plans of action induced by S_i^k, at every step k of the procedure.

Theorem 2 *(Battigalli (1997)) Consider a finite extensive-form game G with perfect information and without relevant ties. Then, G has a unique extensive-form rationalizable outcome, and this outcome coincides with the backward induction outcome.*

In the next section we will provide a complete and formal proof for this result. Our main idea in the proof is to show that, for every finite extensive-form game G with perfect information and without relevant ties, the set of extensive-form rationalizable outcomes does not change by truncating the game G—see Lemma 5. This result then implies Theorem 2 above. To see this, take a game G with perfect information and without relevant ties. Then, by repeatedly truncating the game, we finally obtain a game G^* with only one possible outcome—the BI outcome in G. At the same time, the sequence of games we obtain by repeatedly truncating the game G contains only games without relevant ties. Therefore, Lemma 5 guarantees that the set of extensive-form rationalizable outcomes does not change when we move from one game to another in this sequence. In particular, the set of extensive-form rationalizable outcomes in the original game G must be the same as in the final game G^* in this sequence. But the final game G^* contains only one outcome, which is the BI outcome in G. As such, the only extensive-form rationalizable outcome in G^* is the BI outcome in G. Hence, we can conclude that G contains only one extensive-form rationalizable outcome, which coincides with the BI outcome in G—precisely what we want to show.

But how do we prove that the set of extensive-form rationalizable outcomes does not change when truncating the game? The key will be to prove the following property, which we refer to as the 'truncation lemma'—see Lemma 4. To formally state the lemma, let $S_i^k(G)$ be the set of strategies for player i in G that survive the first k steps in the EFR procedure, and let $N^k(G)$ be the set of decision nodes in G that are reached by some strategy profile in $S^k(G) := \times_{i\in I} S_i^k(G)$. Moreover, when we take subsets $N' \subseteq N(G)$ and $N'' \subseteq N(tr(G))$, we write $N' \subseteq^* N''$ whenever $N' \cap N(tr(G)) \subseteq N''$.

Truncation Lemma: Consider a finite extensive-form game G with perfect information and without relevant ties, and let $G' := tr(G)$ be the truncated game. Then, for all $k \geq 0$:

(a) there is a mapping $f_i^k : S_i^{k+1}(G) \to S_i^k(G')$ such that s_i and $f_i^k(s_i)$ are identical on $N^k(G)$ for every $s_i \in S_i^{k+1}(G)$,
(b) there is a mapping $g_i^k : S_i^k(G') \to S_i^k(G)$ such that σ_i and $g_i^k(\sigma_i)$ are identical on $N^{k-1}(G')$ for every $\sigma_i \in S_i^k(G')$,
(c) $N^{k+1}(G) \subseteq^* N^k(G') \subseteq N^k(G)$.

In part (b), we define $N^{-1}(G') := N(G')$.

So, part (a) states that for every strategy s_i that survives the first $k + 1$ steps of EFR in the original game, we can find a corresponding strategy in the truncated game that (1) survives the first k steps of EFR in the truncated game and (2) coincides with s_i on the nodes in G that can still be reached under $S^k(G)$. In particular, for every strategy profile s that survives the first $k + 1$ steps in the original game, we can find a corresponding strategy profile σ that survives the first k steps in the truncated game, and which coincides with s on the nodes in G that can still be reached under

$S^k(G)$. But since s is in $S^{k+1}(G)$—and hence in $S^k(G)$—it follows that s and σ behave identically on the path of s, and must therefore induce the same outcome. We thus conclude that for every strategy profile s in $S^{k+1}(G)$ there is a corresponding strategy profile σ in $S^k(G')$ which induces the same outcome as s. By letting $k \to \infty$, we obtain that for every extensive-form rationalizable strategy profile $s \in S^\infty(G)$ in the original game, there is a corresponding extensive-form rationalizable strategy profile $\sigma \in S^\infty(G')$ in the truncated game generating the same outcome as s. In other words, every extensive-form rationalizable outcome in the original game G is also an extensive-form rationalizable outcome in the truncated game G'.

In a similar fashion, we can derive from (b) that every extensive-form rationalizable outcome in the truncated game G' is also an extensive-form rationalizable outcome in the original game G. By combining these two insights, we conclude that the extensive-form rationalizable outcomes in G are exactly the same as in the truncated game G', which yields Lemma 5. So, indeed, the truncation lemma implies the result in Lemma 5.

3.2 Example

To illustrate how to construct the "outcome preserving" mappings f_i^k and g_i^k in the truncation lemma, consider the game G in Fig. 1, together with the truncated game G'. By $n_1, ..., n_5$ we denote the decision nodes in G. Note that $N^{last}(G) = n_5$ and $c^*(n_5) = i$. Hence, replacing the decision node n_5 in G by a terminal node with utilities (6, 2, 6) leads to the truncated game G', with decision nodes $n_1, ..., n_4$.

It may be verified that the sets of strategies $S_i^k(G)$ and $S_i^k(G')$, which underly the definition of EFR, are given by Table 1. To see this, let us first concentrate on the original game G. Note that strategy (d, g) can never be optimal for player 2 at n_2, as c gives him 5 for sure, whereas (d, g) yields him at most 4. So, (d, g) is not in $S_2^1(G)$.

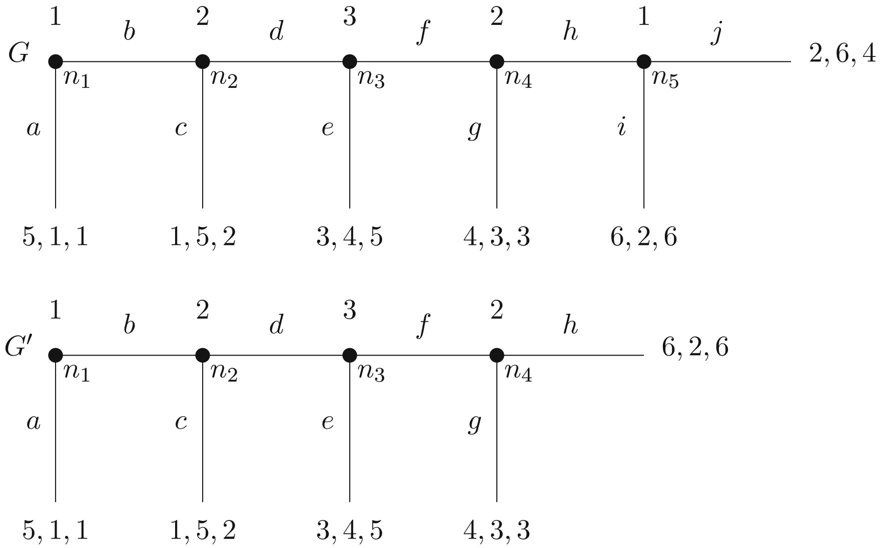

Fig. 1 Illustration of the truncation lemma

Table 1 Sets of strategies S_i^k in the games G and G'

$S_1^0(G) = \{(a,i),(a,j),(b,i),(b,j)\}$	$S_2^0(G) = \{(c,g),(c,h),(d,g),(d,h)\}$	$S_3^0(G) = \{e, f\}$
$S_1^1(G) = \{(a,i),(b,i)\}$	$S_2^1(G) = \{(c,g),(c,h),(d,h)\}$	$S_3^1(G) = \{e, f\}$
$S_1^2(G) = \{(a,i),(b,i)\}$	$S_2^2(G) = \{(c,g)\}$	$S_3^2(G) = \{f\}$
$S_1^3(G) = \{(a,i)\}$	$S_2^3(G) = \{(c,g)\}$	$S_3^3(G) = \{f\}$
$S_1^0(G') = \{a, b\}$	$S_2^0(G') = \{(c,g),(c,h),(d,g),(d,h)\}$	$S_3^0(G') = \{e, f\}$
$S_1^1(G') = \{a, b\}$	$S_2^1(G') = \{(c,g)\}$	$S_3^1(G') = \{e, f\}$
$S_1^2(G') = \{a\}$	$S_2^2(G') = \{(c,g)\}$	$S_3^2(G') = \{e, f\}$

As (b,i) is the only strategy in $S_1^1(G)$ that leads to n_2, player 2 must believe in $B_2^1(G)$, at his decision node n_2, that player 1 will continue with i. But then, player 2 must choose c at n_2 in $S_2^2(G)$. Similarly, (b,i) is the only strategy in $S_1^1(G)$ that leads to n_4, and hence player 2 must believe in $B_2^1(G)$, at his decision node n_4, that player 1 will continue with i. But then, player 2 must choose g at n_4 in $S_2^2(G)$. So, we conclude that $S_2^2(G) = \{(c,g)\}$. Observe also that (b,i) is the only strategy in $S_1^1(G)$ that leads to n_3, and (d,h) is the only strategy in $S_2^1(G)$ that leads to n_3. Therefore, player 3 must in $B_3^1(G)$ believe, at his decision node n_3, that player 1 will continue with i, and that player 2 will continue with h. But then, player 3's only optimal strategy in $S_3^2(G)$ is f. In $B_1^2(G)$, player 1 must therefore believe, at the beginning of the game, that player 2 will choose (c,g)—his only strategy in $S_2^2(G)$. But then, player 1's only optimal strategy in $S_1^3(G)$ is (a,i). Clearly, $S_i^\infty(G) = S_i^3(G)$ for all players i in the game. In particular, player 3's only extensive-form rationalizable strategy in the game is f.

Let us now turn to the truncated game G'. Note that after the truncation, the only optimal choice for player 2 at n_2 is c, and the only optimal choice for player 2 at n_4 is g. Therefore, $S_2^1(G') = \{(c,g)\}$. As there is no player 2 strategy in $S_2^1(G')$ leading to n_3 anymore, player 3 is free to hold any belief at n_3 in $B_3^1(G')$, which means that $S_3^2(G') = \{e, f\}$. Actually, for every $k \geq 2$ there will be no strategy in $S_2^k(G')$ leading to n_3, and hence player 3 is free to hold any belief at n_3 in $B_3^k(G')$ for every $k \geq 2$. This means that $S_3^\infty(G') = \{e, f\}$. Note that in the truncated game, $S_i^\infty(G') = S_i^2(G')$ for all players i. So, we see that player 3 has two extensive-form rationalizable strategies in G'—namely e and f—whereas in the original game he only had one—namely f. However, the extensive-form rationalizable outcome in both games is the same. Namely, as $S_1^\infty(G) = \{(a,i)\}$ and $S_1^\infty(G') = \{a\}$, the extensive-form rationalizable outcome in both games is the outcome which results from player 1 choosing a at the beginning.

We now show how to construct the "outcome preserving" mappings f_i^k and g_i^k in the truncation lemma. By definition, we must construct $f_i^k : S_i^{k+1}(G) \to S_i^k(G')$ such that the strategies $f_i^k(s_i)$ and s_i are always indentical on $N^k(G)$. Similarly, the mapping $g_i^k : S_i^k(G') \to S_i^k(G)$ must be such that the strategies $g_i^k(\sigma_i)$ and σ_i are always identical on $N^{k-1}(G')$. We have summarized the sets $N^k(G)$ and $N^k(G')$ in

Table 2 Sets N^k in games G and G'

$N^0(G) = \{n_1, n_2, n_3, n_4, n_5\}$
$N^1(G) = \{n_1, n_2, n_3, n_4, n_5\}$
$N^2(G) = \{n_1, n_2\}$
$N^3(G) = N^\infty(G) = \{n_1\}$
$N^0(G') = \{n_1, n_2, n_3, n_4\}$
$N^1(G') = \{n_1, n_2\}$
$N^2(G') = N^\infty(G') = \{n_1\}$

Table 3 The "outcome preserving" mappings f_i^k and g_i^k between G and G'

$g_1^0 : S_1^0(G') \to S_1^0(G)$ $g_1^0 : \{a, b\} \to \{(a, i), (a, j), (b, i), (b, j)\}$	$g_2^0 : S_2^0(G') \to S_2^0(G)$ $g_2^0 : \{(c, g), (c, h), (d, g), (d, h)\} \to \{(c, g), (c, h), (d, g), (d, h)\}$	$g_3^0 : S_3^0(G') \to S_3^0(G)$ $g_3^0 : \{e, f\} \to \{e, f\}$
$g_1^0(a) = (a, i)$, $g_1^0(b) = (b, i)$	$g_2^0 = id$	$g_3^0 = id$
$f_1^0 : S_1^1(G) \to S_1^0(G')$ $f_1^0 : \{(a, i), (b, i)\} \to \{a, b\}$	$f_2^0 : S_2^1(G) \to S_2^0(G')$ $f_2^0 : \{(c, g), (c, h), (d, h)\} \to \{(c, g), (c, h), (d, g), (d, h)\}$	$f_3^0 : S_3^1(G) \to S_3^0(G')$ $f_3^0 : \{e, f\} \to \{e, f\}$
$f_1^0(a, i) = a$, $f_1^0(b, i) = b$	$f_2^0 = id$	$f_3^0 = id$
$g_1^1 : S_1^1(G') \to S_1^1(G)$ $g_1^1 : \{a, b\} \to \{(a, i), (b, i)\}$	$g_2^1 : S_2^1(G') \to S_2^1(G)$ $g_2^1 : \{(c, g)\} \to \{(c, g), (c, h), (d, h)\}$	$g_3^1 : S_3^1(G') \to S_3^1(G)$ $g_3^1 : \{e, f\} \to \{e, f\}$
$g_1^1(a) = (a, i)$, $g_1^1(b) = (b, i)$	$g_2^1(c, g) = (c, g)$	$g_3^1 = id$
$f_1^1 : S_1^2(G) \to S_1^1(G')$ $f_1^1 : \{(a, i), (b, i)\} \to \{a, b\}$	$f_2^1 : S_2^2(G) \to S_2^1(G')$ $f_2^1 : \{(c, g)\} \to \{(c, g)\}$	$f_3^1 : S_3^2(G) \to S_3^1(G')$ $f_3^1 : \{f\} \to \{e, f\}$
$f_1^1(a, i) = a$, $f_1^1(b, i) = b$	$f_2^1(c, g) = (c, g)$	$f_3^1(f) = f$
$g_1^2 : S_1^2(G') \to S_1^2(G)$ $g_1^2 : \{a\} \to \{(a, i), (b, i)\}$	$g_2^2 : S_2^2(G') \to S_2^2(G)$ $g_2^2 : \{(c, g)\} \to \{(c, g)\}$	$g_3^2 : S_3^2(G') \to S_3^2(G)$ $g_3^2 : \{e, f\} \to \{f\}$
$g_1^2(a) = (a, i)$	$g_2^2(c, g) = (c, g)$	$g_3^2(e) = g_3^2(f) = f$
$f_1^2 : S_1^3(G) \to S_1^2(G')$ $f_1^2 : \{(a, i)\} \to \{a\}$	$f_2^2 : S_2^3(G) \to S_2^2(G')$ $f_2^2 : \{(c, g)\} \to \{(c, g)\}$	$f_3^2 : S_3^3(G) \to S_3^2(G')$ $f_3^2 : \{f\} \to \{e, f\}$
$f_1^2(a, i) = a$	$f_2^2(c, g) = (c, g)$	$f_3^2(f) = f$

Table 2, and have listed the "outcome preserving" mappings f_i^k and g_i^k in Table 3. In this table, id denotes the identity mapping. So, for instance, $f_2^0 : \{(c, g), (c, h), (d, h)\} \to \{(c, g), (c, h), (d, g), (d, h)\}$ is the identity mapping that maps (c, g) to

(c, g), (c, h) to (c, h) and (d, h) to (d, h). The "critical" transformation in this table is the mapping $g_3^2 : \{e, f\} \to \{f\}$. This transformation maps the strategy e in $S_3^2(G')$ to the different strategy f in $S_3^2(G)$. Note, however, that $N^1(G') = \{n_1, n_2\}$ (see Table 2), and hence the strategies e and $g_3^2(e) = f$ are identical on $N^1(G')$, as they are supposed to be.

The reader may use this example and the associated Tables 1, 2 and 3 as a guideline when going through the proof of the truncation lemma.

3.3 "Redundancies" in Proof

Some critical readers will have noticed that the proof outlined above actually shows more than we really need. To show Battigalli's theorem, it sufficies to prove that when we truncate the game G, then every extensive-form rationalizable outcome in the original game G will also be an extensive-form rationalizable outcome in the truncated game G'. We do not really need the other inclusion. That is, we only need one direction in Lemma 5.

To see this, suppose that for every game G with perfect information and without relevant ties, every extensive-form rationalizable outcome in G will also be an extensive-form rationalizable outcome in the truncated game G' (but not necessarily vice versa). Again, by repeatedly truncating the game G, we eventually arrive at a trivial game G^* with a unique outcome—the BI outcome in G. By repeatedly using the (weaker) result above, we can still conclude that every extensive-form rationalizable outcome in G must be an extensive-form rationalizable outcome in G^*, which can only be the BI outcome in G. Hence, there is only one possible extensive-form rationalizable outcome in G, namely the BI outcome in G. Since we know, by Pearce (1984), that the set of extensive-form rationalizable strategies—and hence the set of extensive-form rationalizable outcomes—is always non-empty, we conclude that G contains exactly one extensive-form rationalizable outcome, namely the BI outcome. So, Battigalli's theorem still follows by this one direction of Lemma 5.

In order to prove this one direction in Lemma 5, it suffices to show part (a) in the truncation lemma above. We do not need part (b) for that purpose. So, the reader may ask why we did not concentrate exclusively on part (a) in the truncation lemma, since that is really all we need. The reason is that we prove part (a) in the truncation lemma by induction on k, and to prove step k of part (a) we need the induction assumption on step $k-1$ of part (b). Or, at least, we do not know how to prove step k of part (a) without using the statement of step $k-1$ of part (b). In other words, part (b) in the truncation lemma helps us to prove part (a)—the part that is really crucial for our purposes.

But, once we have both part (a) and part (b) in the truncation lemma, it not only follows that every extensive-form rationalizable outcome in the original game G is an extensive-form rationalizable outcome in the truncated game G', but also the converse easily follows. That is, it follows that by truncating the game we do not change the set of extensive-form rationalizable outcomes. Since we believe this property to be interesting in its own right, we have stated it as such in Lemma 5, although it states more than we strictly need to prove Battigalli's theorem.

4 Formal Proof

In this section we will provide a complete and formal proof of Battigalli's theorem (Theorem 2) which we stated in the previous section. The keys to proving this result are the two lemmas we state below.

Lemma 3 (Sufficiency principle) *Consider a finite extensive-form game G with perfect information and without relevant ties. For a player i and any $k \geq 0$, consider some belief vector $b_i \in B_i^k$, a decision node $n \in N_i$ and two strategies s_i, s_i' where $u_i(s_i, b_i(n)|n) < u_i(s_i', b_i(n)|n)$. Then, there is some $s_i'' \in S_i^k$ such that $u_i(s_i, b_i(n)|n) < u_i(s_i'', b_i(n)|n)$.*

Proof of Lemma 3 Follows from the Claim on p.54 in Battigalli (1997) and from the proof of Lemma 9.8.3 in Perea (2012). □

From Lemma 3 it follows that, for checking the optimality of a strategy s_i at a decision node $n \in N_i$ against a belief $b_i \in B_i^k$, it is sufficient to check that s_i is optimal among strategies in S_i^k only. For that reason, we have called this lemma "sufficiency principle".

For the next lemma—the "truncation lemma"—we need the following additional definitions and notation. Consider some finite extensive-form game G with perfect information and without relevant ties. Let $N(G)$ be the set of all decision nodes in G. For every $k \geq 0$, denote by $N^k(G)$ the set of decision nodes that are reached by some $s \in S^k(G)$. Remember that $tr(G)$ denotes the truncation of the game G, and that $N^{last}(G)$ is the set of decision nodes in G that are not followed by any other decision node in G. So, $N^{last}(G)$ contains precisely those decision nodes that are in G but not in $tr(G)$. Hence, $N(G) = N(tr(G)) \cup N^{last}(G)$. In the sequel, when we take subsets $N' \subseteq N(G)$ and $N'' \subseteq N(tr(G))$, we write $N' \subseteq^* N''$ whenever $N' \cap N(tr(G)) \subseteq N''$.

Now, take a strategy s_i for player i in G, a strategy σ_i for player i in $tr(G)$, and some subset of decision nodes $\bar{N} \subseteq N(G)$. We say that s_i and σ_i are *identical on* $\bar{N}$ if (a) s_i and σ_i prescribe the same choice at every $n \in \bar{N} \cap N_i(tr(G))$, and (b) s_i prescribes at every $n \in N_i(G) \cap N^{last}(G)$ the unique optimal choice $c^*(n)$. By construction, if s_i and σ_i are identical on $\bar{N}$, then s_i and σ_i always induce the same expected utility on every $n \in N_i(G) \cap \bar{N}$.

The following lemma describes the effect that truncation of a game has on the sets S_i^k that underly the definition of EFR. For that reason we call it the 'truncation lemma'. For the sake of clarity, we write strategies in G as s_i, and strategies in $tr(G)$ as σ_i.

Lemma 4 (Truncation Lemma) *Consider a finite extensive-form game G with perfect information and without relevant ties, and let $G' := tr(G)$ be the truncated game. Then, for all $k \geq 0$:*

(a) there is a mapping $f_i^k : S_i^{k+1}(G) \to S_i^k(G')$ such that s_i and $f_i^k(s_i)$ are identical on $N^k(G)$ for every $s_i \in S_i^{k+1}(G)$,

(b) there is a mapping $g_i^k : S_i^k(G') \to S_i^k(G)$ such that σ_i and $g_i^k(\sigma_i)$ are identical on $N^{k-1}(G')$ for every $\sigma_i \in S_i^k(G')$,
(c) $N^{k+1}(G) \subseteq^* N^k(G') \subseteq N^k(G)$.

In part (b), we define $N^{-1}(G') := N(G')$.

Proof of Lemma 4 To make the reading easier, we write strategies in G as s_i, belief vectors in G as b_i, and utilities in G as u_i, whereas we write strategies in G' as σ_i, belief vectors in G' as β_i, and utilities in G' as v_i. We prove the statements (a), (b) and (c) by induction on k.

Induction start. Let us first consider the case $k = 0$.

(a) Take some strategy $s_i \in S_i^1(G)$. Then, s_i must be optimal for some belief vector $b_i \in B_i^0(G)$, and hence at every last decision node $n \in N_i(G) \cap N^{last}(G)$ it must prescribe the optimal choice $c^*(n)$. Let $f_i^0(s_i)$ simply be the restriction of s_i to the decision nodes in $N_i(G')$. Then, clearly, $f_i^0(s_i) \in S_i^0(G')$, and is identical to s_i on $N^0(G) = N(G)$.
(b) Take some $\sigma_i \in S_i^0(G')$. Let $g_i^0(\sigma_i)$ be the strategy in G that coincides with σ_i on decision nodes in $N_i(G')$, and that at every $n \in N_i(G) \cap N^{last}(G)$ prescribes the optimal choice $c^*(n)$. Then, clearly, $g_i^0(\sigma_i) \in S_i^0(G)$, and is identical to σ_i on $N^{-1}(G') = N(G')$.
(c) By definition, $N^1(G) \subseteq^* N^0(G') \subseteq N^0(G)$.

Induction step. Take now some $k \geq 1$, and assume that (a), (b) and (c) are true for all $k' \leq k - 1$.
(a) We will construct the "outcome preserving" mapping $f_i^k : S_i^{k+1}(G) \to S_i^k(G')$. Take some strategy $s_i \in S_i^{k+1}(G)$. Then, in particular, $s_i \in S_i^1(G)$, and hence s_i prescribes the optimal choice $c^*(n)$ at every $n \in N_i(G) \cap N^{last}(G)$. We define the strategy $f_i^k(s_i)$ in G' as follows. Choose some arbitrary strategy $\sigma_i^* \in S_i^k(G')$. For every decision node $n \in N_i(G')$, define

$$(f_i^k(s_i))(n) := \begin{cases} s_i(n), & \text{if } n \in N^k(G) \\ \sigma_i^*(n), & \text{if } n \notin N^k(G) \end{cases}.$$

First of all, observe that $f_i^k(s_i)$ is identical to s_i on $N^k(G)$. Hence, it remains to show that $f_i^k(s_i) \in S_i^k(G')$. That is, we must show that $f_i^k(s_i)$ is optimal for some belief vector $\beta_i \in B_i^{k-1}(G')$. So, at every $n \in N_i(G')$ we must find some belief $\beta_i(n) \in B_i^{k-1}(G', n)$ such that $f_i^k(s_i)$ is optimal at n for the belief $\beta_i(n)$. Here, $B_i^{k-1}(G', n)$ denotes the set of beliefs that are possible at n in $B_i^{k-1}(G')$. Take some $n \in N_i(G')$. We distinguish two cases: (1) $n \in N^k(G)$ and (2) $n \notin N^k(G)$.

Case 1 Assume that $n \in N^k(G)$.

Since $s_i \in S_i^{k+1}(G)$, strategy s_i is optimal at n for some belief $b_i(n) \in B_i^k(G, n)$. As $n \in N^k(G)$, there is some $s_{-i} \in S_{-i}^k(G)$ that reaches n. But then, by definition of $B_i^k(G)$, we must have that $b_i(n) \in \Delta(S_{-i}^k(G) \cap S_{-i}(G, n))$. We transform the belief $b_i(n)$ into a new belief $\beta_i(n)$ in G' as follows. Let $\beta_i(n) : S_{-i}(G') \to [0, 1]$ be given by

$$\beta_i(n)(\sigma_{-i}) := \sum_{s_{-i} \in S^k_{-i}(G): f^{k-1}_{-i}(s_{-i}) = \sigma_{-i}} b_i(n)(s_{-i}) \tag{1}$$

for all $\sigma_{-i} \in S_{-i}(G')$. Here, for a given $s_{-i} = (s_j)_{j \neq i} \in S^k_{-i}(G)$, we define $f^{k-1}_{-i}(s_{-i})$ as $(f^{k-1}_j(s_j))_{j \neq i}$. Recall that $f^{k-1}_j : S^k_j(G) \to S^{k-1}_j(G')$, which implies that $f^{k-1}_{-i} : S^k_{-i}(G) \to S^{k-1}_{-i}(G')$. Hence, $f^{k-1}_{-i}(s_{-i})$ is well-defined for every $s_{-i} \in S^k_{-i}(G)$.

We show (i) that $\beta_i(n)$ is a probability distribution on $S_{-i}(G')$, (ii) that $\beta_i(n) \in B^{k-1}_i(G', n))$, and (iii) that $f^k_i(s_i)$ is optimal at n for $\beta_i(n)$.

(i) By definition, $\beta_i(n)(\sigma_{-i}) \geq 0$ for all $\sigma_{-i} \in S_{-i}(G')$. It remains to show that $\sum_{\sigma_{-i} \in S_{-i}(G')} \beta_i(n)(\sigma_{-i}) = 1$. By (1),

$$\sum_{\sigma_{-i} \in S_{-i}(G')} \beta_i(n)(\sigma_{-i}) = \sum_{\sigma_{-i} \in S_{-i}(G')} \sum_{s_{-i} \in S^k_{-i}(G): f^{k-1}_{-i}(s_{-i}) = \sigma_{-i}} b_i(n)(s_{-i})$$

$$= \sum_{s_{-i} \in S^k_{-i}(G)} b_i(n)(s_i) = 1,$$

where the latter equality follows from the assumption that $b_i(n) \in \Delta(S^k_{-i}(G))$.

(ii) As $n \in N^k(G)$, we know by induction assumption on (c) that $n \in N^{k-1}(G')$. So, there is some $\sigma_{-i} \in S^{k-1}_{-i}(G')$ that reaches n. Hence, to show that $\beta_i(n) \in B^{k-1}_i(G', n)$, we must prove that $\beta_i(n) \in \Delta(S^{k-1}_{-i}(G') \cap S_{-i}(G', n))$. To this purpose, we first show that $\beta_i(n) \in \Delta(S^{k-1}_{-i}(G'))$, and then we prove that $\beta_i(n) \in \Delta(S_{-i}(G', n))$.

We start by showing that $\beta_i(n) \in \Delta(S^{k-1}_{-i}(G'))$. Suppose that $\beta_i(n)(\sigma_{-i}) > 0$ for some $\sigma_{-i} \in S_{-i}(G')$. Then, by (1), $\sigma_{-i} = f^{k-1}_{-i}(s_{-i})$ for some $s_{-i} \in S^k_{-i}(G)$ with $b_i(n)(s_{-i}) > 0$. So, $s_{-i} = (s_j)_{j \neq i}$ with $s_j \in S^k_j(G)$ for every $j \neq i$. By our induction assumption on (a), we know that $f^{k-1}_j(s_j) \in S^{k-1}_j(G')$ for every $j \neq i$. Hence, $\sigma_{-i} = f^{k-1}_{-i}(s_{-i}) \in S^{k-1}_{-i}(G')$. So, we conclude that $\beta_i(n)(\sigma_{-i}) > 0$ only if $\sigma_{-i} \in S^{k-1}_{-i}(G')$, and hence $\beta_i(n) \in \Delta(S^{k-1}_{-i}(G'))$.

We now show that $\beta_i(n) \in \Delta(S_{-i}(G', n))$. Suppose that $\beta_i(n)(\sigma_{-i}) > 0$ for some $\sigma_{-i} \in S_{-i}(G')$. Then, by (1), $\sigma_{-i} = f^{k-1}_{-i}(s_{-i})$ for some $s_{-i} \in S^k_{-i}(G)$ with $b_i(n)(s_{-i}) > 0$. As $b_i(n) \in \Delta(S^k_{-i}(G) \cap S_{-i}(G, n))$, we know that $s_{-i} \in S_{-i}(G, n)$. So $s_{-i} = (s_j)_{j \neq i}$ with $s_j \in S_j(G, n)$ for all $j \neq i$. That is, s_j chooses at every player j node before n the choice leading to n. By our induction assumption on (a) we know that $f^{k-1}_j(s_j)$ is identical to s_j on $N^{k-1}(G)$. As $n \in N^k(G)$, all player j nodes before n are in $N^{k-1}(G)$, and hence also $f^{k-1}_j(s_j)$ chooses at every player j node before n the choice leading to n. So, $f^{k-1}_j(s_j) \in S_j(G', n)$. But then, $\sigma_{-i} = f^{k-1}_{-i}(s_{-i}) \in S_{-i}(G', n)$. So, we conclude that $\beta_i(n)(\sigma_{-i}) > 0$ only if $\sigma_{-i} \in S_{-i}(G', n)$, and hence $\beta_i(n) \in \Delta(S_{-i}(G', n))$.

We thus conclude that $\beta_i(n) \in \Delta(S^{k-1}_{-i}(G') \cap S_{-i}(G', n))$. This, however, implies that $\beta_i(n) \in B^{k-1}_i(G', n)$, which was to show in (ii).

(iii) We next show that $f_i^k(s_i)$ is optimal at n for $\beta_i(n)$. Suppose not. Then, there is some $\sigma_i' \in S_i(G')$ such that

$$v_i(f_i^k(s_i), \beta_i(n)|n) < v_i(\sigma_i', \beta_i(n)|n),$$

where v_i, as announced above, denotes the utility function for player i in the game G'. As $\beta_i(n) \in B_i^{k-1}(G', n)$, it follows by Lemma 3 that there is some $\sigma_i \in S_i^{k-1}(G')$ such that

$$v_i(f_i^k(s_i), \beta_i(n)|n) < v_i(\sigma_i, \beta_i(n)|n). \tag{2}$$

By our induction assumption on (b) we know that the strategy $g_i^{k-1}(\sigma_i) \in S_i^{k-1}(G)$ is identical to σ_i on $N^{k-2}(G')$. By the induction assumption on (c), $N^{k-1}(G) \subseteq^* N^{k-2}(G')$, and hence $g_i^{k-1}(\sigma_i) \in S_i^{k-1}(G)$ is identical to σ_i on $N^{k-1}(G)$.

Recall that $b_i(n) \in \Delta(S_{-i}^k(G))$. As $g_i^{k-1}(\sigma_i) \in S_i^{k-1}(G)$, it follows that $(g_i^{k-1}(\sigma_i), b_i(n))$ only reaches nodes in $N^{k-1}(G)$. By this, remember, we mean that for every opponents' strategy profile $s_{-i} \in S_{-i}(G)$ with $b_i(n)(s_{-i}) > 0$, the strategy profile $(g_i^{k-1}(\sigma_i), s_{-i})$ only reaches nodes in $N^{k-1}(G)$. As $b_i(n) \in \Delta(S_{-i}^k(G))$, the belief $b_i(n)$ only assigns positive probability to $(s_j)_{j \neq i}$ where $s_j \in S_j^k(G)$. By the induction assumption on (a), we know that every such $s_j \in S_j^k(G)$ is identical to $f_j^{k-1}(s_j)$ on $N^{k-1}(G)$. So, by (1), $b_i(n)$ yields the same expected utility as $\beta_i(n)$ on $N^{k-1}(G)$. As $(g_i^{k-1}(\sigma_i), b_i(n))$ only reaches nodes in $N^{k-1}(G)$, and $g_i^{k-1}(\sigma_i)$ is identical to σ_i on $N^{k-1}(G)$, it follows that

$$v_i(\sigma_i, \beta_i(n)|n) = u_i(g_i^{k-1}(\sigma_i), b_i(n)|n). \tag{3}$$

Moreover, as $s_i \in S_i^{k+1}(G)$, and $b_i(n) \in \Delta(S_{-i}^k(G))$, we have that $(s_i, b_i(n))$ only reaches nodes in $N^k(G)$. We have seen above that $b_i(n)$ yields the same expected utility as $\beta_i(n)$ on $N^{k-1}(G)$, and hence, in particular, $b_i(n)$ yields the same expected utility as $\beta_i(n)$ on $N^k(G)$. As $(s_i, b_i(n))$ only reaches nodes in $N^k(G)$ and s_i is identical to $f_i^k(s_i)$ on $N^k(G)$, we may conclude that

$$v_i(f_i^k(s_i), \beta_i(n)|n) = u_i(s_i, b_i(n)|n). \tag{4}$$

By combining (2), (3) and (4), it follows that

$$u_i(s_i, b_i(n)|n) < u_i(g_i^{k-1}(\sigma_i), b_i(n)|n),$$

which contradicts our assumption that s_i is optimal at n for the belief $b_i(n)$.

Hence, we must conclude that $f_i^k(s_i)$ is optimal at n for $\beta_i(n)$. As $\beta_i(n) \in B_i^{k-1}(G', n)$, we see that $f_i^k(s_i)$ is optimal at n for some belief $\beta_i(n) \in B_i^{k-1}(G', n)$, which completes Case 1.

Case 2 Assume that $n \notin N^k$.

Then, all nodes following n are also not in N^k. Hence, by construction, $f_i^k(s_i)$ coincides with σ_i^* at n and all nodes $n' \in N_i(G')$ following n. Remember that σ_i^* was chosen to be in $S_i^k(G')$. Hence, there is some belief $\beta_i(n) \in B_i^{k-1}(G', n)$ such that σ_i^* is optimal at n for $\beta_i(n)$. But then, as $f_i^k(s_i)$ coincides with σ_i^* at n and all nodes $n' \in N_i(G')$ following n, we also have that $f_i^k(s_i)$ is optimal at n for $\beta_i(n)$. Hence, $f_i^k(s_i)$ is optimal at n for some $\beta_i(n) \in B_i^{k-1}(G', n)$, which completes Case 2.

By combining Case 1 and Case 2, we conclude that for every $n \in N_i(G')$ there is some belief $\beta_i(n) \in B_i^{k-1}(G', n)$ such that $f_i^k(s_i)$ is optimal at n for $\beta_i(n)$. Hence, $f_i^k(s_i)$ is optimal for some belief vector $\beta_i \in B_i^{k-1}(G')$, which means that $f_i^k(s_i) \in S_i^k(G')$. This completes the induction step for (a).

(b) The proof for part (b) is very similar to the proof for part (a). We will construct the "outcome preserving" mapping $g_i^k : S_i^k(G') \to S_i^k(G)$. Take some strategy $\sigma_i \in S_i^k(G')$. We define the strategy $g_i^k(\sigma_i)$ in G as follows. Choose some arbitrary strategy $s_i^* \in S_i^k(G)$. For every decision node $n \in N_i(G)$, define

$$(g_i^k(\sigma_i))(n) := \begin{cases} \sigma_i(n), & \text{if } n \in N^{k-1}(G') \\ s_i^*(n), & \text{if } n \notin N^{k-1}(G') \end{cases}.$$

In particular, at every last decision node $n \in N_i(G) \cap N^{last}(G)$ we have that $(g_i^k(\sigma_i))(n) = s_i^*(n)$ since $n \notin N(G')$. As $s_i^* \in S_i^k(G)$ and $k \geq 1$, we have that $s_i^*(n) = c^*(n)$ at all $n \in N_i(G) \cap N^{last}(G)$, where $c^*(n)$ is the unique optimal choice at n. Hence, $(g_i^k(\sigma_i))(n) = c^*(n)$ for all $n \in N_i(G) \cap N^{last}(G)$. But then, it immediately follows that $g_i^k(\sigma_i)$ is identical to σ_i on $N^{k-1}(G')$. Hence, it remains to show that $g_i^k(\sigma_i) \in S_i^k(G)$. So, at every $n \in N_i(G)$ we must find some belief $b_i(n) \in B_i^{k-1}(G, n)$ such that $g_i^k(\sigma_i)$ is optimal at n for the belief $b_i(n)$. Take some $n \in N_i(G)$. We distinguish two cases: (1) $n \in N^{k-1}(G')$, and (2) $n \notin N^{k-1}(G')$.

Case 1 Assume that $n \in N^{k-1}(G')$.

Since $\sigma_i \in S_i^k(G')$, strategy σ_i is optimal at n for some belief $\beta_i(n) \in B_i^{k-1}(G', n)$. As $n \in N^{k-1}(G')$, there is some $\sigma_{-i} \in S_{-i}^{k-1}(G')$ that reaches n. But then, by definition of $B_i^{k-1}(G')$, we must have that $\beta_i(n) \in \Delta(S_{-i}^{k-1}(G') \cap S_{-i}(G', n))$. We transform the belief $\beta_i(n)$ into a belief $b_i(n)$ in G as follows. Let $b_i(n) : S_{-i}(G) \to [0, 1]$ be given by

$$b_i(n)(s_{-i}) := \sum_{\sigma_{-i} \in S_{-i}^{k-1}(G') : g_{-i}^{k-1}(\sigma_{-i}) = s_{-i}} \beta_i(n)(\sigma_{-i}) \tag{5}$$

for all $s_{-i} \in S_{-i}(G)$. Here, for a given $\sigma_{-i} = (\sigma_j)_{j \neq i} \in S_{-i}^{k-1}(G')$, we define $g_{-i}^{k-1}(\sigma_{-i})$ as $(g_j^{k-1}(\sigma_j))_{j \neq i}$. Recall that $g_j^{k-1} : S_j^{k-1}(G') \to S_j^{k-1}(G)$, which implies that $g_{-i}^{k-1} : S_{-i}^{k-1}(G') \to S_{-i}^{k-1}(G)$. Hence, $g_{-i}^{k-1}(\sigma_{-i})$ is well-defined for every $\sigma_{-i} \in S_{-i}^{k-1}(G')$.

We will show (i) that $b_i(n)$ is a probability distribution on $S_{-i}(G)$, (ii) that $b_i(n) \in B_i^{k-1}(G, n)$, and (iii) that $g_i^k(\sigma_i)$ is optimal at n for $b_i(n)$.

(i) In exactly the same way as in (a) it can be shown that $\sum_{s_{-i} \in S_{-i}(G)} b_i(n)(s_{-i}) = 1$, and hence $b_i(n)$ is a probability distribution on $S_{-i}(G)$.

(ii) As $n \in N^{k-1}(G')$, we know by induction assumption on (c) that $n \in N^{k-1}(G)$. So, there is some $s_{-i} \in S_{-i}^{k-1}(G)$ that reaches n. Hence, to show that $b_i(n) \in B_i^{k-1}(G, n)$, we must prove that $b_i(n) \in \Delta(S_{-i}^{k-1}(G) \cap S_{-i}(G, n))$. To this purpose, we first show that $b_i(n) \in \Delta(S_{-i}^{k-1}(G))$, and then we prove that $b_i(n) \in \Delta(S_{-i}(G, n))$.

We first show that $b_i(n) \in \Delta(S_{-i}^{k-1}(G))$. Suppose that $b_i(n)(s_{-i}) > 0$ for some $s_{-i} \in S_{-i}(G)$. Then, by (5), $s_{-i} = g_{-i}^{k-1}(\sigma_{-i})$ for some $\sigma_{-i} \in S_{-i}^{k-1}(G')$ with $\beta_i(n)(\sigma_{-i}) > 0$. So, $\sigma_{-i} = (\sigma_j)_{j \neq i}$ with $\sigma_j \in S_j^{k-1}(G')$ for every $j \neq i$. By our induction assumption on (b), we know that $g_j^{k-1}(\sigma_j) \in S_j^{k-1}(G)$ for every $j \neq i$. Hence, $s_{-i} = g_{-i}^{k-1}(\sigma_{-i}) \in S_{-i}^{k-1}(G)$. So, $b_i(n)(s_{-i}) > 0$ only if $s_{-i} \in S_{-i}^{k-1}(G)$. We may thus conclude that $b_i(n) \in \Delta(S_{-i}^{k-1}(G))$.

We now show that $b_i(n) \in \Delta(S_{-i}(G, n))$. Suppose that $b_i(n)(s_{-i}) > 0$ for some $s_{-i} \in S_{-i}(G)$. Then, by (5), $s_{-i} = g_{-i}^{k-1}(\sigma_{-i})$ for some $\sigma_{-i} \in S_{-i}^{k-1}(G')$ with $\beta_i(n)(\sigma_{-i}) > 0$. As $\beta_i(n) \in \Delta(S_{-i}^{k-1}(G') \cap S_{-i}(G', n))$, it follows that $\sigma_{-i} \in S_{-i}(G', n)$. Hence, $\sigma_{-i} = (\sigma_j)_{j \neq i}$ with $\sigma_j \in S_j(G', n)$ for all $j \neq i$. That is, σ_j chooses at every player j node before n the choice leading to n. By our induction assumption on (b) we know that $g_j^{k-1}(\sigma_j)$ is identical to σ_j on $N^{k-2}(G')$. As $n \in N^{k-1}(G')$, all player j nodes before n are in $N^{k-2}(G')$, and hence also $g_j^{k-1}(\sigma_j)$ chooses at every player j node before n the choice leading to n. So, $g_j^{k-1}(\sigma_j) \in S_j(G, n)$. But then, $s_{-i} = g_{-i}^{k-1}(\sigma_{-i}) \in S_{-i}(G, n)$. So, we conclude that $b_i(n)(s_{-i}) > 0$ only if $s_{-i} \in S_{-i}(G, n)$, and hence $b_i(n) \in \Delta(S_{-i}(G, n))$.

In total, it follows that $b_i(n) \in \Delta(S_{-i}^{k-1}(G) \cap S_{-i}(G, n))$. This, however, implies that $b_i(n) \in B_i^{k-1}(G, n)$, which was to show in (ii).

(iii) We next show that $g_i^k(\sigma_i)$ is optimal at n for $b_i(n)$. Suppose not. Then, there is some $s_i' \in S_i(G)$ such that

$$u_i(g_i^k(\sigma_i), b_i(n)|n) < u_i(s_i', b_i(n)|n).$$

As $b_i(n) \in B_i^{k-1}(G, n))$, it follows by Lemma 3 that there is some $s_i \in S_i^{k-1}(G)$ such that

$$u_i(g_i^k(\sigma_i), b_i(n)|n) < u_i(s_i, b_i(n)|n). \tag{6}$$

By our induction assumption on (a) we know that the strategy $f_i^{k-2}(s_i) \in S_i^{k-2}(G')$ is identical to s_i on $N^{k-2}(G)$. By induction assumption on (c), $N^{k-2}(G') \subseteq N^{k-2}(G)$, and hence $f_i^{k-2}(s_i) \in S_i^{k-2}(G')$ is identical to s_i on $N^{k-2}(G')$.

Recall that $\beta_i(n) \in \Delta(S_{-i}^{k-1}(G'))$. As $f_i^{k-2}(s_i) \in S_i^{k-2}(G')$, it follows that $(f_i^{k-2}(s_i), \beta_i(n))$ only reaches nodes in $N^{k-2}(G')$. As $\beta_i(n) \in \Delta(S_{-i}^{k-1}(G'))$, the belief $\beta_i(n)$ only assigns positive probability to $(\sigma_j)_{j \neq i}$ where $\sigma_j \in S_j^{k-1}(G')$. By induction assumption on (b), we know that every such $\sigma_j \in S_j^{k-1}(G')$ is identical to

$g_j^{k-1}(\sigma_j)$ on $N^{k-2}(G')$. So, by (5), $\beta_i(n)$ yields the same expected utility as $b_i(n)$ on $N^{k-2}(G')$. As $(f_i^{k-2}(s_i), \beta_i(n))$ only reaches nodes in $N^{k-2}(G')$, and $f_i^{k-2}(s_i)$ is identical to s_i on $N^{k-2}(G')$, it follows that

$$u_i(s_i, b_i(n)|n) = v_i(f_i^{k-2}(s_i), \beta_i(n)|n). \tag{7}$$

Moreover, as $\sigma_i \in S_i^k(G')$, and $\beta_i(n) \in \Delta(S_{-i}^{k-1}(G'))$, we have that $(\sigma_i, \beta_i(n))$ only reaches nodes in $N^{k-1}(G')$. We have seen above that $\beta_i(n)$ yields the same expected utility as $b_i(n)$ on $N^{k-2}(G')$, and hence, in particular, on $N^{k-1}(G')$. As $(\sigma_i, \beta_i(n))$ only reaches nodes in $N^{k-1}(G')$ and σ_i is identical to $g_i^k(\sigma_i)$ on $N^{k-1}(G')$, we may conclude that

$$u_i(g_i^k(\sigma_i), b_i(n)|n) = v_i(\sigma_i, \beta_i(n)|n). \tag{8}$$

By combining (6), (7) and (8), it follows that

$$v_i(\sigma_i, \beta_i(n)|n) < v_i(f_i^{k-2}(s_i), \beta_i(n)|n),$$

which contradicts our assumption that σ_i is optimal at n for the belief $\beta_i(n)$.

Hence, we must conclude that $g_i^k(\sigma_i)$ is optimal at n for $b_i(n)$. As $b_i(n) \in B_i^{k-1}(G, n)$, we see that $g_i^k(\sigma_i)$ is optimal at n for some belief $b_i(n) \in B_i^{k-1}(G, n)$, which completes Case 1.

Case 2 Assume that $n \notin N^{k-1}(G')$.

Then, all nodes following n are also not in $N^{k-1}(G')$. Hence, by construction, $g_i^k(\sigma_i)$ coincides with s_i^* at n and all nodes $n' \in N_i(G)$ following n. Remember that s_i^* was chosen to be in $S_i^k(G)$. Hence, there is some belief $b_i(n) \in B_i^{k-1}(G, n)$ such that s_i^* is optimal at n for $b_i(n)$. But then, as $g_i^k(\sigma_i)$ coincides with s_i^* at n and all nodes $n' \in N_i(G)$ following n, we also have that $g_i^k(\sigma_i)$ is optimal at n for $b_i(n)$. Hence, $g_i^k(\sigma_i)$ is optimal at n for some $b_i(n) \in B_i^{k-1}(G, n)$, which completes Case 2.

By combining Case 1 and Case 2, we conclude that for every $n \in N_i(G)$ there is some belief $b_i(n) \in B_i^{k-1}(G, n)$ such that $g_i^k(\sigma_i)$ is optimal at n for $b_i(n)$. Hence, $g_i^k(\sigma_i)$ is optimal for some belief vector $b_i \in B_i^{k-1}(G)$, which means that $g_i^k(\sigma_i) \in S_i^k(G)$. This completes the induction step for (b).

(c) To show that $N^{k+1}(G) \subseteq^* N^k(G')$, take some $n \in N^{k+1}(G) \backslash N^{last}(G)$. Hence, there is some strategy profile $(s_i)_{i \in I}$ in G reaching n where $s_i \in S_i^{k+1}(G)$ for every i. By part (a), $f_i^k(s_i) \in S_i^k(G')$ and is identical to s_i on $N^k(G)$. As all nodes before n are in $N^k(G)$, it follows that the strategy profile $(f_i^k(s_i))_{i \in I}$ is identical to $(s_i)_{i \in I}$ at all nodes before n, and hence reaches n too. So we see that n is reached by the strategy profile $(f_i^k(s_i))_{i \in I}$, and that $f_i^k(s_i) \in S_i^k(G')$ for all i. But then, $n \in N^k(G')$. So, every $n \in N^{k+1}(G) \backslash N^{last}(G)$ is also in $N^k(G')$, which means that $N^{k+1}(G) \subseteq^* N^k(G')$.

To show that $N^k(G') \subseteq N^k(G)$, take some $n \in N^k(G')$. Hence, there is some strategy profile $(\sigma_i)_{i \in I}$ in G' reaching n where $\sigma_i \in S_i^k(G')$ for every i. By part (b), $g_i^k(\sigma_i) \in S_i^k(G)$ and is identical to σ_i on $N^{k-1}(G')$. As all nodes before n are in $N^{k-1}(G')$, it follows that the strategy profile $(g_i^k(\sigma_i))_{i \in I}$ is identical to $(\sigma_i)_{i \in I}$ at all nodes before n, and hence reaches n too. So we see that n is reached by the strategy profile $(g_i^k(\sigma_i))_{i\in I}$, and that $g_i^k(\sigma_i) \in S_i^k(G)$ for all i. But then, $n \in N^k(G)$. So, every $n \in N^k(G')$ is also in $N^k(G)$, which means that $N^k(G') \subseteq N^k(G)$. This completes the induction step for (c).

By induction on k, the statements (a), (b) and (c) follow for all k, which completes the proof. □

With Lemma 4 at hand, we can now easily show that truncating a game does not change the extensive-form rationalizable outcomes. To formalize this statement, let us denote by $Z^\infty(G)$ the set of extensive-form rationalizable outcomes of the game G.

Lemma 5 (Truncation does not change EFR outcomes) *Consider a finite extensive-form game G with perfect information and without relevant ties, and let $G' = tr(G)$ be the truncated game. Then, for every $z \in Z^\infty(G)$ there is some $z' \in Z^\infty(G')$ which yields the same utilities as z for all players, and vice versa.*

Proof of Lemma 5 Take some extensive-form rationalizable outcome $z \in Z^\infty(G)$ in the original game G. Then, there is some strategy profile $(s_i)_{i\in I} \in S^\infty(G)$ that reaches z. Clearly, there must be some K such that $S_i^K(G) = S_i^\infty(G)$ and $S_i^K(G') = S_i^\infty(G')$ for all players i. As $(s_i)_{i\in I} \in S^{K+1}(G)$, it follows by Lemma 4, part (a), that the strategy profile $(f_i^K(s_i))_{i\in I}$ is in $S^K(G') = S^\infty(G')$, and that it is identical to $(s_i)_{i\in I}$ on $N^K(G)$. As $(s_i)_{i\in I}$ only reaches nodes in $N^K(G)$, we conclude that $(s_i)_{i\in I}$ induces the same utilities for all players as $(f_i^K(s_i))_{i\in I}$. Let z' be the terminal node in G' reached by $(f_i^K(s_i))_{i\in I}$. Then, $z' \in Z^\infty(G')$, and z' induces the same utilities as z.

Take now some extensive-form rationalizable outcome $z' \in Z^\infty(G')$ in the truncated game G'. Then, there is some strategy profile $(\sigma_i)_{i\in I} \in S^\infty(G')$ that reaches z. As $(\sigma_i)_{i\in I} \in S^K(G')$, it follows by Lemma 4, part (b), that the strategy profile $(g_i^K(\sigma_i))_{i\in I}$ is in $S^K(G) = S^\infty(G)$, and that it is identical to $(\sigma_i)_{i\in I}$ on $N^{K-1}(G')$. As $(\sigma_i)_{i\in I}$ only reaches nodes in $N^{K-1}(G')$, we conclude that $(\sigma_i)_{i\in I}$ induces the same utilities for all players as $(g_i^K(\sigma_i))_{i\in I}$. Let z be the terminal node in G reached by $(g_i^K(\sigma_i))_{i\in I}$. Then, $z \in Z^\infty(G)$, and z induces the same utilities as z'. □

We are now in a position to prove Theorem 2.

Proof of Theorem 2 By repeatedly applying the truncation operator to G, we finally end up with a trivial game G^* that only has one terminal node z^*. Now, take some arbitrary extensive-form rationalizable outcome $z \in Z^\infty(G)$ in the original game G. By repeated application of Lemma 5, we conclude that z must yield the same utilities as z^*. By construction, the utilities at z^* are exactly the utilities of the unique BI outcome z' in G. So, the utilities at z must be the same as the utilities of the unique BI outcome z' in G. Since the game G is without relevant ties, z must be equal to the BI outcome in G. So, there is only one extensive-form rationalizable outcome in G, namely the BI outcome. □

5 Two auxiliary results

In this section we state, and prove, two auxiliary results that follow from our proof of Theorem 2. Consider some finite extensive-form game G with perfect information and without relevant ties. For every node n in G—which may be either a decision node or a terminal node—we define $rat(n, G)$ as the highest level k such that n is reached by some strategy profile in $S^k(G)$, and call it the rationality level of node n. More formally:

Definition 6 (*Rationality level of decision nodes*) For every node n in G, the rationality level of n is defined as

$$rat(n, G) = \begin{cases} k < \infty, & \text{if } n \text{ is reached by some } s \in S^k(G), \\ & \text{but not by some } s \in S^{k+1}(G), \\ \infty, & \text{if } n \text{ is reached by some } s \in S^{\infty}(G) \end{cases}.$$

From the truncation lemma above—Lemma 4— we can prove that after truncating the game, the rationality level of every node will either stay the same, or decrease by exactly one. Moreover, if the rationality level in the original game is ∞, then it will stay ∞ in the truncated game. We think this is an interesting property by itself, as it nicely shows how the rationality levels induced by EFR change if we truncate the game. Therefore, we state this property as a formal result here.

Corollary 7 (Truncation and rationality levels) *Let G be some finite extensive-form game with perfect information and without relevant ties. Let $G' := tr(G)$ be the truncated game. Then, for every node n in G' we have that*

$$rat(n, G') = \begin{cases} rat(n, G) \text{ or } rat(n, G) - 1, & \text{if } rat(n, G) < \infty \\ \infty, & \text{if } rat(n, G) = \infty \end{cases}.$$

Proof Take some node n in the truncated game G'. Then, n is a decision node in the original game G. We distinguish two cases, namely when $rat(n, G) < \infty$ and when $rat(n, G) = \infty$.

Suppose first that $rat(n, G) = k < \infty$. Then, by definition, $n \in N^k(G)$. By part (c) in Lemma 4 it follows that $n \in N^{k-1}(G')$, and hence n is reached by some $s \in S^{k-1}(G')$. So, $rat(n, G') \geq k - 1$. Suppose, contrary to what we want to show, that $rat(n, G') \geq k + 1$. Then, n would be reached by some $s \in S^{k+1}(G')$, and hence $n \in N^{k+1}(G')$. But then, by part (c) in Lemma 4 , it would follow that $n \in N^{k+1}(G)$, which would mean that $rat(n, G) \geq k + 1$. This, however, would contradict the assumption that $rat(n, G) = k$. So, we conclude that $rat(n, G') \leq k$. As we already saw that $rat(n, G') \geq k - 1$, it follows that $rat(n, G')$ is either k or $k - 1$. Hence, $rat(n, G')$ is either $rat(n, G)$ or $rat(n, G) - 1$.

Suppose next that $rat(n, G) = \infty$. Then, $n \in N^k(G)$ for all k. Hence, by part (c) in Lemma 4, $n \in N^{k-1}(G')$ for all k, which means that $rat(n, G') = \infty$. This completes the proof. □

We may use Corollary 7 to derive an interesting insight about the relation between BI and EFR. To formally state this insight, we must first define the depth of a decision node n. Within a fixed extensive-form game G with perfect information, the depth of a decision node n is the maximal number of choices between n and a terminal node. In other words, it is the length of a longest path starting at n. The following result states that at every decision node n with depth k, the player at n will choose the BI choice in any strategy profile reaching n that survives the first k steps of the EFR procedure.

At first sight, this result may suggest that the EFR-procedure also corresponds to some kind of BI procedure, in which we start at the last decision nodes (with depth 1), conclude that 1 round of EFR induces the BI choice there, then turn to decision nodes of depth 2, conclude that 2 rounds of EFR induce the BI choice there, and so on. This conclusion is false, however. Namely, for a particular decision node n with depth k, there may be no strategy profile reaching n that survives the first k rounds of the EFR-procedure. In that case, we cannot conclude that k rounds of the EFR-procedure induces the BI choice at n.

Corollary 8 (EFR and BI-choices) *Consider a finite extensive-form game G with perfect information and without relevant ties. Let n be a decision node with depth* $k \geq 1$*, and* $s \in S^k(G)$ *a strategy profile that reaches n. Then, s prescribes the backward induction choice at n.*

Proof Let $c(n)$ be the choice prescribed by s at n, and let n' be the node in G that immediately follows choice $c(n)$ at n. As $s \in S^k(G)$ reaches n, it also reaches n', and hence, by definition, $rat(n', G) \geq k$. Let $G^{k-1} := tr^{k-1}(G)$ be the $(k-1)$-fold truncation of G. Then, n' is a terminal node in G^{k-1}, and n is a last decision node in G^{k-1}, since n has depth k, and n' immediately follows n. By repeated application of Corollary 7, we obtain that

$$rat(n', G^{k-1}) \geq rat(n', G) - (k-1) \geq k - (k-1) = 1.$$

That is, within the game G^{k-1} the terminal node n' is reached by some $s' \in S^1(G^{k-1})$. But then, the choice $c(n)$ that precedes n' must be optimal at the last decision node n within G^{k-1}, since every $s' \in S^1(G^{k-1})$ must prescribe the unique optimal choice at every last decision node in G^{k-1}. By construction, the unique optimal choice in G^{k-1} at the last decision node n is precisely the BI choice within G at n. Hence, $c(n)$ must be the BI choice within G at n. So, the strategy profile $s \in S^k$ must prescribe the BI choice at n. This completes the proof. □

Acknowledgments We would like to thank an associate editor and two referees for very useful comments.

References

Arieli I, Aumann RJ (2012) The logic of backward induction, Working paper

Battigalli P (1996) Strategic rationality orderings and the best rationalization principle. Games Econ Behav 13:178–200

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

Chen J, Micali S (2013) The order independence of iterated dominance in extensive games. Theor Econ 8:125–163

Chen J, Micali S (2011) The robustness of extensive-form rationalizability, Working paper

Church A, Rosser JB (1936) Some properties of conversion. Trans Am Math Soc 39:472–482

Gretlein RJ (1983) Dominance elimination procedures on finite alternative games. Int J Game Theory 12:107–113

Kohlberg E, Mertens J-F (1986) On the strategic stability of equilibria. Econometrica 54:1003–1037
Marx L, Swinkels J (1997) Order independence for iterated weak dominance. Games Econ Behav 18:219–245
Marx L, Swinkels J (2000) Order independence for iterated weak dominance. Games Econ Behav 31:324–329
Pearce D (1984) Rationalizable strategic behavior and the problem of perfection. Econometrica 52:1029–1050
Perea A (2014) Belief in the opponents' future rationality. Games Econ Behav 83:231–254
Perea A (2012) Epistemic game theory: reasoning and choice. Cambridge University Press, Cambridge
Reny PJ (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
Shimoji M, Watson J (1998) Conditional dominance, rationalizability, and game forms. J Econ Theory 83:161–195