Online Appendix of "Non-binding agreements and forward induction reasoning"

Emiliano Catonini*

November 2015

0.1 Complete agreements on subgame perfect equilibria under priority to the agreement.

For complete agreements the equivalence between self-enforceability under priority to the agreement and under priority to rationality does not hold. When the epistemic priority falls on the agreement, a "strict" SPE is self-enforcing. I say that an equilibrium $s \in S$ is *strict* when for every $i \in I$, $r_i(s_{-i}) \subseteq S_i(\zeta(s))$. Strictness yields self-enforceability on top of credibility (of the whole agreement; the path agreement credibility test can give negative response).

Proposition 1 *Fix a strict SPE $s \in S$ of a game with observable actions.*[1] *The corresponding complete agreement is self-enforcing under priority to the agreement.*

Proof:

For every $i \in I$ and $\mu_i \in \Delta_i^e$, $\mu_i(s_{-i}|h^0) = 1$, hence $\rho(\mu_i) \subseteq r(s_{-i}) \subseteq S_i(\zeta(s))$ by strictness. So $S_{i,\Delta^e}^{\infty} \subseteq S_{i,\Delta^e}^{1} \subseteq S_i(\zeta(s))$, yielding self-enforceability in case of credibility.

Fix $i \in I$ and $m \geq 1$. For every $n \leq m$ and $h \in H(S_{i,\Delta^e}^n)$, fix $\mu_i \in \Delta_i^e$ that strongly believes $(S_{-i,\Delta^e}^q)_{q=0}^{n-1}$ such that $\rho(\mu_i)(h) \neq \emptyset$. For every $\widetilde{s}_{-i} \in \mathrm{Supp}\mu_i(\cdot|h)$ and $\widetilde{h} \succeq h$,

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

[1] Battigalli and Friedenberg [1] provide a game without observable actions and with no relevant ties (which implies strictness of equilibria) where a SPE outcome is not induced by any extensive form best response set and so it cannot be delivered by strong-Δ-rationalizability (see next subsection).

$\widetilde{s}_{-i}(\widetilde{h}) = s_{-i}(\widetilde{h})$. Thus by subgame perfection, there exists $\widetilde{s}_i \in \rho(\mu_i)(h) \subseteq S^n_{i,\Delta^e}(h)$ such that for every $\widetilde{h} \succeq h$, $\widetilde{s}_i(\widetilde{h}) = s_i(\widetilde{h})$. Then for every $j \in I$, there exists $\mu_j \in \Delta^e_j$ that strongly believes $(S^q_{-j,\Delta^e})^m_{q=0}$. So $S^{m+1}_{j,\Delta^e} \neq \emptyset$. Inductively $S^{\infty}_{j,\Delta^e} \neq \emptyset$. ■

0.2 Implementability, extensive form best response sets and mutually acceptable courses of actions

Natural candidates for implementability are the sets of outcomes for which there exists an *Extensive Form Best Response Set* (Battigalli and Friedenberg, [1]; henceforth EFBRS) inducing a subset of them. EFBRS's are defined in [1] for the 2-players case. They show that any EFBRS is delivered by strong-Δ-rationalizability under some restrictions, while strong-Δ-rationalizability always delivers an EFBRS. The concept and the result are extended to the N-players incomplete information case by Battigalli and Prestipino [2]. Here I present a simpler version of the two for the framework of this paper.

Definition 1 (EFBRS) *An Extensive Form Best Response Set is a cartesian subset of strategy profiles $Q \subset S$ such that for every $i \in I$ and $s_i \in Q_i$, there exists $\mu_i \in \Delta^H(S_{-i})$ such that:*

1. $s_i \in \rho(\mu_i)$;
2. μ_i *strongly believes* Q_{-i} $(\forall h \in H,\ Q_{-i}(h) \neq \emptyset \Longrightarrow \mu_i(Q_{-i}|h) = 1)$;
3. $\rho(\mu_i) \subseteq Q_i$.

Proposition 2 *Fix $\times_{i\in I} Q_i \subset S$. The following are equivalent:*

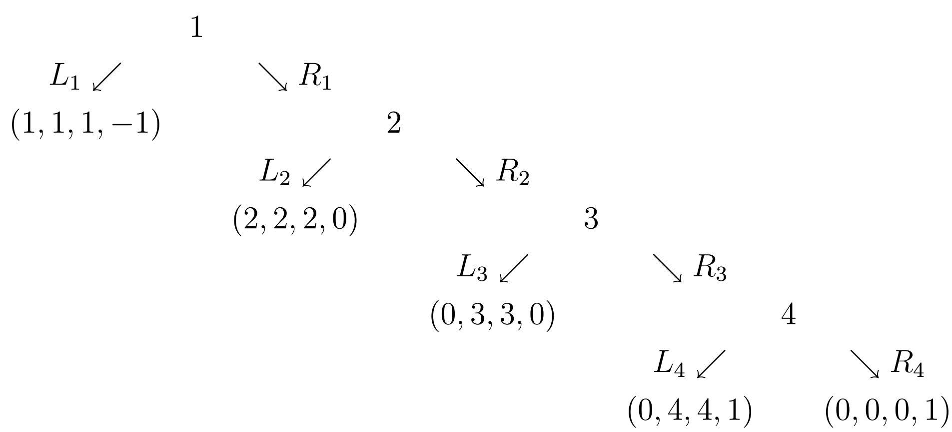
1. *Q is an EFBRS;*
2. *$Q = S^{\infty}_{\Delta}$ for some first-order belief restrictions $(\Delta_i)_{i\in I}$.*

Selective rationalizability can be seen as a special case of strong-Δ-rationalizability by incorporating S3 in the first-order-beliefs restrictions. Hence, also selective rationalizability delivers an EFBRS. Thus, regardless of the epistemic priority assumption, I can claim the following.

Proposition 3 *A set of outcomes $P \subset Z$ is implementable only if there exists an EFBRS Q such that $\zeta(Q) \subseteq P$.*

Proof. If P is implementable, $\zeta(S_e^\infty) \subseteq P$ for some credible $e = (e_i)_{i \in I}$ and by the proposition above S_e^∞ is an EFBRS. ■

The existence of an EFBRS Q such that $\zeta(Q) \subseteq P$ is however not sufficient for P to be implementable, even when the epistemic priority falls on the agreement. The reason is that the first-order belief restrictions delivering Q through strong-Δ-rationalizability could not correspond to any agreement. First, it may be impossible to translate them into *independent* restrictions over *pure* actions. Moreover, when players are more than two, they may need to have different restrictions about a third player, like in this game borrowed from Greenberg, Gupta and Luo [3].



$((R_1, L_2, R_3, L_4), (R_1, L_2, R_3, R_4))$ (inducing outcome R_1, L_2) is an EFBRS. Indeed it is delivered by strong-Δ-rationalizability when the restrictions impose on player 2 the belief that 4 will play R_4 and on player 3 the belief that 4 will play L_4. These restrictions cannot come from an agreement because 2 and 3 have discordant ideas about how 4 will play. In the analysis of [3] the path (R_1, L_2) is instead a *mutually acceptable course of actions*. In their interpretation, (R_1, L_2) can be achieved through an agreement between player 1 and 2. Such agreement does not work here, not because of forward induction (which is not a feature of [3]) but because the definition of implementability is more restrictive. An outcome is implemented if the first-order-belief restrictions *induced by the agreement* suffice to deliver it as the *only* possible one.

0.3 Formal analysis of the examples of Section 2

Formal analysis of example 1 in Section 2.

$P_1 \backslash P_2$	M	R
M	$(6,6)$	$\cdot -$
R	$(0,0)$	$(5,5)$

$- - - \quad - \rightarrow$

$P_1 \backslash P_2$	N	O
A	$(2,0)$	$(1,4)$
S	$(0,10)$	$(2,8)$

Agreement:

$e_1(h^0) = \{M\}$, $e_1((M,R)) = \{A\}$; $e_2(h^0) = \{M\}$, $e_2((M,R)) = \{N\}$.

First-order-belief restrictions:

$\Delta_1^e = \{(\mu_1(\cdot|h^0), \mu_1(\cdot|(M,R))) \in \Delta^H(S_2) : \mu_1((M.N)|h^0) = 1, \mu_1(R.N|(M,R)) = 1\}$;

$\Delta_2^e = \{(\mu_2(\cdot|h^0), \mu_2(\cdot|(M,R))) \in \Delta^H(S_1) : \mu_2((M.A)|h^0) = 1, \mu_2(M.A|(M,R)) = 1\}$.

Note that the two sets are singletons.

Strong-Δ-rationalizability:

$S^1_{1,\Delta^e} = \{M.A\}, S^1_{2,\Delta^e} = \{M.N, M.O\}$, $S^\infty_{1,\Delta^e} = \{M.A\}, S^\infty_{2,\Delta^e} = \{M.N, M.O\}$.

The sequentially rational strategies of player 2 do not allow (M,R); hence, at the second step, player 1 can still believe in $R.N$ after (M,R) and the procedure already comes to convergence. Strongly-Δ-rationalizable strategies comply with the agreement at the reached information sets. The agreement is self-enforcing under priority to the agreement.

Strong rationalizability:

$S_1^1 = \{M.A, M.S, R.A, R.S\}, S_2^1 = \{M.N, M.O, R.N, R.O\}$;

$S_1^\infty = \{M.A, M.S, R.A, R.S\}, S_2^\infty = \{M.N, M.O, R.N, R.O\}$.

each strategy is a sequential best reply to some conjecture: for $s_1 = R.\cdot$ and $s_2 = M.\cdot$ it is enough to put at h^0 probability 1 on, respectively, $R.\cdot$ and $M.A$; for $s_1 = M.A$ and $s_1 = M.S$ on $M.\cdot$ at h^0 and on, respectively, $R.N$ and $R.O$ at (M,R); for $s_2 = R.N$ and $s_2 = R.O$ on $R.\cdot$ at h^0 and on, respectively, $M.S$ and $M.A$ at (M,R).

Selective rationalizability:

$S^1_{1,R\Delta^e} = \{M.A\}, S^1_{2,R\Delta^e} = \{M.N, M.O\}; S^\infty_{1,R\Delta^e} = \{M.A\}, S^\infty_{2,R\Delta^e} = \{M.N, M.O\}$.

The procedure is equivalent to strong-Δ-rationalizability, because all strategies are strongly rationalizable, hence S3 has no bite.

Formal analysis of example 2 in Section 2.

$A\backslash B$	C	D	P
C	$5,5$	$2,6$	$0,2$
D	$6,2$	$3,3$	$0,2$
P	$2,0$	$2,0$	$1,1$

The game is symmetric, so what follows holds for $i = A, B$.

Agreement:

$e_i(h^0) = \{C\}$; $e_i((C,C)) = \{D\}$; $e_i(h) = \{C, D, P\}\ \forall h \neq h^0, (C,C)$.

First-order-belief restrictions:

$\Delta^e_i = \{\mu_i \in \Delta^H(S_{-i}) : \mu_i(S_{-i}((C,C),(D,D)))|h^0) = 1\}$

(the conjectures are restricted also at (C,C) by the chain rule.)

Strong-Δ-rationalizability:

$S^1_{i,\Delta^e} = \{s_i \in S_i : s_i((s_i(h^0), C)) = D \wedge s_i((s_i(h^0), D)) \neq C \neq s_i((s_i(h^0), P))\}$

It is worth deviating in the first stage only if after the expected C, P is expected with sufficiently low probability, so that D is the best reply. A deviation to P is (weakly) optimal if it is expected to induce the co-player to cooperate in the second stage. Moreover, there is no incentive to cooperate in the second stage.

$S^2_{i,\Delta^e} = \{s_i \in S^1_{i,\Delta^e} : s_i(h^0) \neq P \wedge s_i(h^0) = C \Rightarrow s_i((C,D)) = s_i((C,P)) = D\}$

There is no incentive to punish in the first stage since it cannot trigger C in the second. Moreover, if the co-player has deviated and her beliefs are confirmed by observing C, she will not play P, so there is no incentive to react with P to the deviation after having cooperated.

$S^3_{i,\Delta^e} = \{s_i \in S^2_{i,\Delta^e} : s_i(h^0) = D\}$

There is no incentive to cooperate in the first stage since defecting will not trigger the punishment.

$S^4_{i,\Delta^e} = \emptyset$, i.e. the agreement is not credible.

Strong rationalizability:

$S^1_i = \{s_i \in S_i : \forall h \in H(s_i), h \neq h^0, s_i(h) \neq C \wedge s_i(h^0) = C \Rightarrow \exists a_{-i} \in \overline{A}_{-i}, s_i((C, a_{-i})) \neq P\}$

It is not rational to play C in the second stage, or in the first and then play P whatever the co-player has chosen.[2] Excluding to believe in such strategies, among the remaining ones also those that feature C in the first stage may be sequential best replies when one expects P as a reaction otherwise. Thus, we have:

$S^\infty_i = S^1_i$.

Selective rationalizability:

$S^1_{i,R\Delta^e} = S^1_{i,\Delta^e} \cap \{s_i \in S_i : s_i(h^0) \neq P\}$

Since no rational(izable) strategy of the co-player prescribes to cooperate in the second stage, punishing in the first stage cannot induce cooperation in the second.

$S^2_{i,R\Delta^e} = \{s_i \in S^1_{i,R\Delta^e} : s_i(h^0) = C \Rightarrow s_i((C, D)) = D\}$

After cooperating against a defection, there is no incentive to punish. Note that $S^2_{i,R\Delta^e}$ is not a subset of S^2_{i,Δ^e} because having already excluded punishment in the first stage, if punishment occurs there is no constraint to expect defection thereafter. This is due to the epistemic priority difference: if player 1 observes P in the first stage, here she concludes that player 2 does not believe in the agreement, so 2 could expect, say, always P; before, player 1 concluded that player 2 does not believe that player 1 is rational, so 2 could expect C in the second stage.

$S^3_{i,R\Delta^e} = \{s_i \in S^2_{i,R\Delta^e} : s_i(h^0) = D\}$

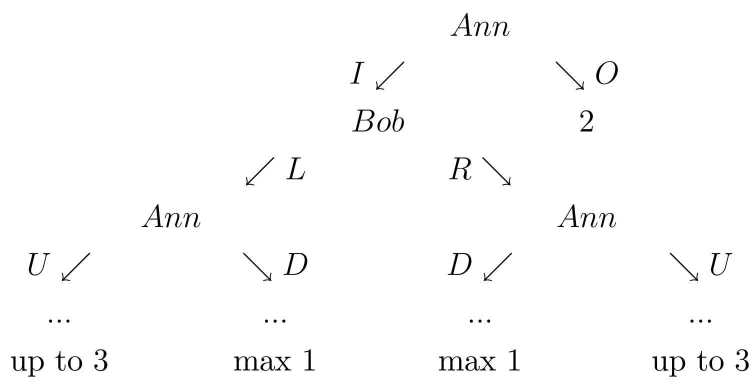
There is no incentive to cooperate in the first stage since defecting will not trigger the punishment.

$S^4_{i,R\Delta^e} = \emptyset$. The agreement is not credible.

[2] With P one can expect at most 2 but under conjectures that make D more profitable. Moreover under every conjecture one can expect at least 1. Thus there is no incentive to give up at least 1 by playing C in the first stage if then one believes that P is the most profitable action in any case.

0.4 Why are players not required to believe in strategies that comply with the agreement at no more reachable histories?

Consider the following partial representation of a game, where Ann obtains a payoff of 2 after Out, at most 1 in the subgames that follow Down, and up to 3 in the subgames that follow Up.



Ann and Bob reach the following agreement: Ann claims to play D at (I, R). At (I, L, D), Bob cannot believe that Ann is rational and would have played D also at (I, R), because any strategy of Ann which prescribes I, D at (I, L), and D at (I, R) is dominated by O. There are instead rational strategies of Ann that allow (I, L, D), those that prescribe U at (I, R) and aim to a payoff of at least 2 afterwards. Thus, if Bob were required by his restrictions to believe at (I, L, D) in strategies of Ann that comply with the agreement, strong-Δ-rationalizability would yield an empty set at the second step. But then, the agreement would be deemed not credible only because *if* a history were reached (and it is not if Ann is rational and plans to comply with the agreement), this would signal the past intention of a player to violate the agreement at a no more reachable history.

Yet "D at (I, R)" should suggest to Bob that Ann will try to get more than 2 after (I, L, U). Will Bob take this into consideration? The answer is yes as long as Ann behaves as he expects. At I, Bob is required to believe that Ann plans to comply with the agreement. If Bob believes also that Ann is rational, he expects Ann to play U after L and aim to a payoff of at least 2 afterwards. This expectation is transfered by CPS-3 to (I, L, U). Yet, if the players reach an information set that Bob does not expect at (I, L, U), Bob is free to believe in Ann's rational strategies that prescribe U at (I, R). Therefore, Bob uses his belief in the agreement only as long as Ann behaves as he expects. However, if at some step n of the rationalizability procedure Ann excludes to play U at (I, R), Bob will believe in strategies of Ann that prescribe D at (I, R) at each information set that the strategies of Ann which survive n steps allow.

0.5 Equilibrium paths that can be upset by a convincing deviation

Take the T-fold repetition G^T of an arbitrary two-players (i and j) static game G with action sets A_i and A_j and payoff function $v_k : A_i \times A_j \to \mathbb{R}$, $k = i, j$. Let b^k and c^k be the first- and second-ranked stage-outcomes of G for player $k = i, j$.

Let $(\overline{a}^1, .., \overline{a}^T)$ be a path of pure Nash equilibria of G. Suppose that there exist $\tau \in \{1, ..., T-1\}$ and $\widehat{a}_i \in A_i$ such that

$$v_i((\widehat{a}_i, \overline{a}_j^\tau)) + v_i(c^i) + (T-\tau-1)v_i(b^i) < \sum_{t=\tau}^{T} v_i(\overline{a}^t) < v_i((\widehat{a}_i, \overline{a}_j^\tau)) + (T-\tau)v_i(b^i) \quad \text{(I)}$$

and

$$(T-\tau)v_j(b^i) > \max_{a_j \in A_j \setminus \{b_j^i\}} v_j(b_i^i, a_j) + (T-\tau-1)v_j(b^j), \quad \text{(J)}$$

where $j \neq i$.

Such a path is called by Osborne [4] "equilibrium path that can be upset by a convincing deviation". In this framework, such paths can be characterized as non credible agreements.

Proposition 4 *Let $\overline{z} = (\overline{a}^1, ..., \overline{a}^T)$ be a path that can be upset by a convincing deviation. The corresponding path agreement is not credible.*

Proof: For $k = i, j$ and every $(a^1, ..., a^T) \in Z$, define $u_k : Z \to \mathbb{R}$ as $u_k((a^1, ..., a^T)) := \sum_{t=1}^{T} v_k(a^t)$. Let $\widehat{h} := (\overline{a}^1, .., (\widehat{a}_i, \overline{a}_j^\tau))$ and $z := (\overline{a}^1, .., (\widehat{a}_i, \overline{a}_j^\tau), b^i, ..., b^i)$.

For every $\mu_i \in \Delta_i^e$, $s_j \in \text{Supp}\mu_i(\cdot|h^0)$ and $s_i' \in S_i(\overline{z})$, $\zeta(s_i', s_j) = \overline{z}$; for every $s_i \in S_i(\widehat{h}) \backslash S_i(z)$, by (I) $u_i(\overline{z}) > u_i(\zeta(s_i, s_j))$, thus $s_i \notin S_{i,\Delta^e}^1$. Yet, there exist $s_j \in S_j$

and $\mu_i \in \Delta_i^e$ such that $\mu_i(s_j|h^0) = 1$ and for every $h = (\widetilde{a}^1, ..., \widetilde{a}^t) \not\succeq \widehat{h}$, $s_j(h) = \overline{a}_j^{t+1}$ and for every $h \succeq \widehat{h}$, $s_j(h) = b_j^i$. By (I) $\rho(\mu_i) \subseteq S_{i,\Delta^e}^1(\widehat{h}) \neq \emptyset$.

Then for every $\mu_j \in \Delta_j^e$ that strongly believes S_{i,Δ^e}^1, $s_i \in \mathrm{Supp}\mu_j(\cdot|\widehat{h})$ and $s_j' \in S_j(z)$, $\zeta(s_i, s_j') = z$; for every $s_j \in S_j(\widehat{h}) \backslash S_j(z)$, by (J) $u_j(z) > u_j(\zeta(s_i, s_j))$, thus $s_j \notin S_{j,\Delta^e}^2$. Yet, there exist $s_i \in S_i$ and $\mu_j \in \Delta_j^e$ that strongly believes S_{i,Δ^e}^1 such that $\mu_j(s_i|h^0) = 1$ and for every $h = (\widetilde{a}^1, ..., \widetilde{a}^t) \in H$, $s_j(h) = \overline{a}_j^{t+1}$. So $\rho(\mu_j) \cap S_j(\overline{z}) \neq \emptyset$ and then $S_{j,\Delta^e}^2(\widehat{h}) \neq \emptyset$.

Then for every $\mu_i \in \Delta_i^e$ that strongly believes S_{j,Δ^e}^2, $\mu_i(S_j(z)|p(\widehat{h})) = 1$, so by (I) $\rho(\mu_i)(\overline{z}) = \emptyset$. Hence $S_{i,\Delta^e}^3(\overline{z}) = \emptyset$.[3]

So, there does not exist $\mu_j \in \Delta_j^e$ such that $\mu_j(S_{i,\Delta^e}^3|h^0) = 1$, thus $S_{j,\Delta^e}^4 = \emptyset$. ■

Other than delivering a class of non credible path agreements, the proposition provides epistemic conditions under which the deviator can confidently upset the path. They are the ones employed in the proof up to the footnote and give rise to the same instances of forward induction reasoning as in the second example of Section 2. That path would indeed fall in the class of paths that can be upset by a convincing deviation by extending the definition to all paths in the natural way.

0.6 Proof of Lemma 1

Lemma 1 *Fix a reduction procedure* $((S_{i,q}^h)_{i\in I})_{q\geq 0}$, $i \in I$, $n \in \mathbb{N}$, $\widehat{h} \in H^h$ *and* $\mu_i^h \in \Delta^{H^h}(S_{-i}^h)$ *t.s.b.* $(S_{-i,q}^h)_{q=0}^{n-1}$ *such that* $\mu_i^h(S_{-i}^h(\widehat{h})|p(\widehat{h})) = 0$. *Fix* $s_i^h \in \rho(\mu_i^h)$, $\mu_i^{\widehat{h}}$ *t.s.b.* $(S_{-i,q}^h(\widehat{h})|\widehat{h})_{q=0}^{n-1}$ *and* $s_i^{\widehat{h}} \in \rho(\mu_i^{\widehat{h}})$.

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

There exists $\widetilde{\mu}_i^h =^{\widehat{h}} \mu_i^{\widehat{h}}$ *t.s.b.* $(S_{-i,q}^h)_{q=0}^{n-1}$ *such that for every* $\widetilde{h} \notin H^{\widehat{h}}$, $\widetilde{\mu}_i^h(\cdot|\widetilde{h}) = \mu_i^h(\cdot|\widetilde{h})$, *and* $\widetilde{s}_i^h \in \rho(\widetilde{\mu}_i^h)$ *(so that* $\rho(\mu_i^h)(\widehat{h}) \neq \emptyset$ *implies* $\rho(\widetilde{\mu}_i^h)(\widehat{h}) \neq \emptyset$*).*

Proof.

Fix a map $\varsigma : S_{-i}^{\widehat{h}} \mapsto S_{-i}^h$ that associates each $s_{-i}^{\widehat{h}} \in S_{-i}^{\widehat{h}}$ with a $s_{-i}^h \in S_{-i,m}^h(\widehat{h})$ such that $s_{-i}^h =^{\widehat{h}} s_{-i}^{\widehat{h}}$, where $m := \max\{q < n : \exists s_{-i}^h \in S_{-i,q}^h(\widehat{h}), s_{-i}^h =^{\widehat{h}} s_{-i}^{\widehat{h}}\}$. Construct an array of additive measures $(\widetilde{\mu}_i^h(\cdot|\widetilde{h}))_{\widetilde{h}\in H^h}$ on S_{-i}^h such that for every $\widetilde{h} \notin H^{\widehat{h}}$, $\widetilde{\mu}_i^h(\cdot|\widetilde{h}) = \mu_i^h(\cdot|\widetilde{h})$, and for every $\widetilde{h} \in H^{\widehat{h}}$ and $s_{-i}^{\widehat{h}} \in S_{-i}^{\widehat{h}}$, $\widetilde{\mu}_i^h(\varsigma(s_{-i}^{\widehat{h}})|\widetilde{h}) = \mu_i^{\widehat{h}}(s_{-i}^{\widehat{h}}|\widetilde{h})$.

[3] Under the beliefs used so far player i is shown to be willing to deviate from the path, confident of having convinced player j to follow her preferred subpath after the deviation.

Since ς is injective, $\widetilde{\mu}_i^h$ satisfies CPS-1. Using the definition of ς, it is immediate to verify that $\widetilde{\mu}_i^h =^{\widehat{h}} \mu_i^{\widehat{h}}$, satisfies CPS-2, and strongly believes $(S_{-i,q}^h)_{q=0}^{n-1}$. Finally, since $\widetilde{\mu}_i^h(S_{-i}(\widehat{h})|p(\widehat{h})) = 0$, $\widetilde{\mu}_i^h$ satisfies CPS-3.

Fix $\widetilde{h} \in H(\widetilde{s}_i^h)\backslash H^{\widehat{h}} = H(\widetilde{s}_i^h)\backslash H^{\widehat{h}}$. If $\widetilde{h} \prec \widehat{h}$, by $\mu_i^h(S_{-i}^h(\widehat{h})|p(\widehat{h})) = 0$ and CPS-3, $\mu_i^h(S_{-i}^h(\widehat{h})|\widetilde{h}) = 0$, and for every $s_{-i}^h \notin S_{-i}^h(\widehat{h})$, $\zeta(s_i^h, s_{-i}^h) = \zeta(\widetilde{s}_i^h, s_{-i}^h)$. If $\widetilde{h} \not\prec \widehat{h}$, for every $s_{-i}^h \in S_{-i}^h(\widetilde{h})$, $\widehat{h} \notin H(s_i^h, s_{-i}^h) \ni \widetilde{h}$, so $\zeta(s_i^h, s_{-i}^h) = \zeta(\widetilde{s}_i^h, s_{-i}^h)$. Hence $s_i^h \in \widehat{r}(\mu_i^h, \widetilde{h})$ implies $\widetilde{s}_i^h \in \widehat{r}(\mu_i^h, \widetilde{h}) = \widehat{r}(\widetilde{\mu}_i^h, \widetilde{h})$. If $\widetilde{s}_i^h, s_i^h \notin S_i^h(\widehat{h})$, the proof is over. Else, fix $\widetilde{h} \in H^{\widehat{h}} \cap H(\widetilde{s}_i^h) = H(s_i^{\widehat{h}})$. For every $s_{-i}^{\widehat{h}} \in S_{-i}^{\widehat{h}}$, $\widetilde{\mu}_i^h(\varsigma(s_{-i}^{\widehat{h}})|\widetilde{h}) = \mu_i^{\widehat{h}}(s_{-i}^{\widehat{h}}|\widetilde{h})$. 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, $\widetilde{s}_i^h|\widehat{h} = s_i^{\widehat{h}} \in \widehat{r}(\mu_i^{\widehat{h}}, \widetilde{h})$ implies $\widetilde{s}_i^h \in \widehat{r}(\widetilde{\mu}_i^h, \widetilde{h})$.
■

0.7 Proof of Lemma 3

Lemma 3 *Fix a red. 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* $H^S := H(\overline{S}^h)$ *and* $D^S := D_l(\overline{S}^h)$*. For every* $i \neq l$*, 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* $s_i^h \in \overline{S}_i^h$*,* $\overline{\mu}_i^h(s_i^h) =^{H^S} \overline{\overline{\mu}}_i^h(s_i^h)$ *strongly bel.* $(\widetilde{S}_{-i,q}^h)_{q=0}^{m-1}$ *and* $\rho(\overline{\mu}_i^h(s_i^h)) \ni \overline{s}_i^h(s_i^h) =^{H^S} s_i^h$*;*
- A2 *for every* $s_i^h \in \overline{S}_i^h$ *and* $\mu_i^h =^{H^S} \overline{\overline{\mu}}_i^h(s_i^h)$ *t.s.b.* $(\widetilde{S}_{-i,q}^h)_{q=0}^{m-1}$*,* $\rho(\mu_i^h) \subseteq \widetilde{S}_{i,m}^h$*.*

Fix μ_l^h *t.s.b.* $(\widetilde{S}_{-l,q}^h)_{q=0}^m$ *and* $\times_{i\neq l}(\overline{s}_i^h(s_i^h))_{s_i^h\in\overline{S}_i^h}$*. Fix* $\widetilde{D} \subseteq D^S$ *and for every* $\widehat{h} \in \widetilde{D}$*, fix* $\widetilde{\mu}_l^{\widehat{h}}$ *t.s.b.* $(\widetilde{S}_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$*. Let* $H^* := H^h \backslash \cup_{\widehat{h}\in\widetilde{D}} H^{\widehat{h}}$*.*

There exists $\widetilde{\mu}_l^h =^{H^*} \mu_l^h$ *t.s.b.* $(\widetilde{S}_{-l,q}^h)_{q=0}^m$ *such that for every* $\widehat{h} \in \widetilde{D}$*,* $\widetilde{\mu}_l^h =^{\widehat{h}} \widetilde{\mu}_l^{\widehat{h}}$*.*

Proof.

For every $\widehat{h} \in \widetilde{D}$ fix a map $\varsigma^{\widehat{h}} : S_{-l}^{\widehat{h}} \mapsto S_{-l}^h$ that associates each $s_{-l}^{\widehat{h}} \in S_{-l}^{\widehat{h}}$ with a $s_{-l}^h \in S_{-l,n}^h(\widehat{h})$ such that $s_{-l}^h =^{\widehat{h}} s_{-l}^{\widehat{h}}$, where $n := \max\{q \leq m : \exists s_{-l}^h \in \widetilde{S}_{-l,q}^h(\widehat{h}), s_{-l}^h =^{\widehat{h}}$

$s^{\widehat{h}}_{-l}\}$. Define an array of additive measures $(\widetilde{\mu}^h_l(\cdot|\widetilde{h}))_{\widetilde{h}\in H^h}$ on S^h_{-l} as follows:

$$\begin{array}{rcl}\widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h}) & = & \mu^h_l(S^{h,s^h_{-l}}_{-l,H^*}|\widetilde{h})\cdot\prod_{\widehat{h}\in\widetilde{D}}\widetilde{\mu}^{\widehat{h}}_l((s^h_{-l}|\widehat{h})|\widetilde{h}),\ \forall\widetilde{h}\in H^S,\forall s^h_{-l}\in S^h_{-l}(\widetilde{h});\\ \widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h}) & = & \widetilde{\mu}^h_l(s^h_{-l}|p(\widehat{h}))/\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h})),\ \forall\widehat{h}\in D^S,\forall\widetilde{h}\in H^{\widehat{h}},\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))\neq 0,\forall s^h_{-l}\in S^h_{-l}(\widetilde{h});\\ \widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h}) & = & \widetilde{\mu}^{\widehat{h}}_l((\varsigma^{\widehat{h}})^{-1}(s^h_{-l})|\widetilde{h}),\ \forall\widehat{h}\in\widetilde{D},\forall\widetilde{h}\in H^{\widehat{h}},\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))=0,\forall s^h_{-l}\in\varsigma^{\widehat{h}}(S^{\widehat{h}}_{-l});\\ \widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h}) & = & 0,\ \forall\widehat{h}\in\widetilde{D},\forall\widetilde{h}\in H^{\widehat{h}},\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))=0,\forall s^h_{-l}\notin\varsigma^{\widehat{h}}(S^{\widehat{h}}_{-l})\wedge\forall\widetilde{h}\in H^h,\forall s^h_{-l}\notin S^h_{-l}(\widetilde{h});\\ \widetilde{\mu}^h_l(\cdot|\widetilde{h}) & = & \mu^h_l(\cdot|\widetilde{h})\text{ else.}\end{array}$$

Fix $\overline{h}\in H^S$ and $s^h_{-l}\in S^h_{-l}(\overline{h})$. For every $\widetilde{s}^h_{-l}\in S^{h,s^h_{-l}}_{-l,H^*}$, $S^{h,\widetilde{s}^h_{-l}}_{-l,H^*}=S^{h,s^h_{-l}}_{-l,H^*}=:\widetilde{S}^h_{-l}$ (▲). So:

$$\widetilde{\mu}^h_l(\widetilde{S}^h_{-l}|\overline{h})=\sum_{\widetilde{s}^h_{-l}\in\widetilde{S}^h_{-l}}\widetilde{\mu}^h_l(\widetilde{s}^h_{-l}|\overline{h})=\sum_{\widetilde{s}^h_{-l}\in\widetilde{S}^h_{-l}}\mu^h_l(\widetilde{S}^h_{-l}|\overline{h})\cdot\prod_{\widehat{h}\in\widetilde{D}}\widetilde{\mu}^{\widehat{h}}_l((\widetilde{s}^h_{-l}|\widehat{h})|\overline{h})=\mu^h_l(\widetilde{S}^h_{-l}|\overline{h}) \quad (\triangle)$$

where the last equality holds because $((\widetilde{s}^h_{-l}|\widehat{h})_{\widehat{h}\in\widetilde{D}})_{\widetilde{s}^h_{-l}\in\widetilde{S}^h_{-l}}=\times_{\widehat{h}\in\widetilde{D}}S^{\widehat{h}}_{-l}$. Fix $\widetilde{z}\succeq\overline{h}$ such that $p(\widetilde{z})\in H^*$. For every $\widetilde{h}\prec\widetilde{z}$, $\widetilde{h}\in H^*$. Then $S^h_{-l}(\widetilde{z})=\cup_{\widetilde{s}^h_{-l}\in S^h_{-l}(\widetilde{z})}S^{h,\widetilde{s}^h_{-l}}_{-l,H^*}$, a union of pairwise identical or disjoint sets by ▲. Then by $\triangle$ $\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{z})|\overline{h})=\mu^h_l(S^h_{-l}(\widetilde{z})|\overline{h})$ (▼). Hence $\widetilde{\mu}^h_l(S^h_{-l}(\overline{h})|\overline{h})=\mu^h_l(S^h_{-l}(\overline{h})|\overline{h})=1$, thus with $\widetilde{\mu}^h_l(S^h_{-l}\backslash S^h_{-l}(\overline{h})|\overline{h})=0$, $\widetilde{\mu}^h_l(\cdot|\overline{h})$ satisfies CPS-1,2. Using ▼ in the last equality of the first equation and the first equation in the second and third equalities of the second equation:

$$\begin{array}{rcl}\widetilde{\mu}^h_l(s^h_{-l}|\overline{h})/\widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h}) & = & \mu^h_l(\widetilde{S}^h_{-l}|\overline{h})/\mu^h_l(\widetilde{S}^h_{-l}|\widetilde{h})=\mu^h_l(S^h_{-l}(\widetilde{h})|\overline{h})=\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|\overline{h}),\ \forall\widetilde{h}\in H^{\overline{h}}\cap H(s^h_{-l})\cap H^S;\\ \widetilde{\mu}^h_l(s^h_{-l}|\overline{h})/\widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h}) & = & \widetilde{\mu}^h_l(s^h_{-l}|\overline{h})\cdot\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))/\widetilde{\mu}^h_l(s^h_{-l}|p(\widehat{h}))=\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))\cdot\widetilde{\mu}^h_l(S^h_{-l}(p(\widehat{h}))|\overline{h})=\\ & = & \widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|\overline{h}),\qquad\forall\widehat{h}\in D^S\cap H^{\overline{h}},\forall\widetilde{h}\in H(s^h_{-l})\cap H^{\widehat{h}},\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))\neq 0;\end{array}$$

while by $\triangle$, since μ^h_l strongly believes $\times_{i\neq l}(\overline{s}^h_i(s^h_i))_{s^h_i\in\overline{S}^h_i}$, $\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|\overline{h})=0$ for all $\widetilde{h}\notin\cup_{\widehat{h}\in D^S}H^{\widehat{h}}\cup H^S$. So $\widetilde{\mu}^h_l$ satisfies CPS-3 when $C=S^h_{-i}(\overline{h})$.

Fix $\widehat{h}\in D^S\backslash\widetilde{D}$ and $\overline{h}\in H^{\widehat{h}}$ such that $\widetilde{\mu}^h_l(S^h_{-l}(\overline{h})|p(\widehat{h}))\neq 0$. Since $\widetilde{\mu}^h_l(\cdot|p(\widehat{h}))$ satisfies CPS-1,2, by definition $\widetilde{\mu}^h_l(\cdot|\overline{h})$ satisfies CPS-1,2 too. Moreover, for every $\widetilde{h}\succ\overline{h}$ such that $\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|\overline{h})\neq 0$, $\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))\neq 0$ too. So for every $s^h_{-l}\in S^h_{-l}(\widetilde{h})$, using that $\widetilde{\mu}^h_l$ satisfies CPS-3 when $C=S^h_{-i}(p(\widehat{h}))$,

$$\widetilde{\mu}^h_l(s^h_{-l}|\overline{h})=\widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h})\cdot\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|p(\widehat{h}))/\widetilde{\mu}^h_l(S^h_{-l}(\overline{h})|p(\widehat{h}))=\widetilde{\mu}^h_l(s^h_{-l}|\widetilde{h})\cdot\widetilde{\mu}^h_l(S^h_{-l}(\widetilde{h})|\overline{h}),$$

i.e., $\widetilde{\mu}_l^h$ satisfies CPS-3 when $C = S^h_{-i}(\overline{h})$. For every $\widetilde{z} \succeq \overline{h}$, using ▼ in the second equality,

$$\widetilde{\mu}_l^h(S^h_{-l}(\widetilde{z})|\overline{h}) = \widetilde{\mu}_l^h(S^h_{-l}(\widetilde{z})|p(\widehat{h}))/\widetilde{\mu}_l^h(S^h_{-l}(\overline{h})|p(\widehat{h})) = \mu_l^h(S^h_{-l}(\widetilde{z})|p(\widehat{h}))/\mu_l^h(S^h_{-l}(\overline{h})|p(\widehat{h})) = \mu_l^h(S^h_{-l}(\widetilde{z})|\overline{h}). \tag{∇}$$

Fix either $\overline{h} \notin \cup_{\widehat{h}\in D^S} H^{\widehat{h}} \cup H^S$ or $\widehat{h} \in D^S \backslash \widetilde{D}$ and $\overline{h} \in H^{\widehat{h}}$ such that $\widetilde{\mu}_l^h(S^h_{-l}(\overline{h})|p(\widehat{h})) = 0$. It holds $\widetilde{\mu}_l^h(\cdot|\overline{h}) = \mu_l^h(\cdot|\overline{h})$. So, $\widetilde{\mu}_l^h(\cdot|\overline{h})$ satisfies CPS-1,2. Moreover, since for every $\widetilde{h} \succ \overline{h}$, $\widetilde{\mu}_l^h(\cdot|\widetilde{h}) = \mu_l^h(\cdot|\widetilde{h})$ too, $\widetilde{\mu}_l^h$ satisfies CPS-3 when $C = S^h_{-i}(\overline{h})$. Finally, together with ▼ and ∇, $\widetilde{\mu}_l^h =^{H^*} \mu_l^h$ (provided $\widetilde{\mu}_l^h$ is a CPS).

Fix $\widehat{h} \in \widetilde{D}$ and $\overline{h} \in H^{\widehat{h}}$. Fix $s^{\widehat{h}}_{-l} \in S^{\widehat{h}}_{-l}(\overline{h})$. For every $\widetilde{s}^{\widehat{h}}_{-l} \in S^{\widehat{h}}_{-l}$, let $S^h(\widetilde{s}^{\widehat{h}}_{-l}) := \{\widetilde{s}^h_{-l} \in S^h_{-l}(\widehat{h}) : \widetilde{s}^h_{-l} =^{\widehat{h}} \widetilde{s}^{\widehat{h}}_{-l}\}$. Note that if $\widehat{s}^{\widehat{h}}_{-l} \neq \overline{s}^{\widehat{h}}_{-l}$, $S^h(\widehat{s}^{\widehat{h}}_{-l}) \cap S^h(\overline{s}^{\widehat{h}}_{-l}) = \emptyset$. So $(S^h(\widetilde{s}^{\widehat{h}}_{-l}))_{\widetilde{s}^{\widehat{h}}_{-l}\in S^{\widehat{h}}_{-l}(\overline{h})}$ is a partition of $S^h_{-l}(\overline{h})$. Then:

$$\begin{aligned}
\widetilde{\mu}_l^h(S^h(s^{\widehat{h}}_{-l})|\overline{h}) &= \widetilde{\mu}_l^h(S^h(s^{\widehat{h}}_{-l})|p(\widehat{h}))/\textstyle\sum_{\widetilde{s}^{\widehat{h}}_{-l}\in S^{\widehat{h}}_{-l}(\overline{h})} \widetilde{\mu}_l^h(S^h(\widetilde{s}^{\widehat{h}}_{-l})|p(\widehat{h})) = \\
&= \frac{\widetilde{\mu}_l^{\widehat{h}}(s^{\widehat{h}}_{-l}|\widehat{h}) \cdot \sum_{s^h_{-l}\in S^h(s^{\widehat{h}}_{-l})} \mu_l^h(S^{h,s^h_{-l}}_{-l,H^*}|p(\widehat{h})) \cdot \prod_{\widetilde{h}\in\widetilde{D}\backslash\{\widehat{h}\}} \widetilde{\mu}_l^{\widetilde{h}}((s^h_{-l}|\widetilde{h})|\widetilde{h})}{\sum_{\widetilde{s}^{\widehat{h}}_{-l}\in S^{\widehat{h}}_{-l}(\overline{h})} \widetilde{\mu}_l^{\widehat{h}}(\widetilde{s}^{\widehat{h}}_{-l}|\widehat{h}) \left(\sum_{s^h_{-l}\in S^h(\widetilde{s}^{\widehat{h}}_{-l})} \mu_l^h(S^{h,s^h_{-l}}_{-l,H^*}|p(\widehat{h})) \cdot \prod_{\widetilde{h}\in\widetilde{D}\backslash\{\widehat{h}\}} \widetilde{\mu}_l^{\widetilde{h}}((s^h_{-l}|\widetilde{h})|\widetilde{h})\right)} = \\
&= \widetilde{\mu}_l^{\widehat{h}}(s^{\widehat{h}}_{-l}|\widehat{h})/\widetilde{\mu}_l^{\widehat{h}}(S^{\widehat{h}}_{-l}(\overline{h})|\widehat{h}) = \widetilde{\mu}_l^{\widehat{h}}(s^{\widehat{h}}_{-l}|\overline{h}) \qquad \text{if } \widetilde{\mu}_l^h(S^h_{-l}(\overline{h})|p(\widehat{h})) \neq 0; \\
\widetilde{\mu}_l^h(S^h(s^{\widehat{h}}_{-l})|\overline{h}) &= \widetilde{\mu}_l^h(\varsigma^{\widehat{h}}(s^{\widehat{h}}_{-l})|\overline{h}) = \widetilde{\mu}_l^{\widehat{h}}(s^{\widehat{h}}_{-l}|\overline{h}) \qquad \text{if } \widetilde{\mu}_l^h(S^h_{-l}(\overline{h})|p(\widehat{h})) = 0;
\end{aligned}$$

because the quantity in brackets does not depend on the particular $\widetilde{s}^{\widehat{h}}_{-l}$. So, like $\widetilde{\mu}_l^{\widehat{h}}$, $\widetilde{\mu}_l^h(\cdot|\overline{h})$ satisfies CPS-1,2 and CPS-3 when $C = S^h_{-i}(\overline{h})$. Hence $\widetilde{\mu}_l^h =^{H^*} \mu_l^h$ is a CPS with $\widetilde{\mu}_l^h =^{\widehat{h}} \widetilde{\mu}_l^{\widehat{h}}$ for all $\widehat{h} \in \widetilde{D}$.

Finally, I show that $\widetilde{\mu}_l^h$ strongly believes $(\widetilde{S}^h_{-l,q})_{q=0}^m$.

Fix $\widetilde{h} \in H^S$, $i \neq l$ and $\widetilde{s}_i^h \in \text{SuppMarg}_{S_i^h}\widetilde{\mu}_l^h(\cdot|\widetilde{h})$. Since $\text{Marg}_{S_i^h}\mu_l^h(S^{h,\widetilde{s}_i^h}_{i,H^*}|\widetilde{h}) \neq 0$ and $\text{Marg}_{S_i^h}\mu_l^h((\overline{s}_i^h(s_i^h))_{s_i^h\in\overline{S}_i^h}|\widetilde{h}) = 1$, there exists $s_i^h \in \overline{S}_i^h$ such that $\overline{s}_i^h(s_i^h) =^{H^*} \widetilde{s}_i^h$. For every $\widehat{h} \in \widetilde{D}$, $\widetilde{s}_i^h|\widehat{h} \in \text{SuppMarg}_{S_i^{\widehat{h}}}\widetilde{\mu}_l^{\widehat{h}}(\cdot|\widehat{h}) \subseteq \widetilde{S}^h_{i,m}(\widehat{h})|\widehat{h} \neq \emptyset$ (by $\widehat{h} \in H((\overline{s}_i^h(s_i^h))_{s_i^h\in\overline{S}_i^h})$), hence there exists $\mu_i^{\widehat{h}}$ t.s.b. $(\widetilde{S}^h_{-i,q}(\widehat{h})|\widehat{h})_{q=0}^{m-1}$ such that $\widetilde{s}_i^h|\widehat{h} \in \rho(\mu_i^{\widehat{h}})$. Since $\overline{\mu}_i^h(s_i^h) =^{H^S} \overline{\overline{\mu}}_i^h(s_i^h)$ which strongly believes $\overline{S}^h_{-i}$, $\overline{\mu}_i^h(s_i^h)(S^h_{-i}(\widehat{h})|p(\widehat{h})) = 0$. Thus, by repeatedly applying Lemma 1, I can find $\mu_i^h =^{H^*} \overline{\mu}_i^h(s_i^h) =^{H^S} \overline{\overline{\mu}}_i^h(s_i^h)$ t.s.b. $(\widetilde{S}^h_{-i,q})_{q=0}^{m-1}$ such that $\widetilde{s}_i^h \in \rho(\mu_i^h)$. By A2 $\rho(\mu_i^h) \subseteq \widetilde{S}^h_{i,m}$. So $\widetilde{\mu}_l^h(\widetilde{S}^h_{-l,m}|\widetilde{h}) = 1$ for all $\widetilde{h} \in H^S$.

Then, for every $\widehat{h} \in D^S$ and $\widetilde{h} \in H^{\widehat{h}}$ such that $\widetilde{\mu}_l^h(S_{-l}^h(\widetilde{h})|p(\widehat{h})) \neq 0$, by construction $\widetilde{\mu}_l^h(\widetilde{S}_{-l,m}^h|\widetilde{h}) = 1$ too. For every $\widehat{h} \in D^S$ and $\widetilde{h} \in H^{\widehat{h}}$ such that $\widetilde{\mu}_l^h(S_{-l}^h(\widetilde{h})|p(\widehat{h})) = 0$, $\widetilde{\mu}_l^h(\varsigma^{\widehat{h}}(\mathrm{Supp}\widetilde{\mu}_l^{\widehat{h}}(\cdot|\widetilde{h}))|\widetilde{h}) = 1$, $\widetilde{\mu}_l^{\widehat{h}}$ strongly believes $(\widetilde{S}_{-l,q}^h(\widehat{h})|\widehat{h})_{q=0}^m$ and if $\mathrm{Supp}\widetilde{\mu}_l^{\widehat{h}}(\cdot|\widetilde{h}) \subseteq \widetilde{S}_{-l,q}^h(\widehat{h})|\widehat{h}$, $\varsigma^{\widehat{h}}(\mathrm{Supp}\widetilde{\mu}_l^{\widehat{h}}(\cdot|\widetilde{h})) \subseteq \widetilde{S}_{-l,q}^h$. Else, $\widetilde{\mu}_l^h(\cdot|\widetilde{h}) = \mu_l^h(\cdot|\widetilde{h})$ and μ_l^h strongly believes $(\widetilde{S}_{-l,q}^h)_{q=0}^m$. So $\widetilde{\mu}_l^h$ strongly believes $(\widetilde{S}_{-l,q}^h)_{q=0}^m$. ■

0.8 Proof of Lemma 5

Lemma 5. *Fix $v \leq n$ and $i \in I$ such that $\widetilde{\sigma}_i^h \in \Delta(r(\widetilde{\sigma}_{-i}^h))$ and for every $\mu_i^h =^* \widetilde{\sigma}_{-i}^h$ t.s.b. $(S_{-i,q}^h)_{q=0}^{v-1}$, $\rho(\mu_i^h) \subseteq S_{i,v}^h$. For every $\widetilde{h} \in D_{-i}$, fix $\widehat{\sigma}_i^{\widetilde{h}} \in \Delta(S_{i,v}^h(\widetilde{h})|\widetilde{h})$. There exists $\widehat{\sigma}_i^h \in \Delta(S_{i,v}^h)$ such that $\widehat{\sigma}_i^h =^* \widetilde{\sigma}_i^h$ and for every $\widetilde{h} \in D_{-i}$, $\widehat{\sigma}_i^h|\widetilde{h} = \widehat{\sigma}_i^{\widetilde{h}}$.*

Proof.

Let $H^* := H^h \backslash (\cup_{\widehat{h} \in D_{-i}} H^{\widehat{h}})$. Define an additive measure $\widehat{\sigma}_i^h$ on S_i^h as:

$$\widehat{\sigma}_i^h(s_i^h) = \widetilde{\sigma}_i^h(S_{i,H^*}^{h,s_i^h}) \cdot \textstyle\prod_{\widehat{h} \in D_{-i}} \widehat{\sigma}_i^{\widehat{h}}(s_i^h|\widehat{h}), \ \ \forall s_i^h \in S_i^h.$$

Fix $s_i^h \in S_i^h$. For every $\widetilde{s}_i^h \in S_{i,H^*}^{h,s_i^h}$, $S_{i,H^*}^{h,\widetilde{s}_i^h} = S_{i,H^*}^{h,s_i^h}$. Then:

$$\widehat{\sigma}_i^h(S_{i,H^*}^{h,s_i^h}) = \textstyle\sum_{\widetilde{s}_i^h \in S_{i,H^*}^{h,s_i^h}} \widetilde{\sigma}_i^h(S_{i,H^*}^{h,\widetilde{s}_i^h}) \cdot \prod_{\widehat{h} \in D_{-i}} \widehat{\sigma}_i^{\widehat{h}}(\widetilde{s}_i^h|\widehat{h}) = \widetilde{\sigma}_i^h(S_{i,H^*}^{h,s_i^h}),$$

where the last equality holds because $((\widetilde{s}_i^h|\widehat{h})_{\widehat{h} \in D_{-i}})_{\widetilde{s}_{-l}^h \in S_{i,H^*}^{h,s_i^h}} = \times_{\widehat{h} \in D_{-i}} S_i^{\widehat{h}}$. Hence $\widehat{\sigma}_i^h \in \Delta(S_i^h)$. Fix $\widetilde{z} \in H^h \cup Z^h$ such that $p(\widetilde{z}) \in H^\sigma$. For every $\widetilde{h} \prec \widetilde{z}$, $\widetilde{h} \in H^*$. Then $S_i^h(\widetilde{z}) = \cup_{\widetilde{s}_i^h \in S_i^h(\widetilde{z})} S_{i,H^*}^{h,\widetilde{s}_i^h}$, a union of pairwise identical or disjoint sets. So $\widehat{\sigma}_i^h(S_i^h(\widetilde{z})) = \widetilde{\sigma}_i^h(S_i^h(\widetilde{z}))$. Thus $\widehat{\sigma}_i^h =^* \widetilde{\sigma}_i^h$.

Fix $\widehat{h} \in D_{-i}$ and $s_i^{\widehat{h}} \in S_i^{\widehat{h}}$. For every $\widetilde{s}_i^{\widehat{h}} \in S_i^{\widehat{h}}$, let $S^h(\widetilde{s}_i^{\widehat{h}}) := \{\widetilde{s}_i^h \in S_i^h(\widehat{h}) : \widetilde{s}_i^h|\widehat{h} = \widetilde{s}_i^{\widehat{h}}\}$. Note that if $\widetilde{s}_i^{\widehat{h}} \neq \overline{s}_i^{\widehat{h}}$, $S^h(\widetilde{s}_i^{\widehat{h}}) \cap S^h(\overline{s}_i^{\widehat{h}}) = \emptyset$. So $(S^h(\widetilde{s}_i^{\widehat{h}}))_{\widetilde{s}_i^{\widehat{h}} \in S_i^{\widehat{h}}}$ is a partition of $S_i^h(\widehat{h})$. Then:

$$(\widehat{\sigma}_i^h|\widehat{h})(s_i^{\widehat{h}}) = \frac{\widehat{\sigma}_i^h(S^h(s_i^{\widehat{h}}))}{\sum_{\widetilde{s}_i^{\widehat{h}} \in S_i^{\widehat{h}}} \widehat{\sigma}_i^h(S^h(\widetilde{s}_i^{\widehat{h}}))} = \frac{\widehat{\sigma}_i^{\widehat{h}}(s_i^{\widehat{h}}) \cdot \sum_{\widetilde{s}_i^h \in S^h(s_i^{\widehat{h}})} \widetilde{\sigma}_i^h(S_{i,H^*}^{h,\widetilde{s}_i^h}) \cdot \prod_{\widetilde{h} \in D_{-i} \backslash \{\widehat{h}\}} \widehat{\sigma}_i^{\widetilde{h}}(\widetilde{s}_i^h|\widetilde{h})}{\sum_{\widetilde{s}_i^{\widehat{h}} \in S_i^{\widehat{h}}} \widehat{\sigma}_i^{\widehat{h}}(\widetilde{s}_i^{\widehat{h}}) \left(\sum_{\widetilde{s}_i^h \in S^h(\widetilde{s}_i^{\widehat{h}})} \widetilde{\sigma}_i^h(S_{i,H^*}^{h,\widetilde{s}_i^h}) \cdot \prod_{\widetilde{h} \in D_{-i} \backslash \{\widehat{h}\}} \widehat{\sigma}_i^{\widetilde{h}}(\widetilde{s}_i^h|\widetilde{h})\right)} = \widehat{\sigma}_i^{\widehat{h}}(s_i^{\widehat{h}}),$$

since the quantity in brackets does not depend on the particular $\widetilde{s}_i^{\widehat{h}}$. So $\widehat{\sigma}_i^h|\widehat{h} = \widehat{\sigma}_i^{\widehat{h}}$.

Finally I show that $\widehat{\sigma}_i^h \in \Delta(S_{i,v}^h)$. Fix $\widehat{s}_i^h \in \mathrm{Supp}\widehat{\sigma}_i^h$. Since $\widetilde{\sigma}_i^h(S_{i,H^*}^{h,\widehat{s}_i^h}) \neq 0$, there exists $\overline{s}_i^h \in \mathrm{Supp}\widetilde{\sigma}_i^h$ such that $\overline{s}_i^h =^{H^*} \widehat{s}_i^h$. Fix $\mu_i^h =^h \widetilde{\sigma}_{-i}^h$ t.s.b. $(S_{-i,q}^h)_{q=0}^{v-1}$. Since $\overline{s}_i^h \in r(\widetilde{\sigma}_{-i}^h)$, by ♣ there exists $s_i^h \in \rho(\mu_i^h)$ such that for every $\widetilde{h} \in H^*$,[4] $s_i^h(\widetilde{h}) = \overline{s}_i^h(\widetilde{h}) = \widehat{s}_i^h(\widetilde{h})$. For every $\widehat{h} \in D_{-i}$, $\widehat{s}_i^h|\widehat{h} \in \mathrm{Supp}\widehat{\sigma}_i^{\widehat{h}} \subseteq S_{i,v}^h(\widehat{h})|\widehat{h}$, so there exists $\mu_i^{\widehat{h}}$ t.s.b. $(S_{-i,q}^h(\widehat{h})|\widehat{h})_{q=0}^{v-1}$ such that $\widehat{s}_i^h|\widehat{h} \in \rho(\mu_i^{\widehat{h}})$. Thus, by repeatedly applying Lemma 1, I can find $\widetilde{\mu}_i^h =^h \widetilde{\sigma}_{-i}^h$ t.s.b. $(S_{-i,q}^h)_{q=0}^{v-1}$ such that $\widehat{s}_i^h \in \rho(\widetilde{\mu}_i^h)$, and by hypothesis $\rho(\widetilde{\mu}_i^h) \subseteq S_{i,v}^h$. ■

References

[1] Battigalli, P. and A. Friedenberg, "Forward induction reasoning revisited", *Theoretical Economics*, **7**, 2012, 57-98.

[2] 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.

[3] Greenberg, J., Gupta, S., Luo, X., "Mutually acceptable courses of action", *Economic Theory*, **40**, 2009, 91-112.

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

[4] The equality at each $\widetilde{h} \in H^* \backslash H^\sigma$ is without loss of generality because $\widetilde{h}$ is not allowed by s_i^h.