Journal of Mathematical Economics 71 (2017) 135–151

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Uncertain information structures and backward induction

Peio Zuazo-Garin

University of the Basque Country (UPV/EHU) and BRiDGE Group, Department of Foundations of Economic Analysis I, Avenida Lehendakari Agirre 83, 48015, Bilbao, Spain

ARTICLE INFO

Article history:
Received 19 June 2016
Received in revised form 23 March 2017
Accepted 24 May 2017

Keywords:
Extensive-form games
Perfect information
Incomplete information
Rationality
Backward induction
Forward induction

ABSTRACT

In everyday economic interactions, it is not clear whether each agent's sequential choices are visible to other participants or not: agents might be deluded about others' ability to acquire, interpret or keep track of data. Following this idea, this paper introduces uncertainty about players' ability to observe each others' past choices in extensive-form games. In this context, we show that monitoring opponents' choices does not affect the outcome of the interaction when every player expects their opponents indeed to be monitoring. Specifically, we prove that if players are rational and there is common strong belief in opponents being rational, having perfect information and believing in their own perfect information, then, the backward induction outcome is obtained regardless of which of her opponents' choices each player observes. The paper examines the constraints on the rationalization process under which reasoning according to Battigalli's (1996) best rationalization principle yields the same outcome irrespective of whether players observe their opponents' choices or not. To this respect we find that the obtention of the backward induction outcome crucially depends on tight higher-order restrictions on beliefs about opponents' perfect information. The analysis provides a new framework for the study of uncertainty about information structures and generalizes the work by Battigalli and Siniscalchi (2002) in this direction.

1. Introduction

1.1. Uncertainty about the information structure: an example

Assumptions regarding common knowledge of the information structure of an economic model can significantly impact predictions. Take for instance the sequential Battle of Sexes with perfect information represented in Fig. 1. Two players, Alexei Ivanovich (A) and Polina Alexandrovna (P) choose first and second respectively between actions *left* and *right*, and obtain utility depending on each history of actions according to the numbers depicted at the bottom of the tree in the picture. By information structure we refer to whether or not Polina observes Alexei's earlier choice before she chooses, which she does in this case of perfect information. The game is played just once, so punishment and reinforcement issues are assumed to be negligible. This description is common knowledge among the players, and we further assume that both of them are rational, and that Alexei believes Polina to be rational. It then seems reasonable to predict that players' choices will lead to the unique backward induction outcome: (2, 1); since Polina is rational and observes Alexei's choice, she will mimic it regardless of whether it is *left* or *right*. Alexei believes all the above, so since he himself is rational too, he will move *left*.

Now consider a commonly known imperfect information situation (Fig. 2): consider the alternative information structure according to which, when her turn arrives, Polina will not have observed Alexei's earlier move. Thus, Polina is uncertain of the outcome her choice will induce. Even if it is additionally assumed that Polina believes both that Alexei is rational and that Alexei believes she is rational, it is easy to see that the above argument justifying outcome (2, 1) finds no defense this time; and that indeed, depending on reciprocal beliefs concerning opponents' choices, every outcome is consistent with rationality and with any assumption about iterated mutual beliefs about rationality.

Consider finally an imperfect information case such as the one represented in Fig. 2, with the following variation: Alexei believes himself to be in a situation like the one in Fig. 1; and Polina believes that Alexei believes himself to be in that situation of perfect information. That is, the information structure of the game is not commonly known this time and, in fact, Alexei happens to be deluded about it. When it is her turn to choose, despite not having observed Alexei's earlier move, Polina can infer that since Alexei believes himself to be in a situation with perfect information, he also believes *left* to be followed by *left* and *right* by *right*, and will therefore choose *left*. Hence, despite not observing Alexei's earlier move, Polina believes that Alexei has chosen left and consequently she chooses *left*.

As the example above illustrates, assumptions regarding common knowledge of the information structure of an economic model

E-mail address: peio.zuazo@ehu.eus.

http://dx.doi.org/10.1016/j.jmateco.2017.05.004
0304-4068/

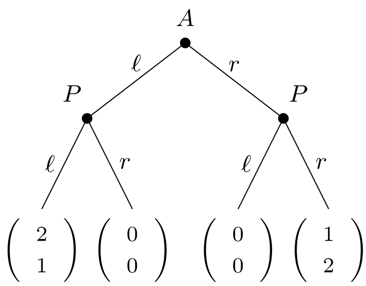

Fig. 1. A game with perfect information.

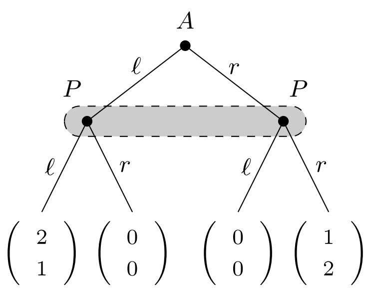

Fig. 2. A game w/o perfect information.

can significantly impact predictions. In order to determine strategic behavior it does not suffice to specify players' ability to observe each others' past choices: careful modeling of the beliefs that players hold about the information structure itself is required too. Consequently, establishing the distinction and exploring the differences in strategic implications between "perfect information" and "common knowledge of perfect information", which refer not only to the way that information flows but also to players' higher-order beliefs about that flow, becomes an interesting issue from a game theoretical perspective. In particular, as the comparison between the first and last situations in the example above suggests, this language enables the class of games for which the backward induction outcome can be considered as a reasonable prediction to be extended to the more general setting of contexts with not necessarily perfect information.

1.2. Information structures and the backward induction outcome: beyond common knowledge of perfect information

Literature on extensive-form games typically assumes that how information sets are distributed along the given game tree in such games is commonly known. This feature can be understood as the information structure of the game being part of the objective rules of the game. However, since information sets describe players' ability to observe, interpret and remember opponents' past behavior, they often depend more on players' personal cognitive abilities than on the rules of the game itself. Thus, since personal cognitive abilities are usually uncertain, it is natural to wonder how predictions in extensive-form games are affected by players facing incomplete information regarding the information structure.

The present paper takes its point of departure from the traditional approach of considering the information structure of an extensive-form game as commonly known, and determines the epistemic assumptions under which the backward induction outcome of the extensive-form is obtained under arbitrary information structures.[1] To that end, we introduce uncertainty about what we call the *information structure* of the extensive-form game. By information structure we refer to how each player's set of histories (i.e. the histories in which it is the player's turn to make a choice) is partitioned into information sets. The information structure can be regarded as players' ability to observe others' past choices, so the uncertainty that we introduce can be read as lack of certainty about whether or not each player is able to observe or remember her opponents' past choices (prior to her turn to choose). To perform our analysis, we first introduce a formal framework that enables incomplete information regarding the information structure of an extensive-form game to be accounted for. In particular, the fact that we allow for each player to face uncertainty about her own information structure means that the traditional notion of information set needs to be broadened to carefully capture the minimum information held by each player whenever it is her turn to make a choice. Next, we present an epistemic framework based on a special kind of conditional belief hierarchies *à la* (Battigalli and Siniscalchi, 1999) (the extensive-form version of Brandenburger and Dekel's (1993) construction of universal type space) that account for uncertainty about information structures. Following this approach, we prove in Theorem 1 that if: (i) players are rational and (ii) there is common strong belief in the event that opponents are rational, have perfect information and strongly believe in their own perfect information, then the backward induction outcome is obtained. Note that we do not assume perfect information: it could be the case that a player does not observe any of her opponents' past choices; still, our common strong belief assumptions enable her to infer what these choices were. Furthermore, we do not impose constraints on each player's beliefs about her own information structure: every player is assumed to believe that her opponents' have perfect information and strongly believe in their own perfect information, but may hold any arbitrary beliefs about her own ability to observe future choices. Still, the obtention of the backward induction outcome crucially hinges on tight assumptions on higher-order beliefs about opponents' perfect information; this illustrates how strong the assumptions on beliefs must be in order for uncertain information structures to not play a role.

The ability of agents involved in some interaction context to obverse each others' choice is often obvious. It might be obvious that there is perfect information: anti-theft devices in a store tell the owner whether a potential thief decided to steal or not. Alternatively, it might be obvious that there is no perfect information: a seller offers a product whose quality he can choose to a buyer; the latter may not necessarily appreciate the quality of the product prior to purchase. This distinction leads to the canonical classification of extensive-form games into those with perfect information and those with imperfect information, in which it is common knowledge that there is perfect and imperfect information, respectively. It turns out that this apparent dichotomy is non-exhaustive, and it is possible to think of situations in which the presence or absence of perfect information is not that obvious: in the first example above the anti-theft device could just be a cheap fake put there to fool potential thieves, while in the second, the buyer might be an expert on the product who is perfectly able to tell the quality of the option offered. Thus, the expected flow of information is sensitive to many aspects surrounding the context of interaction, and it is not clear why agents should not just agree, but *commonly* agree in their appreciation of these aspects and their influence. It is not the aim of this paper to propose a heuristic mechanism that endogenizes the rising of different beliefs about the information structure but rather to point out the possibility of that structure being uncertain, to highlight the relevance of such uncertainty, and to provide conditions in which the assumption of

[1] Thus, we follow the approach by Di Tillio et al. (2014), according to which certain characteristics typically involved in the description of a game, should be treated as individual epistemic attributes of players, and hence, be captured by the usual notion of type.

perfect information being commonly known can be dropped with no significant consequences in terms of behavior.

The epistemic assumptions leading to backward induction in extensive-form games with common knowledge of perfect information have been widely studied in recent years. Despite the apparent simplicity and intuitive appeal of backward induction, this way of reasoning seems unable to capture one crucial aspect of sequential playing: the ability to consistently update beliefs, and in particular, to question whether is plausible for a player who showed erratic behavior in the past actually to behave rationally in the future. Focusing on this aspect of backward induction reasoning, Reny (1992) presents an example of a finite extensive-form game whose unique extensive-form rationalizable (EFR, Pearce, 1984) profile does not coincide with its backward induction profile, and Ben Porath (1997) shows that in Rosenthal's centipede the backward induction outcome is not the only one consistent with initial common belief in rationality. Still, in Reny's example, the outcome induced by both the EFR profile and the backward induction profile is the same, and Battigalli (1997) generalizes this coincidence to the point of proving that for generic finite extensive-form games with perfect information EFR strategy profiles always lead to the unique backward induction outcome.[2]

A series of results follow the identity above: Battigalli and Siniscalchi (2002) introduce the notion of strong belief to formalize forward inducting according to Battigalli's (1996) best rationalization principle, and prove that rationality and common strong belief in rationality induce EFR strategy profiles when the epistemic type structure is complete (i.e., when it is able to represent any possible belief hierarchy that players might hold), and hence, lead to the backward induction outcome. Battigalli and Friedenberg (2012) introduce the notion of extensive-form best-reply sets (EFBRSs), and prove that rationality and common strong belief in rationality induce strategy profiles included in these sets regardless of whether the epistemic type structure is complete or not. However, they present examples where strategy profiles in EFBRSs do not induce the backward induction outcome, so sufficient epistemic conditions for the backward induction outcome for not necessarily complete epistemic type structures remain unclear. Penta (2011) and Perea (2014), exploit the notion of belief in opponents' future rationality and present sufficient conditions for extensive-form games with perfect information and arbitrary epistemic type structures by proving that rationality and common belief in opponents' future rationality induces the backward induction outcome.[3]

A different approach to epistemic analysis in games with perfect information is adopted by Aumann (1995, 1998), who makes use of static partition models which, unlike the models explained above, do not include explicit belief revision.[4] Aumann (1995) proves that ex ante common knowledge of rationality induces the backward induction profile. Samet (2013) modifies this result substituting common knowledge by common belief, and defining rationality in terms of beliefs rather than in terms of knowledge, as done by Aumann. Bonanno (2013) also proves that common belief in rationality induces the backward inductive outcome using belief frames that allow for belief revision and by assuming something analogous to belief in opponents' future rationality. In an earlier paper, Samet (1996) approaches the problem with very rich models that deal with knowledge rather than beliefs, but allow the modeling of hypothetical counterfactual information updates.[5] Literature allowing for incomplete information and exploring its link with rationalization in extensive-form games include Battigalli (2003), Battigalli and Siniscalchi (2007) and Penta (2011, 2012) among others. However, these papers only study incomplete information about players' preferences on the set of outcomes, not about the information structure of the game. Thus, issues regarding beliefs about the latter are not covered.

The rest of the paper is structured as follows: Sections 2 and 3 detail our formalization of extensive-form games and information structures, and the epistemic framework required for the analysis, respectively. Section 4 presents our main finding, which is summarized in Theorem 1, and discusses in detail the relation of the result with the different sources of uncertainty players face. We finish with some remarks and discussion in Section 5. All proofs are relegated to the appendices.

2. Games with uncertain information structure

We consider extensive-form games with incomplete information regarding the information structure of the game. To that end, we formalize two objects: (i) a game tree, similar to the extensive-form games with perfect information in Osborne and Rubinstein (1994, Sect. 6.1) and assumed to be commonly known, and (ii) the set of possible information structures on the given game tree, which is the part of the description of the game that we assume is possibly uncertain. Next, we detail the role of strategies in this context and the way in which they relate to uncertainty about the information structure and outcomes. We finish by adapting the well-known notions of sequential rationality and backward induction outcome to the present set-up and introducing an example to illustrate the definitions. So, we have:

2.1. Game trees

A (finite) *game tree* is an extensive-form game with perfect information as defined by Osborne and Rubinstein (1994). Formally, it consists of a tuple $\mathscr{T} = \langle I, (A_i)_{i\in I}, H, Z, (u_i)_{i\in I}\rangle$, where:

- I is a finite set of *players*.
- For each player i, A_i is a finite set of *actions*. The set of possible actions is denoted by $A = \bigcup_{i\in I} A_i$,[6] and we refer to a pair $(\emptyset, c)$, where c is a finite concatenation of actions, as a *history*. We say that history h *follows* history h' (and that h' *precedes* history h) if it is the case that $h = (h', c)$, where c is a finite concatenation of actions, in which case $h < h'$ is denoted.
- H and Z are finite and disjoint sets of histories. We assume that $(H \cup Z, \leq)$, where $\leq$ is defined in the obvious way, is an oriented tree with root $\emptyset$ and terminal nodes Z. Histories in H and Z are called *partial* and *terminal*, respectively. We also assume that exactly one player chooses at each partial history, that a player never chooses twice in a row and that whenever a player has to choose, at least two actions are

2 While Battigalli's original proof relies on rather intricate mathematics, Heifetz and Perea (2014) present a more intuitive proof that clarifies the logic relating to both outcomes.

3 The papers by Penta (2011) and Perea (2014) are more general indeed: Perea (resp. Penta) proves that in generic extensive-form games with not necessarily perfect information, rationality and common belief in opponents' future rationality (resp. and common belief in opponents' future rationality and in Bayesian updating) induce what he defines as strategy profiles surviving the backward dominance procedure (resp. the backwards rationalizability procedure), which in games with perfect information coincide exactly with the backwards inductive profile. Moreover, Penta, who in addition allows for incomplete information about the payoffs of the game, proves that in his characterization result, the assumptions above can be substituted by common certainty of full rationality and belief persistence. These ideas are applied to robust dynamic implementation in Penta (2015). A non-probabilistic version of belief in opponents' future rationality can be found in Baltag et al. (2009).

4 But, as shown by Perea (2007), do entail important *implicit* assumptions regarding belief revision, namely no revision of beliefs about future behavior.

5 For further references on epistemic game theory focused on extensive-form games, see Perea (2007) and Section 7 in Dekel and Siniscalchi (2015).

6 Not to be confused with the set of action *profiles*.

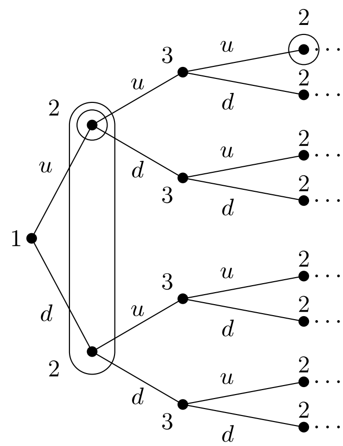

Fig. 3. Different information sets of player 2 precede $\{uuu\}$: $\{u\}$ and $\{u, d\}$.

available to her. Let $A_i(h)$ denote the set of actions available to player i at h,[7] and H_i, the set of partial histories in which it is player i's time to make a choice.[8]

- For each player i, $u_i : Z \to \mathbb{R}$ is player i's *payoff function*. Following Battigalli (1997), we assume that the game has no relevant ties, i.e., that for any player i and any history $h \in H_i$, function u_i is injective when restricted to the set of terminal histories that follow h. Note that this condition holds generically.

2.2. Information structures and information sequences

Following standard terminology, for each player i and subset of histories $v_i \subseteq H_i$, we say that v_i is an *information set* if none of its elements follow one another, and exactly the same actions are available at all of them.[9] Let V_i be a partition of H_i; we say that V_i is an *information partition* for player i if its cells are information sets and it satisfies perfect-recall.[10] Note that we can then denote $A_i(v_i)$ as the actions available at information set v_i with no ambiguity. An *information structure* is then a profile $V = (V_i)_{i \in I}$ of information partitions. For each player i, we denote by $\mathcal{V}_i$ the set of player i's information partitions. Let $\mathcal{V}$ denote the set of information structures.

Notice now that if player i reaches information set v_i then: (S1) she knows that she is at some history in v_i, and (S2) she knows which specific information sets (if any) $(v_i^k)_{k=1}^n$ she previously went through before reaching v_i. In the extensive-form game where information structure V is commonly known statement (S1) implies statement (S2): since i knows that she is at v_i and she also knows that her information structure is V_i, the fact that (by definition of information structures) there exists a unique sequence $(v_i^k)_{k=1}^n \subseteq V_i$ that ends in v_i implies that player i knows that she went through sequence of information sets $(v_i^k)_{k=1}^n$. Obviously, statement (S2) implies statement (S1) as well. Thus, in games with commonly known information structure, statement (S1) and statement (S2) are informationally equivalent.

But remember that we allow for uncertainty about the information structure of the game. In particular, it is possible for a player to face uncertainty about her own information structure.[11] Notice that in this environment statement (S1) does not necessarily imply statement (S2) anymore. The game tree in Fig. 3 illustrates this point. If it is commonly known that the game has perfect information, then, when player 2 reaches information set $\{uuu\}$ we know that the sequence of information sets that she went through is exactly $(\{u\}, \{uuu\})$. In contrast, if we allow for uncertainty about the information structure, when we say that player 2 reached information set $\{uuu\}$ we do not know whether she went through sequence $\sigma_2 = (\{u\}, \{uuu\})$ or sequence $\sigma_2' = (\{u, d\}, \{uuu\})$. Hence, in games with uncertain information structure statement (S2) is more informative than statement (S1).

Thus, in order to study games with uncertain information structure we require an alternative notion that plays the same role in terms of representing information as information sets do in the context of games with commonly known information structure. To define this alternative notion properly, let us introduce some notation and terminology first. For each player i, information partition $V_i \in \mathcal{V}_i$ and information sets $v_i, v_i' \in V_i$, we write $v_i < v_i'$ if there exist some histories $h \in v_i$ and $h' \in v_i'$ satisfying that h' follows h. In such case we say that v_i' *follows* v_i. We say that information set v_i is *minimal* at information partition V_i when $v_i \in V_i$ and v_i follows no $v_i' \in V_i$. Then, we expand the notion of information set in the following way:

Definition 1 (*Information Sequence*)**.** Let $\mathscr{T}$ be a game tree. An information sequence for player i is a concatenation of consecutive information sets of some information partition V_i whose first element is minimal at V_i; i.e., a sequence $\left(v_i^n\right)_{n \le N} \subseteq V_i$, with $V_i \in \mathcal{V}_i$, such that:

(i) $v_i^n < v_i^{n+1}$ for any $n = 1, \dots, N-1$, and there is no $v_i \in V_i$ such that $v_i^n < v_i < v_i^{n+1}$.
(ii) v_i^1 is minimal at V_i.

Let Σ_i denote player i's set of information sequences, and for each $\sigma_i = \left(v_i^n\right)_{n \le N} \in \Sigma_i$, let $v_{\sigma_i} = v_i^N$.

2.3. Strategies and terminal histories

We refer to a pair $\mathscr{G} = \langle \mathscr{T}, \mathcal{V} \rangle$ as a *game with uncertain information structure*.[12] In this context, a strategy is not a description of what action to choose in each history (resp. information set), as in the standard cases of commonly known perfect information (resp. imperfect information), but rather, of what action to choose after any possible information sequence. That is, for each player i, a strategy is a list $s_i \in S_i = \prod_{\sigma_i \in \Sigma_i} A_i\left(v_{\sigma_i}\right)$. We write $S_{-i} = \prod_{j \ne i} S_j$ to represent the set of player i's opponents' strategies. Note that a strategy profile itself does not induce any terminal history; to do

7 Thus, formally, $A_i(h) = \left\{a_i \in A_i \middle| (h, a_i) \in H \cup Z\right\}$. Using this notation, we are requiring that the tree satisfies: (i) $\emptyset \in H$, (ii) for any $(h, a) \in H \cup Z$, $h \in H$, (iii) for any $h \in Z$, $(h, a) \notin Z$ for any $a \in A$, (iv) for any $h \in H$, any $i \in I$ and any $a_i \in A_i$ such that $(h, a_i) \in H \cup Z$, it holds that if $(h, a) \in H \cup Z$ for some $a \in A$, then $a \in A_i$, (v) for any $h \in H$, any $i \in I$, any $a_i \in A_i$ and any $a \in A$ such that $(h, a_i, a) \in H$, $a \notin A_i$, and (vi) for any $h \in H$, if $A_i(h) \ne \emptyset$, then $|A_i(h)| \ge 2$.

8 That is, $H_i = \left\{h \in H \middle| A_i(h) \ne \emptyset\right\}$.

9 That is, for any $h, h' \in v_i$, $h \not< h$, and $A_i(h) = A_i\left(h'\right)$.

10 Denote by $V_i(h)$ the cell in V_i containing history h. Then, perfect recall is satisfied when: (i) for any $h, h' \in H_i$ such that $h' \notin V_i(h)$ and $(h, a_i) < h'$ for some $a_i \in A_i$, for any $h''' \in V_i\left(h'\right)$ there is some $h'' \in V_i(h)$ such that $\left(h'', a_i\right) < h'''$, and (ii) for any $h, h', h'', h''' \in H_i$ such that $h < h''$, $h' < h'''$ and $V_i(h) \ne V_i\left(h'\right)$, $V_i\left(h''\right) \ne V_i\left(h'''\right)$.

11 Uncertainty about own ability to observe opponents' actions is a widespread phenomenon. Consider a buyer who prior to purchase may not be able to appreciate the quality of the product offered by a seller who is able to chose the quality, or, alternatively, a firm which has to make an investment decision whose result depends on the decision of a competing firm, and for which it is not clear whether information regarding the competitor's investment decision will be available on time. Paragraph C in Section 5 addresses why such kind of uncertainty is not at odds with positive introspection and briefly explains how to construct a simpler model where uncertainty about own information structure is precluded.

12 There is a slight redundancy in listing both $\mathscr{T}$ and $\mathcal{V}$: as shown above, the set of all possible information structures ($\mathcal{V}$) is already determined by the game tree ($\mathscr{T}$). However, we keep this notation in order to emphasize the absence of common knowledge assumptions regarding the information structure.

that it is necessary to additionally specify an information structure: for any pair $(s, V) \in S \times \mathcal{V}$ there exists a unique $z\left(s, V \middle| h\right) \in Z$ such that for any player i,

$$z\left(s, V \middle| h\right) \geq \left(h', s_i(\sigma_i)\right) \text{ for any } h' \in H_i \text{ s.t. } h \leq h' \text{ and } h' < z\left(s, V \middle| h\right),$$

where σ_i is player i's unique information sequence satisfying $\sigma_i \subseteq V_i$ and $v_{\sigma_i} = V_i(h')$. Thus, each player's conditional payoffs are naturally determined by conditional terminal histories as follows: $u_i\left(s, V \middle| h\right) = u_i\left(z\left(s, V \middle| h\right)\right)$.

Note that any combination of a strategy profile and an information structure precludes certain information sets being reached, so it is useful to write the following: take player i and information sequence σ_i; then, we denote: (i) by $(S_{-i} \times \mathcal{V})(\sigma_i)$, the set of i's opponents' strategies and information structures that allow for σ_i being reached, (ii) for any $h \in v_{\sigma_i}$, by $(S_{-i} \times \mathcal{V})(\sigma_i, h)$, the set of i's opponents' strategies and information structures that allow for h being reached while σ_i being i's information sequence, and (iii) by $S_i(\sigma_i)$, the set of i's strategies that allow for σ_i being reached.[13] Finally, for any strategy s_i, let $\Sigma_i(s_i) = \left\{\sigma_i \in \Sigma_i \middle| s_i \in S_i(\sigma_i)\right\}$ represent the set of player i's information sequences whose terminal information set might be reached when she plays s_i.

2.4. Conjectures, sequential rationality and backward induction

Throughout the paper, and following the recovery of Renyi's (1955) original notion due to Myerson (1986) and Ben Porath (1997), conditional probability systems will serve as the building block for modeling players' interactive beliefs. First, we recall its definition, and then, apply them to adapt the notion of sequential rationality to games with uncertain information structure. Finally we recall the formalization of the backward induction outcome of a game tree.

2.4.1. Conditional probability systems

A *conditional base space* for player i is defined as a pair $(\mathcal{U}_i, \mathcal{C}_i)$ consisting of: (i) a compact and metrizable *basic uncertainty space* $\mathcal{U}_i$ whose Borel subsets we refer to as *events*, and (ii) a countable family of *conditioning events* $\mathcal{C}_i \subseteq \mathscr{B}(\mathcal{U}_i) \setminus \{\emptyset\}$,[14] whose elements are all both open and closed in $\mathcal{U}_i$. Then, a *conditional probability system* on $(\mathcal{U}_i, \mathcal{C}_i)$ is a map $\mu_i : \mathscr{B}(\mathcal{U}_i) \times \mathcal{C}_i \to [0, 1]$ that satisfies,

(i) $\mu_i\left[\cdot \middle| C\right] \in \Delta(\mathcal{U}_i)$ for any conditioning event C.

(ii) $\mu_i\left[C \middle| C\right] = 1$ for any conditioning event C.

(iii) $\mu_i\left[E \middle| C\right] \cdot \mu_i\left[C \middle| C'\right] = \mu_i\left[E \middle| C'\right]$ for any event E and any two conditioning events C and C' satisfying $E \subseteq C \subseteq C'$.

We denote the set of all possible conditional probability systems on $(\mathcal{U}_i, \mathcal{C}_i)$ by $\Delta^{\mathcal{C}_i}(\mathcal{U}_i)$, which is both compact and metrizable under the topology inherited by endowing $[\Delta(\mathcal{U}_i)]^{\mathcal{C}_i}$ with the product topology (see Lemma 2.1 by Battigalli and Siniscalchi, 1999). For notational convenience the following slight abuse is used throughout the paper: for any basic uncertainty product space $\mathcal{U}_i = X \times Y$, if the family of conditioning events $\mathcal{C}_i$ can be described by indexing its elements using Y,[15] then we denote $\Delta^Y(\mathcal{U}_i) = \Delta^{\mathcal{C}_i}(\mathcal{U}_i)$, and for any conditional probability system μ_i, we write $\mu_i(y)[\,\cdot\,] = \mu_i\left[\cdot \middle| X_y \times \{y\}\right]$ for any index y.

2.4.2. Conjectures and sequential rationality

Conditional probability systems are applied to model the dynamic fact that a player can update beliefs about opponents' choices as the game unfolds. In this sense, a (*correlated*) *conjecture* for player i is a conditional probability system μ_i on conditional base space composed by basic uncertainty space $\mathcal{U}_i = S_{-i} \times \mathcal{V}$ and family of conditioning events $\mathcal{C}_i = \left\{(S_{-i} \times \mathcal{V})(\sigma_i) \middle| \sigma_i \in \Sigma_i^{\emptyset}\right\}$, where $\Sigma_i^{\emptyset} = \{\emptyset\} \cup \Sigma_i$. Conjectures naturally induce conditional expected payoffs: player i's *conditional expected payoff* under conjecture μ_i after information sequence σ_i is given by,

$$U_i\left(\mu_i, s_i \middle| \sigma_i\right) = \sum_{h \in v_{\sigma_i}} \sum_{(s_{-i}, V) \in (S_{-i} \times \mathcal{V})(\sigma_i, h)} \mu_i(\sigma_i)[(s_{-i}, V)] \cdot u_i\left((s_{-i}; s_i), V \middle| h\right),$$

for any $s_i \in S_i$. The notion of sequential rationality is then captured by defining, for each player i, her (*sequential*) *best-reply correspondence* as $\mu_i \mapsto BR_i(\mu_i)$, where,

$$BR_i(\mu_i) = \left\{ s_i \in S_i \middle| s_i \in \bigcap_{\sigma_i \in \Sigma_i(s_i)} \arg\max_{s_i' \in S_i(\sigma_i)} U_i\left(\mu_i, s_i' \middle| \sigma_i\right) \right\},$$

for any conjecture μ_i.[16] Note that each $BR_i(\mu_i)$ is guaranteed to be non-empty due to the game tree being finite and conditional probability systems updating beliefs according to the chain rule. Since S_i is finite, it is immediate to check that BR_i is, in addition, upper-hemicontinuous and closed-valued.

2.4.3. The backward induction outcome

With some abuse of language, we identify the backward induction outcome of the game with uncertain information structure $\mathscr{G}$ with the backward induction outcome of the perfect information game corresponding to its game tree $\mathscr{T}$. For such a game, a strategy profile in terms of histories $((s_i(h))_{h \in H_i})_{i \in I}$ (which is by no means a strategy for game with uncertain information structure $\mathscr{G}$) is called *inductive* if it satisfies that for any player i and any history $h \in H_i$,

$$s_i(h) \in \arg\max_{a_i \in A_i(h)} u_i\left(z\left(s_{-i}; (s_i, a_i) \middle| h\right)\right),$$

where $(s_i, a_i)(h') = a_i$ if $h' = h$ and $(s_i, a_i)(h') = s_i(h')$ otherwise. Clearly, one can interpret an inductive strategy as a backward inductive strategy corresponding to the standard game with (commonly known) perfect information whose game tree is $\mathscr{T}$. For a tree with no relevant ties this profile, which we denote by α, is unique. For each history $h \in H$, let $z_{\mathcal{I}}(h) = z\left(\alpha \middle| h\right)$ denote the inductive outcome of $\mathscr{T}$ conditional on h, and $z_{\mathcal{I}} = z_{\mathcal{I}}(\emptyset)$, the inductive outcome; note that despite α not being a meaningful object in $\mathscr{G}$, each $z_{\mathcal{I}}(h)$ obviously is. For each history h we refer to α_h as the *inductive choice* at history h.

2.5. Stan's Used Car Emporium: Dramatis personae

Let us illustrate the definitions above with an example. Consider the game tree depicted in Fig. 4. Elaine (E) is considering whether

---

[13] Formally, we have these three characterizations:

$$(S_{-i} \times \mathcal{V})(\sigma_i) = \left\{(s_{-i}, V) \in S_{-i} \times \mathcal{V} \middle| \sigma_i \subseteq V_i \text{ and } z((s_{-i}; s_i), V) > h \text{ for some } h \in v_{\sigma_i} \text{ and some } s_i \in S_i\right\},$$

$$(S_{-i} \times \mathcal{V})(\sigma_i, h) = \left\{(s_{-i}, V) \in (S_{-i} \times \mathcal{V})(\sigma_i) \middle| z((s_{-i}, s_i), V) > h \text{ for some } s_i \in S_i\right\},$$

$$S_i(\sigma_i) = \left\{s_i \in S_i \middle| h < z((s_{-i}; s_i), V) \text{ for some } h \in v_{\sigma_i} \text{ and some } (s_{-i}, V) \in (S_{-i} \times \mathcal{V})(\sigma_i)\right\}.$$

[14] For any topological space X let $\mathscr{B}(X)$ and $\Delta(X)$ respectively denote the corresponding Borel σ-algebra and the set of probability measures on $\mathscr{B}(X)$. We endow $\Delta(X)$ with the weak∗ topology, so that if X is compact and metrizable, so is $\Delta(X)$; in particular, every continuous function under this topology will be measurable under $\mathscr{B}(X)$.

[15] i.e., if $\mathcal{C}_i = \left\{X_y \times \{y\} \middle| y \in Y \text{ and } X_y \subseteq X\right\}$.

[16] The fact that players update beliefs according to the chain rule ensures that no dynamic inconsistency issues are present and that the one-shot deviation principle is satisfied.

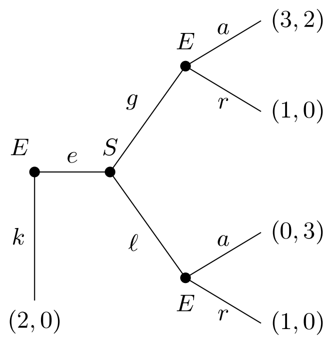

Fig. 4. Stan's Used Car Emporium.

to keep her old car (action k, in the picture) or replace it with a new one. The only place that offers her the latter possibility in her small local area is Stan's Used Car Emporium, where once she gets there (action e), Stan (S) will offer her a car that she can buy for some fixed price. The car offered may be good or a lemon (actions g and ℓ by Stan, respectively), and after a brief inspection, Elaine may decide to accept the offer and pay the price or reject it and leave (actions a and r, respectively). She would be happy to pay for a good car, but prefers her old one to the lemon, and, in any case, Elaine prefers to stay home rather than walk into Stan's and leave empty-handed. Stan makes a profit from either sale, but this profit is obviously higher if the car that Elaine buys is the lemon.

We assume that the reader is familiar with the notions of game tree and CPS, so we only detail the concepts this paper introduces. In regard to information partitions, we have a unique possible one for Stan, $\mathcal{V}_S = \{\{e\}\}$, but two alternatives for Elaine: (i) one in which she is proficient enough in auto-mechanics as to be able to tell a good car from a lemon apart (i.e., in which she has perfect information), $V_E^{PI} = \{\{\emptyset\}, \{g\}, \{\ell\}\}$, and (ii) one in which she knows little about cars and cannot distinguish between the good car and the lemon (i.e., in which she has imperfect information), $V_E^{II} = \{\{\emptyset\}, \{g, \ell\}\}$. Thus, $\mathcal{V}_E = \left\{V_E^{PI}, V_E^{II}\right\}$. Stan only chooses once, so his unique information set can be identified with his unique possible information sequence. On the contrary, we have the following set of information sequences for Elaine: $\Sigma_E = \{\{\emptyset\}, (\{\emptyset\}, \{g\}), (\{\emptyset\}, \{\ell\}), (\{\emptyset\}, \{g, \ell\})\}$. Each of her strategies assign a certain action to each of her information sequences: either k or e to $\{\emptyset\}$ and either a or r to $(\{\emptyset\}, \{g\})$, $(\{\emptyset\}, \{l\})$ and $(\{\emptyset\}, \{g, \ell\})$ (of course, not necessarily the same action to the three of them), for instance, (e, a, r, a). The backward induction outcome of the tree is $z_{\mathcal{I}} = (\emptyset, e, g, a)$. We continue with the example in Section 4.2.

3. Higher-order beliefs, rationality and perfect information

Conjectures as defined in Section 2.4.2 model players' dynamic beliefs about opponents' choices and the information structure. Still, they are not suitable for capturing one essential aspect of strategic reasoning: beliefs about each opponents' beliefs about her opponents' choices, her opponents' beliefs, and so on. That is, higher-order beliefs, or belief hierarchies. In order to properly formalize these ideas we first construct an epistemic framework reminiscent of the one employed by Tan and Werlang (1988), and built upon the original notion of conditional belief hierarchy due to Battigalli and Siniscalchi (1999). Then, under this epistemic framework, we define in Section 3.2 the main epistemic assumption, based on Battigalli and Siniscalchi's (2002) notion of strong belief, which together with rationality leads to our result presented in Section 4.

3.1. Epistemic framework

3.1.1. Epistemic types: canonical construction

We formally represent each player i's higher-order conditional beliefs about opponents' choices and the information structure via epistemic types, or conditional belief hierarchies. Let us recall the construction due to Battigalli and Siniscalchi (1999): consider for each player i, conditional base space $(\mathcal{U}_i, \mathcal{C}_i)$ composed of basic uncertainty space $\mathcal{U}_i = S_{-i} \times \mathcal{V}$ and family of conditioning events $\mathcal{C}_i = \left\{(S_{-i} \times \mathcal{V})(\sigma_i) \,\middle|\, \sigma_i \in \Sigma_i^{\emptyset}\right\}$. Then, set first $X_{i,1} = \mathcal{U}_i$ and $\mathcal{E}_{i,1}^0 = \Delta^{\mathcal{C}_i}(X_{i,1})$, and then, define recursively,

$$X_{i,n+1} = X_{i,n} \times \prod_{j \neq i} \mathcal{E}_{j,n}^0 \text{ and } \mathcal{E}_{i,n+1}^0 = \Delta^{\mathcal{C}_i}\left(X_{i,n+1}\right).$$

Note that conditioning event $C \in \mathcal{C}_i$ can be identified with some information sequence $\sigma_i \in \Sigma_i^{\emptyset}$; thus, in what follows we use the latter to index the former. Now, let $\mathcal{E}_i^0 = \prod_{n \in \mathbb{N}} \mathcal{E}_{i,n}^0$; we refer to each $e_i \in \mathcal{E}_i^0$ as *epistemic type* and we say that e_i is *coherent* if different order hierarchies do not contradict each other, i.e., if $\text{marg}_{X_i^n} e_{i,n+1}(\sigma_i) = e_{i,n}(\sigma_i)$ for any information sequence $\sigma_i \in \Sigma_i^{\emptyset}$ and any natural n. Let $\mathcal{E}_i^1$ denote player i's set of coherent epistemic types, and let $\mathcal{E}_{-i}^0 = \prod_{j \neq i} \mathcal{E}_j^0$. Then define recursively[17]:

$$\mathcal{E}_i^{n+1} = \left\{ e_i \in \mathcal{E}_i^n \middle| e_{i,m}(\sigma_i)\left[\text{Proj}_{X_{i,m}}\left(\mathcal{E}_{-i}^n \times S_{-i} \times \mathcal{V}\right)\right] = 1 \text{ for any } m \in \mathbb{N} \text{ and any } \sigma_i \in \Sigma_i^{\emptyset} \right\},$$

where $\mathcal{E}_{-i}^n = \prod_{j \neq i} \mathcal{E}_j^n$ for any natural n.[18] Each $\mathcal{E}_i^{n+1}$ is capturing the idea that in player i's mind, players hold iterated beliefs up to order n of every player having coherent beliefs. We say that each epistemic type in $\mathcal{E}_i = \bigcap_{n \in \mathbb{N}} \mathcal{E}_i^n$ represents *common certainty in coherence*. It is known from Propositions 2.3 and 2.5 by Battigalli and Siniscalchi (1999) how epistemic types and beliefs about the basic uncertainty space and on opponents' epistemic types relate to each other: there exists a homeomorphism $\psi_i : \mathcal{E}_i^1 \to \Delta^{\mathcal{C}_i^*}\left(\mathcal{E}_{-i}^0 \times S_{-i} \times \mathcal{V}\right)$, where $\mathcal{C}_i^* = \left\{\mathcal{E}_{-i}^0 \times C \middle| C \in \mathcal{C}_i\right\}$, that satisfies both:

(i) $\text{marg}_{X_{i,n}} \psi_i(e_i)(\sigma_i) = e_{i,n}(\sigma_i)$ for any natural n, any information sequence σ_i and any coherent epistemic type e_i.
(ii) Restriction $\psi_{i|_{\mathcal{E}_i}}$ is a homeomorphism between $\mathcal{E}_i$ and $\Delta^{\Sigma_i^{\emptyset}}\left(\mathcal{E}_{-i} \times S_{-i} \times \mathcal{V}\right)$.

3.1.2. States of the world and events

The epistemic analysis is performed in set of *states* (*of the world*) $\Omega = \mathcal{E} \times S \times \mathcal{V}$. For each player i we denote $\Omega_i = \mathcal{E}_i \times S_i$, and for each state ω, we consider the following projections: $\omega_i = \text{Proj}_{\Omega_i}(\omega)$, $e_i(\omega) = \text{Proj}_{\mathcal{E}_i}(\omega)$, $s_i(\omega) = \text{Proj}_{S_i}(\omega)$ and $v(\omega) = \text{Proj}_{\mathcal{V}}(\omega)$. Thus, each state is a description of the information structure, players' choices, and players' belief hierarchies on the previous two contingencies. An *event* is a Borel subset $E \subseteq \Omega$. Note that some events and information sequences are mutually exclusive: for each player i and information sequence σ_i, $E \cap (\mathcal{E} \times (S \times \mathcal{V})(\sigma_i)) = \emptyset$ would imply that player i can only assign null probability to event E at σ_i, regardless of her epistemic type. Thus, for player i and event E, we define player i's set of information sequences *belief-consistent* with E as $\Sigma_i^{\emptyset}(E) = \left\{\sigma_i \in \Sigma_i^{\emptyset} \middle| E \cap (\mathcal{E} \times (S \times \mathcal{V})(\sigma_i)) \neq \emptyset\right\}$. This set represents i's information sequences in which i might assign positive probability to E.

[17] For any subset of a product set $S \subseteq X \times Y$, we denote by $\text{Proj}_Y(S)$ the *projection* of S on Y, i.e., $\text{Proj}_Y(S) = \left\{y \in Y \middle| (x, y) \in S \text{ for some } x \in X\right\}$.
[18] The fact that taking marginals is continuous guarantees that $\mathcal{E}_i^1$ is compact, and from this starting point, it is trivial to check that each $\mathcal{E}_i^n$ is well-defined and compact.

3.2. Rationality, perfect information and forward induction

The main result in Section 4 is based on one behavioral assumption, rationality, and one epistemic assumption, common strong belief that opponents are rational, have perfect information and strongly believe in their own perfect information. The epistemic condition is defined in terms of conditional beliefs, which are defined as follows: player i's *conditional belief operator* after information sequence $\sigma_i \in \Sigma_i^{\emptyset}$ is given by $E \mapsto B_i\left(E\middle|\sigma_i\right)$, where,

$$B_i\left(E\middle|\sigma_i\right) = \left\{\omega \in \Omega\middle|\psi_i\left(e_i\left(\omega\right)\right)\left(\sigma_i\right)\left[\text{Proj}_{\Omega_{-i}\times\mathcal{V}}(E)\right] = 1\right\},$$

for any event $E \subseteq \Omega$.[19] Thus, $B_i\left(E\middle|\sigma_i\right)$ should be interpreted as the event that player i assigns probability 1 to E after σ_i.

3.2.1. Rationality and perfect information

We say that player i is rational when her choices are conditionally optimal given her conditional beliefs. Thus, since states describe both beliefs and strategies they must describe whether players are rational or not. Then, we can formalize the event that player i is *rational* as,[20]

$$R_i = \left\{\omega \in \Omega\middle|s_i(\omega) \in BR_i\left(e_{i,1}(\omega)\right)\right\},$$

and, as usual, we denote $R_{-i} = \bigcap_{j\neq i} R_j$ and $R = \bigcap_{i\in I} R_i$. Similarly, since states describe information structures, they must describe whether players have (not necessarily commonly known) perfect information or not. We say that a player has perfect information if, whenever she has to make a choice, she is informed about everything her opponents previously chose; in other words: when her information sets are singletons. Thus, the event that player i *has perfect information* is defined as,

$$PI_i = \left\{\omega \in \Omega\middle|v_i\left(\omega\right) = \{\{h\}|h \in H_i\}\right\},$$

and, as usual, we denote $PI_{-i} = \bigcap_{j\neq i} PI_j$ and $PI = \bigcap_{i\in I} PI_i$. Notice the event that there is perfect information differs from the standard notion of game with perfect information, in which not only there is perfect information, PI, but furthermore, event PI is commonly known. Obviously, all the sets introduced in this section are closed, and therefore, well-defined events.

3.2.2. Common strong belief in opponents being rational, having perfect information and strongly believing in their own perfect information

Battigalli and Siniscalchi (2002) introduce the concept of *strong belief* in order to formalize the idea of forward induction, i.e., conjecturing about opponents' future behavior depending on the information collected about their past behavior. They define the strong belief operator, which, adapted to the present context, associates each event E with the event that player i conditionally believes E with probability 1 after any information sequence belief-consistent with E. Thus, formally, each player i's *strong belief operator* is given by $E \mapsto SB_i\left(E\right)$, where,

$$SB_i\left(E\right) = \bigcap_{\sigma_i \in \Sigma_i^{\emptyset}(E)} B_i\left(E\middle|\sigma_i\right),$$

for any event $E \subseteq \Omega$. Thus, $SB_i\left(E\right)$ should be read as the event that player i maintains the hypothesis that E is true as long as it is not contradicted by evidence. For our result in Theorem 1 we are interested in the working hypothesis that players believe that their opponents are rational, have perfect information and strongly believe in their own perfect information. Formally, this is represented for each player i by $SB_i(\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j))$. If we iterate these strong belief assumptions, we invoke a special case of Battigalli's (1996) *best rationalization principle*, according to which, players rationalize observed behavior to the highest degree possible. We add the particular feature that this rationalization assumes that opponents have perfect information and strongly believe in their own perfect information. To formalize this idea, for each player i set,

$$CSBORPI_{i,1} = SB_i\left(\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j)\right) \text{ and,}$$

$$CSBORPI_{i,n+1} = CSBORPI_{i,n} \cap SB_i\left(\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n}\right),$$

for any $n \in \mathbb{N}$. Then, set $CSBORPI_i = \bigcap_{n\in\mathbb{N}} CSBORPI_{i,n}$, and denote, as usual, $CSBORPI = \bigcap_{i\in I} CSBORPI_i$. It is important to notice that $CSBORPI$ makes no assumption at all regarding whether or not the game has perfect information: it is consistent with every possible information structure. Furthermore, as mentioned above, this framework enables i to face uncertainty about her own information structure; accordingly, note that $CSBORPI_i$ imposes no constraints on player i's beliefs about what her own information structure is or will be as the game unfolds.[21]

4. Uncertain information structure and the backward induction outcome

Battigalli and Siniscalchi's (2002) Proposition 6 shows that in extensive-form games without uncertainty about the information structure, rationality and common strong belief in opponents' rationality induce Pearce's (1984) *extensive-form rationalizable* (EFR) strategies.[22] In addition, it is known from Battigalli (1997) that in games with (commonly known) perfect information and trees with no relevant ties, EFR strategies induce precisely the unique backward induction outcome of the tree. Our main result in Theorem 1 shows that proper restrictions on the rationalization process enable the perfect information assumption to be dropped with no consequences in terms of outcome. That is, obtaining the backward induction outcome is unrelated to players' ability to observe their opponents' past choices. We present the result in Section 4.1, illustrate its logic in an example in Section 4.2 and discuss the precise relation of the result to the different sources of uncertainty each player faces in Section 4.3.

4.1. Main result

Next, we present the main result of the paper, which shows that perfect information is not required in order to obtain the backward induction outcome if higher-order beliefs about the information structure satisfy certain conditions. An immediate consequence is that the information that players acquire by observing opponents' choices in the perfect information case is found to be irrelevant in terms of behavior. Thus, if we denote by $[z = z_{\mathcal{I}}]$ the event that the backward induction outcome is obtained,[23] we have that:

---

19 There are many ways to define conditional belief operators in dynamic settings; we stick to an adaptation of the one by Dekel and Siniscalchi (2015).

20 For the following characterization remember the best-reply correspondence introduces in Section 2.4.2, and that $e_{i,1}(\omega)$ represents player i's conditional belief in $S_{-i} \times \mathcal{V}$ corresponding to state ω.

21 Although it plays no role in our results, it is convenient to be aware of the fact that opponents' perfect information, PI_{-i}, cannot be falsified in player i's view: she can observe (or not) her opponents' actions, but cannot derive from these actions whether her opponents had perfect information or not. In consequence, player i having strong belief in perfect information, $SB_i(PI_{-i})$, is equivalent to player i conditionally believing in PI_{-i} after *every* information sequence σ_i.

22 Common strong belief in opponents' rationality, $CSBOR$, is defined in a way analogous to $CSBORPI$, merely by replacing each $\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j)$ with just R_{-i}.

23 I.e.,$[z = z_{\mathcal{I}}] = \left\{\omega \in \Omega\middle|z(s(\omega), v(\omega)) = z_{\mathcal{I}}\right\}$.

Theorem 1 (*Sufficiency for the Backward Induction Outcome*)**.** *Let $\mathcal{G}$ be a game with uncertain information structure. Then, under rationality and common strong belief in the event that opponents are rational, have perfect information and strongly believe in their own perfect information, the backward induction outcome of the tree is obtained regardless of the information structure; i.e.,*

$R \cap CSBORPI \subseteq [z = z_{\mathcal{I}}]$.

The main intuition behind the result is detailed in Appendix A.1. Basically, it consists in showing that under the assumptions of the theorem three phenomena arise. First, if player i has missed to observe some of her opponents' past actions she will believe, as long as what she has observed does not contradict it, that her opponents' are playing some rationalizable strategy of the game with commonly known perfect information. Second, because of the latter and the fact that the path consistent with rationalizability is unique, player i will be able to infer the choices she has missed to observe. Third, once player i has inferred which specific history she finds herself at, and given that this specific history will necessarily be in the backward induction path, player i will conclude that the best thing she can due is to continue in the backward induction path.

4.2. Stan's Used Car Emporium: Elaine's reputation and choice

We illustrate the logic behind Theorem 1 by going back to the example in Section 2.5. As mentioned there, intuitively, we can identify Elaine's information structure with her expertise in cars: if she is an expert, she can tell good cars from lemons and has, therefore, perfect information; on the contrary, if she is not an expert and cannot assess the quality of the car offered, she has imperfect information. Then, we can understand Elaine and Stan's belief hierarchies on the information structure as Elaine's reputation regarding expertise in cars. Under this interpretation, let us analyze the interaction in the following three scenarios:

Stan vs. the expert (Common knowledge of perfect information). Assume that Elaine is an expert in cars, and that since both she and Stan live in the same small area, she has a reputation of expertise that is commonly known to them. In that case, not only will Elaine be able to distinguish the good car from the lemon, but her ability to do so will be commonly known to both her and Stan. As mentioned above, the unique backward induction outcome of the game with commonly known perfect information is $z_I = (\emptyset, e, g, a)$, and this is what we obtain under rationality and common strong belief in rationality: a rational Elaine will chose a at $(\emptyset, e, g)$ and r at $(\emptyset, e, \ell)$, where she will believe that Stan takes her for rational and non-rational, respectively. A rational Stan who believes that Elaine is rational will choose g at $(\emptyset, e)$, and a rational Elaine who believes in both Stan's rationality, and Stan's belief that she is rational, will be able to predict all the reasoning above, and will therefore, choose e at $\emptyset$.

Stan vs. the non-expert (Common knowledge of imperfect information). Assume now that Elaine knows little about cars and is unable to distinguish the good car from the bad one, and further assume that this is commonly known. We are now faced with a case of commonly known imperfect information, and appeal to forward induction. There is only one way in which Stan can rationalize seeing Elaine walk into his business: Elaine, who cannot tell the quality of the car she will be offered, must expect to be offered the good one and thus, must have planned to accept the deal regardless of her ignorance of whether it is the good car or the lemon (any other option is worse for Elaine than staying home). Thus, since Elaine apparently plans to accept whatever Stan offers, it becomes optimal for Stan to offer the lemon. Now, before deciding whether to stay home or not, Elaine is able to predict Stan's reasoning (because she believes both that he is rational and that he will rationalize her decision of entering the emporium if she does so). Thus, since she expects to be offered the lemon, Elaine decides to stay home and keep her old car. Hence, we conclude that under rationality and common strong belief in opponents' rationality, the outcome of the game for the commonly known imperfect information case is $z = (\emptyset, k)$.

Stan vs. the fake expert (Theorem 1). Now assume again that Elaine knows little about cars and that, unlike in the previous case, she has a deluded reputation of expertise. This is precisely what is represented by event $\neg PI_E \cap (R \cap CSBORPI)$. Now, as in the first case, Stan expects her to be able to distinguish between good cars and lemons, so if he sees her walk into his emporium, he chooses to offer her the good car. However, despite that, as in the second case, Elaine does not know what car she has in front of her, the fact that she is aware of her deluded reputation of expertise makes her infer that a rational Stan must have offered her the good car. Thus, she accepts the offer. And note that indeed, $R_E \cap CSBORPI_E$ implies that before leaving home, Elaine is able to predict this outcome of her visit to the emporium; thus, she decides to visit Stan's rather than keeping her old car. Consequently, we obtain $z_{\mathcal{I}}$ again, despite the game not having perfect information and Elaine knowing that this is indeed the case.

4.3. Robustness of Theorem 1

As seen in the previous section, in some simple games the logic of why the backward induction should be obtained easily extends to environments where we allow for uncertainty about the information structure but impose common strong belief assumptions on opponents' perfect information. In this section we rely on the analysis of the sensitivity of the result to the different sources of uncertainty players face as a vehicle to observe that, indeed, guaranteeing the obtention of the backward induction outcome under the presence of uncertainty on the information structure necessarily requires tight constraint in players' higher-order beliefs. Specifically, regarding the assumptions of the theorem, despite the result is found to be independent of both information structures and beliefs about own information structures, it does exhibit great sensitivity to departures in the assumptions about beliefs about opponents' information structures. We conjecture that constraints on beliefs about opponents' beliefs about their own information structure are not required for the result to hold[24]; however, as we illustrate in the last paragraph, the intuition behind this conjecture is not completely obvious. Let us take a look at each source of uncertainty separately:

Information structures. The assumptions on Theorem 1 only refer to each player's belief hierarchies (via constraint $CSBORPI_i$) and behavior, or reaction, to these belief hierarchies (R_i). That is, they only put restrictions on the first two component of each state $\omega = (e, s, V)$. Obviously, this results on the assumption on Theorem 1 being consistent with any information structure the game might have.[25] Perfect information is a particular case, but this is not required at all: on the contrary, it could be the case that some player i has not observed any past action by her opponents; Theorem 1 shows that even such player acts exactly as she would in case she was capable of observing every past action by her opponents.

Beliefs about own information structure. As seen, the assumptions on Theorem 1 only refer to each players' belief hierarchies and behavior. We can be more specific though: restrictions on belief hierarchies only affect each player i's beliefs about her opponents'

24 Remember that Theorem 1 assumes that each player i strongly believes each player $j \neq i$ to strongly believes in her own perfect information.

25 Formally, we have that $R \cap CSBORPI \cap [v = V] \neq \emptyset$ for any information structure V.

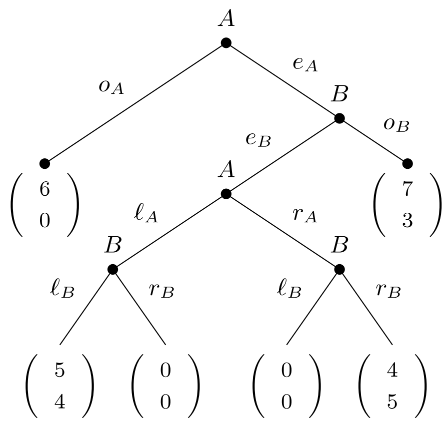

Fig. 5. Uncertainty about own information structure and rationalization.

rationality and information structure, but not about her own information structure. That is, a player might be uncertain about her ability to observe opponents' choices in future stages of the game (or, for instance, certain that she will not observe anything). Theorem 1 shows that this uncertainty is immaterial in terms of outcome: every player will play as if she could observe everything and knew she was going to be able to observe everything. However, as we see next, the way in which players reason off-path is significantly affected, and therefore, reaching the conclusion that their behavior does not depart from the standard case of common knowledge of perfect information is not immediate.

To illustrate this point let us take a look at the game tree in Fig. 5. Both players, Alice (A) and Bob (B) have an outside option (o_A and o_B, respectively), being Alice the first one with the possibility of exercising it. If both players decide to enter the game (i.e., they choose e_A and e_B) they participate in a Battle of Sexes where Alice chooses first.

Let us analyze the standard case of commonly known perfect information first. It is clear that upon observing that Alice entered the game, it becomes strictly dominant for Bob to enter the game too: in case he exercises his outside option his payoff is 3, whereas in case he enters the game he can guarantee a minimum payoff of 4 (because he will be able to observe Alice's action and therefore optimally react to it). In consequence, if at the beginning of the game Alice is rational and expects Bob to be rational, then she will exercise her outside option because she will not expect Bob to exercise his in case she enters the game. Thus, in the case of commonly known perfect information o_A is the only outcome consistent with rationality and common strong belief in opponents' rationality. Still, Bob's only strategy consistent with rationality and common strong in rationality dictates him to enter the game in case he observes that Alice entered the game.

This is in contrast with the case with uncertainty about the information structure, even under the assumptions on Theorem 1. Let us have a look at it. How is Bob going to reason if he observes that Alice entered the game? He cannot discard that she is rational, because he can believe that Alice decided to enter the game expecting him to exercise his outside option. But, as we saw above, if Bob is rational and believes to have perfect information in the continuation game, then he declines to exercise his outside option. In consequence, Alice will not enter the game under the assumption of Theorem 1. We see then that under these assumptions Alice entering the game is a zero probability event for Bob. Thus, if he observes e_A Bob can update his beliefs in an arbitrary way as long as he maintains the belief that Alice will be rational in case he decides to enter the game too. In particular, it is possible for Bob to update his beliefs in the following way: let him assign probability 1 to not having perfect information in the continuation game and probability 1/2 to each of Alice's possible actions in that game. Clearly, these beliefs make optimal for Bob to exercise his outside option. We see then that the reasoning processes of the case of commonly known perfect information and the case with uncertainty about the information structure diverge under the assumptions on Theorem 1; still, the theorem shows that predictions do not.

First-order beliefs about opponents' perfect information. We already observed that the result in Theorem 1 is robust to two aspects described by the state of the world: the information structure and players' beliefs about their own information structure. However, in general, it is not possible to ensure that the result allows for relaxations on the higher-order belief assumptions about opponents' perfect information; indeed, not even for relaxations on first-order belief assumptions. To see it, consider the sequential Battle of Sexes in Section 1.1 and introduce the following variation: Alexei's payoff when he and Polina choose *right* and *left* respectively is 10 instead of 0. In that case, under $R \cap CSBORPI$, Alexei still plays *left*, but as soon as we drop the assumption that he believes Polina to have perfect information, *right* becomes strictly dominant for him, and thus, the backward induction outcome fails to be obtained. This example illustrates the necessity of tight high-order belief assumptions about opponents' perfect information in order to ensure that he backward induction will be obtained for any arbitrary payoff structure of the game tree.

Second-order beliefs about opponents' first-order beliefs about their own information structure. We have already observed that the result in Theorem 1: (i) is independent of both information structures and players' beliefs about their own information structures and (ii) it crucially hinges on tight constraints on higher-order beliefs about opponents' perfect information. The assumption that each player i player strongly believes each opponent $j \neq i$ to strongly believe in her own perfect information turns out to be more involved, as we shall see next.

Consider the modified game between Alice and Bob depicted in Fig. 6. In this case, no matter whether Bob believes Alice to be rational or not, in the game with commonly known perfect information it is dominant for him to enter the game, because he knows that regardless of Alice's next action, he can best respond and therefore guarantee a minimum payoff above the one corresponding to his outside option. Thus, if in the beginning of the game Alice believes Bob to be rational then it becomes optimal for her to enter the game: she expects Bob to also do so and therefore expects a minimum payoff above the one corresponding to her outside option. Thus, players being rational and Alice believing that Bob is rational (in the beginning of the game) is enough to ensure that outcome $(e_A, e_B, \ell_A, \ell_B)$ is obtained.

Let us see next what the impact of uncertainty on the information structure is. In particular, suppose that in the beginning of the game Alice: (i) believes that Bob is rational and has perfect information, (ii) believes that Bob believes that she is rational (and has perfect information; this is trivially true for Alice), but (iii) believes that Bob does not believe that he will have perfect information in the Battle of Sexes game (that is, it is not true that Alice strongly believes in $SB_B(PI_B)$). In such case it is possible for Alice to believe that at information set $\{(\emptyset, e_A)\}$ Bob assigns probability 1/2 to her playing ℓ_A in the Battle of Sexes, and probability 1/2 to her playing ℓ_B. If these were Bob's beliefs upon observing e_A (keep in mind that we are assuming that in Alice's mind Bob expects not to have perfect information in the Battle of Sexes game), it would be optimal for him to exercise his outside option. In consequence, if in the beginning of the game Alice's beliefs about Bob's beliefs at $\{(\emptyset, e_A)\}$ were as described above then it would be optimal for her to exercise her outside option as well. So despite Alice believes Bob to have perfect information and she holds second-order beliefs

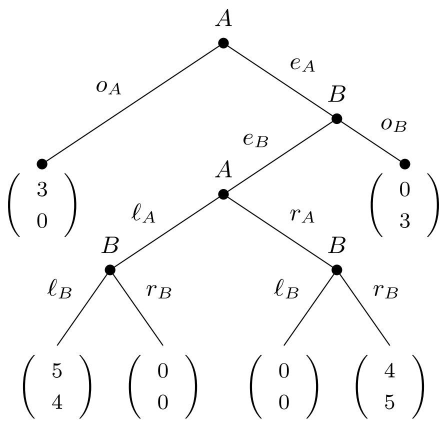

Fig. 6. Impact of lack of strong belief in opponents' strong belief in own perfect information

of mutual rationality, the fact that she does not believe Bob to believe in his own perfect information results in Alice, although rational, playing a strategy that is dominated in the game with commonly known perfect information (whenever she believes in Bob's rationality).

Obviously, if we were assuming (as the theorem does) that Alice believes that Bob believes that he will have perfect information, Alice believing that Bob is rational would suffice for her to decide to enter the game. However, as we argue next, such belief in opponents' belief in own perfect information seems not to be necessary. In addition to assumptions (i)–(iii) in the previous paragraph, suppose that in the beginning of the game Alice believes that Bob believes that she believes that he is rational and has perfect information. Let us see how would Alice reason about Bob's reasoning at $\{(\emptyset, e_A)\}$: despite Bob believes that he will not observe Alice's choice in the Battle of Sexes, he believes that she expects both him to do so and him to best react to whatever she plays. Then, at $\{(\emptyset, e_A)\}$, Bob must expect Alice to play ℓ_A in the Battle of Sexes, and therefore, it becomes optimal for him to enter the game. This is Alice's conclusion about Bob's strategic reasoning at $\{(\emptyset, e_A)\}$. In consequence, despite she does not believe that Bob believes that he will have perfect information, imposing three layers of belief in opponents' rationality and perfect information is enough to lead Alice to entering the game.

In light of the above we conjecture that assumptions on beliefs about opponents' belief about own perfect information are dispensable. However, as the example illustrates, the level of complexity grows fast as these assumptions are abandoned. Specifically, it seems that ensuring behavior that corresponds to kth-order rationalizable behavior in the game with commonly known perfect information (what, following Battigalli and Siniscalchi Battigalli (1997), eventually allows for concluding that the backward induction is obtained) requires more than k iterations of belief in opponents' rationality and perfect information, and that indeed, this difference in iterations is increasing in the length of the game tree. Thus, great difficulties arise in linking behavior in the game with uncertain structure with rationalizable behavior in the game with commonly known perfect information. In consequence, in the absence of higher-order beliefs about opponents' belief in own perfect information, the argument that connects common strong belief in opponents' rationality and perfect information with the backward induction outcome remains unclear.

5. Final remarks

A. Summary. Literature in extensive-form games typically assumes that, in such games, the way in which information sets are distributed along the tree is commonly known. This feature can be understood as the information structure of the game being part of the objective rules of the game. However, since information sets concern players' ability to observe, interpret and remember opponents' past behavior, they often depend more on players' personal cognitive abilities than on the rules of the game itself. Thus, since personal cognitive abilities are usually uncertain, it is natural to wonder how predictions in extensive-form games are affected by players facing incomplete information regarding the information structure. In this context, this paper contributes to the literature in extensive-form games and epistemic game theory by:

(i) Introducing a tractable game-theoretical framework that allows for arbitrary uncertainty regarding the information structure. This is done by defining *games with uncertain information structure*, which consist of a game tree, assumed to be commonly known, and an information structure for the game, assumed to be possibly uncertain.
(ii) Extending, first, the notion of conditional belief hierarchy due to Battigalli and Siniscalchi (1999), and second, the resulting epistemic framework employed by the same authors, so that uncertainty about the information structure can be incorporated.
(iii) Showing that, under suitable restrictions on the rationalization process, the information structure of the game is irrelevant in terms of outcome: players behave exactly as in the case of commonly known perfect information whether or not they observe any of their opponents past choices. Specifically, Theorem 1 proves that, if players are rational (R) and there is common strong belief in opponents being rational, having perfect information and strongly believing in their own perfect information (*CSBORPI*), then the backward induction outcome of the game tree is obtained. Notice that no assumption regarding the information structure of the game is made.

B. Robustness of the result and dependence on the epistemic requirements. Theorem 1 introduces sufficient epistemic conditions for the backward induction outcome for any game tree, regardless of what factual information structure the game happens to have as it is played. These epistemic assumptions differ from those that would represent the standard notion of perfect information game in the absence of uncertainty about the information structure: perfect information, common strong belief in own perfect information and common belief in opponents' perfect information. Thus, Theorem 1 addresses robustness properties of the backward induction outcome in two somewhat opposite ways: (i) it is proved that the backward induction outcome is robust to shocks in the information structure of the game, as long as these shocks do not affect players' higher-order beliefs about their opponents' information structures, and (ii) in general, players' belief in opponents' perfect information plays a crucial role in the backward induction outcome being obtained, so that the latter is found very sensitive, i.e., non-robust, to changes in such beliefs. The sensitivity of the result to the assumption that each player strongly believes that their opponents strongly believe in their own perfect information is inconclusive: we conjecture that the assumption can be dropped, but great difficulties arise when studying an environment where it does not hold.

C. Knowledge of own information structure. The modeling of games with uncertain information structure presented in Section 2 allows for players facing uncertainty not only about their opponents' information structure, but also about their own one. Two observations are worth being made to this respect[26]:

26 Thanks are due to an anonymous referee for calling attention on the following two issues.

(i) A question that naturally arises is whether allowing for uncertainty about own information structure might imply violations of positive introspection, i.e., about a player knowing what she knows. In a standard game with commonly known information structure, as the game unfolds, a player possibly gains information about her opponents' past choices. In a game with uncertain information structure, as the game unfolds, a player gains information both about her opponents' past choices (possibly) and about her own information structure (certainly). That is, the player learns about the possible outcome of the game by discarding those outcomes inconsistent with the actions of her opponents she has observed, and learns about her own information structure by discarding those information structures not consistent with the information sets she already went through. Similarly as learning about opponents' past behavior does not pin down their future behavior and therefore allows for uncertainty about the outcome of the game, partial learning about the information structure (i.e., going through some information sets) does not necessarily pin down what *future* information sets will be like and therefore allows for uncertainty about the information structure (though not, of course, about which the *past* information sets were). As an illustration, consider again the partial tree depicted in Fig. 3. Suppose that Player 2 finds herself at (singleton) information set $\{u\}$. In such case, Player 2 knows that her information structure is not that beginning with information structure $\{u, d\}$ because she knows she has observed Player 1's choice u. Still, this is not enough to conclude that she knows her information structure: she might be uncertain about whether she will observe Player 3's choice or not. For instance, at $\{u\}$, she could assign positive probabilities to the following two information structures:

$$V_2^1 = \{\{u\}, \{d\}, \{uuu, uud\}, \{udu, udd\}, \{duu, dud\}, \{ddu, ddd\}\},$$
$$V_2^2 = \{\{u\}, \{d\}, \{uuu\}, \{uud\}, \{udu\}, \{udd\}, \{duu\}, \{dud\}, \{ddu\}, \{ddd\}\},$$

where V_2^1 is the information structure corresponding to the case in which Player 2 observes Player 3's choice and V_2^2 the one corresponding to the case in which she does not. Note that both are consistent with Player 2 observing Player 1's choice. On the contrary, at $\{u\}$, Player 2 necessarily assigns 0 probability to, for example, information structure:

$$V_2^3 = \{\{u, d\}, \{uuu, uud\}, \{udu, udd\}, \{duu, dud\}, \{ddu, ddd\}\},$$

which would have required Player 2 not having observed Player 1's choice.[27] This would be at odds with her acquired knowledge that she did observe Player 1's choice.

(ii) In cases in which the economic analysis focuses on environments where uncertainty about own information structure can be reasonably excluded, the model can accommodate the assumption that players know their own information structure in the sense that they hold correct beliefs about it. To do so, simply consider the event that each player knows her own information structure together with the event that each player exhibits common belief (i.e., common *everywhere* belief) in the event that opponents' know their own information structure.[28]

However, a more sensible alternative approach to these situations consists on constructing a reduced version of the general framework that losses in generality but gains in tractability and transparency. This can be done by simply envisioning the information structure of each player i as her (privately known) type and then supposing that her set of strategies is conditional on her type/information structure.[29] In this reduced model the way in which each information set can be reached is univocally determined by the type of the player; that is, all the information gained by her during the game is already encoded in her information set and therefore, the notion of information sequence is unnecessary.

D. Initial types. The main result of this paper is formulated in terms of epistemic game theory and restrictions on states of the world. An alternative approach would be to work at a purely game-theoretical level: defining first non-conditional belief hierarchies (or type spaces à la Harsanyi, 1967–1968) which only reflect initial (that is, at history $\emptyset$) uncertainty on the information structure,[30] then introducing some interim solution concept that maps each belief hierarchy to a set of strategies induced by some notion of common strong belief in rationality,[31] and finally, finding what restrictions on belief hierarchies ensure that the backward induction is obtained. This approach may be very interesting, but is problematic in the present set-up. We allow players to have beliefs about their own information structure, but do not restrict these beliefs. This can result in situations in which players update their initial belief hierarchies even when their opponents' behavior did not falsify their initial beliefs (for instance, when player i initially believes that she will not have perfect information, but while on the backward induction path, finds herself at a singleton). After such unexpected own information structure, a player will retain her higher-order beliefs about opponents' rationality, but may arbitrarily update her higher-order beliefs about her opponents' perfect information. In particular, she could update her belief so that she does no longer believe that her opponents will have perfect information in the future. Thus, we cannot ensure that the inductive choice is made.

E. Future belief. An earlier version of this paper presented an alternative analysis of epistemic sufficiency for the backward induction outcome for games with uncertain information structure. This was performed in terms of Perea's (2014) notion of common belief in opponents' future rationality, to which we had to add common belief in opponents' future perfect information. These two assumptions suffice to ensure that players that find themselves at singletons always follow the conditional backward induction path. However, they do not guarantee that players that missed to observe some past action by their opponents will be able to infer in which precise history of their respective information set they find themselves at, and therefore, it is not possible to ensure backward

27 More specifically, this information structure corresponds to a game in which Player 2 does not observe Player 1's choice, and in her second turn she does not observe Player 3's choice but is informed of Player 1's previously unobserved choice.

28 Note that the belief that an opponent holds certain beliefs can never be falsified (unless we additionally impose simultaneous rationality restrictions). Thus, formally, we would set first $K_i^0 = \left\{\omega \in \Omega \middle| SB_i(v_i = v_i(\omega))\right\}$, and then, take the intersection of $\bigcap_{i\in I} K_i^0$ with every $\bigcap_{i\in I} K_i^n$, where $K_i^n = \bigcap_{\sigma_i \in \Sigma_i^{\emptyset}} B_i(\bigcap_{j\neq i} K_j^{n-1} | \sigma_i)$ for any player i.

29 Formally, the set of strategies available to player i given type/information structure V_i would be $S_i(V_i) = \prod_{v_i \in V_i} A_i(v_i)$.

30 Whenever we refer to belief hierarchies in this discussion, we refer always to uncertainty about the information structure, $\mathcal{V}$, not about $S_{-i} \times \mathcal{V}$.

31 A strong belief in rationality version of Penta's (2012) interim sequential rationality, in a set-up where types reflect uncertainty on information structures, not payoffs.

inductive behavior. The earlier version of this paper fixed this problem additionally imposing strong belief in opponents' rationality and perfect information. This last assumption, combined with the previous two, allows players to rationalize the pieces of past behavior observed and infer their exact location in the tree. Still, as accurately pointed out by Andrés Perea to us, the strong belief assumption and the belief in future rationality assumption become mutually exclusive in epistemic type spaces that are sufficiently rich, such as the belief-complete ones we consider in the present version of the paper.

F. ABSENCE OF RELEVANT TIES. The fact that the game trees under consideration have no relevant ties is crucial for Theorem 1. Indeed, if the game tree had more than just one backward induction outcome, it would be impossible to infer opponents' past choices in some non-singleton information sets and therefore to identify which choice is actually the inductive one. Consider an uncertain imperfect information game such as the one in Fig. 2 and modify the payoffs so that every non-null payoff is exactly 1. Assume in addition that the conditions in Theorem 1 are satisfied. In this case these conditions are reduced to: rationality (R), Alexei's belief in Polina's rationality and perfect information (implicit in $CSBORPI_A$) and both Polina's strong belief in Alexei's rationality and her belief in Alexei believing she is rational and has perfect information (implicit in $CSBORPI_P$). Since Alexei believes both that Polina is rational and has perfect information, he believes that any choice of his yields him 1. He is therefore indifferent and may choose either left or *right*. Thus, if it is the case that Alexei is deluded and Polina has no perfect information, there is nothing she can infer from $CSBORPI_P$, and despite being rational, finds no reason to expect *left* or *right* to yield her a higher payoff than the other alternative.

Acknowledgments

A previous version of this manuscript circulated under the title "Imperfect incomplete information and backward induction". Financial support from the (now defunct) Spanish Ministry of Science and Innovation, the Spanish Ministry of Economy and Competitiveness and the Basque Government is acknowledged (grants ECO2009-11213, ECO2012-31346, and IT568-13 and POS-2015-1-0022 respectively). I thank the editor, Sven Rady, and three anonymous referees whose insightful observations and suggestions significantly contributed to shaping the paper. Thanks are due to Elena Iñarra, Annick Laruelle and Fabrizio Germano for inestimable help and advice, and to Bob Aumann, Pierpaolo Battigalli, Gilad Bavly, Eddie Dekel, Aviad Heifetz, Ziv Hellman, Antonio Penta, Andrés Perea, Dov Samet and Marciano Siniscalchi for discussion and valuable comment. I express my gratitude for the immense hospitality of both the Federmann Center for the Study of Rationality at the Hebrew University and Eilon Solan and the Game Theory group in Tel Aviv University, where most of the project was developed during scholar visits. All errors are, from first to last, mine.

A. Preliminaries for the proofs

A.1. Outline of the proof of Theorem 1

The proof of Theorem 1 is conceptually simple but requires some technical intricacies and additional notation.[32] For the sake of expositional clarity, let us present a brief description of the different steps leading to the proof before jumping later to the fully detailed version in Appendix B. The proof is based on the following three intermediate result:

1. Suppose that player i taking part in a game with uncertain information structure: (i) is rational (R_i holds), (ii) has strong belief in his own perfect information ($SB_i(PI_i)$ holds) and (iii) exhibits common strong belief of order n in opponents being rational, having perfect information and having strong belief in their own perfect information ($CSBORPI_{i,n}$ holds). Suppose that s_i is a strategy for player i consistent with (i), (ii) and (iii). We show first that, if player i happens to have perfect information (PI_i holds) then the behavior induced by strategy s_i in the game tree corresponds to a $(n+1)$th-order rationalizable strategy s_i^* (Pearce, 1984) of the game with commonly known perfect information. The proof is performed by induction in Appendix B.2.1; the initial step for $n = 0$ is completed in Lemma 1, and the inductive one, in Lemma 2.
2. We show that the backward induction path of the game tree is consistent with players being rational (R holds), having perfect information (PI holds), having strong belief in their own perfect information ($\bigcap_{i\in I} SB_i(PI_i)$ holds) and exhibiting common strong belief of order n in their opponents being rational, having perfect information and having strong believe in their own perfect information ($\bigcap_{i\in I} CSBORPI_{i,n}$ holds). This follows easily from the result in the previous step. The formal proof is provided in Corollary 1 in Appendix B.2.2.
3. Suppose that player i is taking part in a game with uncertain information structure and that, after every information sequence σ_i consistent with the backward induction path, player i believes that all her opponents are playing some rationalizable strategy of the game with perfect information. We show that then, after every such information sequence σ_i, a rational player i chooses to play the action that corresponds to the backward inductive choice of the (unique) history in which σ_i and the backward induction path intersect. The proof in presented in Lemma 3 in Appendix B.2.3.

It is easy to see how the three intermediate results above interact to give rise to the claim in Theorem 1. For any player i, $CSBORPI_i$ implies that whenever possible, player i believes her opponents to be rational, to have perfect information, to have strong belief in their own perfect information and to each of them exhibit $CSBORPI_j$. It follows from the first step and from the monotonicity of the conditional belief operators that whenever player i can hold these believes, then she also believes her opponents to play according to some rationalizable strategy of the game with perfect information. Hence, it follows from the second step that $CSBORPI_i$ implies that, while player i is not aware of being outside the backward induction path, she believes her opponents to be playing according to some rationalizable strategy of the game with perfect information. As shown in the third step, this makes player i able to infer in which specific history of her information sequence she finds herself at,[33] and conclude that her best possible choice is to stay in the backward induction path.

A.2. Games with (commonly known) perfect information

Given game with uncertain information structure $\mathscr{G} = \langle \mathscr{T}, \mathcal{V}\rangle$, we denote by $\mathscr{G}^*$ the game with perfect information identified with game tree $\mathscr{T}$. Then, in the context of $\mathscr{G}^*$, for each player i we have:

- A set of strategies $S_i^* = \prod_{h\in H_i} A_i(h)$. As usual, $S^* = \prod_{i\in I} S_i^*$ denotes the set of strategy profiles, and $S_{-i}^* = \prod_{j\neq i} S_j^*$, the set of i's opponents' partial strategy profiles. Obviously, for any history $h \in H_i^{\emptyset} = H_i^{\emptyset} \cup \{\emptyset\}$, each strategy profile s^*

32 Specifically, notation regarding games with commonly known perfect information. This is presented in A.2.

33 Remember now that we know from Battigalli (1997) that all rationalizable strategy profiles induce the same path.

induces a unique conditional outcome $z^*(s^*|h) \in Z$; when $h = \emptyset$, we simply write $z^*(s^*)$. For each history $h \in H_i^{\emptyset}$ let $S_{-i}^*(h)$ denote i's opponents' strategies that reach h,[34] and similarly, let $H_i^{\emptyset}(s_i^*)$ denote the set of player i's histories reached by strategy s_i^*.[35]

- A set of conjectures $\Delta^{H_i^{\emptyset}}(S_{-i}^*)$, build upon conditional base space $(\mathcal{U}_i^*, \mathcal{C}_i^*)$ consisting of basic uncertainty space S_{-i}^* and family of conditioning events $\{S_{-i}^*(h)|h \in H_i^{\emptyset}\}$. For any conjecture μ_i, any strategy s_i^* and any history $h \in H_i^{\emptyset}$ we denote player i's conditional expected payoff at h by $U_i^*(\mu_i, s_i^*|h)$.[36] Players i's best-reply correspondence is defined as $BR_i^* : \Delta^{H_i^{\emptyset}}(S_{-i}^*) \rightrightarrows S_i^*$, where,

$$\mu_i \mapsto \left\{ \hat{s}_i^* \in S_i^* \middle| \hat{s}_i^* \in \bigcap_{h \in H_i(\hat{s}_i^*)} \arg\max_{s_i^* \in S_i^*} U_i^*(\mu_i, s_i^*|h) \right\}.$$

Finally, following Pearce (1984), each player i's set of (extensive-form) rationalizable strategies is defined as $\mathrm{EFR}_i^* = \bigcap_{n \geq 0} \mathrm{EFR}_{i,n}^*$, where, first, for each player i, $\mathrm{EFR}_{i,0}^* = S_i^*$, $\mathrm{H}_{i,0}^* = H_i^{\emptyset}$ and $\mathrm{C}_{i,0}^* = \Delta^{H_i^{\emptyset}}(S_{-i}^*(h))$ and then, recursively we set,

$$\begin{aligned}
\mathrm{EFR}_{i,n+1}^* &= \left\{ s_i^* \in \mathrm{EFR}_{i,n}^* \middle| s_i^* \in BR_i^*(\mu_i) \text{ for some } \mu_i \in \mathrm{C}_{i,n}^* \right\}, \\
\mathrm{H}_{i,n+1}^* &= \left\{ h \in \mathrm{H}_{i,n}^* \middle| S_{-i}^*(h) \cap \mathrm{EFR}_{-i,n+1}^* \neq \emptyset \right\}, \\
\mathrm{C}_{i,n+1}^* &= \left\{ \mu_i \in \mathrm{C}_{i,n}^* \middle| \mu_i(h)[\mathrm{EFR}_{-i,n+1}^*] = 1 \text{ for any } h \in \mathrm{H}_{i,n+1}^* \right\},
\end{aligned}$$

for any $n \geq 0$, where $\mathrm{EFR}_{-i,n+1}^* = \prod_{j \neq i} \mathrm{EFR}_{j,n+1}^*$. Notice that finiteness of S^* implies that there exists some $\bar{n}$ such that $\mathrm{EFR}_{i,\bar{n}}^* = \mathrm{EFR}_i^*$ for any player i. For convenience, at some stages we will rely on the result by Shimoji and Watson (1998) that ensures that we can dispense chain-rule updating in the definition of rationalizability. That is, if we initially set $\mathrm{D}_{i,0}^* = \prod_{h \in H_i^{\emptyset}} \Delta(S_{-i}(h))$ and rely on the latter rather than on $\mathrm{C}_{i,0}^*$ to iteratively define rationalizability,[37] the strategy elimination procedure remains exactly the same. Remember that we know from Theorem 4 by Battigalli (1997) that every rationalizable strategy profile induces the backward induction outcome of the game tree.

Behavior in the game with uncertain information structure can be regarded as behavior in the game with commonly known perfect information. We study next how to formalize this idea properly. For each player i let $\sigma_i : H_i^{\emptyset} \times \mathcal{V}_i \to \Sigma_i^{\emptyset}$ be given by $(h, V_i) \mapsto \sigma_i$ where $\sigma_i \subseteq V_i$ and $h \in v_{\sigma_i}$ if $h \neq \emptyset$, and $(h, V_i) \mapsto \emptyset$ otherwise. Map σ_i represents what information would player i have in $\mathcal{G}$ if, given information structure V_i, her history h was reached. Note that σ_i is well-defined, surjective and measurable. When $V_i = V_i^*$ we write simply $\sigma_h = \sigma_i(h, V_i^*)$. Applying these maps we can relate the behaviors corresponding to uncertain information structure and common knowledge of perfect information:

Definition 2 (*Choice Morphism*)**.** Each player i's *choice morphism* is defined as:

$$\begin{aligned}
g_i : \quad S_i \times \mathcal{V}_i &\longrightarrow S_i^* \\
(s_i, V_i) &\mapsto (s_i(\sigma_i(h, V_i)))_{h \in H_i}.
\end{aligned}$$

Map g_i represents what, given strategy s_i and information structure V_i, would player i choose in case her history h was reached.[38] Obviously, g_i is well-defined map and notice that outcomes are related by the following relation: $z^*(g_i(s_i, V_i)_{i \in I}) = z(s, V)$ for any $(s, V) \in S \times \mathcal{V}$. In particular, the latter implies that for any pair $(s_i, V_i) \in S_i \times \mathcal{V}_i$,

$$h \in H_i(g_i(s_i, V_i)) \Longrightarrow \sigma_i(h, V_i) \in \Sigma_i(s_i). \tag{1}$$

Remark 1. Every choice morphism g_i is a well-defined quotient map. Since we are dealing with finite spaces, continuity and closedness are trivial, and thus, all we need to check then is surjectiveness. To see it fix arbitrary strategy $\bar{s}_i \in S_i$ and define map $f_i : S_i^* \to S_i$ by setting for each s_i^*, $f_i(s_i^*)(\sigma_i) = s_i^*(h)$ if $\sigma_i = \sigma_h$ for some $h \in H_i^{\emptyset}$ and $f_i(s_i^*)(\sigma_i) = \bar{s}_i(\sigma_i)$ otherwise. Obviously, f_i satisfies that $g_i(f_i(s_i^*), V_i^*) = s_i^*$ for any $s_i^* \in S_i^*$, and hence, we conclude that g_i is surjective.

B. Proofs

Throughout the proofs we make use of the following notational conventions:

- $s_i^*(\omega)$ denotes player i's behavior in the game tree induced by the strategy and the information structure corresponding state ω.[39]
- V_i^* denotes player i's information structure corresponding to the case of perfect information.[40]
- IC_i denotes the event that after every information sequence consistent with the backward induction path, i chooses the action that would keep the play in such path.[41]
- Let $CSBORPI_{i,0} = \Omega$ and $\Sigma_i^n = \Sigma_i^{\emptyset}(\bigcap_{j \neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n})$ for any $n \geq 0$.

B.1. Main result

Theorem 1 (*Sufficiency for the Backward Induction Outcome*)**.** *Let $\mathcal{G}$ be a game with uncertain information structure. Then, under rationality and common strong belief in the event that opponents are rational, have perfect information and strongly believe in their own perfect information, the backward induction outcome of the tree is obtained regardless of the information structure; i.e.,*

$$R \cap CSBORPI \subseteq [z = z_{\mathcal{I}}].$$

Proof. The proof consists of five inclusions. The first and the fifth hold by definition, and the intermediate three correspond to each of the steps informally detailed in Appendix A.1 and proved in Appendix B.2:

$$\begin{aligned}
R \cap CSBORPI &\subseteq \bigcap_{i \in I} R_i \cap \bigcap_{n=0}^{\bar{n}-1} \bigcap_{\sigma_i \in \Sigma_i^n} B_i \\
&\quad \times \left(\bigcap_{j \neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n} | \sigma_i \right) \\
&\subseteq \bigcap_{i \in I} R_i \cap \bigcap_{\sigma_i \in \Sigma_i^{\bar{n}-1}} B_i(s_{-i}^* \in \mathrm{EFR}_{-i}^* | \sigma_i) && \text{(Step 1)} \\
&\subseteq \bigcap_{i \in I} R_i \cap \bigcap_{\sigma_i : v_{\sigma_i} < z_{\mathcal{I}}} B_i(s_{-i}^* \in \mathrm{EFR}_{-i}^* | \sigma_i) && \text{(Step 2)} \\
&\subseteq \bigcap_{i \in I} IC_i. && \text{(Step 3)}
\end{aligned}$$

The obvious fact that $\bigcap_{i \in I} IC_i \subseteq [z = z_{\mathcal{I}}]$ completes the proof. ■

34 I.e., $S_{-i}^*(h) = \{s_{-i}^* \in S_{-i}^* | z^*(s_{-i}^*; s_i^*) > h \text{ for some } s_i^* \in S_i^*\}$.

35 I.e., $H_i(s_i^*) = \{h \in H_i | z_i^*(s_{-i}^*; s_i^*) > h \text{ for some } s_{-i}^* \in S_{-i}^*\}$.

36 Formally, $U_i^*(\mu_i, s_i^*|h) = \sum_{s_{-i}^* \in S_{-i}^*} \mu_i(h)[s_{-i}^*] \cdot u_i(z^*(s_{-i}^*; s_i^*|h))$.

37 Properly substituting each $\mathrm{C}_{i,n}^*$ by the corresponding $\mathrm{D}_{i,n}^*$ in which updating according to the chain rule is not required.

38 The fact that player i might not know that she is at h ($V_i(h) \neq \{h\}$) does not imply that she does not make a choice at h.

39 That is, $s_i^*(\omega) = g_i(s_i(\omega), v_i(\omega))$.

40 That is, $V_i^* = \{\{h\} | h \in H_i\}$.

41 That is, $IC_i = \{\omega \in \Omega | \omega \in \bigcap_{h < z_{\mathcal{I}}} \bigcap_{\sigma_i \in \Sigma_i(s_i(\omega)) : h \in v_{\sigma_i}} [s_i(\sigma_i) = \alpha_h]\}$.

B.2. Intermediate steps

B.2.1. First step

In this section we show that at each stage n of an induction process, two different things hold. First, that $R_i \cap PI_i \cap SB_i(PI_i) \cap CSBORPI_{i,n}$ implies $(n+1)$th-order rationalizable behavior in the game with perfect information for player i. Second, in order to make the inductive argument work, that if h is a history of player i that in the perfect information game can be reached by some $(n+1)$th-order rationalizable play of her opponents, then, information sequence σ_i consistent with perfect information and ending in $\{h\}$ is consistent with opponents satisfying the conditions above, i.e., with $\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n}$. Lemma 1 proves the claim for $n = 0$ and Lemma 2 inductively extends the claim to arbitrary n.

Lemma 1. *For any player i it holds that:*

(i) $R_i \cap PI_i \cap SB_i(PI_i) \subseteq \left[s_i^* \in \mathrm{EFR}_{i,1}^*\right]$.
(ii) $\sigma_h \in \Sigma_i^0$ *for any history* $h \in \mathrm{H}_{i,1}^*$.

Proof. Suppose that player i has enjoyed local perfect information (i.e., $\sigma_i = \sigma_h$ for some $h \in H_i^{\emptyset}$) and exhibits strong belief in own perfect information. Notice then that player i's conditional expected utilities of the game with uncertain information structure can be regarded as conditional expected payoffs of the perfect information game: for any state $\omega \in SB_i(PI_i)$, any strategy $s_i \in S_i$ and any history $h \in H_i^{\emptyset}$ it holds that:

$$U_i(e_{i,1}(\omega), s_i|\sigma_h) = U_i^*(e_{i,1}^*(\omega), g_i(s_i, V_i^*)|h), \tag{2}$$

where $e_{i,1}^*(\omega) \in (\Delta(S_{-i}^*))^{H_i^{\emptyset}}$ is defined by setting:

$$e_{i,1}^*(\omega)(h)[s_{-i}^*] = e_{i,1}(\omega)(\sigma_h)\left[\prod_{j\neq i} g_j^{-1}(s_j^*) \times \{V_i^*\}\right],$$

for any $s_{-i}^* \in S_{-i}^*$ and any $h \in H_i^{\emptyset}$. In principle, we cannot guarantee that $e_{i,1}^*(\omega)$ is a conditional probability system, because it is not obvious that it updates applying the chain rule. Still, this feature will eventually turn out to be immaterial. To prove the equality in (2) simply notice that:

$$\begin{aligned}
&U_i(e_{i,1}(\omega), s_i|\sigma_h)\\
&= \sum_{h\in v_{\sigma_h}} \sum_{(s_{-i},V)\in(S_{-i}\times\mathcal{V})(\sigma_h,h)} e_{i,1}(\sigma_h)[(s_{-i},V)]\cdot u_i((s_{-i};s_i),V|\sigma_h)\\
&= \sum_{(s_{-i},V)\in S_{-i}\times\mathcal{V}} e_{i,1}(\sigma_h)[(s_{-i},V)]\cdot u_i((s_{-i};s_i),V|\sigma_h)\\
&= \sum_{(s_{-i},V_{-i})\in S_{-i}\times\mathcal{V}_{-i}} e_{i,1}(\sigma_h)[(s_{-i},(V_{-i};V_i^*))]\\
&\quad \cdot u_i((s_{-i};s_i),(V_{-i};V_i^*)|\sigma_h)\\
&= \sum_{s_{-i}^*\in S_{-i}^*} e_{i,1}(\omega)(\sigma_h)\left[\prod_{j\neq i} g_j^{-1}(s_j^*)\times\{V_i^*\}\right]\cdot u_i^*(s_{-i}^*;g_i(s_i,V_i^*)|h)\\
&= \sum_{s_{-i}^*\in S_{-i}^*} e_{i,1}^*(\omega)(h)[s_{-i}^*]\cdot u_i^*(s_{-i}^*;g_i(s_i,V_i^*)|h)\\
&= U_i^*(e_{i,1}^*(\omega)), g_i(s_i,V_i^*|h).
\end{aligned}$$

We now prove each part of the claim separately:

Part (i). Fix player i and state $\omega \in R_i \cap PI_i \cap SB_i(PI_i)$. We are going to check that strategy $s_i^*(\omega) = g_i(s_i(\omega), v_i(\omega))$ is a sequential best-reply to conjecture $e_{i,1}^*(\omega)$. We proceed by contradiction and suppose that $s_i^*(\omega)$ reaches some history $\hat{h}$ in which it fails to be optimal w.r.t. $e_{i,1}^*(\omega)$.[42] Then, there must exist some strategy $\hat{s}_i^* \in S_i^*$ such that $U_i^*(e_{i,1}^*(\omega), \hat{s}_i^*|\hat{h}) > U_i^*(e_{i,1}^*(\omega), s_i^*(\omega)|\hat{h})$. It follows from (2) that $U_i^*(e_{i,1}^*(\omega^*), \hat{s}_i^*|\hat{h}) > U_i(e_{i,1}(\omega), s_i(\omega)|\sigma_{\hat{h}})$. Consider now the following strategy for game with uncertain information structure:

$$\hat{s}_i(\sigma_i) = \begin{cases} \hat{s}_i^*(h) & \text{if } \sigma_i = \sigma_h \text{ for some } h \in H_i,\\ s_i(\omega)(\sigma_i) & \text{otherwise,}\end{cases}$$

for any $\sigma_i \in \Sigma_i$. Clearly, strategy $\hat{s}_i$ induces $\hat{s}_i^*$ when player i has perfect information: $g_i(\hat{s}_i, V_i^*) = \hat{s}_i^*$. Thus, (2) implies that $U_i(e_{i,1}(\omega), \hat{s}_i|\sigma_{\hat{h}}) = U_i^*(e_{i,1}^*(\omega), \hat{s}_i^*|\hat{h})$, and in consequence, $U_i(e_{i,1}(\omega), \hat{s}_i|\sigma_{\hat{h}}) > U_i(e_{i,1}(\omega), s_i(\omega)|\sigma_{\hat{h}})$. But notice that this is a contradiction: since $s_i^*(\omega)$ reaches $\hat{h}$, it follows from (1) that $s_i(\omega)$ reaches $\sigma_{\hat{h}}$,[43] and therefore, since $\omega \in R_i$, strategy $s_i(\omega)$ cannot fail to be optimal w.r.t. conjecture $e_{i,1}(\omega)$ at $\sigma_{\hat{h}}$. Hence, we conclude that $s_i^*(\omega)$ is a sequential best-reply to $e_{i,1}^*(\omega)$. In consequence, $s_i^*(\omega)$ is a first-order rationalizable strategy in the game with perfect information. ♦

Part (ii). For technical convenience, we will prove the following claim, which is slightly more general but also more obscure: for any player i and any first-order rationalizable strategy s_i^*, there exists some state $\omega^i \in R_i \cap PI_i \cap SB_i(PI_i)$ that induces s_i^* in the game with commonly known perfect information, i.e., such that $s_i^*(\omega) = s_i^*$. To see it, fix player i and first-order rationalizable strategy $\hat{s}_i^*$. Since $\hat{s}_i^* \in \mathrm{EFR}_{i,1}^*$ we know that there exists some conjecture $\mu_i \in \mathrm{C}_{i,0}^*$ for which s_i^* is a sequential best-reply. Then:

- For each $j \in I$ pick injective map $f_j : S_j^* \to S_j$ such that $g_j(f_j(s_j^*), V_j^*) = s_j^*$ for every $s_j^* \in S_j^*$. Remember that we know from Remark 1 that such objects exist.
- Fix some arbitrary first-order conditional belief $\bar{e}_{i,1} \in \Delta^{\Sigma_i^{\emptyset}}(S_{-i}\times\mathcal{V})$, and some arbitrary $\bar{s}_i \in BR_i(\bar{e}_{i,1})$.

Now, based on conjecture μ_i, family of maps $(f_j)_{j\neq i}$ and first-order conditional belief $\bar{e}_{i,1}$, we define appropriate first-order conditional belief $\hat{e}_{i,1} \in \Delta^{\Sigma_i^{\emptyset}}(S_{-i}\times\mathcal{V})$ by setting:

$$\hat{e}_{i,1}(\sigma_i)\left[(s_{-i},V)\right] = \begin{cases}\mu_i(h)[(f_j^{-1}(s_j))_{j\neq i}] & \text{if } \sigma_i = \sigma_h \text{ for some } h \in H_i^{\emptyset}, \text{ and } V = V^*,\\ \bar{e}_{i,1}(\sigma_i)[(s_{-j},V)] & \text{otherwise,}\end{cases}$$

for every pair $(s_{-i}, V) \in S_{-i}\times\mathcal{V}$ and every information sequence $\sigma_i \in \Sigma_i^{\emptyset}$. It is routine to check that $\hat{e}_{i,1}$ is a well-defined element of $\Delta^{\Sigma_i^{\emptyset}}(S_{-i}\times\mathcal{V})$,[44] and let $\hat{e}_i$ be some conditional belief hierarchy whose first-order conditional belief in precisely $\hat{e}_{i,1}$. Define now strategy $\hat{s}_i$ as follows:

$$\hat{s}_i(\sigma_i) = \begin{cases}\hat{s}_i^*(h) & \text{if } \sigma_i = \sigma_h \text{ for some } h \in H_i^{\emptyset},\\ \bar{s}_i(\sigma_i) & \text{otherwise,}\end{cases}$$

for any $\sigma_i \in \Sigma_i$. Set now $\omega^i = (\omega_{-i}, \hat{e}_i, \hat{s}_i, V^*)$, where ω_{-i} is some arbitrary element in Ω_{-i}. We claim now that ω^i satisfies the requirements we are asking for:

- $s_i^*(\omega^i) = s_i^*$. This follow from construction of $\hat{s}_i$.
- $\omega^i \in SB_i(PI_i)$. This follow from the fact that $\hat{e}_{i,1}$ assigns full probability to V_i^* whenever possible. The fact that $\omega^i \in PI_i$ is trivially true.
- $\omega^i \in R_i$. It follows from the construction of $\hat{e}_i$ and $\hat{s}_i$ and from (2) that after those information sequences consistent with perfect information, $\hat{s}_i$ is optimal with respect to $\hat{e}_{i,1}$. After information sequences not consistent with perfect information, $\hat{s}_i$ is optimal with respect to $\hat{e}_{i,1}$ because $\bar{s}_i$ is optimal with respect to $\bar{e}_{i,1}$.

42 That is, there exists some $\hat{h} \in H_i(s_i^*(\omega))$ such that $s_i^*(\omega) \notin \mathrm{argmax}_{s_i^*\in S_i^*} U_i^*(e_{i,1}^*(\omega), s_i^*|\hat{h})$.

43 That is, $\hat{h} \in H_i(s_i^*(\omega))$ implies that $\sigma_{\hat{h}} \in \Sigma_i(s_i(\omega))$.

44 One needs to check that $\hat{e}_{i,1}(\sigma_i)[(S_{-i}\times\mathcal{V})(\sigma_i)] = 1$ for any information sequence σ_i, and that $\hat{e}_{i,1}$ updates beliefs applying, when possible, the chain rule.

Now to properly prove part (ii) in the claim of the lemma, fix player i and history $h \in \mathrm{H}^*_{i,1}$. We need to prove that there exist some strategy $s_i \in S_i$ and some state $\omega \in \bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j)$ such that $\sigma_h \subseteq v_i(\omega)$ and $z((s_{-i}(\omega); s_i), v(\omega)) > h$. Fix $s^*_{-i} \in \mathrm{EFR}^*_{i,1}$ and $s^*_i \in S^*_i$ such that $z^*(s^*) > h$, and for each $j \neq i$, pick $\omega^j \in R_j \cap PI_j \cap SB_j(PI_j)$ such that $s^*_j(\omega^j) = s^*_j$. Then, pick arbitrary conditional belief hierarchy e_i and define ω as follows $\omega = ((e_j(\omega^j), s_j(\omega^j))_{j\neq i}, (e_i, s_i), V^*)$, where $s_i = f_i(s^*_i)$ for some map f_i as defined in Remark 1. We have then: that $\sigma_h \subseteq v_i(\omega)$, that $\omega \in \bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j)$ and that $z((s_{-i}(\omega); s_i), v(\omega)) = z^*(s^*) > h$. Thus, we conclude that $\sigma_h \in \Sigma^0_i$. ■

Lemma 2. *For any player i and any $n \geq 0$ it holds that:*

(i) $R_i \cap PI_i \cap SB_i(PI_i) \cap CSBORPI_{i,n} \subseteq \left[s^*_i \in \mathrm{EFR}^*_{i,n+1}\right]$.
(ii) $\sigma_h \in \Sigma^n_i$ *for any history* $h \in \mathrm{H}^*_{i,n+1}$.

Proof. The proof proceeds by induction on n. The initial step ($n = 0$) is implied by Lemma 1, so let us focus in the inductive step. Suppose that $n \geq 0$ is such that both claims holds for every $k = 0, \ldots, n$. Next we check separately that both claims hold for $n + 1$:

Part (i). Fix player i and state $\omega \in R_i \cap PI_i \cap SB_i(PI_i) \cap CSBORPI_{i,n+1}$. Define strategy $s^*_i(\omega)$ and conjecture $e^*_{i,1}(\omega)$ for the game with perfect information exactly as done in the proof of Lemma 1. We know from part (i) of the induction hypothesis that $s^*_i(\omega)$ is a sequential best-reply to $e^*_{i,1}(\omega)$, so all we need to check is that $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,n+1}$.[45] This is verified by showing inductively that $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,k}$ for any $k = 0, \ldots, n+1$. It holds trivially that $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,0}$. Suppose that $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,k}$ for some $k \in \{0, \ldots, n-1\}$; let us prove that then, $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,k+1}$. Pick $h \in \mathrm{H}^*_{i,k+1}$. We know from part (ii) of the induction hypothesis that $\sigma_h \in \Sigma^k_i$, and thus, that:

$$e^*_{i,1}(\omega)(h)\left[\mathrm{EFR}^*_{-i,k+1}\right] = e_{i,1}(\omega)(\sigma_h)\left[\prod_{j\neq i} g_j^{-1}(\mathrm{EFR}^*_{j,k+1}) \times \{V^*_i\}\right]$$
$$\geq \psi_i(e_i(\omega))(\sigma_h) \times \left[\mathrm{Proj}_{\Omega_{-i}\times\mathcal{V}}\left(\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,k}\right)\right] = 1,$$

being the last equality a consequence of $\omega \in SB_i(PI_i) \cap CSBORPI_{i,n}$. Hence, we just learned that conjecture $e^*_{i,1}(\omega)$ assigns probability 1 to $\mathrm{EFR}^*_{-i,k+1}$ at every history reached by partial profiles of strategies in $\mathrm{EFR}^*_{-i,k+1}$. Thus, we conclude that $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,k+1}$. It follows that $e^*_{i,1}(\omega) \in \mathrm{D}^*_{i,n+1}$ and in consequence, that $s^*_i(\omega) \in \mathrm{EFR}^*_{i,n+2}$. ♦

Part (ii). Similarly as done in the proof of Lemma 1, due to technical reasons we will prove first the slightly more general statement: for any player i and any $(n+1)$th-order rationalizable strategy s^*_i, there exists some state $\omega^i \in R_i \cap PI_i \cap SB_i(PI_i) \cap CSBORPI_{i,n}$ that induces s^*_i in the game with commonly known perfect information, i.e., such that $s^*_i(\omega) = s^*_i$. Since we know that the claim is true for $n = 0$ (see proof of Lemma 1), we can proceed inductively: let us suppose that the claim is true for n, and let us prove it for $n+1$. Then, fix player i and $(n+2)$th-order rationalizable strategy s^*_i, and pick conjecture $\mu_i \in \mathrm{C}^*_{i,n+1}$ for which s^*_i is a sequential best-reply. We follow now a constructive process which, despite being similar to that in the proof of part (ii) of Lemma 1 presents some additional technical subtleties. Consider first the following elements:

- First, for each $k = 1, \ldots, n+1$ define:
$$X^k_i(\mu_i) = \{h \in H^{\emptyset}_i | \mu_i(h')[S_{-i}(h)] = 0 \text{ for every } h' < h\} \cap \mathrm{H}^*_{i,k} \setminus \mathrm{H}^*_{i,k+1}.$$
Then, $X^k_i(\mu_i)$ is the set of histories in $\mathrm{H}^*_{i,k} \setminus \mathrm{H}^*_{i,k+1}$ in which μ_i updates without applying the chain rule. It is important to identify set $\bigcup_{k=1}^{n+1} X^k_i(\mu_i)$ because it is telling us in which information sequences consistent with perfect information will we be able to impose beliefs derived from μ_i without spoiling belief updates via the chain rule.[46] It is also useful to identify each particular $X^k_i(\mu_i)$, because it is telling us which the highest degree of higher-order belief in opponents' rationality and perfect information that can be sustained at a given information sequence consistent with perfect information.
- We apply now the induction hypothesis. We know from the latter that for any $k = 0, \ldots, n$, any $h \in X^{k+1}_i(\mu_i)$, any $s^*_{-i} \in \operatorname{supp} \mu_i(h)$ and any $j \neq i$ we can pick some state $\omega^j(s^*_j, k)$ satisfying $s^*_j(\omega^j(s^*_j, k)) = s^*_j$ and $\omega^j(s^*_j, k) \in R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,k}$.
- Fix some arbitrary conditional belief hierarchy $\bar{e}_i$ consistent with $CSBORPI_{i,n+1}$, and pick strategy $\bar{s}_i$ which is a sequential best-reply to $\bar{e}_{i,1}$.

We are going to rely now on the elements introduced above to define some appropriate conditional belief $\hat{\psi}_i \in \Delta^{\Sigma^{\emptyset}_i}(\mathcal{E}_{-i}\times S_{-i}\times\mathcal{V})$. We define it piecewise, depending on the three possible domains each information sequence σ_i can belong to:

- $\sigma_i = \sigma_h$ for some history $h \in X^{k+1}_i(\mu_i)$ and some $k = 0, \ldots, n$. Then, set:
$$\hat{\psi}_i(\sigma_h)[(e_{-i}, s_{-i}, V)] = \begin{cases} \mu_i(h)[s^*_{-i}] & \text{if } V = V^*,\ s^*_{-i} \in \operatorname{supp} \mu_i(h) \\ & \text{and for any } j \neq i, \\ & (e_j, s_j) = (e_j(\omega^j(s^*_j, k)), s_j(\omega^j(s^*_j, k))), \\ \bar{e}_{i,1}(\sigma_i)[(s_{-j}, V)] & \text{otherwise,} \end{cases}$$
for any $(e_{-i}, s_{-i}, V) \in \mathcal{E}_{-i} \times S_{-i} \times \mathcal{V}$. Note that the fact that each $\mu_i(h)$ has finite support implies that every $\hat{\psi}_i(\sigma_h)$ is a well defined element of $\Delta(\mathcal{E}_{-i} \times (S_{-i} \times \mathcal{V})(\sigma_h))$. In addition, we know that by construction it assigns probability 0 to $\mathcal{E}_{-i} \times (S_{-i} \times \mathcal{V})(\sigma_{h'})$ for every $h' \in \bigcup_{k=1}^{n+1} X^k_i(\mu_i)$ such that $h < h'$.
- $\sigma_i = \sigma_h$ for some history $h \notin \bigcup_{k=1}^{n+1} X^k_i(\mu_i)$. Then, pick the unique $h' \in \bigcup_{k=1}^{n+1} X^k_i(\mu_i)$ satisfying that $h' < h$ and $\mu_i(h)'[S^*_{-i}(h)] > 0$ and define $\hat{\psi}_i(\sigma_h)$ applying the chain rule on $\hat{\psi}_i(\sigma_{h'})$.
- $\sigma_i \neq \sigma_h$ for some history $h \in H^{\emptyset}_i$. Then, set $\hat{\psi}_i(\sigma_i) = \psi_i(\bar{e}_i)(\sigma_i)$.

It is then routine to check that $\hat{\psi}_i$ is a well-defined conditional belief in $\Delta^{\Sigma^{\emptyset}_i}(\mathcal{E}_{-i}\times S_{-i}\times\mathcal{V})$.[47] Set then conditional belief hierarchy $\hat{e}_i = \psi_i^{-1}(\hat{\psi}_i)$ and define strategy $\hat{s}_i$ as follows:

$$\hat{s}_i(\sigma_i) = \begin{cases} s^*_i(h) & \text{if } \sigma_i = \sigma_h \text{ for some } h \in H^{\emptyset}_i, \\ \bar{s}_i(\sigma_i) & \text{otherwise,} \end{cases}$$

for any $\sigma_i \in \Sigma_i$. Finally, let $\omega^i = (\omega_{-i}, \hat{e}_i, \hat{s}_i, V^*)$, where ω_{-i} is some arbitrary element in Ω_{-i}. We claim now that ω^i satisfies the requirements we are asking for:

- $s^*_i(\omega^i) = s^*_i$. This follow from construction of $\hat{s}_i$.
- $\omega^i \in CSBORPI_{i,n+1}$. It follows by construction of $\hat{e}_i$. To see it fix first $k \in \{1, \ldots, n\}$ and information sequence $\sigma_i \in \Sigma^k_i$. Now, if $\sigma_i \neq \sigma_h$ for any history $h \in H^{\emptyset}_i$, notice that $\bar{e}_i$ has been chosen ex profeso to guarantee that $\psi_i(\bar{e}_i)(\sigma_i)$ will be assigning full probability to states in the projection of

[45] As mentioned above, in principle we cannot ensure that $e^*_{i,1}(\omega)$ updates beliefs according to the chain rule.

[46] By information sequences consistent with perfect information we mean those $\sigma_i = \sigma_h$ for some history h.

[47] Both common certainty of coherence and update applying the chain rule are satisfied by construction.

$\bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,k}$. Alternatively, if $\sigma_i = \sigma_h$ for some history $h \in H_i^{\emptyset}$, it follows from part (i) of the induction hypothesis that $h \in \mathrm{H}^*_{i,k+1}$. Thus, $\hat{\psi}_i(\sigma_h)$ assigns positive probability only to the projection of states $\omega^j(\cdot, k)$ of the kind obtained via part (ii) of the induction hypothesis, which satisfy that $\omega^j(\cdot, k) \in R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,k}$.

- $\omega^i \in SB_i(PI_i)$. This follow from the fact that $\hat{e}_{i,1}$ assigns full probability to V_i^* whenever possible. The fact that $\omega^i \in PI_i$ is trivially true.
- $\omega^i \in R_i$. It follows from the construction of $\hat{e}_i$ and $\hat{s}_i$ and from (2) that after those information sequences consistent with perfect information, $\hat{s}_i$ is optimal with respect to $\hat{e}_{i,1}$. After information sequences not consistent with perfect information, $\hat{s}_i$ is optimal with respect to $\hat{e}_{i,1}$ because $\bar{s}_i$ is optimal with respect to $\bar{e}_{i,1}$.

Now to properly prove part (ii) in the claim of the lemma, fix player i and history $h \in \mathrm{H}^*_{i,n+2}$. We need to prove that there exist some strategy $s_i \in S_i$ and some state $\omega \in \bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n+1}$ such that $\sigma_h \subseteq v_i(\omega)$ and $z((s_{-i}(\omega); s_i), v(\omega)) > h$. Fix $s^*_{-i} \in \mathrm{EFR}^*_{i,n+2}$ and $s^*_i \in S^*_i$ such that $z^*(s^*) > h$, and for each $j \neq i$, pick $\omega^j \in R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n+1}$ such that $s^*_j(\omega^j) = s^*_j$. Then, pick arbitrary conditional belief hierarchy e_i and define ω as follows $\omega = ((e_j(\omega^j), s_j(\omega^j))_{j\neq i}, (e_i, s_i), V^*)$, where $s_i = f_i(s^*_i)$ for some map f_i as defined in Remark 1. We have then that: $\sigma_h \subseteq v_i(\omega)$, that $\omega \in \bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n+1}$ and that $z((s_{-i}(\omega); s_i), v(\omega)) = z^*(s^*) > h$. Thus, we conclude that $\sigma_h \in \Sigma_i^{n+1}$. ■

B.2.2. Second step

The following result easily follows from the two lemmas in Appendix B.2.1.

Corollary 1. *For any player i it holds that:*

$$\{\sigma_i \in \Sigma_i^{\emptyset} | v_{\sigma_i} < z_{\mathcal{I}}\} \subseteq \bigcap_{n=0}^{\bar{n}-1} \Sigma_i^n.$$

Proof. Fix player i, $n \in \{0, \ldots, \bar{n}-1\}$ and information sequence σ_i preceding the backward induction outcome. Obviously, there exists some history h in v_{σ_i} such that $h < z_{\mathcal{I}}$. Hence, in particular, $h \in \mathrm{H}^*_{i,n+1}$ and thus, we know from part (ii) of Lemma 2 that $\sigma_h \in \Sigma_i^n$. Then, by definition there exist some state $\omega \in \bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n}$ and some strategy $s_i \in S_i$ such that $\sigma_h \subseteq v_i(\omega)$ and $z(s_{-i}(\omega), v_{-i}(\omega), s_i, v_i(\omega)) > h$. Pick now arbitrary information structure such that $\sigma_i \subseteq V_i'$, and define strategy s_i' as follows:

$$s_i(\sigma_i') = \begin{cases} s_i(\sigma_{h'}) & \text{if } h' \leq h \text{ for some } h' \in v_{\sigma_i'}, \\ s_i(\sigma_i') & \text{otherwise,} \end{cases}$$

for any $\sigma_i' \in \Sigma_i$. Set now $\hat{\omega} = (\omega_{-i}, v_{-i}(\omega), e_i(\omega), s_i', V_i')$. Everything we need holds by construction: we have first that $\hat{\omega} \in \bigcap_{j\neq i} R_j \cap PI_j \cap SB_j(PI_j) \cap CSBORPI_{j,n}$, second, that $\sigma_i \subseteq v_i(\hat{\omega})$ and finally, that $z(s_{-i}(\hat{\omega}), v_{-i}(\hat{\omega}), s_i', v_i(\hat{\omega})) > h$ for $h \in v_{\sigma_i}$. Thus, $\sigma_i \in \Sigma_i^n$. ■

B.2.3. Third step

Finally, we present the intermediate result that summarizes how a player infers her exact location in the game tree and optimally reacts to it.

Lemma 3. *For any player i that at every information sequence consistent with the backward induction path believes her opponent to play rationalizable strategies, following the backward induction path is the only optimal choice; i.e.,*

$$R_i \cap \bigcap_{\sigma_i : v_{\sigma_i} < z_{\mathcal{I}}} B_i(s^*_{-i} \in \mathrm{EFR}^*_{-i} | \sigma_i) \subseteq IC_i.$$

Proof. For simplicity, during the proof we write $\sigma_i < z_{\mathcal{I}}$ to mean that $v_{\sigma_i} < z_{\mathcal{I}}$. Then, we are going to check first that, under the assumptions of the claim, at every information sequence consistent with the backward induction path, the corresponding player is able infer in which particular history of the last information set of the information sequence she is at. That is, for any player i, any state ω and any information sequence $\sigma_i \in \Sigma_i(s_i(\omega))$ such that $\sigma_i < z_{\mathcal{I}}$ we have that,

$$B_i(s^*_{-i} \in \mathrm{EFR}^*_{-i} | \sigma_i) \subseteq B_i(s^*_{-i} \in S^*_{-i}(h) | \sigma_i) \tag{3}$$

for the unique $h \in v_{\sigma_i}$ such that $h < z_{\mathcal{I}}$. To see it, proceed by contradiction and suppose that there exists some alternative $h' \in v_{\sigma_i}$ different from h that is reached by some partial profile of strategies $\hat{s}^*_{-i} \in \mathrm{EFR}^*_{-i}$.[48] Since $h < z_{\mathcal{I}}$ we can also pick some rationalizable strategy $\hat{s}^*_i$ that reaches h.[49] Notice though that the fact that $h, h' \in v_{\sigma_i}$ implies that $S^*_i(h) = S^*_i(h')$ and thus, we can conclude that $\hat{s}^*$ is a profile of rationalizable strategies such that $z^*(\hat{s}^*) > h'$ and therefore, such that $z^*(\hat{s}^*) \neq z_{\mathcal{I}}$. We reached a contradiction then, because Theorem 4 by Battigalli (1997) shows that $z^*(s^*) = z_{\mathcal{I}}$ for any profile of rationalizable strategies s^*. Hence, we conclude that $\mathrm{EFR}^*_{-i} \subseteq S^*_{-i}(h)$, from which (3) follows immediately.

Maintain player i, state $\omega \in \bigcap_{\sigma_i < z_{\mathcal{I}}} B_i(s^*_{-i} \in \mathrm{EFR}^*_{-i} | \sigma_i)$ and information sequence $\sigma_i \in \Sigma_i(s_i(\omega))$ such that $\sigma_i < z_{\mathcal{I}}$ fixed. We check next that $U_i(e_{i,1}(\omega), s_i(\omega) | \sigma_i) = u_i(z_{\mathcal{I}})$. Note first that (3) allows for the following characterization of conditional expected utility:

$$U_i(e_{i,1}(\omega), s_i | \sigma_i) = \sum_{V_i \in \mathcal{V}_i} \sum_{s^*_{-i} \in S^*_{-i}} e_{i,1}(\omega)(\sigma_i) \left[\prod_{j\neq i} g_j^{-1}(s^*_j) \times \{V_i\} \right] \cdot u^*_i(s^*_{-i}; g_i(s_i, V_i) | h), \tag{4}$$

for any $s_i \in S_i$. Then, to see that $U_i(e_{i,1}(\omega), s_i(\omega) | \sigma_i) = u_i(z_{\mathcal{I}})$ proceed again by contradiction and consider the two possible situations:

- Suppose that $U_i(e_{i,1}(\omega), s_i(\omega) | \sigma_i) > u_i(z_{\mathcal{I}})$. Then it follows from (4) that there must exist some information structure V_i for player i and some partial profile of rationalizable strategies of i's opponents $\hat{s}^*_{-i}$ that reaches h and such that $u^*_i(\hat{s}^*_{-i}; g_i(s_i(\omega), V_i) | h) > u_i(z_{\mathcal{I}})$. Pick now arbitrary conjecture $\mu_i \in \mathrm{C}^*_i$ and define $\hat{\mu}_i$ as follows:

$$\hat{\mu}_i(h') = \begin{cases} 1_{\{\hat{s}^*_{-i}\}} & \text{if } \hat{s}^*_{-i} \in S^*_{-i}(h'), \\ \mu_i(h') & \text{otherwise,} \end{cases}$$

  for any $h' \in H_i^{\emptyset}$. It is routine to check that $\hat{\mu}_i \in \mathrm{C}^*_i$. Remember now that we know from Theorem 4 by Battigalli (1997) that $z^*(s^*) = z_{\mathcal{I}}$ for any profile of rationalizable strategies s^* and thus, that every rationalizable strategy s^*_i of player i reaches h, the unique history in v_{σ_i} that precedes $z_{\mathcal{I}}$. Then, since $h < z_{\mathcal{I}}$ and every sequential best-reply to $\hat{\mu}_i$ is rationalizable, we know that for any $\hat{s}^*_i \in BR^*_i(\hat{\mu}_i)$ we have both that $U^*_i(\hat{\mu}_i, \hat{s}^*_i | h) = u_i(z_{\mathcal{I}})$ and $\hat{s}^*_i \in \operatorname{argmax}_{s^*_i \in S^*_i} U^*_i(\hat{\mu}_i, s^*_i | h)$. Notice that it follows from the latter that $U^*_i(\hat{\mu}_i, \hat{s}^*_i | h) \geq U^*_i(\hat{\mu}_i, g_i(s_i(\omega), V_i) | h)$, and that:

$$U^*_i(\hat{\mu}_i, g_i(s_i(\omega), V_i) | h) = u^*_i(\hat{s}^*_{-i}, g_i(s_i(\omega), V_i) | h) > u_i(z_{\mathcal{I}}).$$

  Thus, we conclude that $U^*_i(\hat{\mu}_i, \hat{s}^*_i | h) > u_i(z_{\mathcal{I}})$ and $U^*_i(\hat{\mu}_i, \hat{s}^*_i | h) = u_i(z_{\mathcal{I}})$. Hence, we reach a contradiction.

48 I.e., such that $\hat{s}^*_{-i} \in \mathrm{EFR}^*_{-i} \cap S^*_{-i}(h') \neq \emptyset$.

49 I.e., $\hat{s}^*_i \in \mathrm{EFR}^*_i \cap S^*_i(h)$.

- Suppose instead that $U_i(e_{i,1}(\omega), s_i(\omega)|\sigma_i) < u_i(z_{\mathcal{I}})$. Define then strategy $\hat{s}_i$ as follows:

$$\hat{s}_i(\sigma_i') = \begin{cases} \alpha_{h'} & \text{if } \sigma_i' \geq \sigma_i, h' < z_{\mathcal{I}} \text{ and } h' \in v_{\sigma_i'}, \\ s_i(\omega)(\sigma_i) & \text{otherwise,} \end{cases}$$

for any $\sigma_i' \in \Sigma_i$. Clearly, $U_i(e_{i,1}(\omega), \hat{s}_i|\sigma_i) = u_i(z_{\mathcal{I}})$: simply notice that for any partial profile of i's opponents rationalizable strategies $\hat{s}^*_{-i}$ we have that $z^*(\hat{s}^*_{-i}; g_i(\hat{s}_i, V_i)|h) = z_{\mathcal{I}}$ for any player i's information structure V_i. It follows then $U_i(e_{i,1}(\omega), s_i(\omega)|\sigma_i) < U_i(e_{i,1}(\omega), \hat{s}_i|\sigma_i)$, and therefore, $s_i(\omega)$ is not optimal w.r.t. $e_{i,1}(\omega)$ at σ_i. Hence, we reach again a contradiction, because $\sigma_i \in \Sigma_i(s_i(\omega))$ and $\omega \in R_i$.

Thus, it must be true $U_i(e_{i,1}(\omega), s_i(\omega)|\sigma_i) = u_i(z_{\mathcal{I}})$, and therefore, that $s_{\sigma_i}(\omega) = \alpha_h$ for the unique $h \in v_{\sigma_i}$ such that $h < z_{\mathcal{I}}$. In consequence, $\omega \in \bigcap_{h<z_{\mathcal{I}}} \bigcap_{\sigma_i \in \Sigma_i(s_i(\omega)): h \in v_{\sigma_i}} [s_{\sigma_i} = \alpha_h]$, i.e., $\omega \in IC_i$. ■

References

Aumann, Robert J., 1995. Backwards induction and common knowledge of rationality. Games Econom. Behav. 8, 6–19.

Aumann, Robert J., 1998. On the centipede game. Games Econom. Behav. 23, 97–105.

Baltag, Alexandru, Smets, Sonja, Zvesper, Jonathan A., 2009. Keep hoping for rationality: a solution for the backward induction paradox. Synthese 169, 201–303.

Battigalli, Pierpaolo, 1996. Strategic rationality ordorder and the best rationalization principle. Games Econom. Behav. 13, 178–200.

Battigalli, Pierpaolo, 1997. On rationalizability in extensive form games. J. Econom. Theory 74, 40–61.

Battigalli, Pierpaolo, 2003. Rationalizability in infinite, dynamic games with incomplete information. Res. Econ. 53, 1–38.

Battigalli, Pierpaolo, Friedenberg, Amanda, 2012. Forward induction reasoning revisited. Theor. Econ. 7, 57–98.

Battigalli, Pierpaolo, Siniscalchi, Marciano, 1999. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. J. Econom. Theory 88, 188–230.

Battigalli, Pierpaolo, Siniscalchi, Marciano, 2002. Strong belief and forward induction reasoning. J. Econom. Theory 106, 356–391.

Battigalli, Pierpaolo, Siniscalchi, Marciano, 2007. Interactive epistemology in games with payoff uncertainty. Res. Econ. 61, 165–184.

Ben Porath, Elchanan, 1997. Rationality, Nash equilibrium and backward induction in perfect information games. Rev. Econom. Stud. 64, 22–46.

Bonanno, Giacomo, 2013. A dynamic epistemic characterization of backward induction without counterfactuals. Games Econom. Behav. 78, 31–43.

Brandenburger, Adam, Dekel, Eddie, 1993. Hierarchies of beliefs and common knowledge. J. Econom. Theory 59, 189–198.

Dekel, Eddie, Siniscalchi, Marciano, 2015. Epistemic game theory. In: Young, Peyton H., Zamir, Shmuel (Eds.), Handbook of Game Theory with Economic Applications. Elsevier, pp. 619–702.

Di Tillio, Alfredo, Halpern, Joseph Y., Samet, Dov, 2014. Conditional belbelief types. Games Econom. Behav. 87, 253–268.

Harsanyi, John C., 1967–1968. Games with incomplete information played by 'Bayesian' players, I–III. Manage. Sci. 14, 159–182, 320–334, 486–502.

Heifetz, Aviad, Perea, Andrés, 2014. On the outcome equivalence of backward induction and extensive form rationalizability. Internat. J. Game Theory 44, 37–59.

Myerson, Roger B., 1986. Multistage games with communication. Econometrica 54, 323–358.

Osborne, Martin J., Rubinstein, Ariel, 1994. A Course in Game Theory. MIT Press Books.

Pearce, David G., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica 52, 1029–1050.

Penta, Antonio, Bakward induction reasoning in games with incomplete information, University of Winsonsin-Madison, Mimeo, 2011.

Penta, Antonio, 2012. Higher order uncertainty and information: static and dynamic games. Econometrica 80, 631–660.

Penta, Antonio, 2015. Robust dynamic implementation. J. Econom. Theory 160, 280–316.

Perea, Andres, 2007. Epistemic foundations for backward induction: an overview. In: van Benthem, J., Gabbay, D., Löwe, B. (Eds.), Interactive Logic. Proceedings of the 7th Augustus De Morgan Workshop. In: Texts in Logic and Games, vol. 1, Amsterdam University Press, pp. 159–193.

Perea, Andrés, 2014. Belief in the opponent's future rationality. Games Econom. Behav. 81, 235–254.

Reny, Philipp J., 1992. Backward induction, normal form perfection and explicable equilibria. Econometrica 60, 627–649.

Renyi, Alfred, 1955. On a new axiomatic theory of probability. Acta Math. Hungar. 6, 285–335.

Rosenthal, Robert W., 1981. Games of perfect information, predatory pricing and chain-store paradox. J. Econom. Theory 25, 92–100.

Samet, Dov, 1996. Hypothetical knowledge and games with perfect information. Games Econom. Behav. 17, 230–251.

Samet, Dov, 2013. Common belief of rationality in games with perfect information. Games Econom. Behav. 79, 192–200.

Shimoji, Makoto, Watson, Joel, 1998. Conditional dominance, rationalizability, and game forms. J. Econom. Theory 83, 161–1995.

Tan, Tommy Chin-Chiu, Werlang, Sérgio Ribeiro da Costa, 1988. The Bayesian foundation of solution concepts of games. J. Econom. Theory 45, 370–391.