Online appendix to "Forward Induction in a Backward Inductive Manner"

Martin Meier*and Andrés Perea†

This version: December 2023

1 Outline

The outline of this online appendix is as follows: In Section 2 we provide an additional example that illustrates the forward and backward rationalizability procedure. The main difference with the example provided in Section 3.4 of the paper is that the dynamic game is larger, and displays unobserved past choices. In Section 3 we present an epistemic characterization of forward and backward rationalizability, by means of *common backward strong belief in rationality.* In Section 4 we investigate two alternative combinations of forward and backward induction reasoning. In the first concept we start by applying the strong rationalizability procedure, followed by the backwards rationalizability procedure, whereas this order is reversed in the second concept. We compare both concepts to forward and backward rationalizability. In Section 5 we provide the proofs of Section 3.

2 Example

We will now present a new example, with many information sets and unobserved past choices, to illustrate the concept of forward and backward rationalizability. Consider the dynamic game from Figure 1. Note that the information sets h_2, h_3 and h_5 are non-trivial.

In the dynamic game there are six time periods: Period 1 with information set h_1, period 2 with information set h_2, period 3 with information sets h_3 and h_4, period 4 with information sets h_5

E-mail: M.Meier@bath.ac.uk *Affiliation:* University of Bath, Department of Economics, United Kingdom, and IHS Vienna, Austria.

†*E-mail:* a.perea@maastrichtuniversity.nl *Affiliation:* Maastricht University, EpiCenter and Department of Quantitative Economics, The Netherlands.

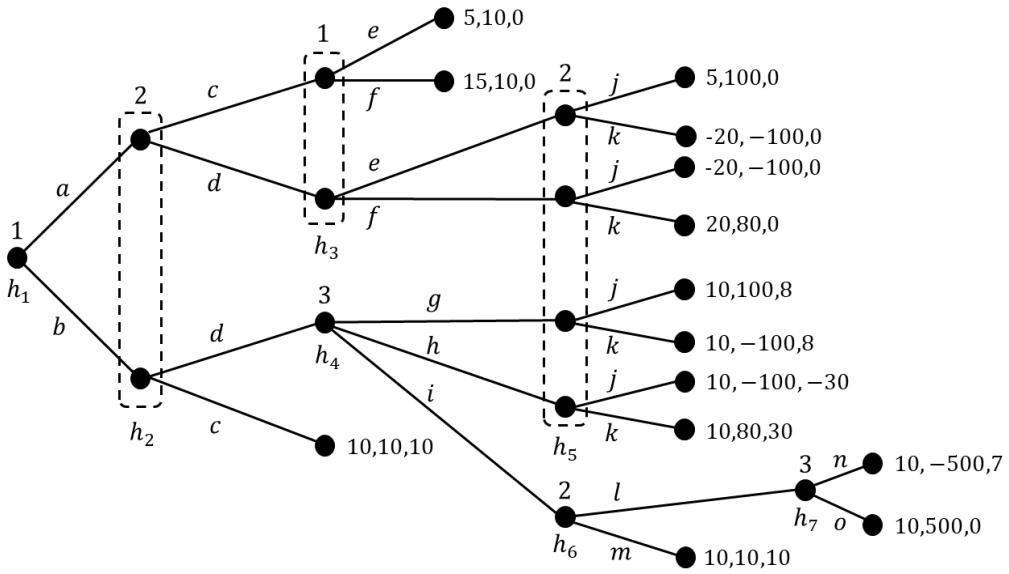

Figure 1: A dynamic game with non-trivial information sets

and h_6, period 5 with information set h_7, and period 6 with no information sets but only terminal histories. To run the forward and backward rationalizability procedure, we thus start at period 5.

Period 5. At information set h_7, player 3's strategy (i, o) is never optimal for any conditional belief, whereas the strategy (i, n) is. Thus, only player 3's strategies g, h and (i, n) are optimal for some conditional belief vector from period 5 onwards. As such,

$$S_3^{5.1} = \{g, h, (i, n)\}.$$

Moreover, we have that $S_1^{5.1} = S_1$ and $S_2^{5.1} = S_2$. Hence,

$$\begin{aligned} B_1^{5.2} &= \{b_1 \in B_1 \mid b_1 \text{ strongly believes } S_2 \times \{g, h, (i, n)\}\} \\ &= \{b_1 \in B_1 \mid b_1(h_1)(S_2 \times \{g, h, (i, n)\}) = b_1(h_3)(S_2 \times \{g, h, (i, n)\}) = 1\}, \end{aligned}$$

and

$$\begin{aligned} B_2^{5.2} &= \{b_2 \in B_2 \mid b_2 \text{ strongly believes } S_1 \times \{g, h, (i, n)\}\} \\ &= \{b_2 \in B_2 \mid b_2(h_2)(S_1 \times \{g, h, (i, n)\}) = \\ & b_2(h_5)(S_1 \times \{g, h, (i, n)\}) = b_2(h_6)(\{b\} \times \{(i, n)\}) = 1\}. \end{aligned}$$

Finally, $B_3^{5.2} = B_3$. As no further restrictions can be derived after this step, the procedure from Period 5 onwards is hereby complete.

Period 4. At h_5, both choices j and k of player 2 are optimal for some conditional belief vector in $B_2^{5.2}$.

At h_6, player 2's conditional belief must be part of $B_2^{5.2}$ above, and hence player 2 must at h_6 assign probability 1 to player 3 choosing n at h_7. But then, player 2's choice l cannot be optimal at h_6 for any conditional belief vector in $B_2^{5.2}$, whereas the choice m can. Hence,

$$S_2^{4.1} = \{(c,(d,j,m),(d,k,m)\}.$$

Moreover, $S_1^{4.1} = S_1$ and $S_3^{4.1} = S_3^{5.1} = \{g, h, (i,n)\}$. We then have that

$$\begin{aligned} B_1^{4.2} &= \{b_1 \in B_1^{5.2} \mid b_1 \text{ strongly believes } \{(c,(d,j,m),(d,k,m)\} \times \{g,h,(i,n)\}\} \\ &= \{b_1 \in B_1 \mid b_1(h_1)(\{(c,(d,j,m),(d,k,m)\} \times \{g,h,(i,n)\}) \\ &= b_1(h_3)(\{(c,(d,j,m),(d,k,m)\} \times \{g,h,(i,n)\}) = 1\} \end{aligned}$$

and

$$\begin{aligned} B_3^{4.2} &= \{b_3 \in B_3^{5.2} \mid b_3 \text{ strongly believes } S_1 \times \{(c,(d,j,m),(d,k,m)\}\} \\ &= \{b_3 \in B_3 \mid b_3(h_4)(\{b\} \times \{(d,j,m),(d,k,m)\} = 1\}, \end{aligned}$$

whereas $B_2^{4.2} = B_2^{5.2}$.

As no further restrictions can be derived after this step, the procedure from Period 4 onwards is hereby complete.

Period 3. At h_3, both of player 1's strategies (a,e) and (a,f) can be optimal for some conditional belief vector in $B_1^{4.2}$. Thus,

$$S_1^{3.1} = S_1.$$

At h_4, player 3's conditional belief must be part of $B_3^{4.2}$, and hence player 3 must believe that player 2 will choose m at h_6. As such, player 3 expects the utility 10 by choosing i at h_4, whereas g gives him only 8. This renders the strategy g suboptimal for player 3 at h_4. Hence,

$$S_3^{3.1} = \{h, (i,n)\}.$$

As player 2 is not active at any information in Period 3, we have that

$$S_2^{3.1} = S_2^{4.1} = \{(c,(d,j,m),(d,k,m)\}.$$

Thus,

$$\begin{aligned} B_1^{3.2} &= \{b_1 \in B_1^{4.2} \mid b_1 \text{ strongly believes } \{(c,(d,j,m),(d,k,m)\} \times \{h,(i,n)\}\} \\ &= \{b_1 \in B_1 \mid b_1(h_1)(\{(c,(d,j,m),(d,k,m)\} \times \{h,(i,n)\}) \\ &= b_1(h_3)(\{(c,(d,j,m),(d,k,m)\} \times \{h,(i,n)\}) = 1\}, \end{aligned}$$

and

$$B_2^{3.2} = \{b_2 \in B_2^{4.2} \mid b_2 \text{ strongly believes } S_1 \times \{h, (i,n)\}$$
$$= \{b_2 \in B_2 \mid b_2(h_2)(S_1 \times \{h, (i,n)\}) =$$
$$b_2(h_5)((\{(a,e),(a,f)\} \times \{h, (i,n)\}) \cup (\{b\} \times \{h\}))$$
$$= b_2(h_6)(\{b\} \times \{(i,n)\}) = 1\},$$

whereas

$$B_3^{3.2} = \{b_3 \in B_3^{4.2} \mid b_3 \text{ strongly believes } S_1 \times \{(c,(d,j,m),(d,k,m)\}\}$$
$$= \{b_3 \in B_3 \mid b_3(h_4)(\{b\} \times \{(d,j,m),(d,k,m)\} = 1\}.$$

Since we can derive no further restrictions after this step, this concludes the procedure from Period 3 onwards.

Period 2. At the information set h_2, each of player 2's strategies in $S_2^{3.1}$ can be optimal for some conditional belief vector in $B_2^{3.2}$. Hence, $S_2^{2.1} = S_2^{3.1}$. As a consequence, the strategy sets and the sets of conditional belief vectors for each of the players remain the same as in Period 3.

Period 1. *Round 1.* At h_1, player 1 can guarantee utility 10 by choosing b. Since the strategy (a,e) yields him at most 5, we conclude that the strategy (a,e) is suboptimal for player 1 at h_1. Therefore,

$$S_1^{1.1} = \{(a,f), b\}.$$

Moreover,

$$S_2^{1.1} = S_2^{3.1} = \{(c,(d,j,m),(d,k,m)\}$$

and

$$S_3^{1.1} = S_3^{3.1} = \{h, (i,n)\}.$$

Round 2. Thus, $B_1^{1.2} = B_1^{3.2}$, and

$$B_2^{1.2} = \{b_2 \in B_2^{3.2} \mid b_2 \text{ strongly believes } \{(a,f), b\} \times \{h, (i,n)\}$$
$$= \{b_2 \in B_2 \mid b_2(h_2)(\{(a,f), b\} \times \{h, (i,n)\}) =$$
$$b_2(h_5)((\{(a,f)\} \times \{h, (i,n)\}) \cup (\{b\} \times \{h\}))$$
$$= b_2(h_6)(\{b\} \times \{(i,n)\}) = 1\},$$

whereas

$$B_3^{1.2} = \{b_3 \in B_3^{3.2} \mid b_3 \text{ strongly believes } \{(a,f), b\} \times \{(c,(d,j,m),(d,k,m)\}\}$$
$$= \{b_3 \in B_3 \mid b_3(h_4)(\{b\} \times \{(d,j,m),(d,k,m)\} = 1\}.$$

Note that at h_5, player 2 can only assign positive probability to the opponents' strategy combinations in $(\{(a,f)\} \times \{h,(i,n)\}) \cup (\{b\} \times \{h\})$. Thus, at h_5 player 2 can only assign positive probability to the second and fourth history. But then, player 2 should definitely choose k at h_5, and we thus have that

$$S_2^{1.2} = \{(c,(d,k,m)\}.$$

Moreover, $S_1^{1.2} = S_1^{1.1}$ and $S_3^{1.2} = S_3^{1.1}$.

Round 3. As such,

$$\begin{gathered} B_1^{1.3} = \{b_1 \in B_1^{1.2} \mid b_1 \text{ strongly believes } \{(c,(d,k,m)\} \times \{h,(i,n)\}\} \\ = \{b_1 \in B_1 \mid b_1(h_1)(\{(c,(d,k,m)\} \times \{h,(i,n)\}) = b_1(h_3)(\{(c,(d,k,m)\} \times \{h,(i,n)\}) = 1\} \end{gathered}$$

and

$$\begin{gathered} B_3^{1.3} = \{b_3 \in B_3^{1.2} \mid b_3 \text{ strongly believes } \{(a,f),b\} \times \{(c,(d,k,m)\}\} \\ = \{b_3 \in B_3 \mid b_3(h_4)(\{b\} \times \{(d,k,m)\} = 1\}, \end{gathered}$$

whereas $B_2^{1.3} = B_2^{1.2}$.

Hence, at h_1 player 1 believes that player 2 chooses either c or (d,k,m). But then, by choosing (a,f) player 1 believes to obtain at least 15. Since by choosing (a,e) he believes to get at most 5, and by choosing b he believes to get 10, the strategies (a,e) and b are suboptimal for player 1 at h_1. Thus,

$$S_1^{1.3} = \{(a,f)\}.$$

At h_4, player 3 believes that player 2 will choose strategy (d,k,m). But then, the only optimal strategy for player 3 at h_4 is h, and hence

$$S_3^{1.3} = \{h\}.$$

Moreover, $S_2^{1.3} = S_2^{1.2}$.

Round 4. We have that

$$\begin{gathered} B_1^{1.4} = \{b_1 \in B_1^{1.3} \mid b_1 \text{ strongly believes } \{(c,(d,k,m)\} \times \{h\}\} \\ = \{b_1 \in B_1 \mid b_1(h_1)(\{(c,(d,k,m)\} \times \{h\}) = b_1(h_3)(\{(c,(d,k,m)\} \times \{h\}) = 1\} \end{gathered}$$

$$\begin{gathered} B_2^{1.4} = \{b_2 \in B_2^{1.3} \mid b_2 \text{ strongly believes } \{(a,f)\} \times \{h\} \\ = \{b_2 \in B_2 \mid b_2(h_2)(\{(a,f)\} \times \{h\}) = b_2(h_5)(\{(a,f) \times \{h\}) = 1\}, \end{gathered}$$

whereas

$$B_3^{1.4} = \{b_3 \in B_3^{1.3} \mid b_3 \text{ strongly believes } \{(a,f)\} \times \{(c,(d,k,m)\}\} = B_3^{1.3}.$$

Hence, at h_2 player 2 must believe that player 1 chooses (a,f). But then, among the strategies in $S_2^{1.3}$, the only optimal strategy for player 2 at h_2 is (d,k,m). Hence,

$$S_2^{1.4} = \{(d,k,m)\}$$

whereas $S_1^{1.4} = S_1^{1.3} = \{(a,f)\}$ and $S_3^{1.4} = S_3^{1.3} = \{h\}$.

Round 5. We finally have that

$$\begin{aligned} B_1^{1.5} &= \{b_1 \in B_1^{1.4} \mid b_1 \text{ strongly believes } \{(d,k,m)\} \times \{h\}\} \\ &= \{b_1 \in B_1 \mid b_1(h_1)(\{(d,k,m)\} \times \{h\}) = b_1(h_3)(\{(d,k,m)\} \times \{h\}) = 1\} \end{aligned}$$

and

$$\begin{aligned} B_2^{1.5} &= \{b_2 \in B_2^{1.4} \mid b_2 \text{ strongly believes } \{(a,f)\} \times \{h\} \\ &= \{b_2 \in B_2 \mid b_2(h_2)(\{(a,f)\} \times \{h\}) = b_2(h_5)(\{(a,f) \times \{h\}) = b_2(h_6)(\{b\} \times \{(i,n)\}) = 1\}, \end{aligned}$$

whereas

$$\begin{aligned} B_3^{1.5} &= \{b_3 \in B_3^{1.4} \mid b_3 \text{ strongly believes } \{(a,f)\} \times \{(d,k,m)\}\} \\ &= \{b_3 \in B_3 \mid b_3(h_4)(\{b\} \times \{(d,k,m)\} = 1\}. \end{aligned}$$

This is where the procedure terminates. In particular, we see that the unique forward and backward rationalizable strategies for the players are (a,f), (d,k,m) and h, respectively.

3 Epistemic Characterization

In this section we investigate what the concept of forward and backward rationalizability entails in terms of reasoning. To this purpose, we offer epistemic conditions on the players' belief hierarchies such that the optimal strategies under these belief hierarchies are precisely the forward and backward rationalizable strategies. Before doing so, we first recall the definition of a (universal) type space for dynamic games, and subsequently formalize the notion of strong belief and optimal choice for types in a type space.

3.1 Type Space

The epistemic conditions we introduce will impose restrictions on the belief hierarchies that the players may have. Such belief hierarchies may conveniently be encoded by means of *types* in a type space. To formalize a type space, we need the following definition and pieces of notation. A topological space $(X, \mathcal{O})$ is called *Polish* if it is separable and completely metrizable. By $\Sigma(X)$ we denote the Borel σ-algebra on X, that is, the smallest σ-algebra that contains all open sets, whereas $\Delta(X)$ denotes the set of all probability measures on $(X, \Sigma(X))$. We endow $\Delta(X)$ with the smallest topology $\mathcal{O}_{\Delta(X)}$ such that each of the sets $\{\mu \in \Delta(X) | \int_X f d\mu \in O\}$ is open in $\Delta(X)$, where f runs over all bounded continuous functions $f : X \to \mathbf{R}$ and O runs over all open subsets of the reals. By Kechris (1995), Theorem 17.23, $(\Delta(X), \mathcal{O}_{\Delta(X)})$ is again a Polish space. We then consider $\Delta(X)$ as a measurable space that is endowed with the Borel σ-algebra (generated by $\mathcal{O}_{\Delta(X)}$). It is a well-known fact that a continuous map between two topological spaces is measurable if both of these spaces are endowed with their respective Borel σ-algebras.

Definition 3.1 (Type space) *A **type space** $\mathcal{T} = ((T_i, \mathcal{O}_i), \beta_i)_{i \in I}$ specifies, for every player i,*

(a) a Polish type space $(T_i, \mathcal{O}_i)$, and

(b) a continuous belief mapping β_i, which assigns to every type $t_i \in T_i$ and information set $h \in H_i$ a probabilistic belief $\beta_i(t_i, h) \in \Delta(S_{-i}(h) \times T_{-i})$.

Moreover, the types must satisfy Bayesian updating whenever possible, that is, for every player i, every type $t_i \in T_i$, and every two information sets $h, h' \in H_i$ where h' follows h and $\beta_i(t_i, h)(S_{-i}(h') \times T_{-i}) > 0$, we have that

$$\beta_i(t_i, h')(\{s_{-i}\} \times E_{-i}) = \frac{\beta_i(t_i, h)(\{s_{-i}\} \times E_{-i})}{\beta_i(t_i, h)(S_{-i}(h') \times T_{-i})}$$

for every $s_{-i} \in S_{-i}(h')$ and every $E_{-i} \in \Sigma(T_{-i})$.

For our epistemic characterization we need to work with a *universal* type space. To explain what it is, we must first introduce the notion of a *type morphism.*

Definition 3.2 (Type morphism) *Consider two type spaces $\mathcal{T} = ((T_i, O_i), \beta_i)_{i \in I}$ and $\mathcal{T}' = ((T_i', O_i'), \beta_i')_{i \in I}$. A **type morphism** from $\mathcal{T}$ to $\mathcal{T}'$ is a tuple $(f_i)_{i \in I}$ of continuous functions $f_i : T_i \to T_i'$ such that, for every player i, every type $t_i \in T_i$, and every information set $h \in H_i$ we have that*

$$\beta_i'(f_i(t_i), h)(\times_{j \neq i}(\{s_j\} \times E_j')) = \beta_i(t_i, h)(\times_{j \neq i}(\{s_j\} \times f_j^{-1}(E_j'))$$

for every opponents' strategy combination $(s_j)_{j \neq i} \in S_{-i}(h)$ and every measurable set $\times_{j \neq i} E_j' \subseteq \times_{j \neq i} T_j'$ of opponents' type combinations.

A type space is then called *universal* if every other type space can be uniquely embedded into it by means of a type morphism.

Definition 3.3 (Universal type space) *A type space $\mathcal{T}$ is* ***universal***[1] *if for every other type space $\mathcal{T}'$ there is a unique type morphism from $\mathcal{T}'$ to $\mathcal{T}$.*

It turns out that every two universal type spaces are isomorphic. As such, we can speak about *the* universal type space. Battigalli and Siniscalchi (1999), Guarino (2022) and Fukuda (2023) have shown that we can always construct a universal type space in our setting.

3.2 Strong Belief

Consider a type space $\mathcal{T} = ((T_i, \mathcal{O}_i), \beta_i)_{i \in I}$, a type $t_i \in T_i$ and an event $E_{-i} \in \Sigma(S_{-i} \times T_{-i})$. Following Battigalli and Siniscalchi (2002), the type t_i is said to *strongly believe* the event E_{-i} if it assigns probability 1 to the event whenever possible. That is,

$$\beta_i(t_i, h)(E_{-i}) = 1 \text{ at all } h \in H_i \text{ where } E_{-i} \cap (S_{-i}(h) \times T_{-i}) \neq \emptyset.$$

3.3 Optimal Choice

Consider a type $t_i \in T_i$, a strategy $s_i \in S_i$ and an information set $h \in H_i(s_i)$ that can possibly be reached by s_i. Then, we denote by

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

the *expected utility* induced by s_i at h for the type t_i. The strategy s_i is *optimal* for the type t_i at information set h if $u_i(s_i, t_i, h) \geq u_i(s_i', t_i, h)$ for all other strategies $s_i' \in S_i(h)$. For a given period m, we say that strategy s_i is optimal for the type t_i *from period m onwards* if for every period $\tau \geq m$, and every information set $h \in H_i(s_i) \cap H^\tau$, the strategy s_i is optimal for type t_i at h.[2] For a given set of types $\hat{T}_i \in \Sigma(T_i)$, we denote by

$$(S_i \times \hat{T}_i)^{rat,m} := \{(s_i, t_i) \in S_i \times \hat{T}_i \mid s_i \text{ is optimal for } t_i \text{ from period } m \text{ onwards}\}$$

the event that player i chooses rationally from period m onwards and that i's type belongs to $\hat{T}_i$.

The following result states that the event of choosing rationally from a certain period onwards is always a "well-behaved" set.

[1] In the literature, such type spaces are sometimes called *terminal.*

[2] Note that if $h \notin H_i(s_i) \cap H^\tau$ for all $\tau \geq m$, then s_i is (vacuously) optimal for every type of player i from period m onwards.

Lemma 3.1 (Rationality is a measurable event) *Suppose that $\hat{T}_i$ is a closed (measurable) subset of T_i. Then, the set $(S_i \times \hat{T}_i)^{rat,m}$ is a closed (measurable) subset of $S_i \times T_i$.*

This result will be important for guaranteeing that the epistemic conditions below are all well-defined. It will also play a key role in the proof of our epistemic characterization.

3.4 Epistemic Characterization

The epistemic conditions we impose on the players' types are as follows: First, we focus on the last period M where players have to move. A player must $(M.1)$ strongly believe in the event that every opponent chooses rationally from period M onwards, $(M.2)$ strongly believe in the event that every opponent chooses rationally from period M onwards and that every opponent satisfies $(M.1)$, and so on. These conditions together yield *common backward strong belief in rationality from period M onwards.* We refer to this event as (M). In fact, since every player moves at most once at period M, event (M) is equivalent to *common belief in rationality at period M.*

We then move to period $M-1$. A player must $(M-1.1)$ strongly believe in the event that every opponent chooses rationally from period $M-1$ onwards and that every opponent satisfies (M). Moreover, a player must $(M-1.2)$ strongly believe in the event that every opponent chooses rationally from period $M-1$ onwards and that every opponent satisfies $(M-1.1)$, and so on. These conditions together yield *common backward strong belief in rationality from period $M-1$ onwards.*

We then continue in this fashion until we reach the beginning of the game. The final restrictions on the types are called *common backward strong belief in rationality.*

Definition 3.4 (Common backward strong belief in rationality) *For every period m, number $k \in \{0, 1, ...\}$ and player i, we define the sets of types $T_i^{m.k}$ that express k-fold backward strong belief in rationality from period m onwards. These sets $T_i^{m.k}$ are inductively defined as follows.*

***Period** M. Set $T_i^{M.0} := T_i$ for every player i. For every $k \geq 1$, inductively define*

$$T_i^{M.k} := \{t_i \in T_i^{M.k-1} \mid t_i \text{ strongly believes } \times_{j \neq i} (S_j \times T_j^{M.k-1})^{rat,M}\}.$$

Set $T_i^M := \cap_{k \geq 0} T_i^{M.k}$ for every player i.

***Period** $m \leq M-1$. Set $T_i^{m.0} := T_i^{m+1}$ for every player i. For every $k \geq 1$, inductively define*

$$T_i^{m.k} := \{t_i \in T_i^{m.k-1} \mid t_i \text{ strongly believes } \times_{j \neq i} (S_j \times T_j^{m.k-1})^{rat,m}\}.$$

Set $T_i^m := \cap_{k \geq 0} T_i^{m.k}$ for every player i.

For a given period m and round k, a type t_i is said to express up to k-fold backward strong belief in rationality from period m onwards if $t_i \in T_i^{m.k}$. The type t_i is said to express common backward strong belief in rationality from period m onwards if $t_i \in T_i^m$. The type t_i is said to express common backward strong belief in rationality if $t_i \in T_i^L$, where L is the first period in the game.

The following result guarantees that the epistemic conditions imposed above lead to "well-behaved" sets.

Lemma 3.2 (Epistemic conditions lead to closed sets) *Each of the sets $T_i^{m.k}$ and T_i^m above is a closed subset of T_i.*

Let us now have a closer look at the epistemic conditions above. The conditions imply that at every information set where a player has to move, he looks for the earliest period m and the highest degree k such that it is possible to believe that (i) every player chooses rationally from period m onwards and expresses common backward strong belief in rationality from period m onwards, and (ii) every player chooses rationally from period $m-1$ onwards and expresses up to k-fold backward strong belief in rationality from period $m-1$ onwards. Moreover, he *will* then believe (i) and (ii). This may be viewed as a *best rationalization principle* for the epistemic concept above.

From this best rationalization principle it is clear that epistemic priority is given to backward induction reasoning: If a player is at an information set, he first looks for the earliest period m such that it is possible to believe that every player chooses rationally from period m onwards and expresses common backward strong belief in rationality from period m onwards. In that case, the player *will* express common backward strong belief in rationality from period m onwards, and hence will believe, in particular, that every opponent will choose rationally from period m onwards. Only afterwards will he think about period $m-1$, and look for the highest degree k such that it is possible to believe that, in addition, every player chooses rationally from period $m-1$ onwards and expresses up to k-fold backward strong belief in rationality from period $m-1$ onwards.

The following result shows that the epistemic conditions in *common backward strong belief in rationality* single out precisely those strategies that are *forward and backward rationalizable.*

Theorem 3.1 (Epistemic characterization) *Consider the universal type space $\mathcal{T} = ((T_i, \mathcal{O}_i), \beta_i)_{i \in I}$. Then, for every player i and strategy $s_i \in S_i$, the following holds:*

(a) strategy s_i is forward and backward rationalizable, if and only if, s_i is optimal from the first period onwards for a type $t_i \in T_i$ that expresses common backward strong belief in rationality,

(b) if $m \leq M-1$ then $s_i \in S_i^{m.0}$, if and only if, s_i is optimal from period $m+1$ onwards for a type $t_i \in T_i^{m+1}$ that expresses common backward strong belief in rationality from period $m+1$ onwards, and

(c) if $k \geq 0$ then $s_i \in S_i^{m.k+1}$, if and only if, s_i is optimal from period m onwards for a type $t_i \in T_i^{m.k}$ that expresses up to k-fold backward strong belief in rationality from period m onwards.

In particular, since we know from the paper that forward and backward rationalizable strategies always exist, it follows that there is always a type that expresses common backward strong belief in rationality. That is, the system of epistemic conditions we offer never leads to logical contradictions.

A major difference with strong rationalizability is that forward and backward rationalizability requires players to do forward induction reasoning from a certain period onwards, in a backward inductive fashion. Strong rationalizability, in contrast, always requires players to do the forward induction reasoning in the whole game, that is, from the first period onwards.

As such, we can also consider a *bounded rationality* version of forward and backward rationalizability in which players only do the forward induction reasoning from period M onwards, from period $M-1$ onwards, until we reach period m. Players would thus not actively reason about choices that are made before period m. Parts (b) and (c) in Theorem 3.1 reveal what has to be imposed, in terms of reasoning, to establish such a bounded rationality variant.

4 Epistemic Priority

One could argue that in the concept of forward and backward rationalizability we give epistemic priority to backward induction reasoning over forward induction reasoning. This may be seen, for instance, from Theorem 3.1: We first impose common strong belief in rationality from the last period onwards, and keep these restrictions when we move to restrictions on reasoning from the penultimate period onwards. In turn, the restrictions on the reasoning from the penultimate period onwards are maintained when restricting the reasoning from earlier periods onwards, and so on. As such, a player will always believe, at every period, that his opponents will choose rationally in the periods that lie ahead – a typical backward induction restriction.

But we could also change the epistemic priority, by first applying the strong rationalizability procedure, and subsequently refining it by the backwards rationalizability procedure. In that alternative procedure we would thus give epistemic priority to forward induction reasoning over backward induction reasoning.

This concept would be different from ours, at least in terms of strategies and beliefs. Consider, for instance, the game from Figure 2, where our concept would yield the strategy c for player 2, whereas the alternative procedure would uniquely select strategy (d, g) for player 2. To see this, note that backwards rationalizability always leads to the unique backward induction strategies in perfect information games without relevant ties, like the one in Figure 2. As player 2's backward induction strategy is c, and our concept is a refinement of backwards rationalizability in terms of strategies, our concept will uniquely select c as well. On the other hand, we have seen in the introduction of the paper that player 2's unique strongly rationalizable strategy is (d, g). As the alternative procedure is a refinement of strong rationalizability in terms of strategies, it will uniquely select (d, g) also.

We do not know at this point whether this alternative procedure always yields the same outcomes as forward and backward rationalizability. The reason is that the alternative procedure does not correspond to a specific elimination order of the strong belief reduction operator. Indeed, after applying the strong rationalizability procedure, which corresponds to recursively applying the strong

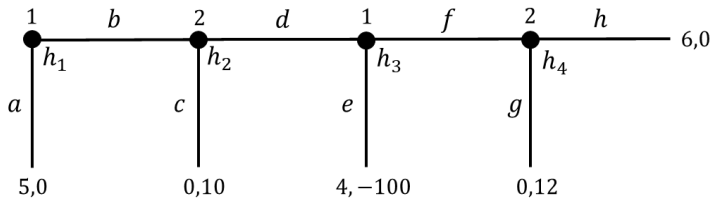

Figure 2: Strong rationalizability may lead to counterintuitive behavior

belief reduction operator at "full speed", the strong belief reduction operator is not able to induce any further eliminations.

For a concept that combines forward and backward induction reasoning, one could also proceed alternatively, by first applying the backwards rationalizability procedure to the whole game, until we can go no further, after which it is refined by the steps in the strong rationalizability procedure. This would correspond to an instance of Δ-*rationalizability* (Battigalli (2003), Battigalli and Siniscalchi (2003)) where Δ consists of the restrictions on beliefs imposed by backwards rationalizability. Like with our procedure, this would also correspond to a scenario where epistemic priority is given to backward induction reasoning, but in a more extreme fashion than we do. Indeed, in the alternative procedure we would first exhaust all the backward induction reasoning in the whole game, after which we exclusively turn to forward induction reasoning in the whole game.

It turns out that this alternative concept may also differ from forward and backward rationalizability in terms of strategies. To see, consider the game from Figure 3. In the alternative procedure, we would start by applying the backwards rationalizability procedure to the whole game. We proceed in a backward inductive fashion here, by first considering the last information set h_4, where nothing can be eliminated. At h_3, we eliminate (In, f) for player 2, after which we can eliminate (In, c) for player 1 at h_2. Finally, we eliminate (In, r) for player 3 at h_1. The backwards rationalizable strategies are thus Out, (In, a) and (In, b) for player 1, Out, (In, d) and (In, e) for player 2, and Out and (In, l) for player 3.

If we take this as an input for the strong rationalizability procedure, then in round 1 of the strong rationalizability procedure we can eliminate (In, b) for player 1 and (In, d) for player 2. Indeed, at h_2 player 1 must believe that player 2 chooses Out, (In, d) or (In, e) and that player 3 chooses (In, l). Hence, player 1 expects at most 1 by choosing (In, b) there. Also, player 2 must believe at h_3 that player 1 will choose (In, a) or (In, b) and that player 3 will choose (In, l). As such, player 2 expects at most 1 by choosing (In, d) there.

In round 2 we can then eliminate, for similar reasons, (In, a) for player 1 and (In, e) for player

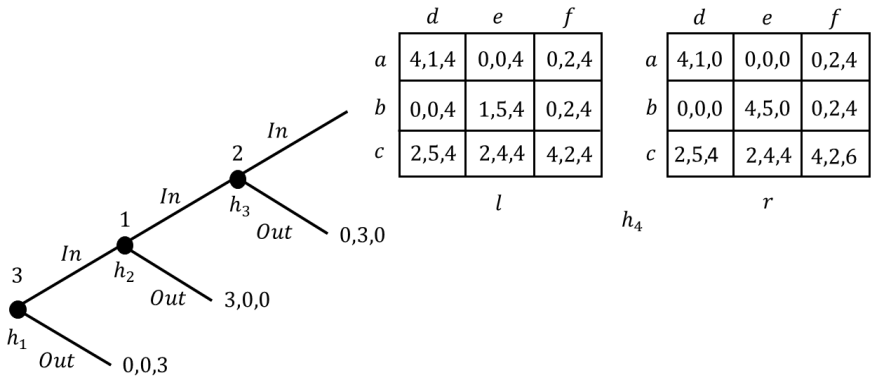

Figure 3: Triple outside option game

2. In round 3 we can finally eliminate (In, l) for player 3. Indeed, player 3 must believe at h_1 that player 1 chooses Out, which yields Out as the only optimal strategy for player 3 at h_1.

The alternative procedure, where we first apply the backwards rationalizability procedure and then strong rationalizability, thus yields the strategy Out for player 1, Out for player 2, and Out for player 3.

Let us now apply our procedure to this game. Applying the strong rationalizability procedure to the subgame starting at h_4 yields no eliminations. If we start at h_3, then we can only eliminate the strategy (In, f) for player 2. If we start at h_2, then in round 1 we eliminate the strategy (In, c) for player 1, as he expects to get at most 2 by choosing (In, c) at h_2. In round 2 we would then eliminate the strategy (In, d) for player 2 and the strategy (In, r) for player 3. Indeed, player 2 expects to get at most 1 by playing (In, d), whereas for player 3 choosing l has become better than choosing r at h_4. In round 3, we would then eliminate (In, a) and (In, b) for player 1, since he expects to obtain no more than 1 by choosing either of these two strategies. Finally, we would move to the game starting at h_1, where we can eliminate the strategy (In, l) for player 3. Our concept would thus yield the strategy Out for player 1, the strategies Out and (In, e) for player 2, and the strategy Out for player 3. As the strategy (In, e) for player 2 was not selected by the alternative procedure, both concepts differ in terms of strategies.

The reason for why our concept allows for player 2's strategy (In, e) but the alternative procedure does not, is the following: The alternative procedure starts by eliminating the strategies (In, f) for player 2, (In, c) for player 1 and (In, r) for player 3. It would then proceed by applying strong rationalizability to the whole game, so that player 2 will conclude at h_3 that player 1 must be

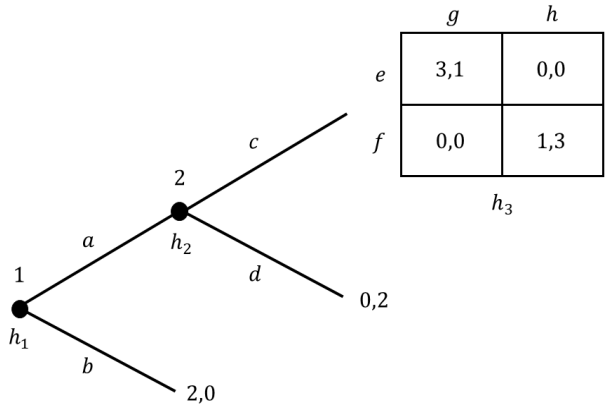

Figure 4: Battle of the sexes with double outside option

choosing (In, a). As a consequence player 2 must choose Out at h_3 according to the alternative procedure.

Our concept proceeds differently: It also starts by eliminating (In, f) for player 2 and (In, c) for player 1. But then, by reasoning from h_2 onwards, we would eliminate (In, d) for player 2 and (In, r) for player 3. If player 1 believes at h_2 that player 2 will no longer choose (In, d) and (In, f), both (In, a) and (In, b) become suboptimal for player 1 at h_1. As (In, c) was already eliminated for player 1 at an earlier stage, player 2 is no longer forced to discriminate between (In, a) and (In, b), and hence player 2 may still believe at h_3 that player 1 chooses (In, b). Hence, player 2 may still rationally choose (In, e) at h_3 according to our concept.

In this example we thus see that the alternative concept is more restrictive, in terms of strategies, than ours. The reverse may also be true, as the example from Figure 4 shows. We have already seen in the introduction and Section 3.4 of the paper that our concept uniquely selects the strategies b for player 1 and (c, h) for player 2.

Suppose now that we would run the alternative procedure. By applying backwards rationalizability first, we would start by eliminating strategy (c, g) for player 2 at h_2, after which we would eliminate (a, e) and (a, f) for player 1 at h_1. Indeed, if player 1 believes at h_1 that player 2 will not choose (c, g), then choosing a can give him at most 1. Hence, the backwards rationalizable strategies are b for player 1, and (c, h) and d for player 2. If we use this as the input for the strong rationalizability procedure, then at h_2 player 2 can no longer discriminate between (a, e) and (a, f) for player 1, and hence player 2 may believe at h_2 that player 1 chooses (a, e) or (a, f). As such, both d and (c, h) can be optimal for player 2 at h_2, which means that the alternative concept would select both

d and (c, h) for player 2.

The reason for this difference is similar to above, but now with the roles of the two concepts reversed: Under our concept, player 2 will certainly believe at h_2 that player 1 chooses (a, f) and not (a, e), and therefore only (c, h) is optimal for player 2. Under the alternative concept, player 2 is no longer forced to discriminate between (a, e) and (a, f), which leaves both d and (c, h) as valid options for player 2.

It can be shown that this alternative procedure corresponds to a specific elimination order of the strong belief reduction operator. But then, it follows in the same way as in the paper that also this procedure is equivalent to strong rationalizability in terms of outcomes. As such, it is also equivalent to forward and backward rationalizability in terms of outcomes.

This alternative procedure thus gives epistemic priority to backward induction reasoning over forward induction reasoning. Similar approaches have been adopted in the equilibrium refinements literature from the eighties and early nineties, where the backward induction concept of sequential equilibrium has been refined by imposing forward induction restrictions. See, for instance, *justifiable sequential equilibrium* (McLennan (1985)), *forward induction equilibrium* (Cho (1987)) and *stable sets of beliefs* (Hillas (1994)).

The issue of *epistemic priority* is explored in depth by Catonini (2019), who proposes the concept of *selective rationalizability.* It proceeds by first applying strong rationalizability, after which it is refined by imposing (common strong belief in) some exogenously given restrictions on beliefs.[3] However, it could happen that these exogenous restrictions are incompatible with the restrictions imposed by strong rationalizability. This will typically be the case when the exogenous restrictions are taken to be the restrictions imposed by backwards rationalizability, because in many dynamic games these restrictions go against the restrictions of strong rationalizability.

The first alternative procedure described above, where we started with the strong rationalizability procedure, and subsequently refined it with the backwards rationalizability procedure, is, strictly speaking, not a selective rationalizability procedure in the sense of Catonini (2019): The restrictions of backwards rationalizability are not being imposed as *exogeneous* restrictions after completing the strong rationalizability procedure, but rather as "optional" restrictions meant to refine, *whenever possible*, the reasoning of strong rationalizability.

5 Proofs

Proof of Lemma 3.1. Consider some information set $h \in H_i$ of player i and let $s_i \in S_i(h)$ be a strategy that allows h to be reached. We first show that the set of types $T_i(s_i, h)$ of player i for whom

[3] Instead of starting with strong rationalizability one could also start with some other concept, and then impose exogenous restrictions on the beliefs. Catonini (2019) uses strong rationalizability as the focal concept here, but his analysis allows it to be replaced by any other rationalizability concept for dynamic games as well.

playing s_i is optimal at h is a closed set of types. To this purpose, we consider for any alternative strategy $r_i \in S_i(h)$, any opponents' strategy combination s_{-i} and any opponents' type combination t_{-i} the utility difference

$$u_i(s_i, s_{-i}, t_{-i}) - u_i(r_i, s_{-i}, t_{-i}) := u_i(z(s_i, s_{-i})) - u_i(z(r_i, s_{-i})).$$

By fixing s_i, r_i, and h, and varying s_{-i} and t_{-i}, we obtain a bounded continuous function

$$u_i(s_i, \cdot, \cdot) - u_i(r_i, \cdot, \cdot) : S_{-i}(h) \times T_{-i} \to \mathbf{R}.$$

This is indeed the case, since we endow $S_{-i}(h)$ with the discrete topology and $u_i(s_i, s_{-i}, t_{-i}) - u_i(r_i, s_{-i}, t_{-i})$ only depends on the S_{-i}-dimension of $S_{-i} \times T_{-i}$. Hence, the set of measures $\mu_i \in \Delta(S_{-i}(h) \times T_{-i})$ such that integrating over this function with respect to μ_i is non-negative is a closed set of measures, which we call $\Delta(S_{-i}(h) \times T_{-i})^{s_i \geq r_i}$. The set of measures $\mu_i \in \Delta(S_{-i}(h) \times T_{-i})$ such that s_i is optimal at h is the intersection

$$\Delta(S_{-i}(h) \times T_{-i})^{s_i, h} := \bigcap_{r_i \in S_i(h)} \Delta(S_{-i}(h) \times T_{-i})^{s_i \geq r_i},$$

which is closed as well. Note that, by construction,

$$T_i(s_i, h) = \{t_i \in T_i \mid \beta_i(t_i, h) \in \Delta(S_{-i}(h) \times T_{-i})^{s_i, h}\}.$$

Since the mapping $\beta_i(\cdot, h) : T_i \to \Delta(S_{-i}(h) \times T_{-i})$ is continuous, we conclude that the set $T_i(s_i, h)$ is closed.

Recall that the set of types t_i such that s_i is optimal at h is precisely $T_i(s_i, h)$. For a given period m, let $H^{\geq m} := H^m \cup H^{m+1} \cup \ldots \cup H^M$ be the collection of information sets from period m onwards. Then, the set of types for which s_i is optimal from period m onwards is

$$T_i^{\geq m}(s_i) := \bigcap_{h \in H_i(s_i) \cap H^{\geq m}} T_i(s_i, h),$$

which is closed in T_i. Note that if s_i does not reach any information set in $H^{\geq m}$, then s_i is automatically optimal from period m onwards for all types in T_i. For each of the finitely many strategies $s_i \in S_i$ of player i, the set $\{s_i\} \times T_i^{\geq m}(s_i)$ is closed in the product topology of $S_i \times T_i$, since it is the product of two closed sets. The set

$$(S_i \times T_i)^{rat, m} = \bigcup_{s_i \in S_i} (\{s_i\} \times T_i^{\geq m}(s_i))$$

is closed in $S_i \times T_i$ since it is the union of finitely many closed sets. If $\hat{T}_i$ is a measurable subset of T_i then $(S_i \times \hat{T}_i)^{rat, m} = (S_i \times T_i)^{rat, m} \cap (S_i \times \hat{T}_i)$ is measurable since it is an intersection of a closed

and a measurable set. If $\hat{T}_i$ is closed, then $S_i \times \hat{T}_i$ is closed and hence $(S_i \times \hat{T}_i)^{rat,m}$ is closed, being the intersection of two closed sets. ■

Proof of Lemma 3.2. We start by proving the following result.

Claim. Let E be a closed subset of $S_{-i} \times T_{-i}$. Then, the set $\{\ t_i \mid t_i$ strongly believes $E\ \}$ is a closed subset of T_i.

Proof of claim. Let $h \in H_i$ be such that $(S_{-i}(h) \times T_{-i}) \cap E \neq \emptyset$. We show that the set of measures in $\Delta(S_{-i}(h) \times T_{-i})$ that assign probability 1 to E is closed set. To this end, let $(\mu_n)_{n\in\mathbb{N}} \to \mu$ be a sequence of probability measures in $\Delta(S_{-i}(h) \times T_{-i})$ converging to $\mu \in \Delta(S_{-i}(h) \times T_{-i})$ such that $\mu_n(E) = 1$ for all $n \in \mathbb{N}$. We have to show that $\mu(E) = 1$. But this follows immediately from the Portemanteau Theorem (Kechris (1995), Theorem 17.20). By continuity, the set $\{t_i \in T_i \mid \beta_i(t_i, h)(E) = 1\}$ is a closed set of types. The set of types t_i that strongly believe E is the finite intersection of such sets of types over all $h \in H_i$ such that $(S_{-i}(h) \times T_{-i}) \cap E \neq \emptyset$. Hence, this is a closed set of types. $\Diamond$

The lemma now follows immediately by iteratively applying the claim and Lemma 3.1. ■

Proof of Theorem 3.1. As a first step we will use the forward and backward rationalizability procedure to build a finite type space. Later we will use this model to prove the theorem. Moreover, we will make sure that the type space is non-reduntant, that is, no two different types of a player induce the same conditional belief hierarchy.

Recall that, for every player i, period m, and round k, the sets $B_i^{m.k}$ and $S_i^{m.k}$ are the collections of conditional belief vectors and strategies, respectively, selected by the forward and backward rationalizability procedure at round k of period m. In particular, $B_i^{L.K_L}$ and $S_i^{L.K_L}$ are the sets of conditional belief hierarchies and strategies, respectively, that survive all rounds at all periods.

For every player i and strategy $s_i \in S_i^{L.K_L}$ choose a conditional belief vector $b_i[s_i] \in B_i^{L.K_L}$ such that s_i is optimal for $b_i[s_i]$ from the first period onwards.

For all other strategies s_i there is a period $m \in \{L, ..., M\}$ and a round k such that $s_i \in S_i^{m.k} \backslash S_i^{m.k+1}$. For such a strategy $s_i \in S_i^{m.k} \backslash S_i^{m.k+1}$ we can then choose a conditional belief vector $b_i[s_i] \in B_i^{m.k}$ such that s_i is optimal for $b_i[s_i]$ from period m onwards if $k \geq 1$, and s_i is optimal from period $m+1$ onwards if $k = 0$. If $m.k = M.0$, then optimality from period $M+1$ onwards means that s_i need not be optimal for $b_i[s_i]$ at all.

Based on these conditional belief vectors $b_i[s_i]$ we will now construct a finite type space $\hat{\mathcal{T}} = ((T_i, \mathcal{O}_i), \beta_i)$ where the sets of types are given by $T_i = \{t_i^{b_i[s_i]} \mid s_i \in S_i\}$, and the belief mappings β_i are such that

$$\beta_i(t_i^{b_i[s_i]}, h)((s_j, t_j)_{j\neq i}) = \begin{cases} b_i[s_i](h)((s_j)_{j\neq i}), & \text{if } t_j = t_j^{b_j[s_j]} \text{ for all } j \neq i \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

for all players i, all strategies s_i, all information sets $h \in H_i$, and all opponents' strategy-type combinations $(s_j, t_j)_{j \neq i} \in S_{-i} \times T_{-i}$. Hence, every type $t_i^{b_i[s_i]}$ has the belief $b_i[s_i](h)$ about the opponents' strategy combinations at every information set $h \in H_i$, and matches, in its belief, every opponent's strategy s_j with the associated type $t_j^{b_j[s_j]}$. It is easy to see that every type in this model satisfies Bayesian updating. Note that $b_i[s_i] = b_i[\hat{s}_i]$ implies that $t_i^{b_i[s_i]} = t_i^{b_i[\hat{s}_i]}$, and hence the type space $\hat{\mathcal{T}}$ is non-redundant by construction.

For every player i and conditional belief vector $b_i \in B_i \backslash \{b_i[s_i] \mid s_i \in S_i\}$ not present in $\hat{\mathcal{T}}$, we add a new type $t_i^{b_i}$ to $\hat{\mathcal{T}}$ whose conditional beliefs are given by

$$\beta_i(t_i^{b_i}, h)((s_j, t_j)_{j \neq i}) = \begin{cases} b_i(h)((s_j)_{j \neq i}), & \text{if } t_j = t_j^{b_j[s_j]} \text{ for all } j \neq i \\ 0, & \text{otherwise} \end{cases} \tag{5.2}$$

The new type space obtained after adding the type $t_i^{b_i}$ to $\hat{\mathcal{T}}$ is denoted by $\hat{\mathcal{T}} \cup \{t_i^{b_i}\}$.

Let $\mathcal{T}$ be a universal type space. Then, by definition, each of the finite type spaces $\hat{\mathcal{T}}$ and $\hat{\mathcal{T}} \cup \{t_i^{b_i}\}$ maps in a unique way to the universal type space $\mathcal{T}$ by a type morphism. Note that for every type $t_i^{b_i[s_i]}$ in $\hat{\mathcal{T}}$ the induced conditional belief hierarchy is the same, no matter whether it is regarded as a type in $\hat{\mathcal{T}}$ or a type in $\hat{\mathcal{T}} \cup \{t_j^{bj}\}$. Since a type morphism always preserves the induced conditional belief hierarchy, the type $t_i^{b_i[s_i]}$ will be mapped to the same type in the universal type space $\mathcal{T}$, no matter whether it is regarded as a type in $\hat{\mathcal{T}}$ or a type in $\hat{\mathcal{T}} \cup \{t_j^{bj}\}$. As each of the type spaces $\hat{\mathcal{T}}$ and $\hat{\mathcal{T}} \cup \{t_i^{b_i}\}$ is non-redundant, every type in these type spaces may be uniquely identified with a type in the universal type space $\mathcal{T}$.

For every player i, period m and number $k \in \{0, 1, ...\}$, we denote by $T_i^{m.k}$ the set of types for player i in the universal type space $\mathcal{T}$ that express k-fold backward strong belief in rationality from period m onwards. Define

$$\hat{B}_i^{m.k} := \{b_i \in B_i \mid \text{there is some } t_i \in T_i^{m.k} \text{ that induces the conditional belief vector } b_i\}$$

and

$$\hat{S}_i^{m.k} := \{s_i \in S_i \mid \text{there is some } t_i \in T_i^{m.k} \text{ with } (s_i, t_i) \in (S_i \times T_i^{m.k})^{rat,m}\}.$$

Here, when we say that "t_i induces the conditional belief vector b_i", we mean that $\text{marg}_{S_{-i}(h)}\beta_i(t_i, h) = b_i(h)$ for every $h \in H_i$. We prove the following result.

Claim. For every period m and number $k \in \{0, 1, ...\}$ it holds that (i) $\hat{B}_i^{m.k} \subseteq B_i^{m.k+1}$, (ii) $B_i^{m.k+1} \subseteq \hat{B}_i^{m.k}$ and for every $b_i \in B_i^{m.k+1}$ we have that $t_i^{b_i} \in T_i^{m.k}$, (iii) $\hat{S}_i^{m.k} \subseteq S_i^{m.k+1}$ and (iv) $S_i^{m.k+1} \subseteq \hat{S}_i^{m.k}$.

Proof of claim. We show the four statements by induction on $m.k$.

We start with $M.0$. Then, $\hat{B}_i^{M.0}$ is, by definition, the set of conditional belief vectors induced by the types in $T_i^{M.0}$. As $T_i^{M.0} = T_i$, this is the set of all conditional belief vectors, and hence $\hat{B}_i^{M.0} = B_i$.

As, by construction, $B_i^{M.1} = B_i$ as well, it follows that $\hat{B}_i^{M.0} = B_i^{M.1}$. Moreover, as $T_i^{M.0} = T_i$, for every $b_i \in B_i^{M.1}$ we have that $t_i^{b_i} \in T_i^{M.0}$. This establishes (i) and (ii).

Moreover, $\hat{S}_i^{M.0}$ contains precisely those strategies that are optimal from period M onwards for some $t_i \in T_i^{M.0}$. As $T_i^{M.0} = T_i$, these are precisely the strategies that are optimal from period M onwards for some conditional belief vector $b_i \in B_i$. By definition, these are precisely the strategies in $S_i^{M.1}$. Hence, we conclude that $\hat{S}_i^{M.0} = S_i^{M.1}$. This establishes (iii) and (iv).

Next, take some $m.k \neq M.0$, and assume that the claim holds for $m.k-1$ if $k \geq 1$, and that the claim holds for any $m+1.k'$ if $k = 0$. We distinguish two cases: (1) $k = 0$, and (2) $k \geq 1$.

Case 1. Suppose that $k = 0$. Then, by definition, there is some round K such that $\hat{B}_i^{m.0} = \hat{B}_i^{m+1.K}$ and $B_i^{m.1} = B_i^{m+1.K+1}$. As, by the induction assumption, $\hat{B}_i^{m+1.K} = B_i^{m+1.K+1}$, we conclude that $\hat{B}_i^{m.0} = B_i^{m.1}$. Moreover, by construction, $\hat{S}_i^{m.0}$ is the set of strategies that are optimal, from period m onwards, for some $b_i \in \hat{B}_i^{m.0}$, whereas $S_i^{m.1}$ is the set of strategies that are optimal, from period m onwards, for some $b_i \in B_i^{m.1}$. Since $\hat{B}_i^{m.0} = B_i^{m.1}$, it follows that $\hat{S}_i^{m.0} = S_i^{m.1}$.

Case 2. Suppose that $k \geq 1$.

(i) We show that $\hat{B}_i^{m.k} \subseteq B_i^{m.k+1}$. Take some $b_i \in \hat{B}_i^{m.k}$. Then, there is some $t_i \in T_i^{m.k}$ that induces b_i. By definition, $T_i^{m.k} \subseteq T_i^{m.k-1}$, and hence $b_i \in \hat{B}_i^{m.k-1}$. By the induction assumption on (i) it follows that $b_i \in B_i^{m.k}$. Hence, we only need to show that b_i strongly believes $S_{-i}^{m.k}$. Let $h \in H_i$ be such that $S_{-i}^{m.k} \cap S_{-i}(h) \neq \emptyset$. We must show that $b_i(h)(S_{-i}^{m.k}) = 1$. By the induction assumption applied to (iii) and (iv) we know that $S_{-i}^{m.k} = \hat{S}_{-i}^{m.k-1}$. Hence, by the definition of $\hat{S}_j^{m.k-1}$ for every $j \neq i$, we know that $\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat,m} \cap (S_{-i}(h) \times T_{-i}) \neq \emptyset$. Therefore, since $t_i \in T_i^{m.k}$, we conclude that $\beta_i(t_i, h)(\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat,m}) = 1$. This, in turn, implies that $b_i(h)(\hat{S}_{-i}^{m.k-1}) = 1$. As, by the induction assumption on (iii) and (iv), $S_{-i}^{m.k} = \hat{S}_{-i}^{m.k-1}$, we conclude that $b_i(h)(S_{-i}^{m.k}) = 1$. Hence, b_i strongly believes $S_{-i}^{m.k}$. Since $b_i \in B_i^{m.k}$, it follows that $b_i \in B_i^{m.k+1}$. As such, $\hat{B}_i^{m.k} \subseteq B_i^{m.k+1}$.

(ii) We show that $B_i^{m.k+1} \subseteq \hat{B}_i^{m.k}$ and for every $b_i \in B_i^{m.k+1}$ we have that $t_i^{b_i} \in T_i^{m.k}$. Take some $b_i \in B_i^{m.k+1}$. Then, in particular, $b_i \in B_i^{m.k}$ and hence we know, by the induction assumption on (ii), that $t_i^{b_i} \in T_i^{m.k-1}$. Thus, to prove that $t_i^{b_i} \in T_i^{m.k}$ it only remains to show that $t_i^{b_i}$ strongly believes $\times_{j \neq i}$ $(S_j \times T_j^{m.k-1})^{rat.m}$. To this end, let $h \in H_i$ be such that $(S_{-i}(h) \times T_{-i}) \cap (\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat,m}) \neq \emptyset$. We must show that $\beta_i(t_i^{b_i}, h)(\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat,m}) = 1$.

By definition, $\hat{S}_{-i}^{m.k-1} = \text{proj}_{S_{-i}}(\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat,m})$. Note that by the induction assumption of (iii) and (iv) we have that $\hat{S}_{-i}^{m.k-1} = S_{-i}^{m.k}$. Therefore, we have $S_{-i}(h) \cap S_{-i}^{m.k} \neq \emptyset$. Since $b_i \in B_i^{m.k+1}$ it follows that $b_i(h)(S_{-i}^{m.k}) = 1$, and hence $b_i(h)(\hat{S}_{-i}^{m.k-1}) = 1$. By the definition of $t_i^{b_i}$ in (5.2) we have $\beta_i(t_i^{b_i}, h)((\hat{S}_{-i}^{m.k-1} \cap S_{-i}(h)) \times T_{-i}) = 1$, and that $\beta_i(t_i, h)$ assigns probability 1 to the set of opponents' strategy-type combinations $\{(s_j, t_j^{b_j[s_j]})_{j \neq i} \mid s_j \in \hat{S}_j^{m.k-1}$ for all $j \neq i\}$. As $\hat{S}_{-i}^{m.k-1} = S_{-i}^{m.k}$ we conclude that $\beta_i(t_i^{b_i}, h)$ assigns probability 1 to the set of opponents' strategy-type combinations $\{(s_j, t_j^{b_j[s_j]})_{j \neq i}$

$\mid s_j \in S_j^{m.k}$ for all $j \neq i\}$.

Consider a type $t_j^{b_j[s_j]}$ where $s_j \in S_j^{m.k}$. Then, we know by the definition of type $t_j^{b_j[s_j]}$ in (5.1) that $t_j^{b_j[s_j]}$ induces the conditional belief vector $b_j[s_j] \in B_j^{m.k}$, and that s_j is optimal for $b_j[s_j]$ from period m onwards. Hence, s_j is optimal for $t_j^{b_j[s_j]}$ from period m onwards. As $b_j[s_j] \in B_j^{m.k}$ we conclude by the induction assumption of (ii) that $t_j^{b_j[s_j]} \in T_j^{m.k-1}$. As s_j is optimal for type $t_j^{b_j[s_j]}$ from period m onwards, it follows that $(s_j, t_j^{b_j[s_j]}) \in (S_j \times T_j^{m.k-1})^{rat,m}$. Recall that $\beta_i(t_i^{b_i}, h)$ assigns probability 1 to the set of opponents' strategy-type combinations $\{(s_j, t_j^{b_j[s_j]})_{j \neq i} \mid s_j \in S_j^{m.k}$ for all $j \neq i\}$. Hence, it follows that $\beta_i(t_i^{b_i}, h)(\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat,m}) = 1$. As such, we conclude that $t_i^{b_i}$ strongly believes $\times_{j \neq i}(S_j \times T_j^{m.k-1})^{rat.m}$.

Since $t_i^{b_i} \in T_i^{m.k-1}$ it follows that $t_i^{b_i} \in T_i^{m.k}$. We thus conclude that for every $b_i \in B_i^{m.k+1}$ we have that $t_i^{b_i} \in T_i^{m.k}$. Since, by (5.2), $t_i^{b_i}$ induces the conditional belief vector b_i, it follows that $b_i \in \hat{B}_i^{m.k}$. Hence, $B_i^{m.k+1} \subseteq \hat{B}_i^{m.k}$.

(iii) We show that $\hat{S}_i^{m.k} \subseteq S_i^{m.k+1}$. Let $s_i \in \hat{S}_i^{m.k}$. Then, in particular, $s_i \in \hat{S}_i^{m.k-1}$. By the induction assumption of (iii) it follows that $s_i \in S_i^{m.k}$. Since $s_i \in \hat{S}_i^{m.k}$, there is a $t_i \in T_i^{m.k}$ such that s_i is optimal for t_i from period m onwards. Let b_i be the conditional belief vector induced by t_i. As the expected utility depends only on first-order beliefs, s_i is optimal for b_i from period m onwards. Since $t_i \in T_i^{m.k}$ it follows, by definition, that $b_i \in \hat{B}_i^{m.k}$. By (i) it then follows that $b_i \in B_i^{m.k+1}$. Hence, s_i is optimal for some $b_i \in B_i^{m.k+1}$ from period m onwards. As we have seen above that $s_i \in S_i^{m.k}$, we conclude that $s_i \in S_i^{m.k+1}$. Hence, $\hat{S}_i^{m.k} \subseteq S_i^{m.k+1}$.

(iv) We finally show that $S_i^{m.k+1} \subseteq \hat{S}_i^{m.k}$. Let $s_i \in S_i^{m.k+1}$. Then, by construction, $b_i[s_i] \in B_i^{m.k+1}$ and s_i is optimal for $b_i[s_i]$ from period m onwards. By (ii) we know that $t_i^{b_i[s_i]} \in T_i^{m.k}$. Moreover, $t_i^{b_i[s_i]}$ induces the conditional belief vector $b_i[s_i]$. Since the expected utility depends only on first-order beliefs, we conclude that s_i is optimal for $t_i^{b_i[s_i]}$ from period m onwards. This implies that $(s_i, t_i^{b_i[s_i]}) \in (S_i \times T_i^{m.k})^{rat.m}$, and hence $s_i \in \hat{S}_i^{m.k}$. Thus, $S_i^{m.k+1} \subseteq \hat{S}_i^{m.k}$. This completes the proof of the claim. $\Diamond$

We are now able to prove the theorem.

(a) Take first a strategy s_i that is forward and backward rationalizable. Then, there is a conditional belief vector $b_i \in B_i^{L.K_L}$ such that s_i is optimal for b_i from the first period onwards. Note that $b_i \in B_i^{L.k+1}$ for all k and hence, by part (ii) of the claim, $t_i^{b_i} \in T_i^{L.k}$ for all k. Therefore, $t_i^{b_i} \in T_i^L$, and hence $t_i^{b_i}$ expresses common backward strong belief in rationality. As $t_i^{b_i}$ induces the conditional belief vector b_i, and s_i is optimal for b_i from the first period onwards, it follows that s_i is optimal for $t_i^{b_i}$ from the first period onwards. As such, s_i is optimal, from the first period onwards, for a type that expresses common backward strong belief in rationality.

Conversely, suppose that s_i is optimal, from the the first period onwards, for a type t_i that expresses common backward strong belief in rationality. Hence, $t_i \in T_i^L$. Suppose that t_i induces the conditional belief vector b_i. Then, s_i is optimal, from the first period onwards, for b_i. Since $t_i \in T_i^{L.k}$ for all k, and t_i induces the conditional belief vector b_i, it follows that $b_i \in \hat{B}_i^{L.k}$ for all k. By part (i) of the claim it follows that $b_i \in B_i^{L.k+1}$ for all k, and hence b_i is forward and backward rationalizable. Since s_i is optimal for b_i from the first period onwards, we conclude that s_i is forward and backward rationalizable.

(b) Take first a strategy $s_i \in S_i^{m.0}$. Then, $s_i \in S_i^{m+1.K_{m+1}}$. Hence, there is a conditional belief vector $b_i \in B_i^{m+1.K_{m+1}}$ such that s_i is optimal for b_i from period $m+1$ onwards. Note that $b_i \in B_i^{m+1.k+1}$ for all k and hence, by part (ii) of the claim, $t_i^{b_i} \in T_i^{m+1.k}$ for all k. Therefore, $t_i^{b_i} \in T_i^{m+1}$, and hence $t_i^{b_i}$ expresses common backward strong belief in rationality from period $m+1$ onwards. As $t_i^{b_i}$ induces the conditional belief vector b_i, and s_i is optimal for b_i from period $m+1$ onwards, it follows that s_i is optimal for $t_i^{b_i}$ from period $m+1$ onwards. As such, s_i is optimal, from period $m+1$ onwards, for a type that expresses common backward strong belief in rationality from period $m+1$ onwards.

Conversely, suppose that s_i is optimal, from period $m+1$ onwards, for a type t_i that expresses common backward strong belief in rationality from period $m+1$ onwards. Hence, $t_i \in T_i^{m+1}$. Suppose that t_i induces the conditional belief vector b_i. Then, s_i is optimal, from period $m+1$ onwards, for b_i. Since $t_i \in T_i^{m+1.k}$ for all k, and t_i induces the conditional belief vector b_i, it follows that $b_i \in \hat{B}_i^{m+1.k}$ for all k. By part (i) of the claim it follows that $b_i \in B_i^{m+1.k+1}$ for all k, and hence $b_i \in B_i^{m.0}$. Since s_i is optimal for b_i from period $m+1$ onwards, we conclude that $s_i \in S_i^{m.0}$.

(c) Take first a strategy $s_i \in S_i^{m.k+1}$. Hence, there is a conditional belief vector $b_i \in B_i^{m.k+1}$ such that s_i is optimal for b_i from period m onwards. By part (ii) of the claim we conclude that $t_i^{b_i} \in T_i^{m.k}$. As $t_i^{b_i}$ induces the conditional belief vector b_i, and s_i is optimal for b_i from period m onwards, it follows that s_i is optimal for $t_i^{b_i}$ from period m onwards. As such, s_i is optimal, from period m onwards, for a type in $T_i^{m.k}$ that expresses k-fold backward strong belief in rationality from period m onwards.

Conversely, suppose that s_i is optimal, from period m onwards, for a type $t_i \in T_i^{m.k}$ that expresses k-fold backward strong belief in rationality from period m onwards. Suppose that t_i induces the conditional belief vector b_i. Then, s_i is optimal, from period m onwards, for b_i. Since $t_i \in T_i^{m.k}$ and t_i induces the conditional belief vector b_i, it follows that $b_i \in \hat{B}_i^{m.k}$. By part (i) of the claim it follows that $b_i \in B_i^{m.k+1}$. Since s_i is optimal for b_i from period m onwards, we conclude that $s_i \in S_i^{m.k+1}$.

This completes the proof. ■

References

[1] Battigalli, P. (2003), Rationalizability in infinite, dynamic games of incomplete information, *Research in Economics* **57,** 1–38.

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

[3] Battigalli, P. and M. Siniscalchi (2002), Strong belief and forward induction reasoning, *Journal of Economic Theory* **106,** 356–391.

[4] Battigalli, P. and M. Siniscalchi (2003), Rationalization and incomplete information, *B.E. Jounal of Theoretical Economics* **3**, 1–46.

[5] Catonini, E. (2019), Rationalizability and epistemic priority orderings, *Games and Economic Behavior* **114,** 101–117.

[6] Cho, I.-K. (1987), A refinement of sequential equilibrium, *Econometrica* **55,** 1367–1389.

[7] Fukuda, S. (2023), The existence of universal qualitative belief spaces, Manuscript.

[8] Guarino, P. (2022), Topology-free type structures with conditioning events, Working paper.

[9] Hillas, J. (1994), Sequential equilibria and stable sets of beliefs, *Journal of Economic Theory* **64,** 78–102.

[10] Kechris, A. (1995), *Classical Descriptive Set Theory,* Springer, Graduate Texts in Mathematics.

[11] McLennan, A. (1985), Justifiable beliefs in sequential equilibria, *Econometrica* **53,** 889–904.