On non-monotonic strategic reasoning.*

Emiliano Catonini†

April 2017

Strong-Δ.Rationalizability (Battigalli 2003, Battigalli and Siniscalchi 2003) is a prominent and widely applied solution concept that introduces first-order belief restrictions in forward induction reasoning. In absence of restrictions, it coincides with Strong Rationalizability (Battigalli and Siniscalchi, 2002). These solution concepts are based on the notion of Strong Belief (Battigalli and Siniscalchi, 2002). The non-monotonicity of Strong Belief implies that the predictions of Strong-Δ-Rationalizability under given restrictions can have empty intersection with the predictions of Strong Rationalizability. Here we show that the set of outcomes predicted by Strong-Δ-Rationalizability actually shrinks as long as (stricter and stricter) restrictions have no bite off-path. So, Strong-Δ-Rationalizability yields a subset of strongly rationalizable outcomes when the restrictions correspond to the belief in a particular path of play. Moreover, under such restrictions, the epistemic priority between belief in rationality and beliefs in the restrictions (Catonini, 2017) is irrelevant for the predicted outcomes: the predictions of Strong-Δ-Rationalizability and Selective Rationalizability (Catonini, 2017) coincide. The workhorse lemma behind these results allows to show also the order independence of the "iterated elimination of never sequential best replies" (of which Strong Rationalizability is the maximal elimination order), and that Strong Rationalizability refines Backward Induction. The outcome equivalence of Strong Rationalizability and Backward Induction in perfect information games with no relevant ties (Battigalli, 1997) follows.

Keywords: Strong-Δ-Rationalizability, Strong Rationalizability, First-order Belief Restrictions, Epistemic Priority, Order Independence, Backward Induction.

*Themain results of this paper appeared already in "Non-binding Agreements and Forward Induction Reasoning" (Catonini, 2015).

†Higher School of Economics, ICEF, emiliano.catonini@gmail.com

1 Introduction

Strong Rationalizability (Battigalli and Siniscalchi, [5]) is a form of extensive-form rationalizability (Pearce, 1984) based on the notion of *Strong Belief*.[1] More precisely, it is the maximal iterated deletion of never *sequential best replies* under strong belief in opponents' strategies that survive the previous steps.[2] Strong-Δ-Rationalizability (Battigalli [3], Battigalli and Siniscalchi [6]) introduces first-order belief restrictions in the same reasoning scheme: only beliefs in an exogenously given set are allowed at all steps.

It is well-known that the introduction of first-order belief restrictions can let the elimination procedure depart completely from Strong Rationalizability. This is due to the non-monotonicity of strong belief: the set of beliefs that display strongly belief in a smaller event is not a subset of the set of beliefs that display strong belief in a larger event. So, for instance, even in a perfect information game without relevant ties, the introduction of first-order belief restrictions can induce completely different outcomes with respect to the only strongly rationalizable one. Are there interesting conditions under which the introduction of first-order belief restrictions simply refines the set of strongly rationalizable outcomes? When such conditions are satisfied, the predictions are reassuringly robust to constrained and unconstrained forward induction reasoning, as captured, respectively, by Strong-Δ-Rationalizability and Strong Rationalizability.

It turns out that in all games with observable actions (i.e. games where, allowing for simultaneous moves, every player knows the current history of the game) Strong-Δ-Rationalizability is monotone in the predicted outcomes with respect to restrictions that "never bite off-path". With this, we refer to restrictions that, at every step of the procedure, exclude beliefs only based on how they look like at histories that all players may allow according to the previous step. So, off-the-path restrictions are responsible for the general non-monotonicity of Strong-Δ-Rationalizability. The intuitive explanation of this phenomenon is that off-the-path restrictions are not "processed" by players for all steps of reasoning, because they bite at histories that should not be reached if players reason for a sufficient number of steps. Beside the theoretical insight, though, this broad condition for monotonicity is of little practical use: one cannot verify it without actually performing Strong-Δ-Rationalizability. Yet, a very important class of restrictions always satisfies this condition: the belief in a path of play.[3]

Then, not surprisingly, under such conditions Strong-Δ-Rationalizability is actually

[1] i.e. belief as long as compatible with the observed behavior.

[2] The epistemic justification of Strong Rationalizability requires, at each step n, strong belief in all the previous steps of the procedure. For the iterated elimination of strategies, strong belief in step $n-1$ suffices.

[3] i.e. belief in opponents' strategies that comply with a particular path as long as no-one deviates from the path. This is different than strong belief in such strategies, which must be maintained also after own deviations from the path. However, the two restrictions are equivalent: see Section 3.

outcome-equivalent to Selective Rationalizability (Catonini [7]), a refinement of Strong Rationalizability with first-order belief restrictions. The conceptual difference between the two procedure is whether the beliefs in rationality are assigned *epistemic priority* with respect to the beliefs in the restrictions or not, at all orders. When, for instance, the observed behavior of the opponents is not compatible with them being rational and best replying to beliefs in their restricted sets, Selective Rationalizability keeps the belief in rationality (if per se possible) and drops the belief in the restrictions, while Strong-Δ-Rationalizability does not keep the belief in rationality.[4]

The same "workhorse lemma" that delivers the two results above can be provide new insight on existing results. First, the "iterated elimination of never sequential best replies" (of which Strong Rationalizability is the maximal elimination order) is order independent in terms of predicted outcomes. Chen and Micali [9] characterize Strong Rationalizability with the iterated elimination of *distinguishably dominated* strategies,[5] and show that the latter is order independent in terms of predicted outcomes. Here, like in the recent work of Perea [14], we work directly with the iterated deletion of never sequential best replies, thus with strong belief and without dominance characterizations. With this, Strong Rationalizability is shown to refine backward induction, which can be seen as a particular, truncated order of elimination of never sequential best replies. As in Chen and Micali [9] and Perea [15], the outcome equivalence of backward induction and forward induction in perfect information games without relevant ties (originally proved by Battigalli [1] and then also by Heifetz and Perea [10]) follows.

Section 2 introduces the formal framework for the analysis. Section 3 defines the elimination procedures of interest and presents the main results. Section 4 provides the results on order-independence. Section 5 provides a sketch of the proof of the central "workhorse lemma", along with an example. The formal proof of the lemma is in the Appendix.

2 Preliminaries

Primitives of the game.[6] Let I be the finite set of *players*. For any profile $(X_i)_{i\in I}$ and any $\emptyset \neq J \subseteq I$, I write $X_J := \times_{j\in J} X_j$, $X := X_I$, $X_{-i} := X_{I\setminus\{i\}}$. Let $(\overline{A}_i)_{i\in I}$ be the finite

[4] Whether the belief in the restrictions is kept or not is immaterial for the procedure, thus Strong-Δ-Rationalizability can be characterized epistemically with or without *transparency* of the restrictions: see Battigalli and Prestipino [4] for details.

[5] By showing the equivalence of the iterated elimination of *distinguishable* and *conditionally* dominated strategies, where the latter was already proved by Shimoji and Watson [17] to be equivalent to Extensive Form Rationalizability (Pearce [13]), which is in turn equivalent to Strong Rationalizability.

[6] The main notation is almost entirely taken from Osborne and Rubinstein [12].

sets of *actions* potentially available to each player. Let $\overline{H} \subseteq \cup_{t=1,...,T}\overline{A}^t \cup \{\emptyset\}$ be the set of histories, where $h^0 := \{\emptyset\} \in \overline{H}$ is the root of the game and T is the finite horizon. For any $h = (a^1, ..., a^t) \in \overline{H}$ and $l < t$, it holds $h' = (a^1, ..., a^l) \in \overline{H}$, and I write $h' \prec h$.[7] Let $Z := \{z \in \overline{H} : \forall h \in \overline{H}, z \nprec h\}$ be the set of terminal histories (henceforth, *outcomes* or *paths*)[8], and $H := \overline{H} \backslash Z$ the set of non-terminal histories (henceforth, just *histories*). For each $i \in I$, let $A_i : H \rightrightarrows \overline{A}_i$ be the correspondence that assigns to each history h, always observed by player i, the set of actions $A_i(h) \neq \emptyset$[9] available at h. Thus, H has the following property: For every $h \in H$, $(h, a) \in \overline{H}$ if and only if $a \in A(h)$. Note that to simplify notation every player is required to play an action at every history: when a player is not truly active at a history, her set of feasible actions consists of just one "wait" action. For each $i \in I$, let $u_i : Z \to \mathbb{R}$ be the *payoff function*. The list $\Gamma = \left\langle I, \overline{H}, (u_i)_{i \in I} \right\rangle$ is a *finite game with complete information and observable actions*.

Derived objects. A strategy of player i is a function $s_i : h \in H \longmapsto s_i(h) \in A_i(h)$. Let S_i denote the set of all strategies of i. A strategy *profile* $s \in S$ naturally induces a unique outcome $z \in Z$. Let $\zeta : S \to Z$ be the function that associates each strategy profile with the induced outcome. For any $h \in \overline{H}$, the set of strategies of i compatible with h is:

$$S_i(h) := \{s_i \in S_i : \exists z \succeq h, \exists s_{-i} \in S_{-i}, \zeta(s_i, s_{-i}) = z\}.$$

For any $(\overline{S}_j)_{j \in I} \subset S$, let $\overline{S}_i(h) := S_i(h) \cap \overline{S}_i$. For any $J \subseteq I$, let $H(\overline{S}_J) := \left\{h \in H : \overline{S}_J(h) \neq \emptyset\right\}$ denote the set of histories compatible with $\overline{S}_J$. For any $h = (h', a) \in \overline{H}$, let $p(h)$ denote the immediate predecessor h' of h.

Since the game has observable actions, each history $h \in H$ is the root of a subgame $\Gamma(h)$. In $\Gamma(h)$, all the objects defined above will be denoted with h as superscript, except for single histories and outcomes, which will be identified with the corresponding history or outcome of the whole game, and not redefined as shorter lists of action profiles. For any $h \in H$, $s_i^h \in S_i^h$, and $\widehat{h} \succ h$, $s_i^h | \widehat{h}$ will denote the strategy $s_i^{\widehat{h}} \in S_i^{\widehat{h}}$ such that $s_i^{\widehat{h}}(\widetilde{h}) = s_i^h(\widetilde{h})$ for all $\widetilde{h} \succeq \widehat{h}$. For any $\overline{S}_i^h \subseteq S_i^h$, $\overline{S}_i^h | \widehat{h}$ will denote the set of all strategies $s_i^{\widehat{h}} \in S_i^{\widehat{h}}$ such that $s_i^{\widehat{h}} = s_i^h | \widehat{h}$ for some $s_i^h \in \overline{S}_i^h$.

Beliefs. In this dynamic framework, beliefs are modeled as Conditional Probability Systems (Renyi, [16]; henceforth, CPS).

Definition 1 *A Conditional Probability System on* $(S_{-i}, (S_{-i}(h))_{h \in H})$ *is a mapping* $\mu(\cdot|\cdot) : 2^{S_{-i}} \times \{S_{-i}(h)\}_{h \in H} \to [0, 1]$ *satisfying the following axioms:*

[7] $\overline{H}$ endowed with the precedence relation $\prec$ is a tree with root h^0.

[8] "Path" will be used with emphasis on the moves, and "outcome" with emphasis on the end-point of the game.

[9] When player i is not truly active at history h, $A_i(h)$ consists of just one "wait" action.

CPS-1 for every $C \in (S_{-i}(h))_{h\in H}$, $\mu(\cdot|C)$ *is a probability measure on* S_{-i}*;*

CPS-2 for every $C \in (S_{-i}(h))_{h\in H}$, $\mu(C|C) = 1$*;*

CPS-3 for every $E \in 2^{S_{-i}}$ *and* $C, D \in (S_{-i}(h))_{h\in H}$, *if* $E \subseteq D \subseteq C$, *then* $\mu(E|C)=\mu(E|D)\mu(D|C)$.

The set of all CPS's on $(S_{-i}, (S_{-i}(h))_{h\in H})$ *is denoted by* $\Delta^H(S_{-i})$.

For brevity, the conditioning events will be indicated with just the information set, which represents all the information acquired by players through observation. For each set $J \subseteq I\backslash\{i\}$ of opponents of player i, and for each set of strategy sub-profiles $\overline{S}_J \subseteq S_J$, I say that a CPS $\mu_i \in \Delta^H(S_{-i})$ *strongly believes* $\overline{S}_J$ if, for all $h \in H(\overline{S}_J)$, $\mu_i(\overline{S}_J \times S_{I\backslash(J\cup\{i\})}|h) = 1$.

Rationality. I consider players who reply rationally to their conjectures. By rationality I mean that players, at every information set, choose an action that maximizes expected utility given the conjecture about how deviators will play and the expectation to reply rationally again in the continuation of the game. This is equivalent (see Battigalli [2]) to playing a *sequential best reply* to the CPS.

Definition 2 *Fix* $\mu_i \in \Delta^H(S_{-i})$. *A strategy* $s_i \in S_i$ *is a sequential best reply to* μ_i *if for every* $h \in H(s_i)$,[10] s_i *is a continuation best reply to* $\mu_i(\cdot|h)$, *i.e. for every* $\widetilde{s}_i \in S_i(h)$,

$$\sum_{s_{-i}\in S_{-i}(h)} u_i(\zeta(s_i, s_{-i}))\mu_i(s_{-i}|h) \geq \sum_{s_{-i}\in S_{-i}(h)} u_i(\zeta(\widetilde{s}_i, s_{-i}))\mu_i(s_{-i}|h).$$

I say that a strategy s_i is *rational* if it is a sequential best reply to some $\mu_i \in \Delta^H(S_{-i})$. The set of sequential best replies to μ_i is denoted by $\rho(\mu_i)$. For each $h \in H$, the set of continuation best replies to $\mu_i(\cdot|h)$ is denoted by $\widehat{r}(\mu_i, h)$. The set of best replies to a conjecture $\nu_i \in \Delta(S_{-i})$ in the normal form of the game is denoted by $r(\nu_i)$.

3 Outcome monotonicity

I provide a very general notion of elimination procedure for a subgame $\Gamma(h)$, which encompasses all the procedure I am ultimately interested in, or that will be needed for the proofs.

Definition 3 *Fix* $h \in H$. *An elimination procedure in* $\Gamma(h)$ *is a sequence* $((S^h_{i,q})_{i\in I})^{\infty}_{q=0}$ *where, for every* $i \in I$,

[10] It would be totally immaterial to require s_i to be optimal also at the histories precluded by itself.

EP1 $S_{i,0}^h = S_i^h$*;*

EP2 $S_{i,n-1}^h \supseteq S_{i,n}^h$ *for all* $n \in \mathbb{N}$*;*

EP3 for every $s_i^h \in S_{i,\infty}^h = \cap_{n\in\mathbb{N}} S_{i,n}^h$*, there exists* μ_i^h *that strongly believes* $(S_{-i,q}^h)_{q=0}^{\infty}$ *such that* $s_i^h \in \rho(\mu_i^h) \subseteq S_{i,\infty}^h$*.*

Lemma 1 *For every elimination procedure* $((S_{i,q}^h)_{i\in I})_{q=0}^{\infty}$ *and every* $\widehat{h} \succ h$*,* $((S_{i,q}^h(\widehat{h})|\widehat{h})_{i\in I})_{q=0}^{\infty}$ *is an elimination procedure.*

Proof. EP1 and EP2 are obvious. To prove EP3, note the following. For every $i \in I$ and $s_i^{\widehat{h}} \in S_{i,\infty}^h(\widehat{h})|\widehat{h}$, there exists $s_i^h \in S_{i,\infty}^h$ such that $s_i^h|\widehat{h} = s_i^{\widehat{h}}$. By EP3 for $((S_{i,q}^h)_{i\in I})_{q=0}^{\infty}$, there exists μ_i^h that strongly believes $(S_{-i,q}^h)_{q=0}^{\infty}$ such that $s_i^h \in \rho(\mu_i^h) \subseteq S_{i,\infty}^h$. Thus, the pushforward $\mu_i^{\widehat{h}}$ of $(\mu_i^h(\cdot|\widetilde{h}))_{\widetilde{h}\in H^{\widehat{h}}}$ through the map $s_{-i}^h \longmapsto s_{-i}^h|\widehat{h}$ strongly believes $(S_{-i,q}^h(\widehat{h})|\widehat{h})_{q=0}^{\infty}$. Clearly $s_i^{\widehat{h}} \in \rho(\mu_i^{\widehat{h}})$. Finally, fix $\overline{s}_i^{\widehat{h}} \in \rho(\mu_i^{\widehat{h}})$. Define $\overline{s}_i^h$ as $\overline{s}_i^h(\widetilde{h}) = s_i^h(\widetilde{h})$ for all $\widetilde{h} \not\succeq \widehat{h}$ and $\overline{s}_i^h|\widehat{h} = \overline{s}_i^{\widehat{h}}$ for all $\widetilde{h} \succeq \widehat{h}$. Clearly $\overline{s}_i^h \in \rho(\mu_i^h)$. Thus, $\overline{s}_i^{\widehat{h}} \in S_{i,\infty}^h(\widehat{h})|\widehat{h}$. ■

Indeed, elimination procedures have been defined purposedly to encompass the implications in the subgames of traditional elimination procedures for the whole game. In a subgame, substrategies can be eliminated "exogenously" and not because they are not sequential best replies to any valid conjecture in the subgame. On the other hand, substrategies can survive even if the opponents do not reach the subgame anymore. Note that the elimination can stop for some steps and then resume: for this reason, EP2 allows a weak inclusion at all steps

Now I specialize Definition [16] for the procedures in the whole game I am ultimately interested in.

Definition 4 *An elimination procedure* $((S_{i,q})_{i\in I})_{q=0}^{\infty}$ *is "unconstrained" when for every* $n \in \mathbb{N}$*,* $i \in I$*, and* μ_i *that strongly believes* $(S_{-i,q})_{q=0}^{n-1}$*,* $\rho(\mu_i) \subseteq S_{i,n}$*.*

Definition 5 *An elimination procedure* $((S_{i,q}^h)_{i\in I})_{q=0}^{\infty}$ *is "maximal" when for every* $n \in \mathbb{N}$*,* $i \in I$*, and* $s_i \in S_{i,n}$*,* $s_i \in \rho(\mu_i)$ *for some* μ_i *that strongly believes* $(S_{-i,q})_{q=0}^{n-1}$*.*

Definition 6 *Strong Rationalizability (Battigalli and Siniscalchi, [5]) is the unique unconstrained and maximal elimination procedure. Let* $((S_i^q)_{i\in I})_{q=0}^{\infty}$ *denote it, and let* M *be the* $n \in \mathbb{N}$ *such that* $S^{n-1} \neq S^n = S^{n+1}$*.*

Definition 7 *For each* $i \in I$*, fix* $\Delta_i \subseteq \Delta^H(S_{-i}^h)$*. Strong-*$\Delta$*-Rationalizability (Battigalli [3], Battigalli and Siniscalchi [6]) is the elimination procedure* $((S_{i,\Delta}^q)_{i\in I})_{q=0}^{\infty}$ *such that, for every* $n \in \mathbb{N}$*,* $i \in I$*, and* $s_i \in S_i$*,* $s_i \in S_{i,n}$ *if and only if* $s_i \in \rho(\mu_i)$ *for some* $\mu_i \in \Delta_i$ *that strongly believes* $(S_{-i,\Delta}^q)_{q=0}^{n-1}$*.*

Definition 8 *For each $i \in I$, fix $\Delta_i \subseteq \Delta^H(S^h_{-i})$. Selective Rationalizability (Catonini [7]) is the elimination procedure $((S^q_{i,R\Delta})_{i\in I})^\infty_{q=0}$ such that $(S^q_{R\Delta})^M_{q=0} = (S^q)^M_{q=0}$ and for every $n > M$, $i \in I$, and $s_i \in S_i$, $s_i \in S^n_{i,R\Delta}$ if and only if $s_i \in \rho(\mu_i)$ for some $\mu_i \in \Delta_i$ that strongly believes $(S^q_{-i,R\Delta})^{n-1}_{q=0}$.*[11]

The main technical result of the paper is the outcome inclusion between two elimination procedures with the following feature. Take the final output of the first procedure and fix beliefs that justify the surviving strategies. Consider all the beliefs that, along the paths predicted by the first procedure, assign the same probability distribution over such paths as one of the fixed beliefs. Suppose that, in both procedures, the sequential best replies to these beliefs always survive. Then, the final output of the second procedure predicts all such paths.

Lemma 2 *Fix $h \in H$, two elimination procedures $((\overline{S}^h_{i,q})_{i\in I})^\infty_{q=0}$, $((S^h_{i,q})_{i\in I})^\infty_{q=0}$, and, for every $i \in I$, a map $\overline{\overline{\mu}}^h_i : \overline{S}^h_{i,\infty} \longrightarrow \Delta^{H^h}_i(S^h_{-i})$ such that $\overline{\overline{\mu}}^h_i(s^h_i)$ strongly believes $(\overline{S}^h_{-i,q})^\infty_{q=0}$ and $s^h_i \in \rho(\overline{\overline{\mu}}^h_i(s^h_i)) \subseteq \overline{S}^h_{i,\infty}$ for all $s^h_i \in \overline{S}^h_{i,\infty}$. Suppose that for every $i \in I$, $s^h_i \in \overline{S}^h_{i,\infty}$, $m \in \mathbb{N}$, and μ^h_i that strongly believes $(S^h_{-i,q})^{m-1}_{q=0}$ (resp., $(\overline{S}^h_{-i,q})^{m-1}_{q=0}$) with $\mu^h_i(S_{-i}(z)|\widetilde{h}) = \overline{\overline{\mu}}^h_i(s^h_i)(S_{-i}(z)|\widetilde{h})$ for all $\widetilde{h} \in H(\overline{S}^h_\infty)$ and $z \in Z^{\widetilde{h}} \cap \zeta(\overline{S}^h_\infty)$, $\rho(\mu^h_i) \subseteq S^h_{i,m}$ (resp., $\rho(\mu^h_i) \subseteq \overline{S}^h_{i,m}$). Then $\zeta(\overline{S}^h_\infty) \subseteq \zeta(S^h_\infty)$.*

Section 5 contains a sketch of the proof of the lemma. Now I focus on the implications of the lemma for the elimination procedures of interest.

Consider first-order belief restrictions $(\Delta_i)_{i\in I}$ with the following characteristic: for each player i and CPS μ_i, only the beliefs at the strongly-Δ-rationalizable histories about the strongly-Δ-rationalizable paths matter to determine whether μ_i belongs to Δ_i or not. Then, Strong-Δ-Rationalizability satifies the hypotheses of Lemma 2 as first elimination procedure, whereas Strong Rationalizability, being an unconstrained procedure, satisfies the hypotheses of Lemma 2 as second elimination procedure. The desired outcome inclusion result with respect to belief restrictions that "do not end up off-path" obtains.

Theorem 1 *Fix $(\Delta_i)_{i\in I} \subseteq \times_{i\in I}\Delta^H(S_{-i})$. Suppose that for each $i \in I$ and $\mu_i, \mu'_i \in \Delta^H(S_{-i})$, if $\mu_i \in \Delta_i$ and $\mu'_i(S_{-i}(z)|\widetilde{h}) = \mu'_i(S_{-i}(z)|\widetilde{h})$ for all $\widetilde{h} \in H(S^\infty_\Delta)$ and $z \in \zeta(S^\infty_\Delta)$, then $\mu'_i \in \Delta_i$. Then, $\zeta(S^\infty_\Delta) \subseteq \zeta(S^\infty)$.*

[11] Selective Rationalizability is defined in [7] under the more restrictive assumption of *independent rationalization*. That is, a valid μ_i is required to strongly believe $(S^q_{j,R\Delta})^{n-1}_{q=0}$ for all $j \neq i$, in place of just $(S^q_{-i,R\Delta})^{n-1}_{q=0}$. However, this assumption is immaterial for the result on Selective Rationalizability of this paper (Theorem 3).

Proof. For each $i \in I$ and $s_i \in S^{\infty}_{i,\Delta}$, fix any $\overline{\overline{\mu}}^h_i(s^h_i) \in \Delta_i$ that strongly believes $(S^q_{-i,\Delta})^{\infty}_{q=0}$ such that $s_i \in \rho(\mu_i)$. By hypothesis of this theorem, the hypothesis of Lemma 2 obtains. For every $m \in \mathbb{N}$ and μ_i that strongly believes $(S^q_{-i})^{m-1}_{q=0}$, $\rho(\mu_i) \in S^m_i$. Thus, by Lemma 2, $\zeta(S^{\infty}_{\Delta}) \subseteq \zeta(S^{\infty})$. ■

As discussed in the Introduction, Theorem 1 provides insight on what can determine the non monotonicity of predictions with respect to belief restrictions: the presence of off-the-path belief restrictions. Yet, it is of little help in determining ex-ante which belief restrictions preserve common strong belief in rationality and which do not. This is because whether restrictions are off-path or not has to be assessed with respect of the final output of Strong-Δ-Rationalizability itself.

Consider now first-order belief restrictions that correspond to the belief in a specific path $z \in Z$. That is, at the beginning of the game, players believe that the opponents will play compatibly with the path. By CPS-3, this belief is maintained as long as no deviation from the path occurs. Moreover, assume that if a player deviates from the path, the opponents keep believing that the other players were not planning to deviate. This is coherent with the notion of *belief in the (path) agreement* adopted in [7]. All this coincides with assuming that every player i strongly believes in $S_j(z)$ for all $j \neq i$. Preliminarly, I show that this is equivalent to the belief in $S_{-i}(z)$ on path only.

Lemma 3 *Fix $z \in Z$. For each $i \in I$, let Δ_i be the set of all μ_i's such that $\mu_i(S_{-i}(z)|h) = 1$ for all $h \prec z$, and let Δ^*_i be the set of all μ_i's that strongly believe $S_j(z)$ for all $j \neq i$. Then, $S^{\infty}_{\Delta} = S^{\infty}_{\Delta^*}$ and $S^{\infty}_{R\Delta} = S^{\infty}_{R\Delta^*}$.*

Proof. Fix $n \geq 0$ and suppose to have shown that for each $m \leq n$, $S^m_{\Delta} = S^m_{\Delta^*}$ ($S^0_{\Delta} = S^0_{\Delta^*}$.trivially holds). If $S^n_{\Delta} = \emptyset$, $S^{n+1}_{\Delta} = S^{n+1}_{\Delta^*} = \emptyset$. Else, for each $i \in I$, there exists $\overline{\mu}_i \in \Delta_i$ that strongly believes $(S^q_{-i,\Delta})^{n-1}_{q=0}$ such that $\rho(\overline{\mu}_i) \cap S_i(z) \neq \emptyset$. Fix $i \in I$ and $s_i \in S_i \backslash S_i(z)$. Let $m := \max\left\{q \leq n : s_i \in S^q_{i,\Delta}\right\}$. If $m > 0$, there exists $\mu_i \in \Delta_i$ that strongly believes $(S^q_{-i,\Delta})^{m-1}_{q=0}$ such that $s_i \in \rho(\mu_i)$. Fix $\mu^*_i \in \Delta_i$ that strongly believes $(S^q_{-i,\Delta})^{m-1}_{q=0}$ such that $\mu^*_i(\cdot|h) = \overline{\mu}_i(\cdot|h)$ for all $h \prec z$ and $\mu^*_i(\cdot|\widetilde{h}) = \mu_i(\cdot|\widetilde{h})$ for all $\widetilde{h} \in H(S_i(z)) \backslash H(S_{-i}(z))$ (it is compatible with CPS-3 because $\overline{\mu}_i(S_{-i}(\widetilde{h})|h) = 0$ for all $h \prec z$ and $\widetilde{h} \in H(S_i(z)) \backslash H(S_{-i}(z))$). Then, there exists $s^*_i \in \rho(\mu^*_i)(z) \subseteq S^m_{i,\Delta}$ such that for all $\widetilde{h} \in H(s_i) \cap H(S_i(z)) \backslash H(S_{-i}(z))$, $s^*_i(\widetilde{h}) = s_i(\widetilde{h})$. If $m = 0$, fix the unique $s^*_i \in S_i(z)$ such that for all $\widetilde{h} \not\prec z$, $s^*_i(\widetilde{h}) = s_i(\widetilde{h})$. For each $h \in H(S_i(z))$, let $\eta^h(s_i) := s^*_i$. For each $h \notin H(S_i(z))$, let $\eta^h(s_i) := s_i$. For all $s_i \in S_i(z)$ and $h \in H$, let $\eta^h(s_i) := s_i$.

Fix now $i \in I$ and $\mu_i \in \Delta_i$ that strongly believes $(S^q_{-i,\Delta})^n_{q=0}$. Note that for each $s_i \in S_i$ and $h \in H$, if $s_i \in S_i(h)$, $\eta^h(s_i) \in S_i(h)$, and if $h \in H(S_i(z))$, $\eta^h(s_i) \in S_i(z)$. Thus, I can construct $\mu^*_i \in \Delta^*_i$ that strongly believes $(S^q_{-i,\Delta})^n_{q=0} = (S^q_{-i,\Delta^*})^n_{q=0}$ as, for all $h \in H$,

$\mu_i^*((s_j)_{j\neq i}|h) = \mu_i(((\eta^h)^{-1}(s_j))_{j\neq i}|h)$. For each $h \prec z$, since $\mu_i(S_{-i}(z)|h) = 1$, $\mu_i^*(\cdot|h) = \mu_i(\cdot|h)$, and for each $h \not\prec z$ and $\widetilde{z} \succ h$, by construction, $\mu_i^*(S_{-i}(\widetilde{z})|h) = \mu_i(S_{-i}(\widetilde{z})|h)$. Hence, $\rho(\mu_i) = \rho(\mu_i^*)$. So, $S_\Delta^{n+1} \subseteq S_{\Delta^*}^{n+1}$. By $\Delta_i^* \subseteq \Delta_i$ and $(S_{-i,\Delta}^q)_{q=0}^n = (S_{-i,\Delta^*}^q)_{q=0}^n$, $S_{\Delta^*}^{n+1} \subseteq S_\Delta^{n+1}$.

The proof can be repeated for Selective Rationalizability with $n \geq M$ in place of $n \geq 0$, where $(S_{R\Delta}^q)_{q=0}^M = (S_{R\Delta^*}^0)_{q=0}^M$ holds by definition. ■

If the belief restrictions on $S_{-i}(z)$ only along z end up off the paths predicted at some intermediate step of Strong-Δ-Rationalizability, the procedure yields an empty set at the following step. Otherwise, Theorem 1 can be easily applied and monotonocity of strategic reasoning with respect to path restrictions obtains.

Theorem 2 *Fix $z \in Z$. Let Δ_i^* be the set of all μ_i's that strongly believe $S_j(z)$ for all $j \neq i$. Then $\zeta(S_{\Delta^*}^\infty) \subseteq \zeta(S^\infty)$.*

Proof. For each $i \in I$, let Δ_i be the set of all μ_i's such that $\mu_i(S_{-i}(z)|h) = 1$ for all $h \prec z$. If $S_\Delta^\infty = \emptyset$, $\zeta(S_\Delta^\infty) \subseteq \zeta(S^\infty)$ is trivially true, so suppose $S_\Delta^\infty \neq \emptyset$. For each $i \in I$, and $s_i \in S_{i,\Delta}^\infty$, $s_i \in \rho(\overline{\mu}_i)$ for some $\overline{\mu}_i \in \Delta_i$. For each $\overline{\mu}_i \in \Delta_i$ and μ_i with $\mu_i(S_{-i}(z)|h) = \overline{\mu}_i(S_{-i}(z)|h)$ for all $h \prec z$, $\mu_i \in \Delta_i$. Thus, the hypotheses of Theorem 1 hold, and $\zeta(S_\Delta^\infty) \subseteq \zeta(S^\infty)$. Then, by Lemma 3, $\zeta(S_{\Delta^*}^\infty) \subseteq \zeta(S^\infty)$. ■

Also Selective Rationalizability eventually saves only strategies that are sequential best replies to beliefs in the restricted sets. Therefore, for path restrictions, Lemma 2 holds with Selective Rationalizability and Strong-Δ-Rationalizability regardless of the roles assigned to the two procedures. Then, the outcome equivalence of the two procedures under path restrictions obtains.

Theorem 3 *Fix $z \in Z$. Let Δ_i^* be the set of all μ_i's that strongly believe $S_j(z)$ for all $j \neq i$. Then $\zeta(S_{\Delta^*}^\infty) = \zeta(S_{R\Delta^*}^\infty)$.*

Proof. For each $i \in I$, let Δ_i be the set of all μ_i's such that $\mu_i(S_{-i}(z)|h) = 1$ for all $h \prec z$. First I show that $\zeta(S_\Delta^\infty) \subseteq \zeta(S_{R\Delta}^\infty)$. If $S_\Delta^\infty = \emptyset$ it is trivially true, so suppose $S_\Delta^\infty \neq \emptyset$. For each $i \in I$, and $s_i \in S_{i,\Delta}^\infty$, $s_i \in \rho(\overline{\mu}_i)$ for some $\overline{\mu}_i \in \Delta_i$. For each $\overline{\mu}_i \in \Delta_i$ and μ_i with $\mu_i(S_{-i}(z)|h) = \overline{\mu}_i(S_{-i}(z)|h)$ for all $h \prec z$, $\mu_i \in \Delta_i$. Thus, the hypotheses of Lemma 2 hold. So, $\zeta(S_\Delta^\infty) \subseteq \zeta(S_{R\Delta}^\infty)$. The same proof can be repeated for $\zeta(S_\Delta^\infty) \supseteq \zeta(S_{R\Delta}^\infty)$. Hence $\zeta(S_\Delta^\infty) = \zeta(S_{R\Delta}^\infty)$. Then, by Lemma 3, $\zeta(S_{\Delta^*}^\infty) = \zeta(S_{R\Delta^*}^\infty)$. ■

The last two theorems clearly hold with strong belief in $S_{-i}(z)$ instead of $(S_j(z))_{j\neq i}$.

4 Order independence

In absence of belief restrictions, that is with unconstrained elimination procedures, the hypotheses of Theorem 2 clearly hold. An unconstrained elimination procedure is what I refered to in the Introduction as an order of iterated elimination of never sequential best replies. Thus, using Theorem 2 in both directions with the maximal unconstrained elimination procedure and any non maximal one, the order independence of iterated elimination of never sequential best replies in terms of predicted outcomes obtains.

Theorem 4 *For any unconstrained elimination procedure* $((S_{i,q})_{i\in I})_{q=0}^{\infty}$, $\zeta(S_{\infty}) = \zeta(S^{\infty})$.

Proof. Any two uncontrained elimination procedures, taken in both orders, obviously satisfy the hypotheses of Lemma 2. ■

Backward Induction is an elimination procedure of actions where an action of a player at a history is eliminated when it is not optimal against any belief about the surviving current and future actions of the opponents. An outcome equivalent elimination procedure of strategies deletes all the strategies that reach the history and prescribe such action. These strategies are clearly not optimal under strong belief in the surviving strategies of the opponents. A strategy, instead, may be never sequential best reply as a whole, but still survive Backward Induction because, at each history, it prescribes an action which is part of some optimal continuation plan.

Definition 9 *Backward Induction is a sequence* $((S_{i,B}^{q})_{i\in I})_{q=0}^{\infty}$ *where, for every* $i \in I$,

BI1 $S_{i,B}^{0} = S_i$;

BI2 for each $n \in \mathbb{N}$ *and* $s_i \in S_i$, $s_i \in S_{i,B}^{n}$ *if and only if* $s_i \in S_{i,B}^{n-1}$ *and for each* $h \in H(s_i)$, *there exist* μ_i *that strongly believes* $S_{-i,B}^{n-1}$ *and* $\widetilde{s}_i \in S_i(h)$ *such that* $\widetilde{s}_i \in \widehat{r}(\mu_i, h)$ *and* $\widetilde{s}_i(h) = s_i(h)$.[12]

Backward induction can be seen as part of non-maximal, unconstrained elimination procedure.

Lemma 4 *Let* N *be the smallest* n *such that* $S_B^n = S_B^{n+1}$. *Let* $((\widetilde{S}_i^q)_{i\in I})_{q=0}^{N} := ((S_{i,B}^{q})_{i\in I})_{q=0}^{N}$ *and, for every* $n > N$, $i \in I$, *and* $s_i \in S_i$, *let* $s_i \in \widetilde{S}_i^n$ *if and only if there exists* μ_i *that strongly believes* $(\widetilde{S}_{-i}^q)_{q=0}^{n-1}$ *such that* $s_i \in \rho(\mu_i)$. *Thus,* $((\widetilde{S}_i^q)_{i\in I})_{q=0}^{\infty}$ *is an unconstrained elimination procedure.*

[12] Note that for any $h, h' \in H$ with $h \not\preceq h' \not\preceq h$ and $S_{i,B}^{n}(h) \cap S_{i,B}^{n}(h') \neq \emptyset$, and for any $s_i \in S_{i,B}^{n}(h)$ and $s_i' \in S_{i,B}^{n}(h')$, there exists $s_i'' \in S_{i,B}^{n}(h) \cap S_{i,B}^{n}(h')$ such that $s_i''|h = s_i$ and $s_i''|h' = s_i'$. Thus, all combinations of backward induction moves survive and the use of strategies in only for coherence with the framework of this paper.

Proof. EP1 is satisfied by BI1. EP3 is satisfied by finiteness of the game. EP2 is satisfied for all $n > N$ by construction. It remains to show that EP2 is satisfied for $n \leq N$. Fix $i \in I$, μ_i that strongly believes $(\widetilde{S}^q_{-i})^{n-1}_{q=0}$, and $s_i \in \rho(\mu_i)$. Then, for all $h \in H(s_i)$, $s_i \in \widehat{r}(\mu_i, h)$. Thus, by BI2, $s_i \in \widetilde{S}^n_i$. ■

Being an unfinished, unconstrained elimination procedure, the backward induction procedure predicts a superset of the outcomes predicted by Strong Rationalizability.

Theorem 5 *Every strongly rationalizable outcome is a backward induction outcome.*

Proof. Immediate from Lemma 4 and Theorem 4. ■

Since in perfect information games without relevant ties the backward induction outcome is unique, the following obtains.

Corollary 6 (Battigalli, [1]) *In every perfect information game without relevant ties, Strong Rationalizability and Backward Induction yield the same unique outcome.*

5 Proof of the main lemma.

The proof proceeds as follows. For simplicity, assume that there are two players, i and j; the argument extends immediately to games with more than 2 players. We show by induction that for every $\overline{s}^h_i \in \overline{S}^h_{i,\infty}$ and $n \in \mathbb{N}$, there are: (1) $\overline{\mu}^h_i(\overline{s}^h_i)$ that strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$ and, along the paths induced by $\overline{S}^h_\infty$ (henceforth, just "paths"), mimicks[13] a CPS $\overline{\overline{\mu}}^h_i(\overline{s}^h_i)$ that justifies that $\overline{s}^h_i \in \overline{S}^h_{i,\infty}$; (2) $s^h_i \in \rho(\overline{\mu}^h_i(\overline{s}^h_i))$ that mimicks $\overline{s}^h_i$ along the paths; and the same for j. By the assumption about $((S^h_{k,q})_{k\in I})^\infty_{q=0}$, $s^h_i \in \overline{S}^h_{i,n}$. All such s^h_i's allow to construct at step $n+1$ a CPS $\overline{\mu}^h_j(\overline{s}^h_j)$ as in (1) for each $\overline{s}^h_j \in \overline{S}^h_{j,\infty}$.

Now, suppose by contradiction that for some $\overline{s}^h_j \in \overline{S}^h_{j,\infty}$, every such $\overline{\mu}^h_j(\overline{s}^h_j)$ does not justify any strategy s^h_j that mimicks $\overline{s}^h_j$ along the paths. For each history $\widehat{h}$ that immediately follows a unilateral deviation of player j from the paths, consider the most pessimistic belief of j over $S^h_{i,n}(\widehat{h})|\widehat{h}$. For each $\overline{s}^h_i \in \overline{S}^h_{i,\infty}$, by induction hypothesis there is $s^h_i \in S^h_{i,n}$ that mimicks $\overline{s}^h_i$ along the paths and is a sequential best reply to a belief, $\overline{\mu}^h_i(\overline{s}^h_i)$, that assigns probability zero to each deviation of j until it occurs. Then (by Lemma 5), $\overline{\mu}^h_i(\overline{s}^h_i)$ can be combined with any beliefs after j's deviations, and s^h_i can be combined with any reactions to such beliefs. This is proved by Lemma 6 So, player j can have a belief μ^h_j that mimicks

[13] In the sense of the statement of the lemma.

$\overline{\overline{\mu}}_j^h(\overline{s}_j^h)$ along the paths and, at the same time, features the most pessimistic belief after each deviation. By the initial assumption of this paragraph, player j will deviate under μ_j^h. Let $\widehat{h}$ be a history that may immediately follow the deviation under μ_j^h. So, player j will deviate towards $\widehat{h}$ also when μ_j is constructed with a less pessimistic belief over $S_{i,n}^h(\widehat{h})|\widehat{h}$. This is proved by Lemma 7 As said, player i can be surprised by the deviation. Thus, she can allow $\widehat{h}$ and, at the same time, have any belief thereafter. So, with the assumption on $((S_{i,q}^h)_{i\in I})_{q=0}^{\infty}$, $S_n^h(\widehat{h})|\widehat{h}$ features all the sequential best replies to CPS's $\mu_i^{\widehat{h}}$ that strongly believe $(S_{-i,q}^{\widehat{h}})_{q=0}^{n}$ for all $i \in I$.

Refine $S_n^h(\widehat{h})|\widehat{h}$ by iteratively eliminating strategies that are not sequential best replies to any $\mu_i^{\widehat{h}}$ that strongly believes $(S_{-i,q}^{\widehat{h}})_{q=0}^{m}$. Then, we obtain an elimination procedure $((\overline{S}_{k,q}^{\widehat{h}})_{k\in I})_{q=0}^{\infty}$ with non-empty $\overline{S}_{\infty}^{\widehat{h}}$ that satisfies the assumption of the lemma. That the deviation is profitable against all beliefs over $\overline{S}_{i,\infty}^{\widehat{h}}$ w.r.t. remaining on the paths against $\overline{\overline{\mu}}_j^h(s_j^h)$ implies that also the elimination procedure $((S_{k,q}^{\widehat{h}})_{k\in I})_{q=0}^{\infty} := ((\overline{S}_{k,q}^h(\widehat{h})|\widehat{h})_{k\in I})_{q=0}^{\infty}$ satisfies the assumption of the lemma. Note the inversion of the roles of the two procedures with respect to the original procedures from which they have been derived. If the lemma holds in the subgame $\Gamma(\widehat{h})$, we have the desired contradiction: $S_{\infty}^{\widehat{h}}$ is non-empty too, hence $\widehat{h} \in H(\overline{S}_{\infty}^h)$, but $\widehat{h}$ follows a unilateral deviation from the paths induced by $\overline{S}_{\infty}^h$. Proceeding by induction on the depth of subgames and observing that the lemma clearly holds for subgames of depth 1, the proof is complete.

Now we follow the sketch above on an example. Consider the following game.

$A\backslash B$	W	E
N	$2,2$	$\cdot -$
S	$0,0$	$1,1$

$\longrightarrow$

$A\backslash B$	L	C	R
U	$1,1$	$1,0$	$0,0$
M	$0,0$	$0,1$	$1,0$
D	$0,0$	$0,0$	$0,3$

Let $((S_{i,q}^{h^0})_{i\in I})_{q=0}^{\infty}$ be Strong Rationalizability: $((S_i^q)_{i\in I})_{q=0}^{\infty}$. At the first step, Ann eliminates $N.D$. At the second step, Bob eliminates $E.R$. At the third step, Ann eliminates $N.M$. At the fourth step, Bob eliminates $E.C$. The final output is $S^{\infty} = (S, N.U) \times (W, E.L)$. Strong Rationalizability trivially satisfies the assumption of the lemma.

For each player $i = A, B$, let Δ_i be the set of CPS's that strongly believe in opponents' strategies that comply with the path (N, W):

$$\Delta_i := \left\{\mu_i \in \Delta_i^H(S_{-i}) : \mu_i(S_{-i}((N,W))|h^0) = 1\right\},\ i = A, B.$$

Let $((\overline{S}_{i,q}^{h^0})_{i\in I})_{q=0}^{\infty}$ be Strong-Δ-Rationalizability: $((S_{i,\Delta}^q)_{i\in I})_{q=0}^{\infty}$. At the first step, Ann eliminates S and $N.D$, and Bob eliminates $E.L$ and $E.C$. At the second step, Ann eliminates

$N.U$ and Bob eliminates $E.R$. The final output is: $S^\infty_\Delta = \{(N.M, W)\}$.

Let $\overline{s}_A = N.M$, $\overline{s}_B = W$, $\overline{\overline{\mu}}^h_A(N.M) = (\delta_W, \delta_{E.R})$, and $\overline{\overline{\mu}}^h_B(W) = (\delta_{N.M}, \delta_{N.M})$, where δ_s indicates a Dirac measure on s. For every $n \in \mathbb{N}$, $i = A, B$, and μ_i that strongly believes $S^{n-1}_{-i,\Delta}, ..., S^0_{-i,\Delta}$ with $\mu_i(S_{-i}((N,W))|h^0) = \overline{\overline{\mu}}^h_i(S_{-i}((N,W))|h^0) = 1$, $\rho(\mu_i) \subseteq S^n_{i,\Delta}$. So, $((S^q_{i,\Delta})_{i\in I})^\infty_{q=0}$ satisfies the assumption of the lemma. Indeed, $\zeta(S^\infty_\Delta) = \{(N,W)\} \subseteq \zeta(S^\infty)$, but $S^\infty_{A,\Delta} \cap S^\infty_A = \emptyset$.

Now we follow the sketch above. Fix $n \in \mathbb{N}$ and suppose to have shown that for each $i \in A, B$, there exist:

1. $\overline{\mu}_i(\overline{s}_i)$ that strongly believes $S^{n-1}_{-i}, ..., S^0_{-i}$ with $\overline{\mu}_i(\overline{s}_i)(S_{-i}((N,W))|h^0) = \overline{\overline{\mu}}^h_i(S_{-i}((N,W))|h^0) = 1$;

2. $s_i \in \rho(\overline{\mu}_i(\overline{s}_i)) \subseteq S^n_i$ with $s_i(h^0) = N$ if $i = A$, $s_i(h^0) = W$ if $i = B$.

Suppose that (♠) for every μ_B that strongly believes $S^n_A, ..., S^0_A$ with $\mu_B(S_A((N,W))|h^0) = \overline{\overline{\mu}}^h_B(S_A((N,W))|h^0) = 1$, $\rho(\mu_B) \cap S_B((N,W)) = \emptyset$. For each $a \in S^n_A((N,E))|(N,E)$, fix $s_A \in S^n_A((N,W))$ with $s_A|(N,E) = a$; there exists μ_B that strongly believes $S^n_A, ..., S^0_A$ with $\mu_B(s_A|h^0) = 1$, and by (♠), $\rho(\mu_B) \subseteq S_B((N,E))$. For each $b \in S^n_B((N,E))|(N,E)$, fix $s_B \in S^n_B((N,E))$ with $s_B|(N,E) = B$; there exists μ_A that strongly believes $S^{n-1}_B, ..., S^0_B$ with $\mu_A(\cdot|h^0) = \overline{\mu}_i(\overline{s}_A)(\cdot|h^0)$ and $\mu_A(s_B|(N,E)) = 1$, and $\rho(\mu_A) \subseteq S^n_A((N,E))$.

Let $(\overline{S}^{(N,E)}_q)^\infty_{q=0} = ((S^q((N,E))|(N,E))^n_{q=0}(\overline{S}^{(N,E)}_q)^\infty_{q=n+1})$, where for each $m \geq n+1$, $i = A, B$, and μ_i that strongly believes $(\overline{S}^{(N,E)}_{-i,q})^{m-1}_{q=0}$, $\rho(\mu_i) \subseteq \overline{S}^{(N,E)}_{i,m}$. Since $S^n((N,E))|(N,E)$ is non-empty (by induction hypothesis and by (♠)) and features all best replies to beliefs in the set, $(\overline{S}^{(N,E)}_q)^\infty_{q=0}$ is an elimination procedure with $\overline{S}^{(N,E)}_\infty \neq \emptyset$. Let $(S^{(N,E)}_q)^\infty_{q=0} = ((S^q_\Delta((N,E))|(N,E))^\infty_{q=0}$. For each $a \in \overline{S}^{(N,E)}_{A,\infty}$ and $q \in \mathbb{N}$, if $a \in S^{(N,E)}_{A,q}$, there is $s_A \in S^q_{A,\Delta}((N,W))$ with $s_A|(N,E) = a$; thus, there exists μ_B that strongly believes $S^q_{B,\Delta}, ..., S^0_{B,\Delta}$ with $\mu_B(s_A|h^0) = 1$, and by the incentives given by (♠), $\rho(\mu_B) \subseteq S^n_{B,\Delta}((N,E))$. So, the best replies to a are in $S^{(N,E)}_{B,q+1}$. For each $b \in \overline{S}^{(N,E)}_{B,\infty}$ and $q \in \mathbb{N}$, if $b \in S^{(N,E)}_{B,q}$, there is $s_B \in S^q_{B,\Delta}((N,E))$ with $s_B|(N,E) = b$; thus, there exists μ_A that strongly believes $S^q_{A,\Delta}, ..., S^0_{A,\Delta}$ with $\mu_A(\cdot|h^0) = \overline{\overline{\mu}}_i(\overline{s}_A)(\cdot|h^0)$ and $\mu_A(s_B|(N,E)) = 1$, and $\rho(\mu_A) \subseteq S^n_{A,\Delta}((N,E))$. So, the best replies to b are in $S^{(N,E)}_{A,q+1}$. Then, since $\overline{S}^{(N,E)}_\infty$ is a set with the best reply property, $\overline{S}^{(N,E)}_\infty \subseteq S^{(N,E)}_\infty$, which contradicts $S^\infty_\Delta((N,E)) = \emptyset$.

6 Appendix

Formal proof of Lemma 2.

We need additional notation. For any $h \in H$, $\widehat{h} \succeq h$, $(s_j^h)_{j\in I} \in S^h$, $(s_j^{\widehat{h}})_{j\in I} \in S^{\widehat{h}}$, $\mu_i^h \in \Delta^{H^h}(S_{-i}^h)$, $\mu_i^{\widehat{h}} \in \Delta^{H^{\widehat{h}}}(S_{-i}^{\widehat{h}})$, $\widehat{Z} \subseteq Z^{\widehat{h}}$, and $J \subseteq I$, let:

- $s_J^h =^{\widehat{Z}} s_J^{\widehat{h}}$ if for each $z \in \widehat{Z}$ and $\widehat{h} \preceq \widetilde{h} \prec z$, $s_J^h(\widetilde{h}) = s_J^{\widehat{h}}(\widetilde{h})$;
- $\mu_i^h =^{\widehat{Z}} \mu_i^{\widehat{h}}$ if for each $z \in \widehat{Z}$ and $\widehat{h} \preceq \widetilde{h} \prec z$, $\mu_i^h(S_{-i}^h(z)|\widetilde{h}) = \mu_i^{\widehat{h}}(S_{-i}^{\widehat{h}}(z)|\widetilde{h})$;
- $s_J^h =^{\widehat{h}} s_J^{\widehat{h}}$ and $\mu_i^h =^{\widehat{h}} \mu_i^{\widehat{h}}$ if, respectively, $s_J^h =^{Z^{\widehat{h}}} s_J^{\widehat{h}}$ and $\mu_i^h =^{Z^{\widehat{h}}} \mu_i^{\widehat{h}}$;
- $\widehat{r}(\mu_i^h, \widehat{h})$ is the set of continuation best replies to $\mu_i^h(\cdot|\widehat{h})$.

Moreover, for any $\overline{S}^h = \times_{i\in I}\overline{S}_i^h \subseteq S^h$, define the set of histories that follow a unilateral deviation by player i from the paths induced by $\overline{S}^h$ as:

- $D_i(\overline{S}^h) := \{\widetilde{h} \in H\backslash H(\overline{S}^h) : p(\widetilde{h}) \in H(\overline{S}^h) \wedge \widetilde{h} \in H(\overline{S}_{-i}^h)\}$.

The first two lemmata claim the survival of strategies, or conjectures over such strategies, which combine substrategies that have survived by assumption. The reason why such lemmata are needed is merely the following. Fix $\widehat{s}_i^h, \overline{s}_i^h \in S_{i,n}^h$ and $\widehat{h}, \overline{h} \in H(\widehat{s}_i^h) \cap H(\overline{s}_i^h)$ such that $\overline{h} \not\succeq \widehat{h} \not\succeq \overline{h}$: there needs not exist $s_i^h \in S_{i,n}^h(\widehat{h}) \cap S_{i,n}^h(\overline{h})$ such that $s_i^h|\widehat{h} = \overline{s}_i^h|\widehat{h}$ and $s_i^h|\overline{h} = \widehat{s}_i^h|\overline{h}$. The intuitive reason is the following: player i may allow $\widehat{h}$ and $\overline{h}$ either because she is confident that $\widehat{h}$ will be reached and she has appropriate expectations after $\widehat{h}$, or because she is confident that $\overline{h}$ will be reached and she has appropriate expectations after $\overline{h}$. If $\widehat{s}_i^h$ is best reply to the first conjecture and $\overline{s}_i^h$ is best reply to the second conjecture, $\widehat{s}_i^h|\overline{h}$ and $\overline{s}_i^h|\widehat{h}$ may be "emergency plans" for an unpredicted contingency, after which the expectations would not have justified the choice to allow $\overline{h}$ and $\widehat{h}$ in the first place. Here is an example. The following is a simplified version of the game in Figure 4 in Battigalli [1], provided by Gul and Reny. The payoffs are of player 1.

$$\begin{array}{ccccccccc}
 & & 2 & \leftarrow o- & \mathbf{1} & & & & \\
 & & & & \downarrow i & & & & \\
1 & \leftarrow a- & \mathbf{1} & \leftarrow l- & \mathbf{2} & -r \rightarrow & \mathbf{1} & -a' \rightarrow & 1 \\
 & & \downarrow b & & & & b' \downarrow & & \\
0 & \leftarrow c- & \mathbf{3} & & & & \mathbf{3} & -c' \rightarrow & 0 \\
 & & \downarrow d & & & & d' \downarrow & & \\
 & & 3 & & & & 3 & &
\end{array}$$

Player 1 can rationally play $i.a.b'$ (if she expects r and d' but not d), $i.b.a'$ (if she expects l and d but not d'), but not $i.a.a'$. If one starts from $i.a.b'$, you cannot modify b' into a' because $i.a.b'$ is a sequential best reply only to CPS's that assign initial positive probability to r, therefore the belief at (i, r) cannot be modified without modifying the initial belief, hence the previous choices. Instead, $i.a.b'$ can be modified into $i.b.b'$ because $i.a.b'$ is rational under zero probability to l.

Lemma 5 *Fix an elimination procedure $((S^h_{i,q})_{i\in I})_{q\geq 0}$, $n \in \mathbb{N}$, $i \in I$, $\widehat{h} \in H^h$, and μ^h_i that strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$ such that $\mu^h_i(S^h_{-i}(\widehat{h})|p(\widehat{h})) = 0$. Fix $s^h_i \in \rho(\mu^h_i) \cap S^h_i(\widehat{h})$, $\mu^{\widehat{h}}_i$ that strongly believes $(S^h_{-i,q}(\widehat{h})|\widehat{h})^{n-1}_{q=0}$, and $s^{\widehat{h}}_i \in \rho(\mu^{\widehat{h}}_i)$.*

Consider the unique $\widetilde{s}^h_i =^{\widehat{h}} s^{\widehat{h}}_i$ such that for every $\widetilde{h} \notin H^{\widehat{h}}$, $\widetilde{s}^h_i(\widetilde{h}) = s^h_i(\widetilde{h})$.

There exists $\widetilde{\mu}^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$ that strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$ such that $\widetilde{\mu}^h_i(\cdot|\widetilde{h}) = \mu^h_i(\cdot|\widetilde{h})$ for all $\widetilde{h} \notin H^{\widehat{h}}$, and $\widetilde{s}^h_i \in \rho(\widetilde{\mu}^h_i)$ (so, $\rho(\mu^h_i)(\widehat{h}) \neq \emptyset$ implies $\rho(\widetilde{\mu}^h_i)(\widehat{h}) \neq \emptyset$).

Proof.

Fix a map $\varsigma : S^{\widehat{h}}_{-i} \to S^h_{-i}$ such that for each $s^{\widehat{h}}_{-i} \in S^{\widehat{h}}_{-i}$, $\varsigma(s^{\widehat{h}}_{-i}) =^{\widehat{h}} s^{\widehat{h}}_{-i}$ and $\varsigma(s^{\widehat{h}}_{-i}) \in S^h_{-i,m}(\widehat{h})$ for all $m \geq 0$ with $s^{\widehat{h}}_{-i} \in S^h_{-i,m}(\widehat{h})|\widehat{h}$. Since ς is injective, we can construct an array of probability measures $\widetilde{\mu}^h_i = (\widetilde{\mu}^h_i(\cdot|\widetilde{h}))_{\widetilde{h}\in H^h}$ on S^h_{-i} as $\widetilde{\mu}^h_i(\cdot|\widetilde{h}) = \mu^h_i(\cdot|\widetilde{h})$ for all $\widetilde{h} \notin H^{\widehat{h}}$ and $\widetilde{\mu}^h_i(\varsigma(s^{\widehat{h}}_{-i})|\widetilde{h}) = \mu^{\widehat{h}}_i(s^{\widehat{h}}_{-i}|\widetilde{h})$ for all $\widetilde{h} \in H^{\widehat{h}}$ and $s^{\widehat{h}}_{-i} \in S^{\widehat{h}}_{-i}$. Thus, $\widetilde{\mu}^h_i$ satisfies CPS-1. It is immediate to verify that $\widetilde{\mu}^h_i$ satisfies CPS-2, strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$, $\widetilde{\mu}^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$ Finally, since $\widetilde{\mu}^h_i(S_{-i}(\widehat{h})|p(\widehat{h})) = 0$, $\widetilde{\mu}^h_i$ satisfies CPS-3.

Fix $\widetilde{h} \in H(\widetilde{s}^h_i)\backslash H^{\widehat{h}} = H(s^h_i)\backslash H^{\widehat{h}}$. If $\widetilde{h} \prec \widehat{h}$, by $\mu^h_i(S^h_{-i}(\widehat{h})|p(\widehat{h})) = 0$ and CPS-3, $\mu^h_i(S^h_{-i}(\widehat{h})|\widetilde{h}) = 0$, and for every $s^h_{-i} \notin S^h_{-i}(\widehat{h})$, $\zeta(s^h_i, s^h_{-i}) = \zeta(\widetilde{s}^h_i, s^h_{-i})$. If $\widetilde{h} \not\prec \widehat{h}$, for every $s^h_{-i} \in S^h_{-i}(\widetilde{h})$, $\widehat{h} \notin H(s^h_i, s^h_{-i})$, so $\zeta(s^h_i, s^h_{-i}) = \zeta(\widetilde{s}^h_i, s^h_{-i})$. Hence $s^h_i \in \widehat{r}(\mu^h_i, \widetilde{h})$ implies $\widetilde{s}^h_i \in \widehat{r}(\mu^h_i, \widetilde{h}) = \widehat{r}(\widetilde{\mu}^h_i, \widetilde{h})$. Fix $\widetilde{h} \in H(\widetilde{s}^h_i) \cap H^{\widehat{h}} = H(s^{\widehat{h}}_i)$. For every $s^{\widehat{h}}_{-i} \in S^{\widehat{h}}_{-i}$, $\widetilde{\mu}^h_i(\varsigma(s^{\widehat{h}}_{-i})|\widetilde{h}) = \mu^{\widehat{h}}_i(s^{\widehat{h}}_{-i}|\widetilde{h})$. For every $\widehat{s}^h_i \in S^h_i(\widehat{h})$, $\zeta(\widehat{s}^h_i|\widehat{h}, s^{\widehat{h}}_{-i}) = \zeta(\widehat{s}^h_i, \varsigma(s^{\widehat{h}}_{-i}))$. So, $\widetilde{s}^h_i|\widehat{h} = s^{\widehat{h}}_i \in \widehat{r}(\mu^{\widehat{h}}_i, \widetilde{h})$ implies $\widetilde{s}^h_i \in \widehat{r}(\widetilde{\mu}^h_i, \widetilde{h})$. ■

Lemma 6 *Fix an elimination procedure $((\widetilde{S}^h_{i,q})_{i\in I})_{q\geq 0}$, subsets of strategies $(\overline{S}^h_i)_{i\in I}$, $m \in \mathbb{N}$, and $l \in I$. Let $Z^S := \zeta(\overline{S}^h)$. For every $i \in I$, suppose that there exists a map $\overline{\overline{\mu}}^h_i : \overline{S}^h_i \to \Delta^{H^h}(S^h_{-i})$ such that for all $s^h_i \in \overline{S}^h_i$, $\overline{\overline{\mu}}^h_i(s^h_i)$ strongly believes $\overline{S}^h_{-i}$, and:*

A1 there exist maps $\overline{\mu}^h_i : \overline{S}^h_i \to \Delta^{H^h}(S^h_{-i})$ and $\overline{s}^h_i : \overline{S}^h_i \to S^h_i$ such that for all $s^h_i \in \overline{S}^h_i$, $\overline{\mu}^h_i(s^h_i) =^{Z^S} \overline{\overline{\mu}}^h_i(s^h_i)$ strongly believes $(\widetilde{S}^h_{-i,q})^{m-1}_{q=0}$ and $\rho(\overline{\mu}^h_i(s^h_i)) \ni \overline{s}^h_i(s^h_i) =^{Z^S} s^h_i$;

A2 for every $s^h_i \in \overline{S}^h_i$ and $\mu^h_i =^{Z^S} \overline{\overline{\mu}}^h_i(s^h_i)$ that strongly believes $(\widetilde{S}^h_{-i,q})^{m-1}_{q=0}$, $\rho(\mu^h_i) \subseteq \widetilde{S}^h_{i,m}$.

Fix $l \in I$ and $s^h_l \in \overline{S}^h_l$. Let $D^S := D_l(\overline{S}^h)$. For every $\widehat{h} \in D^S$, fix $\widetilde{\mu}^{\widehat{h}}_l$ that strongly believes $(\widetilde{S}^h_{-l,q}(\widehat{h})|\widehat{h})^m_{q=0}$.

There exists $\widetilde{\mu}^h_l =^{Z^S} \overline{\overline{\mu}}^h_l(s^h_l)$ that strongly believes $(\widetilde{S}^h_{-l,q})^m_{q=0}$ such that $\widetilde{\mu}^h_l =^{\widehat{h}} \widetilde{\mu}^{\widehat{h}}_l$ for all $\widehat{h} \in D^S$.

Proof.

We show that for every $i \neq l$ and $s_i^h \in \overline{S}_i^h$, and for every map $\varsigma : \widehat{h} \in D^S \mapsto s_i^{\widehat{h}} \in \widetilde{S}_{i,m}^h(\widehat{h})|\widehat{h}$, there exists $\widetilde{s}_i^h \in \widetilde{S}_{i,m}^h$ such that $\widetilde{s}_i^h =^{Z^S} \overline{s}_i^h(s_i^h)$ and $\widetilde{s}_i^h =^{\widehat{h}} \varsigma(\widehat{h})$ for all $\widehat{h} \in D^S$. The map ς is well defined because for each $\widehat{h} \in D^S$, by A1 $\widehat{h} \in H(\overline{s}_i^h(\widehat{s}_i^h))$ for some $\widehat{s}_i^h \in \overline{S}_i^h$, and by A2, $\overline{s}_i^h(\widehat{s}_i^h) \in \widetilde{S}_{i,m}^h$. Using all such $\widehat{s}_i^h$'s, it is easy to construct the desired $\widetilde{\mu}_l^h$.

By A1, there exists $\overline{\mu}_i^h(s_i^h) =^{Z^S} \overline{\overline{\mu}}_i^h(s_i^h)$ that strongly believes $(\widetilde{S}_{-i,q}^h)_{q=0}^{m-1}$ such that $\overline{s}_i^h(s_i^h) \in \rho(\overline{\mu}_i^h(s_i^h))$. Fix $\widehat{h} \in D^S \cap H(s_i^h)$. Since $\overline{\mu}_i^h(s_i^h) =^{Z^S} \overline{\overline{\mu}}_i^h(s_i^h)$ and $\overline{\overline{\mu}}_i^h(s_i^h)$ strongly believes $\overline{S}_{-i}^h$, $\overline{\mu}_i^h(s_i^h)(S_{-i}^h(\widehat{h})|p(\widehat{h})) = 0$. Since $\varsigma(\widehat{h}) \in \widetilde{S}_{i,m}^h(\widehat{h})|\widehat{h}$, there exists $\mu_i^{\widehat{h}}$ that strongly believes $(\widetilde{S}_{-i,q}^h(\widehat{h})|\widehat{h})_{q=0}^{m-1}$ such that $\varsigma(\widehat{h}) \in \rho(\mu_i^{\widehat{h}})$. Thus, by Lemma 5, there exist $\widetilde{\mu}_i^h =^{\widehat{h}} \mu_i^{\widehat{h}}$ that strongly believes $(\widetilde{S}_{-i,q}^h)_{q=0}^{m-1}$ such that $\widetilde{\mu}_i^h(\cdot|\widetilde{h}) = \overline{\mu}_i^h(s_i^h)(\cdot|\widetilde{h})$ for all $\widetilde{h} \notin H^{\widehat{h}}$, and $\widetilde{s}_i^h \in \rho(\widetilde{\mu}_i^h)$ such that $\widetilde{s}_i^h =^{\widehat{h}} \varsigma(\widehat{h})$ and $\widetilde{s}_i^h(\widetilde{h}) = \overline{s}_i^h(s_i^h)(\widetilde{h})$.for all $\widetilde{h} \notin H^{\widehat{h}}$. Iterating for each $\widehat{h} \in D^S$, we obtain $\widetilde{\mu}_i^h =^{Z^S} \overline{\mu}_i^h(s_i^h)$ that strongly believes $(\widetilde{S}_{-i,q}^h)_{q=0}^{m-1}$ such that $\widetilde{\mu}_i^h =^{\widehat{h}} \mu_i^{\widehat{h}}$ for all $\widehat{h} \in D^S$, and $\widetilde{s}_i^h \in \rho(\widetilde{\mu}_i^h)$ such that $\widetilde{s}_i^h =^{Z^S} s_i^h$ and $\widetilde{s}_i^h =^{\widehat{h}} \varsigma(\widehat{h})$.for all $\widehat{h} \in D^S$. By A2, $\widetilde{s}_i^h \in \widetilde{S}_{i,m}^h$. ■

Lemma 7 *Fix two elimination procedures $((\overline{S}_{i,q}^h)_{i\in I})_{q\geq 0}$ and $((S_{i,q}^h)_{i\in I})_{q\geq 0}$. For every $i \in I$, let $\overline{S}_i^h := \overline{S}_{i,\infty}^h$ and let $\overline{\overline{\mu}}_i^h : \overline{S}_i^h \to \Delta^{H^h}(S_{-i}^h)$ be a map such that for every $s_i^h \in \overline{S}_i^h$, $\overline{\overline{\mu}}_i^h(s_i^h)$ strongly believes $(\overline{S}_{-i,q}^h)_{q=0}^{\infty}$ and $s_i^h \in \rho(\overline{\overline{\mu}}_i^h(s_i^h))$. Let $Z^S := \zeta(\overline{S}^h)$. Fix $n \in \mathbb{N}$, $l \in I$, and $\widehat{s}_l^h \in \overline{S}_l^h$ such that:*[14]

A3 for every $i \in I$ and $m \leq n$, $(S_q^h)_{q\geq 0}$ satisfies A1;

A4 for every $i \in I$ and $m \in \mathbb{N}$, $(S_q^h)_{q\geq 0}$ satisfies A2;

A5 for every $i \in I$ and $m \in \mathbb{N}$, $(\overline{S}_q^h)_{q\geq 0}$ satisfies A2;

A6 for every $s_l^h =^{Z^S} \widehat{s}_l^h$ and $\mu_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$ that strongly believes $(S_{-l,q}^h)_{q=0}^n$, $s_l^h \notin \rho(\mu_l^h)$.

Let $D^S := D_l(\overline{S}^h)$. For every $\widehat{h} \in D^S$ and $m \in \mathbb{N}$, call $M_m^{\widehat{h}}$ (resp., $\overline{M}_m^{\widehat{h}}$) the set of all $\mu_l^{\widehat{h}}$ that strongly believe $(S_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$ (resp., $(\overline{S}_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$) for which there exists $\widehat{\mu}_l^{\widehat{h}}$ that strongly believes $(S_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^n$ such that $\mu_l^{\widehat{h}}(S_{-i}(z)|\widehat{h}) = \widehat{\mu}_l^{\widehat{h}}(S_{-i}(z)|\widehat{h})$ for all $z \in \zeta(\widehat{r}(\widehat{\mu}_l^{\widehat{h}}, \widehat{h}) \times \mathrm{Supp}\widehat{\mu}_l^{\widehat{h}}(\cdot|\widehat{h}))$.[15]

Thus, there exists $\widehat{h} \in D^S$ such that:

1. *for every $m \leq n$ and $\mu_l^{\widehat{h}} \in M_m^{\widehat{h}}$, there exists $\mu_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$ that strongly believes $(S_{-l,q}^h)_{q=0}^m$ such that $\mu_l^h =^{\widehat{h}} \mu_l^{\widehat{h}}$ and $\rho(\mu_l^h)(\widehat{h}) \neq \emptyset$;*

[14] A3, A4 and A5 need not hold for $i = l$ to recall Lemma 6 and prove this lemma. However, l has been included to reuse A3, A4 and A5 in the final proof of Lemma 2.

[15] Note: $\widehat{\mu}_l^{\widehat{h}}$ refers to the second procedure even when $\mu_l^{\widehat{h}}$ refers to the first.

2. for every $p \in \mathbb{N}$ and $\widetilde{\mu}_l^{\widehat{h}} \in \overline{M}_p^{\widehat{h}}$, there exists $\widetilde{\mu}_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$ that strongly believes $(\overline{S}_{-l,q}^h)_{q=0}^p$ such that $\widetilde{\mu}_l^h =^{\widehat{h}} \widetilde{\mu}_l^{\widehat{h}}$ and $\rho(\widetilde{\mu}_l^h)(\widehat{h}) \neq \emptyset$.[16]

Proof.

Suppose by contraposition that there is a partition $(D, \overline{D})$ of D^S such that for every $\widehat{h} \in D$, there exist $m(\widehat{h}) \leq n$ and $\mu_l^{\widehat{h}} \in M_{m(\widehat{h})}^{\widehat{h}}$ that violate 1, and for every $\widehat{h} \in \overline{D}$ there exist $m(\widehat{h}) \in \mathbb{N}$ and $\mu_l^{\widehat{h}} \in \overline{M}_{m(\widehat{h})}^{\widehat{h}}$ that violate 2. For each $\widehat{h} \in D^S$, fix corresponding $\widehat{\mu}_l^{\widehat{h}}$. Let $\overline{\overline{\mu}}_l^h := \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$. By Lemma 6, there exists $\widetilde{\mu}_l^h =^{Z^S} \overline{\overline{\mu}}_l^h$ that strongly believes $(S_{-l,q}^h)_{q=0}^n$ such that for every $\widehat{h} \in D^S$, $\widetilde{\mu}_l^h =^{\widehat{h}} \widehat{\mu}_l^{\widehat{h}}$. We want to show that there exists $s_l^h \in \rho(\widetilde{\mu}_l^h)$ such that $s_l^h =^{Z^S} \widehat{s}_l^h$, violating A6.

Fix $\widehat{h} \in D$. Substitute $\widehat{\mu}_l^{\widehat{h}}$ with $\mu_l^{\widehat{h}}$ in the construction of $\widetilde{\mu}_l^h$ and obtain a new $\mu_l^h =^{\widehat{h}} \mu_l^{\widehat{h}}$ that strongly believes $(S_{-l,q}^h)_{q=0}^{m(\widehat{h})}$ with $\mu_l^h(S_{-l}(z)|\widetilde{h}) = \widetilde{\mu}_l^h(S_{-l}(z)|\widetilde{h})$ for all $\widetilde{h} \notin H^{\widehat{h}}$ and $z \notin Z^{\widehat{h}}$. By definition of $M_m^{\widehat{h}}$, player l expects a non higher payoff against $\widehat{\mu}_l^{\widehat{h}}$ than against $\mu_l^{\widehat{h}}$. Thus, $\rho(\mu_l^h)(\widehat{h}) \neq \emptyset$ (by the contrapositive hypothesis) implies $\rho(\widetilde{\mu}_l^h)(\widehat{h}) \neq \emptyset$. So, $H(\rho(\widetilde{\mu}_l^h)) \cap D = \emptyset$.

Write $\overline{D} = \{h^1, ..., h^k\}$ where $m(h^1) \geq ... \geq m(h^k)$. Note that $(\overline{S}_q^h)_{q\geq 0}$ satisfies A1 with $\overline{\mu}_i^h(\cdot) = \overline{\overline{\mu}}_i^h(\cdot)$ and the identity function for $\overline{s}_i^h(\cdot)$. Then, by Lemma 6,[17] for each $j = 1, ..., k$, there exists $\mu_{l,j}^h =^{Z^h \setminus \cup_{t=1}^j Z^{h^t}} \overline{\overline{\mu}}_l^h$ that strongly believes $(\overline{S}_{-l,q}^h)_{q=0}^{m(h^j)}$ such that $\mu_{l,j}^h =^{h^t} \mu_l^{h^t}$ for all $1 \leq t \leq j$. Let $\mu_{l,0}^h := \overline{\overline{\mu}}_l^h$. Fix $j = 1, ..., k$ and suppose to have shown that $\rho(\mu_{l,j-1}^h) = \rho(\overline{\overline{\mu}}_l^h)$. Then $\rho(\mu_{l,j-1}^h) \cap S_l^h(h^j) = \emptyset$. By the contrapositive hypothesis, $\rho(\mu_{l,j}^h) \cap S_l^h(h^j) = \emptyset$. For all $\widetilde{h} \notin H^{h^j}$ and $z \notin Z^{h^j}$, $\mu_{l,j}^h(S_{-l}(z)|\widetilde{h}) = \mu_{l,j-1}^h(S_{-l}(z)|\widetilde{h})$. Then, $\rho(\mu_{l,j}^h) = \rho(\mu_{l,j-1}^h)$. Inductively, $\rho(\mu_{l,k}^h) = \rho(\overline{\overline{\mu}}_l^h) \ni \widehat{s}_l^h$.

Fix $\widetilde{h} \in H(\widehat{s}_l^h) \cap H^S \cap H(\rho(\widetilde{\mu}_l^h))$. By $\widetilde{\mu}_l^h =^{Z^S} \overline{\overline{\mu}}_l^h =^{Z^S} \mu_{l,k}^h$, $\widetilde{\mu}_l^h(S_{-l}(z)|\widetilde{h}) = \mu_{l,k}^h(S_{-l}(z)|\widetilde{h})$ for all $z \in Z^{\widetilde{h}} \cap Z^S$. Then, since $\overline{\overline{\mu}}_l^h$ strongly believes $\overline{S}_{-l}^h$, $\widehat{s}_l^h$, as well as any other $\widetilde{s}_l^h \in S_l^h$ with $H(\widetilde{s}_l^h) \cap D^S = \emptyset$, induces the same outcome distribution against $\widetilde{\mu}_l^h(\cdot|\widetilde{h})$ and $\mu_{l,k}^h(\cdot|\widetilde{h})$. Moreover, $H(\rho(\widetilde{\mu}_l^h)) \cap D = \emptyset$. Finally, for all $\widehat{h} \in \overline{D}$, by definition of $\overline{M}_m^{\widehat{h}}$, player l expects a non higher payoff against $\widehat{\mu}_l^{\widehat{h}}$ than against $\mu_l^{\widehat{h}}$, and recall that $\widetilde{\mu}_l^h =^{\widehat{h}} \widehat{\mu}_l^{\widehat{h}}$ and $\mu_{l,k}^h =^{\widehat{h}} \mu_l^{\widehat{h}}$. So, $\widehat{s}_l^h \in \widehat{r}(\mu_{l,k}^h, \widetilde{h})$ implies $\widehat{s}_l^h \in \widehat{r}(\widetilde{\mu}_l^h, \widetilde{h})$. Proceeding from the root of the game, this implies $H(\widehat{s}_l^h) \cap H^S \subseteq H(\rho(\widetilde{\mu}_l^h)) \cap H^S$. Thus, there exists $s_l^h \in \rho(\widetilde{\mu}_l^h)$ such that $s_l^h(\widetilde{h}) = \widehat{s}_l^h(\widetilde{h})$ for all $\widetilde{h} \in H^S$. ■

Proof of Lemma 2.

[16] Since $\widehat{h} \notin H^S$, the statement must hold vacously for some $p \in \mathbb{N}$ (i.e. $\overline{M}_p^{\widehat{h}} = \emptyset$).

[17] Using the identity function for $\overline{s}_i^h(\cdot)$ in the proof of the lemma and without iterating at histories $\widehat{h} \in D^S \setminus \{h^1, ..., h^j\}$, the constructed $\mu_{l,j}^h$ clearly has the desired features.

Recall that the depth of a game is the length of the longest terminal history of the game. Suppose that $\Gamma(h)$ has depth $k \in \mathbb{N}$ and, if $k > 1$, that the lemma holds for games of depth $1, ..., k-1$. Let $\overline{S}^h_\infty \neq \emptyset$, otherwise the lemma trivially holds.

We prove by induction that $\zeta(\overline{S}^h_\infty) \subseteq \zeta(S^h_\infty)$. Note first that A4 and A5 hold by hypothesis of the lemma.

Induction Hypothesis (n): $(S^h_q)^\infty_{q=0}$ satisfies A3 at n (so by A4 $\zeta(S^h_n) \supseteq \zeta(\overline{S}^h_\infty)$).

Basis step (1): for all $i \in I$, the Inductive Hypothesis holds with $\overline{\mu}^h_i(\cdot) = \overline{\overline{\mu}}^h_i(\cdot)$.

Inductive step (n+1).

Suppose by contradiction that the Inductive Hypothesis does not hold at $n+1$. Then A6 holds for some $l \in I$ and $\widehat{s}^h_l \in \overline{S}^h_{l,\infty}$. Lemma 7 yields $\widehat{h} \in D_l(\overline{S}^h_\infty)$. If $\Gamma(h)$ has depth 1, $D_l(\overline{S}^h_\infty) = \emptyset$, so we have the desired contradiction. Else, define $((\overline{S}^{\widehat{h}}_{i,q})_{i\in I})_{q\geq 0}$ as follows: for every $i \in I$ and $m \leq n$, $\overline{S}^{\widehat{h}}_{i,m} = S^h_{i,m}(\widehat{h})|\widehat{h}$; for every $m > n$, $s^{\widehat{h}}_i \in \overline{S}^{\widehat{h}}_{i,m}$ if and only if there exists $\mu^{\widehat{h}}_i$ that strongly believes $(\overline{S}^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$ such that $s^{\widehat{h}}_i \in \rho(\mu^{\widehat{h}}_i)$.

For every $i \neq l$, since $\widehat{h} \in D_l(\overline{S}^h_\infty)$, $\overline{S}^h_{i,\infty}(\widehat{h}) \neq \emptyset$. So, fix $\widehat{s}^h_i \in \overline{S}^h_{i,\infty}(\widehat{h})$. For every $m \leq n$, the Induction Hypothesis provides $\overline{s}^h_i(\widehat{s}^h_i) \in S^h_{i,m}(\widehat{h}) \neq \emptyset$ and $\overline{\mu}^h_i(\widehat{s}^h_i) =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_i(\widehat{s}^h_i)$ that strongly believes $(S^h_{-i,q})^{m-1}_{q=0}$ such that $\overline{\mu}^h_i(\widehat{s}^h_i)(S^h_{-i}(\widehat{h})|p(\widehat{h})) = 0$. Hence, by Lemma 5, for every $\mu^{\widehat{h}}_i$ that strongly believes $(\overline{S}^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$, there exists $\mu^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$ that strongly believes $(S^h_{-i,q})^{m-1}_{q=0}$ such that $\mu^h_i =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_i(\widehat{s}^h_i)$ and $\rho(\mu^h_i)(\widehat{h}) \neq \emptyset$. By A4, $\rho(\mu^h_i) \subseteq S^h_{i,m}$. So, $\rho(\mu^{\widehat{h}}_i) \subseteq \overline{S}^{\widehat{h}}_{i,m}$.

Fix $\mu^{\widehat{h}}_l$ that strongly believes $(\overline{S}^{\widehat{h}}_{-l,q})^n_{q=0}$: trivially $\mu^{\widehat{h}}_l \in M^{\widehat{h}}_n$. Hence, by Lemma 7.(1), there exists $\widetilde{\mu}^h_l =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_l(\widehat{s}^h_l)$ that strongly believes $(S^h_{-l,n})^n_{q=0}$ such that $\widetilde{\mu}^h_l =^{\widehat{h}} \mu^{\widehat{h}}_l$ and $\rho(\widetilde{\mu}^h_l)(\widehat{h}) \neq \emptyset$. By A4, $\rho(\widetilde{\mu}^h_l) \subseteq S^h_{l,n}$. So $\rho(\mu^{\widehat{h}}_l) \subseteq \overline{S}^{\widehat{h}}_{l,n} \neq \emptyset$.

Hence, for every $i \in I$ and $\mu^{\widehat{h}}_i$ that strongly believes $(\overline{S}^{\widehat{h}}_{-i,q})^n_{q=0}$, $\rho(\mu^{\widehat{h}}_i) \subseteq \overline{S}^{\widehat{h}}_{i,n} \neq \emptyset$. So, $\overline{S}^{\widehat{h}}_{i,n} \supseteq \overline{S}^{\widehat{h}}_{i,n+1}$ and $((\overline{S}^{\widehat{h}}_{i,q})_{i\in I})_{q\geq 0}$ is an elimination procedure with $\overline{S}^{\widehat{h}}_\infty \neq \emptyset$.

For every $m \leq n$, $\widehat{\mu}^{\widehat{h}}_l$ that strongly believes $(\overline{S}^{\widehat{h}}_{-l,q})^\infty_{q=0}$, and $\mu^{\widehat{h}}_l =^{\zeta(\overline{S}^{\widehat{h}}_\infty)} \widehat{\mu}^{\widehat{h}}_l$ that strongly believes $(\overline{S}^{\widehat{h}}_{-l,q})^{m-1}_{q=0}$, $\mu^{\widehat{h}}_l \in M^{\widehat{h}}_m$.[18] Thus, by Lemma 7.(1) there exists $\widetilde{\mu}^h_l =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_l(\widehat{s}^h_l)$ that strongly believes $(S^h_{-l,q})^{m-1}_{q=0}$ such that $\widetilde{\mu}^h_l =^{\widehat{h}} \mu^{\widehat{h}}_l$ and $\rho(\widetilde{\mu}^h_l)(\widehat{h}) \neq \emptyset$. By A4, $\rho(\widetilde{\mu}^h_l) \subseteq S^h_{l,m}$. So $\rho(\mu^{\widehat{h}}_l) \subseteq \overline{S}^{\widehat{h}}_{l,m}$.

Then, for every $m \in \mathbb{N}$, $i \in I$, $\widehat{\mu}^{\widehat{h}}_i$ that strongly believes $(\overline{S}^{\widehat{h}}_{-i,q})^\infty_{q=0}$ and $\mu^{\widehat{h}}_i =^{\zeta(\overline{S}^{\widehat{h}}_\infty)} \widehat{\mu}^{\widehat{h}}_i$ that strongly believes $(\overline{S}^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$, $\rho(\mu^{\widehat{h}}_i) \subseteq \overline{S}^{\widehat{h}}_{i,m}$. Thus, $((\overline{S}^{\widehat{h}}_{i,q})_{i\in I})_{q\geq 0}$ satisfies the hypothesis of Lemma 2.

[18] Note that $\widehat{\mu}^{\widehat{h}}_l$ strongly believes $(\overline{S}^{\widehat{h}}_{-l,q})^n_{q=0} = (S^h_{-l,q}(\widehat{h})|\widehat{h})^n_{q=0}$, and that $\rho(\widehat{\mu}^{\widehat{h}}_l) \times \overline{S}^{\widehat{h}}_{-l,\infty} \subseteq \overline{S}^{\widehat{h}}_\infty$, so $\mu^{\widehat{h}}_l =^{\zeta(\overline{S}^{\widehat{h}}_\infty)} \widehat{\mu}^{\widehat{h}}_l$ verifies the definition of $M^{\widehat{h}}_m$ in the statement of Lemma 7.

Define now $((S_{i,q}^{\widehat{h}})_{i\in I})_{q\geq 0}$ as $((\overline{S}_{i,q}^{h}(\widehat{h})|\widehat{h})_{i\in I})_{q\geq 0}$. By Remark 1 it is an elimination procedure.

For every $i\neq l$, $m\in\mathbb{N}$, and $\mu_i^{\widehat{h}}$ that strongly believes $(S_{-i,q}^{\widehat{h}})_{q=0}^{m-1}$, by Lemma 5 there exists $\widetilde{\mu}_i^{h} =^{\widehat{h}} \mu_i^{\widehat{h}}$ that strongly believes $(\overline{S}_{-i,q}^{h})_{q=0}^{m-1}$ such that for every $\widetilde{h}\notin H^{\widehat{h}}$, $\widetilde{\mu}_i^{h}(\cdot|\widetilde{h}) = \overline{\overline{\mu}}_i^{h}(\widehat{s}_i^{h})(\cdot|\widetilde{h})$ and $\rho(\widetilde{\mu}_i^{h})(\widehat{h})\neq\emptyset$. By A5, $\rho(\widetilde{\mu}_i^{h})\subseteq\overline{S}_{i,m}^{h}$.

For every $m\in\mathbb{N}$, $\widehat{\mu}_l^{\widehat{h}}$ that strongly believes $(\overline{S}_{-l,q}^{\widehat{h}})_{q=0}^{\infty}$, and $\mu_l^{\widehat{h}} =^{\zeta(\overline{S}_{\infty}^{\widehat{h}})} \widehat{\mu}_l^{\widehat{h}}$ that strongly believes $(S_{-l,q}^{\widehat{h}})_{q=0}^{m-1}$, $\mu_l^{\widehat{h}}\in\overline{M}_m^{\widehat{h}}$.[19] Thus, by Lemma 7.(2) there exists $\widetilde{\mu}_l^{h} =^{\zeta(\overline{S}_{\infty}^{h})} \overline{\overline{\mu}}_l^{h}(\widehat{s}_l^{h})$ that strongly believes $(\overline{S}_{-l,q}^{h})_{q=0}^{m-1}$ such that $\widetilde{\mu}_l^{h} =^{\widehat{h}} \mu_l^{\widehat{h}}$ and $\rho(\widetilde{\mu}_l^{h})(\widehat{h})\neq\emptyset$. By A5 $\rho(\widetilde{\mu}_l^{h})\subseteq\overline{S}_{l,m}^{h}$.

Then, for every $m\in\mathbb{N}$, $i\in I$, $\widehat{\mu}_i^{\widehat{h}}$ that strongly believes $(\overline{S}_{-i,q}^{\widehat{h}})_{q=0}^{\infty}$ and $\mu_i^{\widehat{h}} =^{\zeta(\overline{S}_{\infty}^{\widehat{h}})} \widehat{\mu}_i^{\widehat{h}}$ that strongly believes $(S_{-i,q}^{\widehat{h}})_{q=0}^{m-1}$, $\rho(\mu_i^{\widehat{h}})\subseteq S_{i,m}^{\widehat{h}}$. Thus, $((S_{i,q}^{\widehat{h}})_{i\in I})_{q\geq 0}$ satisfies the hypothesis of Lemma 2

Since $\Gamma(\widehat{h})$ has strictly lower depth than $\Gamma(h)$, Lemma 2 holds. Hence, $\zeta(S_{\infty}^{\widehat{h}})\supseteq\zeta(\overline{S}_{\infty}^{\widehat{h}})\neq\emptyset$. But this contradicts $\widehat{h}\in D_l(\overline{S}_{\infty}^{h})$. ■

References

[1] Battigalli, P., "On rationalizability in extensive games", *Journal of Economic Theory*, **74**, 1997, 40-61.

[2] Battigalli, P., "Dynamic Consistency and Imperfect Recall", *Games and Economic Behavior*, **20(1)**, 1997, 31-50.

[3] Battigalli, P., "Rationalizability in Infinite, Dynamic Games of Incomplete Information," *Research in Economics,* **57**, 2003, 1-38.

[4] Battigalli, P. and A. Prestipino, "Transparent Restrictions on Beliefs and Forward Induction Reasoning in Games with Asymmetric Information", *The B.E. Journal of Theoretical Economics* (Contributions), **13**, 2013, Issue 1.

[5] Battigalli, P. and M. Siniscalchi, "Strong Belief and Forward Induction Reasoning," *Journal of Economic Theory*, **106**, 2002, 356-391.

[6] Battigalli P. and M. Siniscalchi, "Rationalization and Incomplete Information," *The B.E. Journal of Theoretical Economics*, **3(1)**, 2003, 1-46.

[7] Catonini, E., "Rationalizability and Epistemic Priority Orderings", working paper, 2017.

[19] See the previous footnote with $\overline{M}_m^{\widehat{h}}$ in place of $M_m^{\widehat{h}}$.

[8] Catonini, E., “Self-Enforcing Agreements and Forward Induction Reasoning”, working paper, 2017.

[9] Chen, J., and S. Micali, “The order independence of iterated dominance in extensive games”, *Theoretical Economics*, **8**, 2013, 125-163.

[10] Heifetz, A., and A. Perea, “On the Outcome Equivalence of Backward Induction and Extensive Form Rationalizability”, *International Journal of Game Theory*, **44**, 2015, 37–59.

[11] Kohlberg, E. and J.F. Mertens, “On the Strategic Stability of Equilibria”, *Econometrica*, **54**, 1986, 1003-1038.

[12] Osborne, M. J. and A. Rubinstein, “A Course in Game Theory”, 1994, Cambridge, Mass.: MIT Press.

[13] Pearce, D., “Rational Strategic Behavior and the Problem of Perfection”, *Econometrica*, **52**, 1984, 1029-1050.

[14] Perea, A., “Order Independence in Dynamic Games”, working paper, 2017.

[15] Perea, A., “Why Forward Induction leads to the Backward Induction outcome: a new proof for Battigalli’s theorem”, working paper, 2017.

[16] Renyi, A., “On a New Axiomatic Theory of Probability”, *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, 1955, 285-335.

[17] Shimoji, M. and J. Watson, “Conditional dominance, rationalizability, and game forms”, *Journal of Economic Theory*, **83(2)**, 1998, 161-195.