# On non-monotonic strategic reasoning.*

**Emiliano Catonini**†

May 2018

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.**

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

†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 [15]) based on the notion of *Strong Belief*.[1] Concretely, it is the maximal iterated deletion of "never sequential best replies" under belief systems that assign probability 1, as long as possible, to opponents' strategies that survive the previous step of the procedure.[2] Strong-$\Delta$-Rationalizability (Battigalli [3], Battigalli and Siniscalchi [6]) introduces first-order belief restrictions in the same reasoning scheme: only belief systems 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: strong belief in a smaller event does not imply 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 (see, e.g., the introductory example of Catonini [7]). Are there interesting conditions under which the introduction of first-order belief restrictions refines the set of strongly rationalizable outcomes? When such conditions are satisfied, the predictions are reassuringly robust to "restricted" and "unrestricted" forward induction reasoning, as captured, respectively, by Strong-$\Delta$-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) the set of outcomes predicted by Strong-$\Delta$-Rationalizability shrinks as more and more restrictions that "never bite off-path" are imposed. With this, I refer to restrictions that exclude belief systems only based on what they predict about opponents' behavior along the paths that survive all steps of Strong-$\Delta$-Rationalizability. So, off-the-path restrictions are responsible for the general non-monotonicity of Strong-$\Delta$-Rationalizability. The reason is the following. Suppose that at some step of reasoning, the behavior of an opponent that is object of a player's restricted beliefs ends up off-path. Then, the player will not verify whether such restricted beliefs are compatible with higher orders of rationality, because the off-path behavior of the opponent is no more refined by the elimination procedure.

Beside the theoretical insight, though, this broad condition for outcome monotonicity is of little practical use: one cannot verify it without actually performing Strong-$\Delta$-Rationalizability. Yet, a very important class of restrictions always satisfies this condition:

[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.

strong belief in a path of play. This class of restrictions is very important both for theory and practise. Pratically, agreements among real players often specify only the outcome to achieve (i.e., a path) and fall through if a player deviates (i.e., they not specify off-the-path behavior). Theoretically, path restrictions can be used to shed light on some forward induction refinements in the equilibrium literature, stemming from the seminal contribution of Kohlberg and Mertens [12], where deviations from an equilibrium path are interpreted as an attempt of the deviatior to improve upon her equilibrium payoff. Examples of this are the epistemic justifications of the Iterated Intuitive Criterion (Cho and Kreps [10]) by Battigalli and Siniscalchi [6] and of "equilibrium paths that cannot be upset by a convincing deviation" (Osborne [13]) by Catonini [8]. This approach is generalized in Catonini [8] to capture in a transparent way forward induction reasoning under this interpretation of deviations from a path prescribed by a non-binding agreement among players.

With the same methodology, I prove that under path restrictions the *epistemic priority* between beliefs in the path and beliefs in rationality does not matter for the predicted outcomes. When a player displays behavior that cannot be rational under her belief restrictions, Strong-$\Delta$-Rationalizability assumes that the opponents drop the belief that this player is rational.[3] Selective Rationalizability (Catonini [7]), a refinement of Strong Rationalizability with first-order belief restrictions, assumes instead that opponents keep believing that the player is rational (if per se compatible with the observed behavior) and drop the belief that the player has beliefs in the restricted set. The predictions under path restrictions have the further advantage of being robust to this epistemic priority choice.

The same workhorse lemma that yields the main results also yields and provides new insight on the following result, already proven by Perea [17]: the iterated deletion 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,[4] and show its order independence in terms of outcomes. Here, like in the recents work of Perea ([18], [17]), I work directly with the iterated deletion of never sequential best replies, thus with strong belief and without dominance characterizations. This is not the only similarity between Perea's methodology and mine. Perea bases his results on the "monotonicity on reachable histories" of the strong belief operator, which shares with my methodology

[3] Whether the belief that the player has beliefs in the restricted set is kept or not is immaterial for the procedure, thus Strong-$\Delta$-Rationalizability can be characterized epistemically with or without *transparency* of the restrictions: see Battigalli and Prestipino [4] for details.

[4] By showing the equivalence of the iterated elimination of *distinguishable* and *conditionally* dominated strategies, where the latter was already proved by Shimoji and Watson [20] to be equivalent to Extensive Form Rationalizability (Pearce [15]), which is in turn equivalent to Strong Rationalizability.

the following intuition: strong belief in a smaller set of strategies justifies fewer possible behaviors along the paths induced by this set with respect to strong belief in a larger set. Monotonicity on reachable histories further claims that this intuition goes through also when the second set is richer than the first only in terms of behavior along the paths induced by the first if the sets have been generated through an elimination order of the strong belief operator (which coincides with the "iterated elimination of never sequential best replies" defined here). The workhorse lemma of this paper allows instead to compare two elimination procedures with (nested) belief restrictions by showing directly the following: although the more permissive procedure can actually become more restrictive after deviations from the paths induced by the other procedure, this will not induce players to abandon these paths in absence of belief restrictions after each potential deviation.

As in Chen and Micali [9], Strong Rationalizability is also shown to refine backward induction, here captured by Backwards Extensive Form Rationalizability (Penta [16]), which can be seen as a particular, unfinished elimination order of never sequential best replies. 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 Perea [18] and Heifetz and Perea [11] in a more transparent way) follows.

Section 2 introduces the formal framework for the analysis. Section 3 defines elimination procedures and introduces the workhorse lemma. Section 4 presents the results on outcome monotonicity with respect to first-order belief restrictions and outcome equivalence with respect to the epistemic priority order. Section 5 presents the results on order independence and backward induction. Section 6 provides a sketch of the proof of the workhorse lemma, along with an example. The Appendix provides the formal proof.

## 2 Preliminaries

**Primitives of the game.**[5] Let $I$ be the finite set of *players*. For any profile of sets $(X_i)_{i\in I}$ and any subset of players $\emptyset \neq J \subseteq I$, 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 sets of *actions* potentially available to each player. Let $\overline{H} \subseteq \cup_{t=1,...,T}\overline{A}^t \cup \{h^0\}$ be the set of histories, where $h^0 \in \overline{H}$ is the empty, initial history and $T$ is the finite horizon. The set $\overline{H}$ must have the following properties. First property: 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 $h' \prec h$.[6] Let $Z := \{z \in \overline{H} : \nexists h \in \overline{H}, z \prec h\}$ be the set of terminal histories (henceforth, *outcomes*

[5] The main notation is almost entirely taken from Osborne and Rubinstein [14].

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

or *paths*)[7], and $H := \overline{H} \backslash Z$ the set of non-terminal histories (henceforth, just *histories*). Second property: For every $h \in H$, there exists a non-empty set $A_i(h) \subseteq \overline{A}_i$ for each $i \in I$,[8] such that $(h, a) \in \overline{H}$ if and only if $a \in A(h)$. 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 an element of $\times_{h \in H} A_i(h)$. 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 subset of player $J \subseteq I$ and any $\overline{S}_J \subset S_J$, let $\overline{S}_J(h) := S_J(h) \cap \overline{S}_J$. 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 $\Gamma(h)$. If $h \neq h^0$, all the objects in $\Gamma(h)$ will be denoted with $h$ as superscript, except for each history $h' \succeq h$ and outcome $z \succ h$, 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 $h \in H$, $s_i^h \in S_i^h = \times_{h' \succeq h} A_i(h')$, and $\widehat{h} \in H^h = \{h' \in H : h \preceq h'\}$, $s_i^h | \widehat{h}$ 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, [19]; henceforth, CPS).

**Definition 1** *Fix $i \in I$. An array of probability measures $(\mu_i(\cdot|h))_{h \in H}$ over co-players' strategies $S_{-i}$ is a Conditional Probability System if for all $h \in H$, $\mu_i(S_{-i}(h)|h) = 1$, and for all $h' \succ h$ and $\overline{S}_{-i} \subseteq S_{-i}(h')$,*

$$\mu_i(\overline{S}_{-i}|h) = \mu_i(S_{-i}(h')|h) \cdot \mu_i(\overline{S}_{-i}|h').$$

*The set of all CPS's on $S_{-i}$ 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 subset of

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

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

opponents' strategies $\overline{S}_{-i} \subseteq S_{-i}$, I say that a CPS $\mu_i \in \Delta^H(S_{-i})$ *strongly believes* $\overline{S}_{-i}$ if, for all $h \in H(\overline{S}_{-i})$, $\mu_i(\overline{S}_{-i}|h) = 1$. I fix the following convention: $H(\emptyset) = \emptyset$. With this, the empty set is always strongly believed, because the condition is vacuously satisfied.

**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 payoff given the belief about how opponents 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 $\mu_i$ if for every $h \in H(s_i)$,*[9] *$s_i$ 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 $\mu_i$ 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 Elimination procedures and the main lemma

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 $i \in I$,*

*EP1* $S^h_{i,0} = S^h_i$;

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

*EP3 for every* $s^h_i \in S^h_{i,\infty} = \cap_{n \in \mathbb{N}} S^h_{i,n}$, *there exists* $\mu^h_i$ *that strongly believes* $(S^h_{-i,q})^{\infty}_{q=0}$ *such that* $s^h_i \in \rho(\mu^h_i) \subseteq S^h_{i,\infty}$.

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

Definition 3 allows $S^h_{i,n} = \emptyset$ for some $n \in \mathbb{N}$, which implies $S^h_{i,m} = \emptyset$ for all $m > n$, but does not imply $S^h_{j,n+1} = \emptyset$ for any $j \neq i$: as already established, the empty set is always strongly believed, hence EP3 can be satisfied by a non-empty $S^h_{j,n+1}$. Moreover, EP2 allows an equality for all players also at steps before "convergence". All this allows Definition 3 to encompass the implications in a subgame of an elimination procedure for a larger subgame (which I will call "truncation" of the elimination procedure in the subgame).

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

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

For some $j \neq i$, we can have $S^h_{j,n+1}(\widehat{h})|\widehat{h} \neq \emptyset$ although $S^h_{i,n}(\widehat{h})|\widehat{h} = \emptyset$. Moreover, we can have $S^h_{j,n}(\widehat{h})|\widehat{h} = S^h_{j,n+1}(\widehat{h})|\widehat{h}$ for all $j \in I$, but $S^h_{j,n+1}(\widehat{h})|\widehat{h} \supset S^h_{j,n+2}(\widehat{h})|\widehat{h}$. This is because strategies in $((S^h_{i,q}(\widehat{h})|\widehat{h})_{i\in I})^\infty_{q=0}$ can be eliminated "exogeneously", due to eliminations from $((S^h_{i,q})_{i\in I})^\infty_{q=0}$ that affect $((S^h_{i,q}(\widehat{h})|\widehat{h})_{i\in I})^\infty_{q=0}$ at step $n+2$ and not at step $n+1$, and not because they are not sequential best replies to any valid conjecture in the subgame. For this reason, and to encompass elimination procedures with first-order-belief restrictions, an elimination procedure does not impose to save all the strategies that are sequential best replies to some $\mu_i$ that strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$. This makes an elimination procedure more general than an order of elimination of the "strong belief operator", defined in Perea [**?**]. Like for an order of elimination of the strong belief operator, instead, an elimination procedure allows to "forget" to eliminate strategies that are not sequential best replies to any $\mu_i$ that strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$ at all steps $n$ before convergence: for this reason, EP3 only refers to the final output of the procedure.

The workhorse lemma of the paper claims the outcome inclusion between two elimination procedures, $((\overline{S}^h_{i,q})_{i\in I})^\infty_{q=0}$ and $((S^h_{i,q})_{i\in I})^\infty_{q=0}$, with the following feature. Take the final output $\overline{S}^h_\infty$ of the first procedure and, for each player $i$ and each strategy $s^h_i \in \overline{S}^h_{i,\infty}$, fix a CPS $\overline{\overline{\mu}}^h_i(s^h_i)$ that satisfies EP3, i.e., it strongly believes $(\overline{S}^h_{-i,q})^\infty_{q=0}$ and justifies $s^h_i$: $s^h_i \in \rho(\overline{\overline{\mu}}^h_i(s^h_i)) \subseteq \overline{S}^h_{i,\infty}$. Consider now a CPS $\mu^h_i$ that, along the paths predicted by the first procedure, that is at every history $\widetilde{h} \in H(\overline{S}^h_\infty)$, assigns the same probability as $\overline{\overline{\mu}}^h_i(s^h_i)$ to the fact that the opponents will play compatibly with each of these paths $z \in \zeta(\overline{S}^h_\infty)$: $\mu^h_i(S_{-i}(z)|\widetilde{h}) = \overline{\overline{\mu}}^h_i(s^h_i)(S_{-i}(z)|\widetilde{h})$. Suppose now that, for each $m \in \mathbb{N}$, if $\mu^h_i$ strongly believes

$(\overline{S}^h_{-i,q})^{m-1}_{q=0}$, then $\rho(\mu^h_i) \subseteq \overline{S}^h_{i,m}$, and if $\mu^h_i$ strongly believes $(S^h_{-i,q})^{m-1}_{q=0}$, then $\rho(\mu^h_i) \subseteq S^h_{i,m}$. The lemma claims that then the second procedure predicts a superset $\zeta(S^h_\infty) \supseteq \zeta(\overline{S}^h_\infty)$ of the paths predicted by the first.

**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 $\overline{\overline{\mu}}^h_i : \overline{S}^h_{i,\infty} \to \Delta^{H^h}_i(S^h_{-i})$ such that, for each $s^h_i \in \overline{S}^h_{i,\infty}$, $\overline{\overline{\mu}}^h_i(s^h_i)$ strongly believes $(\overline{S}^h_{-i,q})^\infty_{q=0}$ and $s^h_i \in \rho(\overline{\overline{\mu}}^h_i(s^h_i)) \subseteq \overline{S}^h_{i,\infty}$. Suppose that for every $i \in I$, $s^h_i \in \overline{S}^h_{i,\infty}$, $m \in \mathbb{N}$, and $\mu^h_i$ that strongly believes $(S^h_{-i,q})^{m-1}_{q=0}$ (resp., $(\overline{S}^h_{-i,q})^{m-1}_{q=0}$),*

$$\left(\forall \widetilde{h} \in H(\overline{S}^h_\infty), \forall Z^{\widetilde{h}} \cap \zeta(\overline{S}^h_\infty), \mu^h_i(S_{-i}(z)|\widetilde{h}) = \overline{\overline{\mu}}^h_i(s^h_i)(S_{-i}(z)|\widetilde{h})\right) \Rightarrow \left(\rho(\mu^h_i) \subseteq S^h_{i,m}\right)$$
$$\text{(resp., } \rho(\mu^h_i) \subseteq \overline{S}^h_{i,m}\text{)}.$$

*Then, $\zeta(\overline{S}^h_\infty) \subseteq \zeta(S^h_\infty)$.*

Section 6 contains a sketch of the proof of the lemma, while the Appendix contains the formal proof. Now I focus on the implications of the lemma for the elimination procedures of interest.

# 4 Belief-restrictions and monotonicity

In this section, I am going to consider the following elimination procedures (for the whole game).

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

**Definition 5** *An elimination procedure $((S_{i,q})_{i\in I})^\infty_{q=0}$ 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 $\mu_i$ that strongly believes $(S_{-i,q})^{n-1}_{q=0}$.*

**Definition 6** *Strong Rationalizability (Battigalli and Siniscalchi, [5]) is the unique unconstrained and maximal elimination procedure. Let $((S^q_i)_{i\in I})^\infty_{q=0}$ 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 $\Delta_i \subseteq \Delta^H(S^h_{-i})$. Strong-$\Delta$-Rationalizability (Battigalli [3], Battigalli and Siniscalchi [6]) is the elimination procedure $((S^q_{i,\Delta})_{i\in I})^\infty_{q=0}$ 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 $(S^q_{-i,\Delta})^{n-1}_{q=0}$.*

**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 $(S^q_{-i,R\Delta})^{n-1}_{q=0}$.*[10]

Consider first-order belief restrictions $(\Delta_i)_{i\in I}$ with the following characteristic: for each player $i$ and CPS $\mu_i$, all that matters to determine whether $\mu_i$ belongs to $\Delta_i$ are the probabilities assigned by $\mu_i$ at each strongly-$\Delta$-rationalizable history $h \in H(S^{\infty}_{\Delta})$ to the fact that opponents will play compatibly with each strongly-$\Delta$-rationalizable path $z \in \zeta(S^{\infty}_{\Delta})$: $\mu_i(S_{-i}(z)|h)$. Then, Strong-$\Delta$-Rationalizability satisfies the hypotheses of Lemma 2 as first elimination procedure, whereas Strong Rationalizability, being an unconstrained procedure, obviously satisfies the hypotheses of Lemma 2 as second elimination procedure. The desired outcome inclusion with respect to belief restrictions that "do not end up off-path" obtains.

**Theorem 1** *For each $i \in I$, fix a set of CPS's $\Delta_i \subseteq \Delta^H(S_{-i})$. Suppose that for each $i \in I$, $\mu_i \in \Delta_i$, and $\mu'_i \in \Delta^H(S_{-i})$,*

$$\left(\forall \widetilde{h} \in H(S^{\infty}_{\Delta}), \forall z \in \zeta(S^{\infty}_{\Delta}), \mu'_i(S_{-i}(z)|\widetilde{h}) = \mu_i(S_{-i}(z)|\widetilde{h})\right) \Rightarrow \left(\mu'_i \in \Delta_i\right).$$

*Then, $\zeta(S^{\infty}_{\Delta}) \subseteq \zeta(S^{\infty})$.*

**Proof.** For each $i \in I$ and $s_i \in S^{\infty}_{i,\Delta}$, fix any $\overline{\mu}^h_i(s^h_i) \in \Delta_i$ 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 $\mu_i$ 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 predicted outcomes with respect to belief restrictions: the presence of off-the-path restrictions. Yet, it is of little help in determining ex-ante which belief restrictions preserve the behavioral consequences of 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-$\Delta$-Rationalizability itself.

Consider now first-order belief restrictions that correspond to the belief in a specific path $z \in Z$ along the path itself. This is what I call "path restrictions". I note preliminarly that this is equivalent to strong belief in $S_j(z)$ for all $j \neq i$ (the proof is in the Appendix).[11]

[10] Selective Rationalizability is defined in [7] under the more restrictive assumption of *independent rationalization*. That is, a valid $\mu_i$ 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).

[11] This corresponds to *belief in the (path) agreement* in [7].

The reason is that after a deviation from the path by a player different than $j$, believing that $j$ would have kept complying with the path is not restrictive for the expected behavior of $j$ after the deviation.

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

Path restrictions hold at histories that precede $z$. Therefore, if such restrictions ends up off-path, it means that some player has abandoned the path, so the opponents cannot believe in $S_{-i}(z)$ from the start anymore, and Strong-$\Delta$-Rationalizability yields the empty set. Otherwise, Theorem 1 can be applied and, via Lemma 3, monotonocity of strategic reasoning under strong belief in a path obtains.

**Theorem 2** *Fix $z \in Z$. Let $\Delta_i^*$ be the set of all $\mu_i \in \Delta^H(S_{-i})$ 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 $\Delta_i$ be the set of all $\mu_i$'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 $\mu_i$ 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 under strong belief in the path. Therefore, for path restrictions, Lemma 2 holds with Selective Rationalizability and Strong-$\Delta$-Rationalizability regardless of the roles assigned to the two procedures. Then, via Lemma 3, the outcome equivalence of the two procedures under strong belief in a path obtains.

**Theorem 3** *Fix $z \in Z$. Let $\Delta_i^*$ be the set of all $\mu_i \in \Delta^H(S_{-i})$ 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 $\Delta_i$ be the set of all $\mu_i$'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 $\mu_i$ 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 $(S_j(z))_{j\neq i}$.

## 5 Order independence and backward induction

In absence of belief restrictions, that is, for unconstrained elimination procedures, the hypotheses of Theorem 2 clearly hold. An unconstrained elimination procedure is what we referred to in the Introduction as an order of iterated elimination of never sequential best replies, and in Perea [18] it is called an elimination order of the strong belief operator. 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. ■

In games with observable actions, the well-known backward induction procedure for games with perfect information has been generalized as follows (see, for instance, Chen and Micali [9]). Starting from the bottom of game, an action of a player at a history is eliminated when it is not "folding-back optimal" against any conjecture over the surviving actions of the opponents at the same history and at the future histories, that is, it is not optimal given the already computed optimal actions at the future histories. Penta [16] has translated backward induction for games with observable actions in the language of extensive-form rationalizability, i.e., as a procedure of elimination of strategies that are not sequentially optimal for any appropriate conditional probability system. Penta's Backwards Extensive-Form Rationalizability is simplified here for games with complete information.

**Definition 9** *Backwards Extensive-Form Rationalizability is a sequence* $((S_{i,BR}^q)_{i\in I})_{q=0}^{\infty}$ *where, for every* $i \in I$,

*BR1* $S_{i,BR}^0 = S_i$;

*BR2 for each* $n \in \mathbb{N}$ *and* $s_i \in S_i$, $s_i \in S_{i,BR}^n$ *if and only if* $s_i \in S_{i,B}^{n-1}$ *and there exists* $\mu_i \in \Delta^H(S_{-i})$ *such that, for each* $h \in H$,

- **(i)** *there is* $\widetilde{s}_i \in S_i(h)$ *such that* $\widetilde{s}_i(\widetilde{h}) = s_i(\widetilde{h})$ *for each* $\widetilde{h} \succeq h$ *and* $\widetilde{s}_i \in \widehat{r}(\mu_i, \widetilde{h})$;
- **(ii)** *for each* $\widetilde{s}_{-i} \in S_{-i}(h)$ *with* $\mu_i(\widetilde{s}_{-i}|h) > 0$, *there is* $s_{-i} \in S_{-i,BR}^{n-1}$ *such that* $\widetilde{s}_{-i}(\widetilde{h}) = s_{-i}(\widetilde{h})$ *for each* $\widetilde{h} \succeq h$.

Condition BR2.(i) requires $s_i$ to be a continuation best reply not only at each $h \in H(s_i)$, as for sequential best replies of Definition 2, but also at each $h \notin H(s_i)$. Then, condition

BR2.(ii) requires to keep refining beliefs also at histories that cannot be reached anymore. So, Backwards Extensive-Form Rationalizability can be stricter, in terms of strategies, than an unconstrained elimination procedure. Yet, given that realization-equivalent classes are all that matters for elimination procedures and that such refinement of beliefs is off-path, it turns out that Backwards Extensive-Form Rationalizability is outcome-equivalent to an *unfinished*, unconstrained elimination procedure.

**Lemma 4** *Let $N$ be the smallest $n$ such that $S^n_{BR} = S^{n+1}_{BR}$. There exists an unconstrained elimination procedure $((S_{i,q})_{i\in I})_{q=0}^{\infty}$ such that for each $n \leq N$,*

$$S_n = \left\{ s \in S : \exists s' \in S^n_{BR}, \forall h \in H(S^n_{BR}), s(h) = s'(h) \right\}.$$

**Proof.** Define $\left((S_{i,n})_{i\in I}\right)_{n=0}^{N}$ as above, and for each $n > N$ and $i \in I$, let $s_i \in S_{i,n}$ if and only if there exists $\mu_i$ that strongly believes $(S_{-i,q})_{q=0}^{n-1}$ such that $s_i \in \rho(\mu_i)$. It is immediate to check that $((S_{i,q})_{i\in I})_{q=0}^{\infty}$ is an elimination procedure. To show that it is unconstrained, fix $n \leq N$ and suppose by way of induction that for each $m < n$, $i \in I$, and $\mu_i$ that strongly believes $(S_{-i,q})_{q=0}^{m-1}$, we have $\rho(\mu_i) \subseteq S_{i,m}$ (it is vacously true for $m = 0$). Fix $\mu_i$ that strongly believes $(S_{-i,q})_{q=0}^{n-1}$. I show that $\rho(\mu_i) \subseteq S_{i,n}$. By definition of $S_{-i,n-1}$, I can construct $\mu_i'$ that satisfies BR2.(ii) such that for all $h \in H(S_{n-1})$ and $z \in \zeta(S_{n-1})$, $\mu_i(S_{-i}(z)|h) = \mu_i'(S_{-i}(z)|h)$. For each $s_i' \in \rho(\mu_i')$, there is a realization equivalent $s_i''$ that satisfies BR2.(i), so that $s_i'' \in S^n_{i,BR} \subseteq S^{n-1}_{i,BR}$. For each $s_i \in \rho(\mu_i)$, by the induction hypothesis we have $s_i \in S_{i,n-1}$. Then, we have $\zeta(\rho(\mu_i') \times S_{-i,n-1}), \zeta(\rho(\mu_i) \times S_{-i,n-1}) \subseteq \zeta(S_{n-1}) = \zeta\left(S^{n-1}_{BR}\right)$. Thus, for each $s_i \in \rho(\mu_i)$, there is $s_i' \in \rho(\mu_i')$ such that $s_i(h) = s_i'(h)$ for all $h \in H(S_{n-1})$. Since there is $s_i'' \in S^n_{i,BR}$ realization equivalent to $s_i'$, so that $s_i''(h) = s_i'(h) = s_i(h)$ for all $h \in H(S_{n-1}) \cap H(s_i)$, by definition of $S_{i,n}$ we have $s_i \in S_{i,n}$. Thus, $\rho(\mu_i) \subseteq S_{i,n}$. ■

Being outcome-equivalent to an unfinished, unconstrained elimination procedure, Backwards Extensive Rationalizability predicts a superset of the outcomes predicted by Strong Rationalizability.

**Theorem 5** *Every strongly rationalizable outcome is also a backwards extensive-form rationalizable 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.*

## 6 Proof of the main lemma.

The rough intuition for the proof of the main lemma is the following. Take the paths induced by the first procedures. If the game had only these paths, they would survive also the second procedure, for the following two reasons. First, the fact that they survive the first procedure indicates that for every player there are beliefs over these paths that justify allowing each of them. Second, all these beliefs are allowed also under the second procedure by assumption. Then, the only way one of these paths can be eliminated along the second procedure is that at some step $n$, some player finds a deviation *outside* of these paths more profitable, however she believes the opponents will react to the deviation. Since the opponents may be surprised by the deviation (all the paths survived until step $n-1$), they can react with any continuation plan that survives until step $n$. So, we have that both the deviator and the opponents can play any sequential best reply to any belief in the subgame that follows the deviation. This allows to generate an auxiliary elimination procedure for the subgame that refines the continuation plans that survive the second procedure for the whole game until step $n$, and terminates with a non-empty set. Take the subpaths induced by this auxiliary procedure. We want to show that they would have survived also the truncation of the first procedure for the whole game, which implies that the subgame is reached at the end of the procedure and contradicts that the subgame follows a deviation from the paths predicted by the procedure. If the subgame is a static game, i.e., it has depth 1, this is easy to see: the auxiliary procedure generates a best response set where all the beliefs the deviator can have induce a deviation from the original path, which is actually available until the end of the first procedure (it can be sustained by other surviving actions of the opponents in the subgame). If the subgame has depth higher than 1, then suppose by induction that the lemma is true in games of that depth, so that the truncation of the first procedure in the subgame induces a superset of the paths induced by the auxiliary procedure, which leads to the same contradiction.

I refine now this intuition briefly illustrating with mathematical notation the precise arguments of the formal proof. For simplicity, assume that there are two players, $i$ and $j$; the argument extends immediately to games with more than 2 players. I argue by induction that for every $\overline{s}_i^h \in \overline{S}_{i,\infty}^h$ and $n \in \mathbb{N}$, there are: (1) $\overline{\mu}_i^h(\overline{s}_i^h)$ that strongly believes $(S_{j,q}^h)_{q=0}^{n-1}$ and assigns the same probabilities to the (opponents playing compatibly with) the paths induced by $\overline{S}_\infty^h$ (henceforth, just "paths") as some $\overline{\overline{\mu}}_i^h(\overline{s}_i^h)$ that justifies $\overline{s}_i^h \in \overline{S}_{i,\infty}^h$; (2) $s_i^h \in \rho(\overline{\mu}_i^h(\overline{s}_i^h))$ that mimicks $\overline{s}_i^h$ along the paths. Then, by the assumption on $(S_{i,q}^h, S_{j,q}^h)_{q=0}^\infty$, we have $s_i^h \in S_{i,n}^h$. All such $s_i^h$'s allow to construct at step $n+1$ a CPS $\overline{\mu}_j^h(\overline{s}_j^h)$ as in (1) for each $\overline{s}_j^h \in \overline{S}_{j,\infty}^h$.

Now, suppose by contradiction that for some $\overline{s}_j^h \in \overline{S}_{j,\infty}^h$, every such $\overline{\mu}_j^h(\overline{s}_j^h)$ does not

justify any strategy $s_j^h$ that mimicks $\overline{s}_j^h$ along the paths. For each history $\widehat{h}$ that immediately follows a unilateral deviation of player $j$ from the paths, that is, that follows a history along the paths where $i$ takes an action compatible with some of the paths and $j$ does not, consider the most pessimistic belief of $j$ over $S_{i,n}^h(\widehat{h})|\widehat{h}$. For each $\overline{s}_i^h \in \overline{S}_{i,\infty}^h$, by induction hypothesis there is $s_i^h \in S_{i,n}^h$ that mimicks $\overline{s}_i^h$ along the paths and is a sequential best reply to a belief $\overline{\mu}_i^h(\overline{s}_i^h)$ as in (1), thus which assigns probability zero to each deviation of $j$ until it occurs. Then, the beliefs specified by $\overline{\mu}_i^h(\overline{s}_i^h)$ along the paths can be combined with any beliefs after $j$'s deviations in a new CPS $\mu_i^h$ that satisfies (1). Clearly, there is a sequential best reply to $\mu_i^h$ which mimicks $s_i^h$ along the paths and prescribes any sequentially rational reaction to the chosen beliefs after $j$'s deviations. This is proved by Lemma 6. So, at step $n+1$, player $j$ can have a belief $\mu_j^h$ that mimicks $\overline{\overline{\mu}}_j^h(\overline{s}_j^h)$ along the paths (in the sense of (1)) and, at the same time, features the most pessimistic belief after each deviation. By the initial assumption of this paragraph, player $j$ will still deviate under $\mu_j^h$ and, calling $\widehat{h}$ a history that immediately follows this deviation, also under $\widetilde{\mu}_j^h$ constructed like $\mu_j^h$ except for a less pessimistic belief over $S_{i,n}^h(\widehat{h})|\widehat{h}$. This is proved by Lemma 7. So, $S_{j,n+1}^h(\widehat{h})|\widehat{h}$ features all the sequential best replies to CPS's $\mu_j^{\widehat{h}}$ that strongly believe $(S_{i,q}^{\widehat{h}})_{q=0}^n$, thus $S_{j,n}^h(\widehat{h})|\widehat{h} \supseteq S_{j,n+1}^h(\widehat{h})|\widehat{h}$ as well. The same holds with $i$ in place of $j$, for the argument exposed above.

Refine $S_n^h(\widehat{h})|\widehat{h}$ by iteratively eliminating strategies that are not sequential best replies to any $\mu_k^{\widehat{h}}$, $k = i, j$, that strongly believes in the previous steps. 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 main lemma. That the deviation is profitable against all beliefs over $\overline{S}_{i,\infty}^{\widehat{h}}$ with respect to remaining on the paths under $\overline{\overline{\mu}}_j^h(s_j^h)$ 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 main 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: $S_{\infty}^{\widehat{h}} = \overline{S}_{k,\infty}^h(\widehat{h})|\widehat{h}$ is non-empty too, hence $\widehat{h} \in H(\overline{S}_{\infty}^h)$, but $\widehat{h}$ follows a 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.

Finally, I am going to follow the sketch above on an example. Consider the following game.

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

$\rightarrow$

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

Take Strong Rationalizability, $((S_i^q)_{i\in I})_{q=0}^{\infty}$, as second procedure in the statement of the main lemma. 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 $S^{h^0}_{i,\infty} = (S, N.U) \times (W, E.L)$. Strong Rationalizability trivially satisfies the assumption of the lemma.

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

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

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

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 $\overline{\mu}_i(\overline{s}_i)(S_{-i}(z)|h^0) = \overline{\overline{\mu}}^h_i(S_{-i}(z)|h^0) = 1$;
2. $s_i \in \rho(\overline{\mu}_i(\overline{s}_i)) \subseteq S^n_i$ with $s_i(h^0) = N$ for $i = A$, $s_i(h^0) = W$ for $i = B$.

Suppose by contradiction the following:

(♠) For every $\mu_B$ that strongly believes $S^n_A, ..., S^0_A$ with $\mu_B(S_A(z)|h^0) = \overline{\overline{\mu}}^h_B(S_A(z)|h^0) = 1$, $\rho(\mu_B) \cap S_B(z) = \emptyset$.[12]

Let $\widehat{h} := (N, E)$. For each $a \in S^n_A(\widehat{h})|\widehat{h}$ (non-empty by the induction hypothesis), fix $s_A \in S^n_A(z)$ with $s_A|\widehat{h} = a$;[13] there exists $\mu_B$ that strongly believes $S^n_A, ..., S^0_A$ with $\mu_B(s_A|h^0) = 1$, so by (♠) $\rho(\mu_B) \subseteq S^n_B(\widehat{h})$. For each $b \in S^n_B(\widehat{h})|\widehat{h}$, fix $s_B \in S^n_B(\widehat{h})$ with $s_B|\widehat{h} = b$; there exists $\mu_A$ that strongly believes $S^n_B, ..., S^0_B$ with $\mu_A(W|h^0) = \overline{\mu}_A(\overline{s}_A)(W|h^0) = 1$ ($W \in S^n_B$ by the induction hypothesis) and $\mu_A(s_B|\widehat{h}) = 1$, so $\rho(\mu_A) \subseteq S^n_A(\widehat{h})$. Hence, $S^n(\widehat{h})|\widehat{h}$ features all best replies to beliefs in the set.

Let $(\overline{S}^{\widehat{h}}_q)^\infty_{q=0} = ((S^q(\widehat{h})|\widehat{h})^n_{q=0}, (\overline{S}^{\widehat{h}}_q)^\infty_{q=n+1})$, where for each $m \geq n+1$, $i = A, B$, and $\mu_i$ that strongly believes $(\overline{S}^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$, $\rho(\mu_i) \subseteq \overline{S}^{\widehat{h}}_{i,m}$. Since $S^n(\widehat{h})|\widehat{h}$ features all best replies to beliefs in the set, $\overline{S}^{\widehat{h}}_n \supseteq \overline{S}^{\widehat{h}}_{n+1}$ and thus $(\overline{S}^{\widehat{h}}_q)^\infty_{q=0}$ is an elimination procedure with $\overline{S}^{\widehat{h}}_\infty \neq \emptyset$. Let

[12] For Ann, it is obvious that this cannot hold, as $S$ is not optimal against $W$.

[13] It obviously exists because any strategy of Ann that allows $(N, E)$ must prescribe $N$.

$(S_q^{\widehat{h}})_{q=0}^{\infty} = ((S_\Delta^q(\widehat{h})|\widehat{h})_{q=0}^{\infty}$. For each $a \in \overline{S}_{A,\infty}^{\widehat{h}}$ and $q \in \mathbb{N}$, if $a \in S_{A,q}^{\widehat{h}}$, there is $s_A \in S_{A,\Delta}^q(z)$ with $s_A|\widehat{h} = a$; thus, there exists $\mu_B$ that strongly believes $S_{B,\Delta}^q, ..., S_{B,\Delta}^0$ with $\mu_B(s_A|h^0) = 1$, and by the incentives given by (♠), $\rho(\mu_B) \subseteq S_{B,\Delta}^n(\widehat{h})$. So, the best replies to $a$ are in $S_{B,q+1}^{\widehat{h}}$. For each $b \in \overline{S}_{B,\infty}^{\widehat{h}}$ and $q \in \mathbb{N}$, if $b \in S_{B,q}^{\widehat{h}}$, there is $s_B \in S_{B,\Delta}^q(\widehat{h})$ with $s_B|\widehat{h} = b$; thus, there exists $\mu_A$ that strongly believes $S_{A,\Delta}^q, ..., S_{A,\Delta}^0$ with $\mu_A(\cdot|h^0) = \overline{\overline{\mu}}_i(\overline{s}_A)(\cdot|h^0)$ and $\mu_A(s_B|\widehat{h}) = 1$, and $\rho(\mu_A) \subseteq S_{A,\Delta}^n(\widehat{h})$. So, the best replies to $b$ are in $S_{A,q+1}^{\widehat{h}}$. Then, since $\overline{S}_\infty^{\widehat{h}}$ is a set with the best reply property, $\emptyset \neq \overline{S}_\infty^{\widehat{h}} \subseteq S_\infty^{\widehat{h}}$, which contradicts $S_\Delta^\infty(\widehat{h}) = \emptyset$.

# 7 Appendix

**Proof of Lemma 3.** Fix $n \geq 0$ and suppose to have shown that for each $m \leq n$, $S_\Delta^m = S_{\Delta^*}^m$ ($S_\Delta^0 = S_{\Delta^*}^0$.trivially holds). If $S_\Delta^n = \emptyset$, $S_\Delta^{n+1} = S_{\Delta^*}^{n+1} = \emptyset$. Else, for each $i \in I$, there exists $\overline{\mu}_i \in \Delta_i$ that strongly believes $(S_{-i,\Delta}^q)_{q=0}^{n-1}$ 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 $m := \max\left\{q \leq n : s_i \in S_{i,\Delta}^q\right\}$. If $m > 0$, there exists $\mu_i \in \Delta_i$ that strongly believes $(S_{-i,\Delta}^q)_{q=0}^{m-1}$ such that $s_i \in \rho(\mu_i)$. Fix $\mu_i^* \in \Delta_i$ that strongly believes $(S_{-i,\Delta}^q)_{q=0}^{m-1}$ 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 $s_i^* \in \rho(\mu_i^*)(z) \subseteq S_{i,\Delta}^m$ 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 $(S_{-i,\Delta}^q)_{q=0}^n$. 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 $(S_{-i,\Delta}^q)_{q=0}^n = (S_{-i,\Delta^*}^q)_{q=0}^n$ 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. ■

**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.

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

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 $\mu^h_i$ 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 $\varsigma$ 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 $s^h_i\in\overline{S}^h_i$, and for every map $\varsigma:\widehat{h}\in D^S\mapsto s^{\widehat{h}}_i\in\widetilde{S}^h_{i,m}(\widehat{h})|\widehat{h}$, there exists $\widetilde{s}^h_i\in\widetilde{S}^h_{i,m}$ such that $\widetilde{s}^h_i=^{Z^S}\overline{s}^h_i(s^h_i)$ and $\widetilde{s}^h_i=^{\widehat{h}}\varsigma(\widehat{h})$ for all $\widehat{h}\in D^S$.

The map $\varsigma$ 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 $\widetilde{s}_i^h$'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 $z \in \zeta(\widehat{r}(\widehat{\mu}_l^{\widehat{h}}, \widehat{h}) \times \text{Supp}\widehat{\mu}_l^{\widehat{h}}(\cdot|\widehat{h}))$.*[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 $\rho(\mu_l^h)(\widehat{h}) \neq \emptyset$;*

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]

[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.

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

**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.**

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 $\overline{S}_\infty^h \neq \emptyset$, otherwise the lemma trivially holds.

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

[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.

**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.

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

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

[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.

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.

[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.


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

[10] Cho I.K. and D. Kreps, "Signaling Games and Stable Equilibria", *Quarterly Journal of Economics*, **102**, 1987, 179-222.

[11] 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.

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

[13] Osborne, M., "Signaling, Forward Induction, and Stability in Finitely Repeated Games", *Journal of Economic Theory*, **50**, 1990, 22-36.

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

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

[16] Penta, A., "Backward Induction Reasoning in Games with Incomplete Information", 2011, working paper.

[17] Perea, A., "Order Independence in Dynamic Games", working paper, 2017.

[18] Perea, A., "Why Forward Induction leads to the Backward Induction outcome: a new proof for Battigalli's theorem", *Games and Economic Behavior*, **110**, 2018, 120–138.

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

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