# On non-monotonic strategic reasoning.*

**Emiliano Catonini**†

January 2017

Strong Rationalizability (Battigalli and Siniscalchi, 2002) is a prominent and well understood solution concept for dynamic games, based on the notion of strong belief. Yet, the non-monotonicity of strong belief induces interpretative difficulties when strategic reasoning departs from Strong Rationalizability in two ways. In one direction, the introduction of exogenous belief restrictions (in the fashion of Strong-$\Delta$.Rationalizability; Battigalli 2003, Battigalli and Siniscalchi 2003) can drive the predictions surprisingly far from the strongly rationalizable outcomes. In the other direction, a slower pace of elimination of strategies typically modifies the output of the procedure. In this paper, I shed light and partially neutralize these two difficulties. In the first direction, I show that, reassuringly, belief restrictions that never end up off-path cannot induce non strongly rationalizable outcomes. Moreover, for belief restrictions on a specific path of play, the epistemic priority choice (Catonini, 2017) between the restrictions and rationality is immaterial for the predicted outcomes. In the second direction, I show that Strong Rationalizability is order independent with respect to the predicted outcomes. Since a truncated order of elimination corresponds to Backward Induction, I obtain 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 Rationalizability, Belief Restrictions, Epistemic Priority, Order Independence, Backward Induction.**

---

*A special thanks goes to Pierpaolo Battigalli, this paper would not exist without his mentoring. Thank you also to Adam Brandenburger, Shurojit Chatterji, Yi-Chun Chen, Alfredo Di Tillio, Amanda Friedenberg, Amanda Jakobsson, Atsushi Kajii, Mattia Landoni, Xiao Luo, Elena Manzoni, Andres Perea, Burkhard Schipper, Madhav Shrihari Aney, Satoru Takahashi, Elias Tsakas, and Dimitrios Tsomocos for precious suggestions.

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

# 1 Introduction

Strong Rationalizability (Battigalli and Siniscalchi, 2002) is the iterated deletion of never *sequential best replies* under *strong belief*[1] in opponents' strategies that survive the previous step(s).[2] Thus, the first step of Strong Rationalizability captures sequential rationality; the second step adds strong belief in the opponents' rationality; and so forth, towards *common strong belief in rationality.* These lines of strategic reasoning can be modified in two opposite directions, while seemingly maintaining the same spirit.

In one direction, the elimination procedure can be made more restrictive. A way to do so is by introducing, beside strong belief in the previous step(s), exogenous belief restrictions. Strong-$\Delta$-Rationalizability (Battigalli [3], Battigalli and Siniscalchi [6]) introduces the belief restrictions in the most intuitive way; that is, from the first step. However, doing so, the predictions can depart dramatically from Strong Rationalizability: even an outcome that is not compatible with the mere strong belief in rationality can be strongly-$\Delta$-rationalizable. Why is this the case? Are there relevant conditions under which belief restrictions do bring to a restriction of the predicted outcomes?

In the other direction, Strong Rationalizability can be made, at least temporarily, more permissive. At some step, some strategies may not be eliminated. If some strategy is anyway eliminated until only strategies that *cannot* be eliminated remain, I will simply talk of a different *elimination order* for Strong Rationalizability itself. Different elimination orders have an algorthimic interest, as players, for instance, may be tempted to compute the impact on the own choices of the elimination of an opponent's strategy, before considering which other strategies of the opponents can be eliminated at the same stage of reasoning. However, different elimination orders do in general modify the output of the procedure. Can this have an impact on the predicted outcomes?

The main goal of this paper is to shed light on these phenomena and answer the questions above in all finite dynamic games with observable actions.[3] It is well known that, in abstract terms, non-monotonicity with respect to belief restrictions and order dependence are effects of the non-monotonicity of strong belief. In both cases, a reduction/expansion of the strategy sets at one step does not necessarily lead to their reduction/expansion at the following step. Reasoning on the material implications of the non-monotonicity of strong belief provides indeed satisfactory insight on order dependence. "Forgetting" to eliminate a

[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] Games where every player always knows the current history of the game, i.e. - allowing for truly simultaneous moves - information sets are singletons. For instance, all repeated games with perfect monitoring are games with observable actions.

strategy may expand the set of reached history of a player and induce an opponent to refine further her strategies at that history at the next step. In my view, the non-monotonicity of strong belief does not provide as good intuition of the non-monotonicity with respect to belief restrictions. By contrast, a better and more useful understanding of the matter could be obtained by identifying conditions under which belief restrictions do restrict predictions.

Thus, the first result of the paper identifies a sufficient condition under which Strong-$\Delta$-Rationalizability delivers a subset of strongly rationalizable outcomes. Such condition is the absence of off-the-path belief restrictions with respect to the strongly-$\Delta$-rationalizable paths. Hence, predictions depart from Strong Rationalizability when the belief restrictions end up having bite off-path. This provides the following intuition: when a player abandons a history where belief restrictions have a bite, they stop being processed by the the opponents, who will not increase the level of strategic sophistication of the player upon reaching that history. This may induce a departure from the spirit of Strong Rationalizability, i.e. from common strong belief in rationality. Put down in this way, the sufficient condition seems to have little practical use, since Strong-$\Delta$-Rationalizability has to be performed anyway, thus revealing any possible non-monotonicity, in order to verify it. However, there are very interesting restrictions that always verify the condition: the belief in a specific path of play. Under such restrictions, Strong-$\Delta$-Rationalizability delivers either an empty set, or a set of paths that include the path itself, thus leaving no bite to belief restrictions off-path.

Selective Rationalizability ([7]), instead, introduces the restrictions after that Strong Rationalizability has been performed. In this way, the beliefs in rationality are given *epistemic priority* with respect to the beliefs in the restrictions and the spirit of Strong Rationalizability is preserved. Then, one could expect that for path restrictions, which preserve common strong belief in rationality also under Strong-$\Delta$-Rationalizability, the two procedures predict the same outcomes. I prove that this is actually the case: players do not have to bother about the epistemic priority issue when they agree on a path of play.

As for order dependence, I find that it is actually innocuous in terms of predicted outcomes. In light of the results of Chen and Micali [9], this comes as no surprise. Chen and Micali characterize Strong Rationalizability with the iterated elimination of *distinguishably dominated* strategies,[4] and show that the latter is order independent in terms of predicted outcomes. However, this does not imply order independence of Strong Rationalizability itself. Here I prove the order independence of Strong Rationalizability is terms of predicted outcomes, working directly with the iterated deletion of never sequential best replies, thus with strong belief and without dominance characterizations.

[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 [15] to be equivalent to Extensive Form Rationalizability (Pearce [13]), which is in turn equivalent to Strong Rationalizability.

A way to start eliminating iteratively never sequential best replies is actually Backward Induction. First, one can iteratively eliminate only strategies that are not optimal against any valid belief at pre-terminal histories of the game. Then, one can move to histories of depth 2, and so on. Once Backward Induction is completed, the elimination procedure can continue until all strategies are sequential best replies to valid beliefs, and the strongly rationalizable outcomes are obtained. But then, Backward Induction yields a superset of those outcomes. This result was already proved by Chen and Micali, but again via their dominance characterization. Since in perfect information games without relevant ties the backward induction outcome is unique, the outcome equivalence of backward and forward induction in this class of games due to Battigalli [1] obtains.[5]

Technically, all the result of the paper rely on one key outcome inclusion result. First, I define a vast class of *elimination procedures* where, at the end of the procedure, each surviving strategy is optimal under strong belief in all the steps of the procedure. Consider the output of an elimination procedure and a set of beliefs that justify it. Fix another elimination procedure. Suppose that, in both procedures, a strategy survives a step of elimination if (but not necessarily only if) it is optimal against a belief that mimicks a belief in the set along the paths induced by the first procedure *and* strongly believes in the previous steps of the procedure. Then, the second elimination procedure delivers a superset of the outcomes predicted by the first procedure.

Section 2 introduces the formal framework for the analysis. Section 3 defines elimination procedures and provides the key lemma and the final results. Section 4 is dedicated to the proof of the technical lemma.

## 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 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 $Z := \{z \in \overline{H} : \forall h \in \overline{H}, z \not\prec h\}$ be the set of terminal histories (henceforth, *outcomes* or

[5] The corollary was also noticed by Chen and Micali, and the result was also proved algorithmically by Heifetz and Perea [10] (whereas Battigalli relied on stability arguments a la Kohlberg and Mertens [11]).

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

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

*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, [14]; 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:*

*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* $\mu(E|C) = \mu(E|D)\mu(D|C)$.

[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.

*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 $\mu_i$ 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 $\mu_i$ 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 Main results

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* $s^h_i\in\rho(\mu^h_i)\subseteq S^h_{i,\infty}$.

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

**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 $s^h_{-i}\longmapsto s^h_{-i}|\widehat{h}$ 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, $\overline{s}^{\widehat{h}}_i\in S^h_{i,\infty}(\widehat{h})|\widehat{h}$. ■

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 [14] for the procedures in the whole game I am ultimately interested in.

**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 $((S^h_{i,q})_{i\in I})^\infty_{q=0}$ 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 $(S^q_{-i,R\Delta})^{n-1}_{q=0}$.*[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 $\mu^h_i$ 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 $\zeta(\overline{S}^h_\infty) \subseteq \zeta(S^h_\infty)$.*

While Lemma 2 may seem rather intuitive, it is not immediate to prove it for the following reason. Off the paths predicted by the first procedure (henceforth, just "paths"), the two procedures can depart in terms of predicted sub-paths. This might create the incentive for deviations from the paths along the second procedure. Yet, this does not happen. Suppose that, at some step of the second procedure, for some belief over the paths $\overline{\overline{\mu}}^h_i(s^h_i)$, no mimicking belief $\mu^h_i$ as above was able anymore to discourage a deviation from the paths. This means that there exists a unilateral deviation from the paths that the deviator finds profitable for any belief over the continuation plans of the opponents thereafter (Lemma 7). The opponents may be surprised by the deviation, hence they may reach the post-deviation history for any post-deviation belief too (Lemma 5). Keep refining separately the continuation plans after the deviation, as in an unconstrained elimination procedure. The final output provides a set of subpaths that all justify the deviation. Thus, all these subpaths survive under the first procedure, a contradiction. This is given by Lemma 2 in the subgame that follows the deviation, inverting the roles of the first procedure (taking just its implications after the deviation) and the second procedure (taking just the surviving continuation plans at the step of the deviation and refining them separately). The iterative

[11] 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).

procedure continues until the deviation is followed by a subgame of depth 1. There, no further deviation is available, so Lemma 2 holds trivially. Section 4 is dedicated to prove Lemma 2. Now I focus on the implications of Lemma 2 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 $\mu_i$, only the beliefs at the strongly-$\Delta$-rationalizable histories about the strongly-$\Delta$-rationalizable paths matter to determine whether $\mu_i$ belongs to $\Delta_i$ or not. Then, Strong-$\Delta$-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 $\widetilde{h} \in H(S_\Delta^\infty)$ and $z \in \zeta(S_\Delta^\infty)$, then $\mu_i' \in \Delta_i$. Then, $\zeta(S_\Delta^\infty) \subseteq \zeta(S^\infty)$.*

**Proof.** For each $i \in I$ and $s_i \in S_{i,\Delta}^\infty$, fix any $\overline{\overline{\mu}}_i^h(s_i^h) \in \Delta_i$ that strongly believes $(S_{-i,\Delta}^q)_{q=0}^\infty$ 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 $(S_{-i}^q)_{q=0}^{m-1}$, $\rho(\mu_i) \in S_i^m$. Thus, by Lemma 2, $\zeta(S_\Delta^\infty) \subseteq \zeta(S^\infty)$. ■

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-$\Delta$-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 $\Delta_i$ be the set of all $\mu_i$'s such that $\mu_i(S_{-i}(z)|h) = 1$ for all $h \prec z$, and let $\Delta_i^*$ be the set of all $\mu_i$'s that strongly believe $S_j(z)$ for all $j \neq i$. Then, $S_\Delta^\infty = S_{\Delta^*}^\infty$ and $S_{R\Delta}^\infty = S_{R\Delta^*}^\infty$.*

**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, $S^{n+1}_\Delta \subseteq S^{n+1}_{\Delta^*}$. By $\Delta^*_i \subseteq \Delta_i$ and $(S^q_{-i,\Delta})^n_{q=0} = (S^q_{-i,\Delta^*})^n_{q=0}$, $S^{n+1}_{\Delta^*} \subseteq S^{n+1}_\Delta$.

The proof can be repeated for Selective Rationalizability with $n \geq M$ in place of $n \geq 0$, where $(S^q_{R\Delta})^M_{q=0} = (S^0_{R\Delta^*})^M_{q=0}$ 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-$\Delta$-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 $\Delta^*_i$ be the set of all $\mu_i$'s that strongly believe $S_j(z)$ for all $j \neq i$. Then $\zeta(S^\infty_{\Delta^*}) \subseteq \zeta(S^\infty)$.*

**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 $S^\infty_\Delta = \emptyset$, $\zeta(S^\infty_\Delta) \subseteq \zeta(S^\infty)$ is trivially true, so suppose $S^\infty_\Delta \neq \emptyset$. For each $i \in I$, and $s_i \in S^\infty_{i,\Delta}$, $s_i \in \rho(\overline{\mu}_i)$ 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 $\zeta(S^\infty_\Delta) \subseteq \zeta(S^\infty)$. Then, by Lemma 3, $\zeta(S^\infty_{\Delta^*}) \subseteq \zeta(S^\infty)$. ■

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-$\Delta$-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 $\Delta_i^*$ be the set of all $\mu_i$'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 $\Delta_i$ 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 $(S_j(z))_{j \neq i}$.

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 Strong Rationalizability. Thus, using Theorem 2 in both directions with the maximal unconstrained elimination procedure and any non maximal one, the order independence of Strong Rationalizability 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 $\mu_i$ 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]

[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.

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 $\mu_i$ 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.*

**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 $i \in I$, $\mu_i$ that strongly believes $(\widetilde{S}_{-i}^q)_{q=0}^{n-1}$, 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, $s_i \in \widetilde{S}_i^n$. ■

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.*

# 4 Proof of the main lemma.

Additional notation is needed. For any $h \in H$, $\widehat{h} \succeq h$, $(s_j^h)_{j\in I} \in S^h$, $(s_j^{\widehat{h}})_{j\in I} \in S^{\widehat{h}}$, $\mu_i^h \in \Delta^{H^h}(S_{-i}^h)$, $\widetilde{\mu}_i^{\widehat{h}} \in \Delta^{H^{\widehat{h}}}(S_{-i}^{\widehat{h}})$, $\widehat{Z} \subseteq Z^{\widehat{h}}$, and $J \subseteq I$, let:

- $s_J^h =^{\widehat{Z}} s_J^{\widehat{h}}$ if for each $z \in \widehat{Z}$ and $\widehat{h} \preceq \widetilde{h} \prec z$, $s_J^h(\widetilde{h}) = s_J^{\widehat{h}}(\widetilde{h})$;
- $\mu_i^h =^{\widehat{Z}} \widetilde{\mu}_i^{\widehat{h}}$ if for each $z \in \widehat{Z}$ and $\widehat{h} \preceq \widetilde{h} \prec z$, $\mu_i^h(S_{-i}^h(z)|\widetilde{h}) = \widetilde{\mu}_i^{\widehat{h}}(S_{-i}^{\widehat{h}}(z)|\widetilde{h})$;
- $s_J^h =^{\widehat{h}} s_J^{\widehat{h}}$ and $\mu_i^h =^{\widehat{h}} \widetilde{\mu}_i^{\widehat{h}}$ if, respectively, $s_J^h =^{Z^{\widehat{h}}} s_J^{\widehat{h}}$ and $\mu_i^h =^{Z^{\widehat{h}}} \widetilde{\mu}_i^{\widehat{h}}$.

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:

- $D_i(\overline{S}^h) := \{\widetilde{h} \in H \backslash H(\overline{S}^h) : p(\widetilde{h}) \in H(\overline{S}^h) \wedge \widetilde{h} \in H(\overline{S}^h_{-i})\}$.[13]

The next 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 $\widehat{s}^h_i, \overline{s}^h_i \in S^h_{i,n}$ and $\widehat{h}, \overline{h} \in H(\widehat{s}^h_i) \cap H(\overline{s}^h_i)$ such that $\overline{h} \not\preceq \widehat{h} \not\preceq \overline{h}$: there needs not exist $s^h_i \in S^h_{i,n}(\widehat{h}) \cap S^h_{i,n}(\overline{h})$ such that $s^h_i|\widehat{h} = \overline{s}^h_i|\widehat{h}$ and $s^h_i|\overline{h} = \widehat{s}^h_i|\overline{h}$. 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 certain expectations after $\widehat{h}$, or because she is confident that $\overline{h}$ will be reached and she has certain expectations after $\overline{h}$. If $\widehat{s}^h_i$ is best reply to the first conjecture and $\overline{s}^h_i$ is best reply to the second conjecture, $\widehat{s}^h_i|\overline{h}$ and $\overline{s}^h_i|\widehat{h}$ 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.

Consider a player who may be surprised by history $\widehat{h}$, in the sense that she may play a strategy that allows $\widehat{h}$ while believing that the deviators do not until $\widehat{h}$ is actually reached. This player can keep the same beliefs and the same strategy out of $\Gamma(\widehat{h})$, whatever she believes the deviators would do and hence however she may play after $\widehat{h}$.

**Lemma 5** *Fix an elimination procedure* $((S^h_{i,q})_{i \in I})_{q \geq 0}$*,* $i \in I$*,* $n \in \mathbb{N}$*,* $\widehat{h} \in H^h$*, and* $\mu^h_i$ *t.s.b.* $(S^h_{-i,q})^{n-1}_{q=0}$ *such that* $\mu^h_i(S^h_{-i}(\widehat{h})|p(\widehat{h})) = 0$*. Fix* $s^h_i \in \rho(\mu^h_i) \cap S^h_i(\widehat{h})$*,* $\mu^{\widehat{h}}_i$ *t.s.b.* $(S^h_{-i,q}(\widehat{h})|\widehat{h})^{n-1}_{q=0}$ *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* $\widetilde{\mu}^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$ *t.s.b.* $(S^h_{-i,q})^{n-1}_{q=0}$ *such that for every* $\widetilde{h} \notin H^{\widehat{h}}$*,* $\widetilde{\mu}^h_i(\cdot|\widetilde{h}) = \mu^h_i(\cdot|\widetilde{h})$*, 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, I 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 $\widetilde{\mu}^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$, satisfies CPS-2, and strongly believes $(S^h_{-i,q})^{n-1}_{q=0}$. 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 $s^{\widehat{h}}_{-i} \in S^{\widehat{h}}_{-i}$, $\widetilde{\mu}^h_i(\varsigma(s^{\widehat{h}}_{-i})|\widetilde{h}) =$

[13] Set of histories that follow a unilateral deviation by player $i$ from the histories induced by $\overline{S}^h$.

$\mu_i^{\widehat{h}}(s_{-i}^{\widehat{h}}|\widetilde{h})$. For every $\widehat{s}_i^h \in S_i^h(\widehat{h})$, $\zeta(\widehat{s}_i^h|\widehat{h}, s_{-i}^{\widehat{h}}) = \zeta(\widehat{s}_i^h, \varsigma(s_{-i}^{\widehat{h}}))$. So, $\widehat{s}_i^h|\widehat{h} = s_i^{\widehat{h}} \in \widehat{r}(\mu_i^{\widehat{h}}, \widetilde{h})$ implies $\widehat{s}_i^h \in \widehat{r}(\widetilde{\mu}_i^h, \widetilde{h})$. ■

Lemma 6 exploits Lemma 5 from the perspective of a deviator. If the opponents may be surprised by different deviations from the same predicted behavior, the deviator can expect any combination of reactions. The lemma is less general to target the particular setting in which it will be used.

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

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

*A2 for every* $s_i^h \in \overline{S}_i^h$ *and* $\mu_i^h =^{Z^S} \overline{\overline{\mu}}_i^h(s_i^h)$ *t.s.b.* $(\widetilde{S}_{-i,q}^h)_{q=0}^{m-1}$, $\rho(\mu_i^h) \subseteq \widetilde{S}_{i,m}^h$.

*Fix* $l \in I$ *and* $s_l^h \in \overline{S}_l^h$. *Let* $D^S := D_l(\overline{S}^h)$. *For every* $\widehat{h} \in D^S$, *fix* $\widetilde{\mu}_l^{\widehat{h}}$ *t.s.b.* $(\widetilde{S}_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$. *There exists* $\widetilde{\mu}_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(s_l^h)$ *t.s.b.* $(\widetilde{S}_{-l,q}^h)_{q=0}^m$ *such that* $\widetilde{\mu}_l^h =^{\widehat{h}} \widetilde{\mu}_l^{\widehat{h}}$ *for all* $\widehat{h} \in D^S$.

**Proof.**

I show that for every $i \neq l$, $s_i^h \in \overline{S}_i^h$, and $\varsigma : \widehat{h} \in D^S \mapsto s_i^{\widehat{h}} \in \widetilde{S}_{i,m}^h(\widehat{h})|\widehat{h}$, 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 $\varsigma$ is well defined because by A1 and A2, $H^S \subseteq H(\widetilde{S}_{-l,m}^h)$, thus $D^S \subseteq H(\widetilde{S}_{-l,m}^h)$ as well. 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 $\mu_i^{\widehat{h}}$ t.s.b. $(\widetilde{S}_{-i,q}^h(\widehat{h})|\widehat{h})_{q=0}^{m-1}$ such that $\varsigma(\widehat{h}) \in \rho(\mu_i^{\widehat{h}})$. Thus, by Lemma 5, there exist $\widetilde{\mu}_i^h =^{\widehat{h}} \mu_i^{\widehat{h}}$ t.s.b. $(S_{-i,q}^h)_{q=0}^{m-1}$ 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 $\widehat{h} \in D^S$, I obtain $\widetilde{\mu}_i^h =^{Z^S} \overline{\mu}_i^h(s_i^h)$ 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, $\widetilde{s}_i^h \in \widetilde{S}_{i,m}^h$. ■

Fix a set of strategy profiles $\overline{S}^h$ delivered by an elimination procedure. Suppose that until step $n$, each player $i$ is willing to play strategies that mimic those in $\overline{S}_i^h$ along the paths induced by $\overline{S}^h$ while expecting the deviators to do the same. At step $n+1$, instead, some player $l$ stops playing any strategy of hers that mimics a strategy $\widehat{s}_l^h$ in $\overline{S}_l^h$. Since at $n$ the deviators may be surprised by any deviation, player $l$ might expect them to play any combination of substrategies that survive $n$ steps after the potential deviations. Hence, there must exist one particular deviation that player $l$ prefers to mimicking $\widehat{s}_l^h$ whatever she may conjecture thereafter.

**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* $i \in I$ *call* $\overline{S}_i^h := \overline{S}_{i,\infty}^h$ *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* $\mu_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$ *t.s.b.* $(S_{-l,q}^h)_{q=0}^n$, $s_l^h \notin \rho(\mu_l^h)$.

*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* $\mu_l^{\widehat{h}}$ *t.s.b.* $(S_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$ *(resp.* $(\overline{S}_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$*) such that* $\mu_l^{\widehat{h}}(S_{-i}(z)|\widehat{h}) = \widehat{\mu}_l^{\widehat{h}}(S_{-i}(z)|\widehat{h})$ *for some* $\widehat{\mu}_l^{\widehat{h}}$ *t.s.b.* $(S_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^n$ *and 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] *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* $\mu_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$ *t.s.b.* $(S_{-l,q}^h)_{q=0}^m$ *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* $\widetilde{\mu}_l^h =^{Z^S} \overline{\overline{\mu}}_l^h(\widehat{s}_l^h)$ *t.s.b.* $(\overline{S}_{-l,q}^h)_{q=0}^p$ *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 $\widetilde{\mu}_l^h =^{Z^S} \overline{\overline{\mu}}_l^h$ t.s.b. $(S_{-l,q}^h)_{q=0}^n$ such that for every $\widehat{h} \in D^S$, $\widetilde{\mu}_l^h =^{\widehat{h}} \widehat{\mu}_l^{\widehat{h}}$. I 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.

[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 the theorems.

[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. $\overline{M}_p^{\widehat{h}} = \emptyset$).

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 $\mu_l^h =^{\widehat{h}} \mu_l^{\widehat{h}}$ t.s.b. $(S_{-l,q}^h)_{q=0}^{m(\widehat{h})}$ 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, $\rho(\mu_l^h)(\widehat{h}) \neq \emptyset$ 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 $\mu_{l,j}^h =^{Z^h \setminus \cup_{t=1}^j Z^{h^t}} \overline{\overline{\mu}}_l^h$ t.s.b. $(\overline{S}_{-l,q}^h)_{q=0}^{m(h^j)}$ 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 obviously $\rho(\mu_{l,j-1}^h) \cap S_l^h(h^j) = \emptyset$. 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 $\overline{S}_{-l}$, $\widehat{s}_l^h$, 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 $\mu_l^{\widehat{h}}$. Then, $\widehat{s}_l^h \in \widehat{r}(\mu_{l,k}^h, \widetilde{h})$ 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 $\widetilde{h} \in H(\widehat{s}_l^h) \cap H^S$. ■

**Proof of Lemma 2.**

Recall that the depth of a game is the length of the longest terminal history of the game. The lemma trivially holds for games of depth 1, i.e. simultaneous moves games. Thus, suppose that the lemma holds for games of depth $1, ..., k-1$. I show that the lemma holds for games of depth $k$. Let $\overline{S}_\infty^h \neq \emptyset$, otherwise the inclusion is trivially verified.

I 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.

**Induction Hypothesis (n):** $(S_q^h)_{q=0}^\infty$ satisfies A3 at $n$ (so by A4 $\zeta(S_n^h) \supseteq \zeta(\overline{S}_\infty^h)$).

**Basis step (1):** for every $i \in I$, the I.H. holds with $\overline{\mu}_i^h(\cdot) = \overline{\overline{\mu}}_i^h(\cdot)$.

**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 $\widehat{s}_l^h \in \overline{S}_{l,\infty}^h$. Lemma 7 yields $\widehat{h} \in D_l(\overline{S}_\infty^h)$. Define $((\overline{S}_{i,q}^{\widehat{h}})_{i \in I})_{q \geq 0}$ as follows: for every $i \in I$ and $m \leq n$, $\overline{S}_{i,m}^{\widehat{h}} = S_{i,m}^h(\widehat{h})|\widehat{h}$; for every $m > n$, $s_i^{\widehat{h}} \in \overline{S}_{i,m}^{\widehat{h}}$ if and only if there exists $\mu_i^{\widehat{h}}$ t.s.b. $(\overline{S}_{-i,q}^{\widehat{h}})_{q=0}^{m-1}$ such that $s_i^{\widehat{h}} \in \rho(\mu_i^{\widehat{h}})$.

[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.

For every $i \neq l$, since $\widehat{h} \in D_l(\overline{S}^h_\infty)$, $\emptyset \neq \overline{S}^h_{i,\infty}(\widehat{h})$. 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 $\overline{\mu}^h_i(\widehat{s}^h_i) =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_i(\widehat{s}^h_i)$ t.s.b. $(S^h_{-i,q})^{m-1}_{q=0}$ 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 $\mu^{\widehat{h}}_i$ t.s.b. $(\overline{S}^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$, there exists $\mu^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$ t.s.b. $(S^h_{-i,q})^{m-1}_{q=0}$ 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 $\mu^{\widehat{h}}_l$ t.s.b. $(\overline{S}^{\widehat{h}}_{-l,q})^n_{q=0}$: trivially $\mu^{\widehat{h}}_l \in M^{\widehat{h}}_n$. Hence by Lemma 7.(1) there exists $\widetilde{\mu}^h_l =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_l(\widehat{s}^h_l)$ t.s.b. $(S^h_{-l,n})^n_{q=0}$ 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 $\mu^{\widehat{h}}_i$ t.s.b. $(\overline{S}^{\widehat{h}}_{-i,q})^n_{q=0}$, $\rho(\mu^{\widehat{h}}_i) \subseteq \overline{S}^{\widehat{h}}_{i,n} \neq \emptyset$. 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 $m \leq n$, $\widehat{\mu}^{\widehat{h}}_l$ t.s.b. $(\overline{S}^{\widehat{h}}_{-l,q})^\infty_{q=0}$, and $\mu^{\widehat{h}}_l =^{\zeta(\overline{S}^{\widehat{h}}_\infty)} \widehat{\mu}^{\widehat{h}}_l$ t.s.b. $(\overline{S}^{\widehat{h}}_{-l,q})^{m-1}_{q=0}$, $\mu^{\widehat{h}}_l \in M^{\widehat{h}}_m$;[18] thus by Lemma 7.(1) there exists $\widetilde{\mu}^h_l =^{\zeta(\overline{S}^h_\infty)} \overline{\overline{\mu}}^h_l(\widehat{s}^h_l)$ t.s.b. $(S^h_{-l,q})^{m-1}_{q=0}$ 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 $m \in \mathbb{N}$, $i \in I$, $\widehat{\mu}^{\widehat{h}}_i$ t.s.b. $(\overline{S}^{\widehat{h}}_{-i,q})^\infty_{q=0}$ and $\mu^{\widehat{h}}_i =^{\zeta(\overline{S}^{\widehat{h}}_\infty)} \widehat{\mu}^{\widehat{h}}_i$ t.s.b. $(\overline{S}^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$, $\rho(\mu^{\widehat{h}}_i) \subseteq \overline{S}^{\widehat{h}}_{i,m}$. 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 Lemma 1 it is an elimination procedure.

For every $i \neq l$, $m \in \mathbb{N}$, and $\mu^{\widehat{h}}_i$ t.s.b. $(S^{\widehat{h}}_{-i,q})^{m-1}_{q=0}$, by Lemma 5 there exists $\widetilde{\mu}^h_i =^{\widehat{h}} \mu^{\widehat{h}}_i$ t.s.b. $(\overline{S}^h_{-i,q})^{m-1}_{q=0}$ 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, $\rho(\widetilde{\mu}^h_i) \subseteq \overline{S}^h_{i,m}$.

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

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

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

[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.

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

## 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.

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

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