Rationalizability and Epistemic Priority Orderings*

Emiliano Catonini
ICEF
Higher School of Economics
Moscow, Russia
emiliano.catonini@gmail.com

At the beginning of a dynamic game, players may have exogenous theories about how the opponents are going to play. Suppose that these theories are commonly known. Then, players will refine their first-order beliefs, and challenge their own theories, through strategic reasoning. I develop and characterize epistemically a new solution concept, Selective Rationalizability, which accomplishes this task under the following assumption: when the observed behavior is not compatible with the beliefs in players' rationality and theories of all orders, players keep the orders of belief in rationality that are per se compatible with the observed behavior, and drop the incompatible beliefs in the theories. Thus, Selective Rationalizability captures Common Strong Belief in Rationality (Battigalli and Siniscalchi, 2002) and refines Extensive-Form Rationalizability (Pearce, 1984; BS, 2002), whereas Strong-Δ-Rationalizability (Battigalli, 2003; Battigalli and Siniscalchi, 2003) captures the opposite epistemic priority choice. Selective Rationalizability can be extended to encompass richer epistemic priority orderings among different theories of opponents' behavior. This allows to establish a surprising connection with strategic stability (Kohlberg and Mertens, 1986).

Keywords: Forward induction, Strong Belief, Strong Rationalizability, Strong-Δ-Rationalizability, Strategic Stability.

1 Introduction

Consider the following dynamic game with perfect information.

$$
\begin{array}{cccccc}
 & Ann & & & & \\
N \swarrow & & \searrow B & & & \\
0,0 & & & Bob & & \\
 & & R \swarrow & & \searrow A & \\
 & & -2,0 & & & Ann \\
 & & & & P \swarrow & & \searrow I \\
 & & & & -1,-3 & & 1,1
\end{array}
$$

Ann can try to *B*ribe Bob, a public officer, or *N*ot. If she does, Bob can *A*ccept or *R*eport her, so that Ann loses two utils. If Bob accepts, Ann can *I*mplement her plan, achieving the Pareto dominating outcome, or repent (P) and speak with a prosecutor, harming both Bob and herself.

Suppose that Ann is rational[1] and, at the beginning of the game, believes with probability 1 that Bob would play R after B. I call this belief "(first-order belief) restriction". Then, she plays N. Suppose that Bob is rational and believes that Ann is rational and that the restriction holds. Then, he expects Ann to

*This is a short version of the paper for this volume. The full version of the paper can be found online on the author's institutional webpage.

[1] i.e. subjective expected utility maximizer.

J. Lang (Ed.): TARK 2017
EPTCS 251, 2017, pp. 102–117, doi:10.4204/EPTCS.251.8

play N. So, what would Bob believe after observing B? He cannot believe at the same time that Ann is rational and that the restriction holds: the two things are at odds given B. Which of the two beliefs will Bob keep? This is the epistemic priority issue. Suppose that he keeps the belief that the restriction holds. So, he drops the belief that Ann is rational. Then he can also expect Ann to play P after (B,A) and so play R. If Ann believes that Bob reasons in this way, she can keep her restriction and then play N.

These lines of strategic reasoning are captured by Strong-Δ-Rationalizability (Battigalli, [5]; Battigalli and Siniscalchi, [10]). In this process, the faith in the restrictions is so strong that Bob is ready to deem Ann irrational after B. This could be the case if, for instance, the belief that Bob would play R is suggested by a commonly believed social convention that always holds in context of the game (see Battigalli and Friedenberg [6]). Suppose instead that in the context of the game, public officers are not commonly believed to be incorruptible. However, Bob declares that he would play R in case of B. If Bob observes that Ann plays B anyway, he might think that Ann has not taken his words seriously, rather than thinking that Ann is irrational. Then, Bob would expect Ann to play I after A, hence he would play A instead of R. If Ann believes that Bob is rational and keeps believing that she is rational after B, she must believe that Bob will play A, differently than what the restriction suggests. Hence, under this reasoning scheme, such restriction to first-order beliefs cannot hold.

Note that opposite conclusions were reached without any uncertainty about payoffs: the two situations do not represent two different types of Bob, but only two different strengths of the belief that he would report Ann.

In Section 3, I construct a rationalizability procedure, Selective Rationalizability, that captures these instances of forward induction reasoning in dynamic games with perfect recall.[2] Selective Rationalizability refines a notion of Extensive-Form Rationalizability (Pearce, [23], Battigalli and Siniscalchi, [9]), which I will call "Rationalizability" for brevity. Thus, Selective Rationalizability represents a natural way for players to refine their beliefs through (*partial*) coordination and consequent forward induction considerations when lone strategic reasoning about rationality does not pin down a unique plan of actions. As above, Selective Rationalizability delivers an empty set when the "tentative" first-order belief restrictions of a player are at odds with strategic reasoning. In this case, Selective Rationalizability is agnostic as to whether players will fall back on some merely rationalizable strategy, or will still refine their beliefs with the restrictions, up to some feasible order.

Note that strong-Δ-Rationalizability, instead, does not refine Rationalizability: in the example, N is not a rationalizable outcome.[3] It is worth noting that Selective Rationalizability can also be seen as an instance of Strong-Δ-Rationalizability, where the restrictions are the conjunction of the original restrictions and the rationalizable first-order beliefs. However, keeping the two separate has both conceptual and technical advantages. The separation allows to investigate the epistemic priority issue between the two different sources of belief restrictions, and to compare Strong-Δ-Rationalizability and Selective Rationalizability for the *same* restrictions. In general, one could expect Selective Rationalizability to always yield a subset of the outcomes predicted by Strong-Δ-Rationalizability. Two counterexamples in the full version of the paper show that, (i) opposite to the example above, Selective Rationalizability can yield non-empty predictions when Strong-Δ-Rationalizability rejects the first-order belief restrictions; and (ii) Selective Rationalizability and Strong-Δ-Rationalizability can even yield non-empty disjoint

[2]For notational simplicity, here the focus is kept on complete information games. Just like Strong-Δ-Rationalizability, Selective Rationalizability can be easily extended to games with incomplete information.

[3]The game has no simultaneous moves and no relevant ties. Therefore, as shown by Battigalli [4] first and Heifetz and Perea [19] later, backward induction and Extensive-Form Rationalizability predict the same unique outcome.

predictions. However, as I show in [12], Selective Rationalizability and Strong-Δ-Rationalizability are outcome-equivalent when the belief restrictions correspond to a specific path of play.

In Section 4, I clarify with an epistemic characterization the strategic reasoning hypotheses that motivate Selective Rationalizability. To simplify the epistemic analysis, the game is assumed to have a finite set of non-terminal histories, hence finite horizon, although Selective Rationalizability can be applied to all games with a countable set of non-terminal histories, hence possibly infinite horizon. Selective Rationalizability captures the behavior of rational players who restrict their beliefs about opponents' behavior for some exogenous reason. Moreover, at the beginning of the game, players believe that opponents are rational and have their own restrictions; that opponents believe that everyone else is rational and has precisely the own restrictions; and so on. These beliefs are tentative because at some information set of a player, the observed behavior of one opponent may be incompatible, say, with the opponent being rational and, at the same time, having beliefs in her restricted set. In this case, the player will drop the belief that the opponent has such restrictions, rather than dropping the belief that the opponent is rational. More generally, players always keep all orders of belief in rationality that are per se compatible with the observed behavior, and drop all orders of belief in the restrictions that are at odds with them. I call this choice *epistemic priority to rationality*. Strong-Δ-Rationalizability predicts instead the behavior of players who assign epistemic priority to the beliefs in the restrictions, and drop the incompatible beliefs in rationality. Thus, Selective Rationalizability captures a version of Common Strong Belief in Rationality (Battigalli and Siniscalchi, [9]), whereas Strong-Δ-Rationalizability does not.

In Section 5, I extend the analysis to finer epistemic priority orderings. Each player can have multiple theories, say two, about opponents' behavior: a weaker theory and a stronger theory (in the sense of more restrictive). Players reason according to everyone's weaker theory like under Selective Rationalizability. On top of this, as long as compatible with strategic reasoning about the weaker theories, players reason according to the stronger theories. So, when a player displays behavior which is not compatible with strategic reasoning about both theories, the opponents keep believing that the player is reasoning according to the weaker theories, and drop the belief that the opponent is reasoning according to the stronger ones.[4] In this short version of the paper, I consider two theories that correspond to an equilibrium path and an equilibrium strategy profile. This allows to establish a surprising connection with strategic stability (Kohlberg and Mertens [20]).

Since players' theories of opponents' behavior are assumed to be commonly known,[5] the most natural application of Selective Rationalizability is explicit, pre-play coordination among players. Since a non-binding agreement is purely cheap talk, if a player displays behavior which is not compatible with rationality and belief in the agreement, the opponents are, in my view, more likely to abandon the belief that the player believes in the agreement, rather than the belief that the opponent is rational. As in the example, the agreement can also be interpreted as a set of public announcements.[6] Thus, Selective Rationalizability seems to be an appropriate tool to combine strategic reasoning and equilibrium play, especially when the motivation for equilibrium is explicit coordination. The application of Selective

[4]Note that, by non-monotonicity of strong belief, strategic reasoning about the stronger theories can potentially lead to behavior that cannot be rationalized under the weaker theories. For this reason, the epistemic priority issue arises.

[5]In the sense that players know what theories the opponents are supposed to have.

[6]Or, extending Selective Rationalizability to incomplete information games, the restrictions can model public news about a state of nature. For instance, in a financial market, players can tentatively believe that everyone is reasoning according to the same public information about a state of nature. Yet, if a player does not behave accordingly, the opponents may believe that the player has different information rather than deeming the player irrational.

Rationalizability to agreements and its relationship with equilibrium are deeply investigated in [14]. In particular, the outcomes that Selective Rationalizability uniquely pins down for some restrictions do not include and are not included in the set of subgame perfect equilibrium outcomes. However, I show in [13] that there always exists a subgame perfect equilibrium in behavioral strategies whose possible outcomes are delivered by Selective Rationalizability for particular restrictions. It is worth noting that the flexibility of Selective Rationalizability, which allows to model incomplete coordination instead of coordination on full strategy profiles, can be crucial to induce an outcome of the game (see [14] for details).

2 Preliminaries

Consider a finite dynamic game with complete information and perfect recall. Some notation:

- I is the finite set of *players*, and for any *profile* $(X_i)_{i\in I}$ and any $\emptyset \neq J \subseteq I$, I write $X_J := \times_{j\in J} X_j$, $X := X_I$, $X_{-i} := X_{I\setminus\{i\}}$, $X_{-i,j} := X_{I\setminus\{i,j\}}$;
- H_i is the set of *information sets* of player i, endowed with the precedence relation $\prec$;
- Z is the set of *terminal histories*;
- $u_i : Z \to \mathbb{R}$ is the *payoff function* of player i.

A *strategy* is a function $s_i : h \in H_i \mapsto s_i(h) \in A_i(h)$, where $A_i(h)$ is the set of available actions of player i at information set h. The set of all strategies is denoted by S_i. A strategy profile clearly induces one and only one terminal history; let $\zeta : S \to Z$ denote the map which associates each strategy profile $s \in S$ with the induced terminal history $z \in Z$. The set of strategies of player i which allow to reach an information set h (not necessarily of player i!) is

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

For any $(\overline{S}_j)_{j\in I} \subset S$, let $\overline{S}_i(h) := S_i(h) \cap \overline{S}_i$. If $h \in H_i$, $S_{-i}(h)$ represents the partial observation by player i of opponents' strategies up to h. For any $J \subseteq I$, $H_i(\overline{S}_J) := \left\{h \in H_i : \overline{S}_J(h) \neq \emptyset\right\}$ is the set of information sets of i compatible with $\overline{S}_J$.

Players update their beliefs about opponents' strategies and beliefs as the game unfolds. A Conditional Probability System (Renyi, [24]; henceforth CPS) assigns to each information set a belief, conditional on the observed opponents' behavior. Here I define CPS's over the opponents' state space $\Omega_{-i} := \times_{j\neq i}(S_j \times T_j)$, where epistemic type spaces $(T_j)_{j\in I}$ will be defined in Section 4.

Definition 1 *A Conditional Probability System on $(\Omega_{-i}, (T_{-i} \times S_{-i}(h))_{h\in H_i})$, with Borel sigma algebra $B(\Omega_{-i})$, is a mapping $\mu(\cdot|\cdot) : B(\Omega_{-i}) \times (T_{-i} \times S_{-i}(h))_{h\in H_i} \to [0,1]$ satisfying the following axioms:*

CPS-1. for every $C \in (T_{-i} \times S_{-i}(h))_{h\in H_i}$, $\mu(C|C) = 1$;

CPS-2. for every $C \in (T_{-i} \times S_{-i}(h))_{h\in H_i}$, $\mu(\cdot|C)$ is a probability measure on Ω_{-i};

CPS-3. for every $E \in B(\Omega_{-i})$ and $B, C \in (T_{-i} \times S_{-i}(h))_{h\in H_i}$, if $E \subseteq B \subseteq C$ then $\mu(E|B)\mu(B|C) = \mu(E|C)$.

The set of all CPS's of player i is denoted by $\Delta^{H_i}(\Omega_{-i})$.[7] For brevity, conditioning events will be indicated with just the information set.

CPS's on strategies are defined by replacing Ω_{-i} with S_{-i} and $(T_{-i} \times S_{-i}(h))_{h\in H_i}$ with $(S_{-i}(h))_{h\in H_i}$. For any $J \subseteq I\setminus\{i\}$ and $\overline{S}_J \subseteq S_J$, I say that $\mu_i \in \Delta^{H_i}(S_{-i})$ *strongly believes* (Battigalli and Siniscalchi, [9])[8]

[7] If each Ω_i is compact metrizable, endowing the set $\Delta(\Omega_{-i})$ of Borel probability measures on Ω_{-i} with the topology of weak convergence and $(\Delta(\Omega_{-i}))^{H_i}$ with the product topology, Battigalli and Siniscalchi [8] prove that $\Delta^{H_i}(\Omega_{-i})$ is a compact metrizable subset of $(\Delta(\Omega_{-i}))^{H_i}$.

[8] Battigalli and Siniscalchi make a stricter use of the term strong belief, by referring only to Borel subsets of Ω_{-i} or S_{-i}.

$\overline{S}_J$ if $\mu_i(\overline{S}_J \times S_{I\setminus(J\cup\{i\})}|h) = 1$ for all $h \in H_i(\overline{S}_J)$.

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 their belief about how the opponents will play and the expectation to choose rationally again in the continuation of the game. This is equivalent (see Battigalli [3]) to playing a *sequential best reply* to the CPS.

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

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

The set of sequential best replies to a CPS $\mu_i \in \Delta^{H_i}(S_{-i})$ is denoted by $\rho(\mu_i)$.

3 Selective Rationalizability

Before defining Selective Rationalizability, I have to pin down the behavior of players when they only reason about rationality. This task has already been accomplished in the literature under different assumptions. Pearce [23] defines Extensive-Form Rationalizability under *structural consistency* (an underlying feature also of sequential equilibrium). Battigalli [2] assumes *strategic independence*, which requires players to maintain the first-order belief about each opponent whenever her individual behavior does not contradict them. Battigalli and Siniscalchi [9] remove any assumption of independence and require players to maintain each order of belief in rationality only until *none* of the opponents contradict it. Then, they give to the resulting elimination procedure, *Strong Rationalizability*, an epistemic characterization based on the notion of *strong belief*. For this reason, I adopt Strong Rationalizability as a starting point, but I amend it by introducing *independent rationalization*: players maintain an order of belief in rationality of an opponent as long as her individual behavior does not contradict it. The motivation for this choice is two-fold. First, it is coherent with the emphasis on the persistence of beliefs in rationality. Second, there is an important motivation for the adoption of independent rationalization in Selective Rationalizability, which will be explained later. As far as Strong Rationalizability is concerned, it is easy to observe that independent rationalization is immaterial for the predicted outcomes, since it kicks in at an information set only when it is not reached anymore by some opponent. Instead, I do not adopt strategic independence. This is not in contradiction with independent rationalization: there can be correlations[9] also among the choices of players with different orders of belief in rationality (and actually, players do commonly believe in rationality along the rationalizable paths). However, assuming strategic independence would complicate the notation but not alter the results. For brevity and to distinguish it from the original notion of Strong Rationalizability, I will call this version simply ”Rationalizability”.

Definition 3 (Rationalizability) *Consider the following procedure.*

(Step 0) *For each $i \in I$, let $S_i^0 = S_i$.*

(Step n>0) *For each $i \in I$ and $s_i \in S_i$, let $s_i \in S_i^n$ if and only if there is $\mu_i \in \Delta^{H_i}(S_{-i})$ such that:*

R1 $s_i \in \rho(\mu_i)$*;*

[9] For instance, a player can believe that a sunny day will induce more optimistic beliefs in two opponents (regardless of their strategic sophistication); see, for instance, Aumann [1] and Brandenburger and Friedenberg [11].

R2 μ_i *strongly believes* S^q_j *for all* $j \neq i$ *and* $q < n$.

Finally let $S^\infty_i = \cap_{n\geq 0} S^n_i$. *The profiles in* S^∞ *are called rationalizable.*

Strong-Δ-Rationalizability is defined exactly like Strong Rationalizability, except that at each step only beliefs μ_i in a restricted set of CPS's $\Delta_i \subset \Delta^{H_i}(S_{-i})$ are allowed.

Selective Rationalizability refines Rationalizability in the following way. Each player has an exogenous theory of opponents' behavior and refines the rationalizable first-order beliefs according to this theory. The theory of player i is represented by a set of CPS's $\Delta_i \subseteq \Delta^{H_i}(S_{-i})$ over opponents' strategies. Players are aware of the theories of everyone else. Therefore, players can also expect *each* opponent to refine her first-order beliefs according to the own theory. This expectation towards an opponent is maintained as long as the opponent herself is not observed making a move that contradicts it. Moreover, players expect *each* opponent to reason about everyone else in the same way. Also this expectation is maintained as long as the opponent herself does not make a move that contradicts it. And so on. Thus, Selective Rationalizability is defined under independent rationalization. This allows better comparability with the equilibrium literature. Without independent rationalization, if a player deviates from the agreed-upon equilibrium path, each opponent is free to believe that any other opponent is not going to implement her threat. In this way, no coordination of threats would be required. These issues are widely discussed in [14]. Note however that independent rationalization is immaterial for the message of this paper and for the analysis of all the examples: players are only two in all games except for the game of Section 5, where independent rationalization plays no role anyway.

Definition 4 (Selective Rationalizability) *Fix a profile* $(\Delta_i)_{i\in I}$ *of compact subsets of CPS's. Let* $((S^m_i)_{i\in I})^\infty_{m=0}$ *denote the Rationalizability procedure. Consider the following procedure.*

(Step 0) *For each* $i \in I$, *let* $S^0_{i,R\Delta} = S^\infty_i$.

(Step n>0) *For each* $i \in I$ *and* $s_i \in S_i$, *let* $s_i \in S^n_{i,R\Delta}$ *if and only if there is* $\mu_i \in \Delta_i$ *such that:*

S1 $s_i \in \rho(\mu_i)$;

S2 μ_i *strongly believes* $S^q_{j,R\Delta}$ *for all* $j \neq i$ *and* $q < n$;

S3 μ_i *strongly believes* S^q_j *for all* $j \neq i$ *and* $q \in \mathbb{N}$.

Finally, let $S^\infty_{i,R\Delta} = \cap_{n\geq 0} S^n_{i,R\Delta}$. *The profiles in* $S^\infty_{R\Delta}$ *are called selectively-rationalizable.*

Step 0 initializes Selective Rationalizability with the rationalizable strategy profiles. This is only to stress that Selective Rationalizability refines Rationalizability: S3 already implies that players strongly believe in the rationalizable strategies of each opponent, and that the strategies surviving Step 1 are rationalizable. Indeed, Selective Rationalizability can also be seen as an extension of Rationalizability, in a unique elimination procedure where the first-order belief restrictions kick in once no more strategies can be eliminated otherwise.

Selective Rationalizability can be simplified in different ways according to the structure of the restrictions. S3 can be substituted by the requirement that strategies be rationalizable when first-order beliefs are not restricted at the non-rationalizable information sets.

Definition 5 *I say that* $\Delta_i \subseteq \Delta^{H_i}(S_{-i})$ *is rationalizable if* $\mu^*_i \in \Delta_i$ *whenever there exists* $\mu_i \in \Delta_i$ *such that* $\mu^*_i(\cdot|h) = \mu_i(\cdot|h)$ *for all* $h \in H_i(S^\infty)$.

Proposition 1 *Suppose that for every $i \in I$, Δ_i is rationalizable. Then, S3 can be substituted by $s_i \in S_i^0 = S_i^\infty$ in the definition of Selective Rationalizability.*

Proposition 2 *Fix $(\Delta_i)_{i\in I} \subseteq \times_{i\in I}\Delta^{H_i}(S_{-i})$ with $S_{R\Delta}^\infty \neq \emptyset$. There exists a profile $(\Delta_i^*)_{i\in I}$ of rationalizable*[10] *subsets of CPS's such that $\zeta(S_{R\Delta^*}^\infty) = \zeta(S_{R\Delta}^\infty)$.*

Thus, the class of rationalizable restrictions suffices to yield all the possible behavioral implications of Selective Rationalizability.

Selective Rationalizability and Strong-Δ-Rationalizability can yield the empty set. This happens when at some step there is no $\mu_i \in \Delta_i$ that satisfies S2 and S3, or the equivalent of S2 for Strong-Δ-Rationalizability. This means that the restrictions are not compatible with strategic reasoning about rationality and the restrictions themselves.

4 Epistemic framework and characterization theorem

I adopt the epistemic framework of Battigalli and Prestipino [7], dropping the incompleteness of information dimension. Players' beliefs over strategies of all orders are given an implicit representation through a compact, complete and continuous *type structure* $(\Omega_i, T_i, g_i)_{i\in I}$,[11] where for every $i \in I$, $\Omega_i = S_i \times T_i$, T_i is a compact metrizable space of *epistemic types*, and $g_i = (g_{i,h})_{h\in H_i} : T_i \to \Delta^{H_i}(\Omega_{-i})$ is a continuous and onto[12] *belief map*. I will call "events" the elements of the Borel sigma-algebras on each Ω_i, and of the product sigma algebras on the Cartesian spaces $\Omega_J := \times_{i\in J\subseteq I}\Omega_i$.

The first-order belief map of player i, $f_i = (f_{i,h})_{h\in H_i} : T_i \to \Delta^{H_i}(S_{-i})$, is defined as $f_{i,h}(t_i) = \mathrm{Marg}_{S_{-i}} g_{i,h}(t_i)$ for all $i \in I$ and $h \in H_i$, so it inherits continuity from g_i. The event in Ω_i where the restrictions of player i hold is

$$[\Delta_i] := \{(s_i, t_i) \in \Omega_i : f_i(t_i) \in \Delta_i\};$$

$[\Delta_i]$ is compact because Δ_i is compact and f_i is continuous. The cartesian set where the restrictions of all players hold is $[\Delta] := \times_{i\in I}[\Delta_i]$.

From now on, fix a Cartesian (across players) event $E = \times_{i\in I}E_i \subseteq \Omega$. The closed[13] event where player i believes in E_{-i} at an information set $h \in H_i$ is defined as

$$B_{i,h}(E_{-i}) := \{(s_i, t_i) \in \Omega_i : g_{i,h}(t_i)(E_{-i}) = 1\}.$$

The closedness of $B_{i,h}(E_{-i})$ implies the closedness of all the following belief events. The event where i believes in E_{-i} at every information set is $B_i(E_{-i}) := \cap_{h\in H_i} B_{i,h}(E_{-i})$.

[10] Although it is not formally shown, it is straightforward to observe that if each Δ_i is compact, each Δ_i^* constructed in the proof is compact too. This shows that the epistemic characterization of Selective Rationalizability holds under $(\Delta_i^*)_{i\in I}$; it is however immaterial for the results of this section.

[11] Friedenberg [16] proves that in static games, such a type structure contains all hierarchies of beliefs about strategies. Although this result has not been formally extended to dynamic games, to the best of my knowledge, no counterexample to this extension has ever been found. However, the canonical type structure for CPS of Battigalli and Siniscalchi [8] is compact, complete, and continuous (and it contains all collectively coherent hierarchies of beliefs by construction).

[12] This imposes to choose type spaces with the cardinality of the continuum.

[13] See Battigalli and Prestipino [7].

If $\text{Proj}_S E = S$, E is an *epistemic event*. Else, it could be impossible for player i to believe in E_{-i} at some information set $h \in H_i$, because $\text{Proj}_{S_{-i}} E_{-i} \cap S_{-i}(h) = \emptyset$. However, player i may want to believe in E_{-i} as long as not contradicted by observation. The event where this persistency of the belief holds is:

$$\overline{SB}_i(E_{-i}) := \bigcap_{h \in H_i : \text{Proj}_{S_{-i}} E_{-i} \cap S_{-i}(h) \neq \emptyset} B_{i,h}(E_{-i}).$$

The *strong belief* operator $\overline{SB}_i$ is non-monotonic: if $E_{-i} \subset F_{-i}$, it needs not be the case that $\overline{SB}_i(E_{-i}) \subset \overline{SB}_i(F_{-i})$. This will explain why Strong-Δ-Rationalizability is not a refinement of Strong Rationalizability, and Selective Rationalizability, for the same restrictions, is not a refinement of Strong-Δ-Rationalizability.

Suppose now that, for each opponent j, player i believes that the true pair (s_j, t_j) is in E_j, as long as this is not contradicted by observation. Then I say that i strongly believes in E_j for all $j \neq i$. Formally, I define the *independent strong belief* operator as

$$SB_i(E_{-i}) := \cap_{j \neq i} \overline{SB}_i(E_j \times \Omega_{-j,i}).$$

Note that $(s_i, t_i) \in SB_i(E_{-i})$ if and only if, for each $j \neq i$, $g_i(t_i)$ strongly believes in E_j, i.e. $g_{i,h}(t_i)(E_j \times \Omega_{-i,j}) = 1$ for all $h \in H_i$ with $\text{Proj}_{S_j} E_j \cap S_j(h) \neq \emptyset$.

Let $B(E) := \times_{i \in I} B_i(E_{-i})$, $SB(E) := \times_{i \in I} SB_i(E_{-i})$, and $CSB_i(E) := E_i \cap SB_i(E_{-i})$. The correct and mutual strong belief in the event E is denoted by:

$$CSB(E) := \times_{i \in I} CSB_i(E) = E \cap SB(E).$$

Let $B^0(E) := E = CSB^0(E)$. For all $n \in \mathbb{N}$, define the following n-th order belief operators: $B^{n+1}(E) := B(B^n(E))$ and $CSB^{n+1}(E) := CSB(CSB^n(E))$. An epistemic event E is *transparent* when it holds and is believed by all players at every information set and at every order. The corresponding event is $B^*(E) := \cap_{n \geq 0} B^n(E)$. If E is not an epistemic event, I will be interested in the event $CSB^\infty(E) := \cap_{n \geq 0} CSB^n(E)$.

First-order beliefs and higher-order beliefs have no bite in terms of behavior and predictions over opponents' behavior without rationality and beliefs in rationality. The "rationality of player i" event is denoted by

$$R_i := \{(s_i, t) \in \Omega_i : s_i \in \rho(f_i(t_i))\},$$

and it is closed whenever $\rho \circ f_i$, as assumed here, is upper-hemicontinuous.[14] The *rationality* event is $R := \times_{i \in I} R_i$.

Here I consider rational players who keep, as the game unfolds, the highest order of belief in rationality of each opponent that is consistent with her observed behavior. Players further refine their first-order beliefs through the own restrictions. All this is captured by the event $[\Delta] \cap CSB^\infty(R)$. The event "rationality and common independent strong belief in rationality" $CSB^\infty(R)$ characterizes Rationalizability.[15] Furthermore, players believe, as long as not contradicted by observation, that each opponent: (1) reasons in the same way; (2) believes, as long as not contradicted by observation, that everyone else reasons in the same way; and so on. The n-th order of this belief is captured by the event $CSB^n([\Delta] \cap CSB^\infty(R))$, and it characterizes the $n+1$-th step of Selective Rationalizability. The event $CSB^\infty([\Delta] \cap CSB^\infty(R))$ captures all the steps of reasoning at once.

[14] Finiteness suffices for upper-hemicontinuity.

[15] Analogously, "rationality and common strong belief in rationality" characterizes Strong Rationalizability (Battigalli and Siniscalchi [9]).

Theorem 1 *Fix a profile $\Delta = (\Delta_i)_{i \in I}$ of compact subsets of CPS's. Then, for every $n \geq 0$,*

$$S^{n+1}_{R\Delta} = \mathrm{Proj}_S CSB^n([\Delta] \cap CSB^\infty(R)),$$

and

$$S^{\infty}_{R\Delta} = \mathrm{Proj}_S CSB^\infty([\Delta] \cap CSB^\infty(R)).$$

The comparison between the characterization of Selective Rationalizability and the characterization of Strong-Δ-Rationalizability proposed by Battigalli and Prestipino [7] clarifies the epistemic priority difference behind the two solution concepts. In the event $\overline{CSB}^\infty(R \cap B^*([\Delta])) \subset B^*([\Delta])$ that characterizes Strong-Δ-Rationalizability,[16] players keep at every information set every order of belief in the restrictions. In the event $CSB^\infty([\Delta] \cap CSB^\infty(R)) \subset CSB^\infty(R)$, players keep at every information set the highest order of belief in each opponent's rationality which is per se compatible with her observed behavior.

5 Finer epistemic priority orderings

Consider the following game, where after I Cleo chooses the matrix.

Cleo — $I \longrightarrow$

$\downarrow O$

$A \backslash B$	W	E
N	$2,2,3.6$	$0,0,0$
S	$0,0,0$	$2,2,4$

$M1$	L	R
U	$1,1,3.3$	$0,0,3.3$
D	$0,0,3.3$	$1,1,3.9$
$M2$	L	R
U	$0,0,0$	$1,1,8.1$
D	$1,1,8.1$	$0,0,0$

All strategies are rationalizable. Suppose that players have theories of opponents' behavior that come from an equilibrium or an incomplete agreement among all players. An incomplete agreement or an equilibrium[17] align any two players' beliefs about a third player's moves. Are there restrictions of this kind under which Selective Rationalizability yields outcome $(O,(S,E))$? Yes. It is sufficient that Cleo expects Ann and Bob to play (S,E) after O and, for instance, (U,L) after I. Then, upon observing I, Ann and Bob drop the belief that Cleo has the aforementioned first-order belief restrictions. Thus, they can expect Cleo to pick any of the two matrices. If they believe that Cleo picks matrix $M1$, Ann may play U when she believes that Bob will play L, and vice versa.

Suppose now instead that Ann and Bob have an alternative theory to rationalize Cleo's move. They believe that Cleo believed that they would have coordinated on (S,E) after O, but does not believe that they will play (U,L) after I. If Ann and Bob rationalize the move of Cleo under this light, they expect Cleo to pick $M2$, because $(I,M1)$ is not rational given the belief in (S,E). Under $M2$, Ann and Bob cannot coordinate on (U,L).

Suppose now that Cleo expects Ann and Bob to play (N,W) after O and (U,L) after I. Upon observing I, as above, Ann and Bob believe the Cleo believed in (N,W) after O, but does not believe in (U,L) after I. But this does not exclude that Cleo would play $M1$, hoping in (D,R). Thus, Ann may play U when she believes that Bob will play L, and vice versa. So, Cleo's initial restrictions are compatible with the belief that Ann and Bob have the same restrictions about each other's moves, and will rationalize I

[16] $\overline{CSB}^\infty$ is defined like CSB^∞ starting from $\overline{SB}_i$ instead of SB_i.

[17] With the notable exception of self-confirming equilibrium (Fudenberg and Levine [17]).

under her belief in (N,W) only. The restrictions yield outcome $(O,(N,W))$ as unique prediction not just under Selective Rationalizability, but also under the additional strategic reasoning hypotheses.

Note a paradoxical fact: to convince Cleo to play O, Ann and Bob must promise to play (N,W), which yields Cleo a payoff of 3.6, instead of (S,E), which yields Cleo a payoff of 4. The intuitive explanation is that a higher expectation of Cleo after O allows her to take a convincing position of power after I.

Two important questions arise now. First: does the exclusion of $(O,(S,E))$ and not of $(O,(N,W))$ correspond to some existing equilibrium refinement? Note that both outcomes are induced by a subgame perfect equilibrium in (extensive-form/strongly) rationalizable strategies. Second, and most importantly: can the strategic reasoning above be modeled as an epistemic priority order between different theories of opponents behavior and be captured by a solution concept analogous to Selective Rationalizability?

The answer to the first question is yes: strategic stability a la Kohlberg and Mertens [20].[18]

Definition 6 (Kohlberg and Mertens [20]) *For each $i \in I$, let Σ_i be the set of mixed strategies of i, i.e. the set of probability distributions over S_i. A closed set of mixed equilibria $\widehat{\Sigma} \subseteq \Sigma$ is stable if it is minimal with respect to the following property: for any $\varepsilon > 0$, there exists $\delta_0 > 0$ such that for any completely mixed $(\sigma_i)_{i\in I} \in \Sigma$ and $(\delta_i)_{i\in I}$ with $\delta_i < \delta_0$ for all $i \in I$, the perturbed game where for every $i \in I$, every $s_i \in S_i$ is substituted by $(1-\delta_i)s_i + \delta_i\sigma_i$ has a mixed equilibrium ε-close to $\widehat{\Sigma}$.*

Consider first a set of two mixed equilibria $\widehat{\Sigma} = \{(\sigma_i)_{i\in I}, (\sigma_i')_{i\in I}\}$ inducing outcome $(O,(N,W))$, where $\sigma_C(O) = \sigma_C'(O) = 1$, $\sigma_A(N.D) = \sigma_B(W.R) = 1/\sqrt{2}$, and $\sigma_A'(N.D) = \sigma_B'(W.R) = 2/3$. Under σ, Cleo is actually indifferent between O and $I.M1$, while under σ', she is indifferent between O and $I.M2$. I show that $\widehat{\Sigma}$ is stable. Fix any completely mixed $(\widetilde{\sigma}_i)_{i\in I} \in \Sigma$, an arbitrarily small δ_0, and $(\delta_i)_{i\in I}$ with $\delta_i < \delta_0$ for all $i \in I$. Consider the game perturbed as in Definition 6 and indicate with tilde the perturbed strategies. If $\widetilde{\sigma}_A(I.M1) > \widetilde{\sigma}_A(I.M2)$ (resp., $\widetilde{\sigma}_A(I.M1) < \widetilde{\sigma}_A(I.M2)$), assign small probability to $\widetilde{I.M1}$ (resp., $\widetilde{I.M2}$) and the complementary probability to $\widetilde{O}$ in such a way that $I.M1$ and $I.M2$ are played with probability $1/2$. Then, after I, Ann and Bob are indifferent between their actions regardless of the belief about the action of the other. Thus, since all strategies are perturbed in the same way, Ann and Bob are indifferent between $\widetilde{N.U}$ and $\widetilde{N.D}$, and between $\widetilde{W.L}$ and $\widetilde{W.R}$. Assign probability to these strategies in such a way that Cleo is indifferent between $\widetilde{O}$ and $\widetilde{I.M1}$ (resp., $\widetilde{I.M2}$).[19] For any $\varepsilon > 0$, by picking a small enough δ_0, we have an equilibrium in the perturbed strategies where the induced probabilities over the original strategies are ε-close to those assigned by σ (resp., σ').

Instead, there is no stable set of equilibria inducing $(O,(S,E))$: any perturbation of O that gives negligible probability to $I.M1$ with respect to $I.M2$ cannot be compensated by giving positive probability to $\widetilde{I.M1}$, because $\widetilde{I.M1}$ cannot be optimal under belief in (S,E) (albeit perturbed). Thus, Ann and Bob must play a (perturbed) equilibrium of matrix $M2$, which cannot discourage a deviation to $\widetilde{I.M2}$.

This is not the first time that a connection between equilibrium refinements a la strategic stability and rationalizability is established. In signaling games, Battigalli and Siniscalchi [10] show that

[18] Strategic stability has been chosen over Forward Induction equilibria of Govindan and Wilson [18] or Man [21] because the latter do not refine extensive-form rationalizability, hence do not capture all orders of strong belief in rationality. Strategic stability, instead, refines iterated admissibility, which in generic games corresponds to extensive-form rationalizability (Shimoji, [25]).

[19] Since the perturbed strategies assign positive probability to S and E, the expected payoff of Cleo after O is lower than 3.6. Thus, $N.U$ and $N.L$ (resp., $N.D$ and $N.R$) must be assigned probability higher than $1/\sqrt{2}$ (resp., than $2/3$).

when an equilibrium outcome satisfies the Iterated Intuitive Criterion (Cho and Kreps [15]), Strong-Δ-Rationalizability yields a non-empty set for the corresponding restrictions (i.e. the belief that opponents play compatibly with the path). In [14] I prove that Selective Rationalizability yields the empty set for a class of non strategically stable equilibrium paths: those that *can be upset by a convincing deviation* (Osborne [22]). So, one could think that strategic stability simply requires non-emptiness of Selective Rationalizability under the belief in the equilibrium path. This is false. In the example above, Selective Rationalizability yields a non-empty set under the belief in $(O,(S,E))$ (but does not yield $(O,(S,E))$ as unique prediction). Thus, there is no incompatibility between the belief in the path and the rationalization of deviations based on it (unlike for equilibrium paths that can be upset by a convincing deviation). The problem is the incompatibility between the rationalization of deviations based on the belief in the path and the threats that sustain the path in equilibrium. This calls for a rationalizability procedure that takes both into account in a given *epistemic priority ordering*; in particular the "theory" that players comply with the path will be assigned higher epistemic priority with respect the "theory" that players implement also the equilibrium threats. In the full version of the paper, I construct and characterize epistemically such rationalizability procedure. The scope is expanded to an arbitrary number of theories of opponents' behavior, of an arbitrary nature (i.e. not just path versus full equilibrium behavior). Without the ambition to perfectly characterize strategic stability, the application of this rationalizability procedure to an equilibrium path and profile captures in a general and transparent way the spirit of the strategic reasoning stories in the background of strategic stability and related refinements.

6 Appendix

For each $i \in I$ and $h \in H_i$, let $p(h)$ be the immediate predecessor of h.

Lemma 1 *Fix a profile of rationalizable subsets of CPS's* $(\Delta_i)_{i\in I}$.
For every $n \geq 0$, $i \in I$, $h \in H_i(S^n_{i,R\Delta})\backslash H_i(S^\infty)$ *with* $p(h) \in H_i(S^\infty)$, *and* $s_i \in S^\infty_i(h)$, *there exists* $s^*_i \in S^n_{i,R\Delta}$ *such that* $s^*_i(h') = s_i(h')$ *for all* $h' \succeq h$.

Proof. By $S^0_{i,R\Delta} = S^\infty_i$, the result trivally holds for $n = 0$. Fix $n > 0$ and suppose to have proved the result for all $q < n$. Fix $i \in I$, $h \in H_i(S^n_{i,R\Delta})\backslash H_i(S^\infty)$ with $p(h) \in H_i(S^\infty)$, and $s_i \in S^\infty_i(h)$. Fix $\mu_i \in \Delta_i$ that satifies S2 and S3 with $\rho(\mu_i)(h) \neq \emptyset$ (it exists by $h \in H_i(S^n_{i,R\Delta})$) and μ'_i that satisfies S3 with $s_i \in \rho(\mu'_i)$. For each $j \neq i$ and $s_j \in S^\infty_j(h)$, letting $m := \max\left\{q < n : S^q_{j,R\Delta}(h) \neq \emptyset\right\}$, by the Induction Hypothesis there exists $s^*_j \in S^m_{j,R\Delta}(h)$ such that $s^*_j(h') = s_j(h')$ for all $h' \succeq h$. Let $\eta^h_j(s_j) := s^*_j$. For each $s_j \in S_j(h)\backslash S^\infty_j(h)$, let $\eta^h_j(s_j) := s_j$. Since μ_i strongly believes $S^0_{-i,R\Delta} = S^\infty_{-i}$, $h \notin H_i(S^\infty_{-i})$, and $p(h) \in H_i(S^\infty_{-i})$, $\mu_i(S_{-i}(h)|p(h)) = 0$. Thus, I can construct μ^*_i that satisfies S2 and S3 as (i) $\mu^*_i(\cdot|h') = \mu_i(\cdot|h')$ for all $h' \not\succeq h$, and (ii) $\mu^*_i(s_{-i}|h') = \mu'_i(\times_{j\neq i}(\eta^h_j)^{-1}(s_j)|h')$ for all $h' \succeq h$ and $s_{-i} = (s_j)_{j\neq i} \in S_{-i}(h)$. By (i) and rationalizability of Δ_i, $\mu^*_i \in \Delta_i$. By (i) and (ii), there exists $s^*_i \in \rho(\mu^*_i)(h) \subseteq S^n_{i,R\Delta}$ such that $s^*_i(h') = s_i(h')$ for all $h' \succeq h$. ■

Proof of Proposition 1. Fix $n \in \mathbb{N}$, $i \in I$, $\mu_i \in \Delta_i$ that satifies S2 at n, and $s_i \in \rho(\mu_i) \cap S^\infty_i$. Fix μ'_i that satisfies S3 with $s_i \in \rho(\mu'_i)$. Fix $h \in H_i(s_i)\backslash H_i(S^\infty)$ with $p(h) \in H_i(S^\infty)$. For each $j \neq i$ and $s_j \in S^\infty_j(h)$, letting $m := \max\left\{q < n : S^q_{j,R\Delta}(h) \neq \emptyset\right\}$, by Lemma 1 there exists $s^*_j \in S^m_{j,R\Delta}(h)$ such that $s^*_j(h') = s_j(h')$ for all $h' \succeq h$. Let $\eta^h_j(s_j) := s^*_j$. For each $s_j \in S_j(h)\backslash S^\infty_j(h)$, let $\eta^h_j(s_j) := s_j$. Since μ_i strongly believes

$S^0_{-i,R\Delta} = S^\infty_{-i}$, $h \notin H_i(S^\infty_{-i})$, and $p(h) \in H_i(S^\infty_{-i})$, $\mu_i(S_{-i}(h)|p(h)) = 0$. Thus, there exists μ_i^* that satisfies S2 and S3 such that (i) $\mu_i^*(\cdot|h) = \mu_i(\cdot|h)$ for all $h \in H_i(S^\infty)$, and (ii) $\mu_i^*(s_{-i}|h') = \mu_i'(\times_{j\neq i}(\eta_j^h)^{-1}(s_j)|h')$ for all $h \in H_i(s_i)\backslash H_i(S^\infty)$ with $p(h) \in H_i(S^\infty)$, $h' \succeq h$, and $s_{-i} = (s_j)_{j\neq i} \in S_{-i}(h)$. By (i) and rationalizability of Δ_i, $\mu_i^* \in \Delta_i$. By (i) and (ii), $s_i \in \rho(\mu_i) \subseteq S^n_{i,R\Delta}$. ■

Proof of Proposition 2. For each $i \in I$, let $\overline{\Delta}_i$ be the set of all $\mu_i \in \Delta_i$ that satisfy S3 and S2 under $(\Delta_j)_{j\in I}$ for all $n \in \mathbb{N}$. By finiteness,[20] (1) $S^\infty_{R\overline{\Delta}} = S^1_{R\overline{\Delta}} = \times_{i\in I}\rho(\overline{\Delta}_i) = S^\infty_{R\Delta}$. Let $\mu_i \in \Delta_i^*$ if and only if there exists $\overline{\mu}_i \in \overline{\Delta}_i$ such that $\mu_i(\cdot|h) = \overline{\mu}_i(\cdot|h)$ for all $h \in H_i(S^\infty)$. Obviously, $\overline{\Delta}_i \subseteq \Delta_i^*$ for all $i \in I$; thus, $S^1_{R\overline{\Delta}} \subseteq S^1_{R\Delta^*}$.

I show first that Δ_i^* is rationalizable. Fix μ_i and $\overline{\mu}_i \in \Delta_i^*$ such that $\mu_i(\cdot|h) = \overline{\mu}_i(\cdot|h)$ for all $h \in H_i(S^\infty)$. Then, there exists $\overline{\overline{\mu}}_i \in \overline{\Delta}_i$ such that $\overline{\overline{\mu}}_i(\cdot|h) = \overline{\mu}_i(\cdot|h) = \mu_i(\cdot|h)$ for all $h \in H_i(S^\infty)$, so $\mu_i \in \Delta_i^*$.

Fix $i \in I$ and $s_i \in S^1_{i,R\Delta^*}$. Fix $\mu_i \in \Delta_i^*$ such that $s_i \in \rho(\mu_i)$. By definition of Δ_i^*, there exists $\overline{\mu}_i \in \overline{\Delta}_i \subseteq \Delta_i^*$ such that $\mu_i(\cdot|h) = \overline{\mu}_i(\cdot|h)$ for all $h \in H_i(S^\infty)$. Thus, there exists $\overline{s}_i \in \rho(\overline{\mu}_i) \subseteq \rho(\overline{\Delta}_i) = S^1_{i,R\overline{\Delta}}$ such that $\overline{s}_i(h) = s_i(h)$ for all $h \in H_i(S^\infty) \supseteq H_i\left(S^1_{R\Delta^*}\right)$. Hence, by $S^1_{R\overline{\Delta}} \subseteq S^1_{R\Delta^*}$, (★) $H_i(S^1_{j,R\Delta^*}) \cap H_i(S^\infty) = H_i(S^1_{j,R\overline{\Delta}}) \cap H_i(S^\infty)$ for all $j \neq i$, and by $\zeta(S^1_{R\Delta^*}) \subseteq \zeta(S^\infty)$, (2) $\zeta(S^1_{R\Delta^*}) = \zeta(S^1_{R\overline{\Delta}})$.

Since $\overline{s}_i \in S^2_{i,R\overline{\Delta}} = S^1_{i,R\overline{\Delta}}$, there exists $\overline{\mu}_i \in \overline{\Delta}_i$ that strongly believes $(S^1_{j,R\overline{\Delta}})_{j\neq i}$ such that $\overline{s}_i \in \rho(\overline{\mu}_i)$. For each $h \in H_i(S^\infty)$ and $s_{-i} \in S_{-i}$ with $\overline{\mu}_i(s_{-i}|h) > 0$, since $\overline{\mu}_i$ strongly believes in S^∞_{-i}, $\overline{s}_i \in S^\infty_i$, and $\overline{s}_i(h') = s_i(h')$ for all $h' \in H_i(S^\infty)$, $\zeta(\overline{s}_i, s_{-i}) = \zeta(s_i, s_{-i})$. Thus, by $\overline{s}_i \in \rho(\overline{\mu}_i)$, s_i is a continuation best reply to $\overline{\mu}_i(\cdot|h)$ too. Fix μ_i' that satisfies S3 such that $s_i \in \rho(\mu_i')$. Fix $h \in H_i(s_i)\backslash H_i(S^\infty)$ with $p(h) \in H_i(S^\infty)$. For each $j \neq i$ and $s_j \in S^\infty_j(h)$, letting $m := \max\left\{q = 0, 1 : S^q_{j,R\Delta^*}(h) \neq \emptyset\right\}$, by Lemma 1 there exists $s_j^* \in S^m_{j,R\Delta^*}(h)$ such that $s_j^*(h') = s_j'(h')$ for all $h' \succeq h$. Let $\eta_j^h(s_j) := s_j^*$. For each $s_j \in S_j(h)\backslash S^\infty_j(h)$, let $\eta_j^h(s_j) := s_j$. Since $\overline{\mu}_i$ strongly believes S^∞_{-i}, $h \notin H_i(S^\infty_{-i})$, and $p(h) \in H_i(S^\infty_{-i})$, $\overline{\mu}_i(S_{-i}(h)|p(h)) = 0$. For all $j \neq i$, by $S^1_{R\overline{\Delta}} \subseteq S^1_{R\Delta^*}$ and (★), $\overline{\mu}_i(S^1_{j,R\Delta^*} \times S_{-i,j}|h) = 1$ for all $h \in H_i(S^1_{j,R\Delta^*}) \cap H_i(S^\infty)$. Thus, there exists μ_i^* that satisfies S3 and strongly believes $(S^1_{j,R\Delta^*})_{j\neq i}$ such that (i) $\mu_i^*(\cdot|h) = \overline{\mu}_i(\cdot|h)$ for all $h \in H_i(S^\infty)$, and (ii) $\mu_i^*(s_{-i}|h') = \mu_i'(\times_{j\neq i}(\eta_j^h)^{-1}(s_j)|h')$ for all $h \in H_i(s_i)\backslash H_i(S^\infty)$ with $p(h) \in H_i(S^\infty)$, $h' \succeq h$, and $s_{-i} = (s_j)_{j\neq i} \in S_{-i}(h)$. By (i) and rationalizability of Δ_i^*, $\mu_i^* \in \Delta_i^*$. By (i) and (ii), $s_i \in \rho(\mu_i) \subseteq S^2_{i,R\Delta^*}$. Thus, (3) $S^\infty_{R\Delta^*} = S^1_{R\Delta^*}$. By 1-2-3, $\zeta(S^\infty_{R\Delta^*}) = \zeta(S^\infty_{R\Delta})$. ■

PROOF OF THEOREM 1.

First, I prove a generalized version of Theorem 1. Applying this generalized version to Rationalizability yields the hypotheses to apply it to Selective Rationalizability and prove Theorem 1.

Consider this generalized rationalizability procedure:

Definition 7 *Fix a profile of compact subsets of CPS's $(\Delta_i)_{i\in I}$. Fix another profile of compact subsets of CPS's $(\Delta_i^G)_{i\in I}$. Fix $n \geq 1$ and, if $n > 1$, suppose to have defined $((S^q_{i,G})_{i\in I})^{n-1}_{q=1}$. For every $i \in I$ and $s_i \in S_i$, let $s_i \in S^n_{i,G}$ if and only if there exists $\mu_i \in \Delta_i$ such that:*

G1 $s_i \in \rho(\mu_i)$;

G2 μ_i *strongly believes* $S^q_{j,G}$ *for all* $j \neq i$ *and* $q < n$;

[20] Or other milder conditions which guarantee that every $s_i \in S^\infty_{i,R\Delta}$ is a sequential best reply to some belief μ_i that strongly believes $((S^q_{j,R\Delta})_{j\in I})^\infty_{q=0}$.

G3 $\mu_i \in \Delta_i^G$.

Call $\Delta_i^{n,G}$ *the set of all* $\mu_i \in \Delta_i$ *that satisfy G2 and G3.*

Finally, let $S_{i,G}^{\infty} = \cap_{n\geq 1} S_{i,G}^n$ *and* $\Delta_i^{\infty,G} = \cap_{n\geq 1}\Delta_i^{n,G}$.

Consider the following property for a Cartesian event $E = \times_{i\in I} E_i \subseteq \Omega$.

Definition 8 *A Cartesian event* $E = \times_{i\in I} E_i$ *satisfies the "completeness property" if for every* $i \in I$, $t_i \in \mathrm{Proj}_{T_i} E_i$, $s_i \in \rho(f_i(t_i))$, *and maps*[21] $(\tau_j)_{j\neq i}$ *with* $\tau_j : \bar{s}_j \in \mathrm{Proj}_{S_j} E_j \mapsto (\bar{s}_j, t_j) \in E_j$ *for all* $j \neq i$, *there exists* $t_i' \in T_i$ *such that* $(s_i, t_i') \in E_i$, $f_i(t_i') = f_i(t_i)$, *and* $g_{i,h}(t_i')\left[\tau_j(s_j) \times \Omega_{-i,j}\right] = f_{i,h}(t_i)\left[s_j \times S_{-i,j}\right]$ *for all* $h \in H_i$, $j \neq i$, *and* $s_j \in \mathrm{Proj}_{S_j} E_j$.

Now I can state a generalized characterization theorem.[22]

Lemma 2 *Fix a closed, Cartesian event* $E = \times_{i\in I} E_i \subseteq R$ *with the completeness property such that for each* $i \in I$, $f_i(\mathrm{Proj}_{T_i} E) = \Delta_i \cap \Delta_i^G$ *(which implies* $S_G^1 = \mathrm{Proj}_S E$*).*[23]
Then, for every $n \in \mathbb{N}$, $CSB^{n-1}(E)$ *has the completeness property and for each* $i \in I$, $f_i(\mathrm{Proj}_{T_i} CSB^{n-1}(E)) = \Delta_i^{n,G}$ *if* $n = 1$, *and* $f_i(\mathrm{Proj}_{T_i} CSB_i(CSB^{n-2}(E))) = \Delta_i^{n,G}$ *if* $n > 1$ *(which implies* $S_G^n = \mathrm{Proj}_S CSB^{n-1}(E)$*).*[24]

Moreover, $CSB^{\infty}(E)$ *has the completeness property and for each* $i \in I$, $f_i(\mathrm{Proj}_{T_i} CSB^{\infty}(E)) = \Delta_i^{\infty,G}$ *(which implies* $S_G^{\infty} = \mathrm{Proj}_S CSB^{\infty}(E)$*).*[25]

Proof. For finite n, the proof is by induction.

Induction Hypothesis (n=1,...,m): the Lemma holds for $n = 1, \dots m$.

Basis step (n=1): the Lemma holds for $n = 1$ by hypothesis.

Inductive step (n=m+1): Let $F = \times_{i\in I} F_i := CSB^{m-1}(E)$ and $G = \times_{i\in I} G_i := CSB^m(E)$, where for all $i \in I$, $F_i = CSB_i(CSB^{m-2}(E))$ and $G_i = CSB_i(F)$

Fix $i \in I$ and $\mu_i \in \Delta_i^{m+1,G} \subseteq \Delta_i^{m,G}$. Then, by the Induction Hypothesis, there exists $t_i \in \mathrm{Proj}_{T_i} F_i$ such that $f_i(t_i) = \mu_i$. Fix maps $(\tau_j)_{j\neq i}$ with $\tau_j : \bar{s}_j \in \mathrm{Proj}_{S_j} F_j \mapsto (\bar{s}_j, t_j) \in F_j$ for all $j \neq i$. By the Induction Hypothesis, F has the completeness property. So, there exists $(s_i', t_i') \in F_i$ such that $f_i(t_i') = f_i(t_i) = \mu_i$, and for every $h \in H_i$, $j \neq i$, and $s_j \in \mathrm{Proj}_{S_j} F_j$, $g_{i,h}(t_i')\left[\tau_j(s_j) \times \Omega_{-i,j}\right] = f_{i,h}(t_i)\left[s_j \times S_{-i,j}\right]$. Then, since $f_i(t_i) = \mu_i \in \Delta_i^{m+1,G}$ strongly believes $S_{j,G}^m = \mathrm{Proj}_{S_j} F_j$ (by the Induction Hypothesis), $g_i(t_i')$ strongly believes F_j. So, $(s_i', t_i') \in SB_i(F_{-i}) \cap F_i$. Thus, $(s_i', t_i') \in G_i$.

Fix $i \in I$ and $t_i \in \mathrm{Proj}_{T_i} G_i$. Since $t_i \in \mathrm{Proj}_{T_i} F_i$, by the Induction Hypothesis $f_i(t_i) \in \Delta_i^{m,G}$. Since $t_i \in \mathrm{Proj}_{T_i} SB_i(F)$, $g_i(t_i)$ strongly believes F_j for all $j \neq i$, hence $f_i(t_i)$ strongly believes $\mathrm{Proj}_{S_j} F_j$. By the Induction Hypothesis $\mathrm{Proj}_{S_j} F_j = S_j^m$. So $f_i(t_i) \in \Delta_i^{m+1,G}$.

Now I show that G has the completeness property. Fix $i \in I$, $t_i \in \mathrm{Proj}_{T_i} G_i \subseteq \mathrm{Proj}_{T_i} F_i$, $s_i \in \rho(f_i(t_i))$, and maps $(\tau_j)_{j\neq i}$ with $\tau_j : \bar{s}_j \in \mathrm{Proj}_{S_j} G_j \mapsto (\bar{s}_j, t_j) \in G_j \subseteq F_j$ for all $j \neq i$. Extend each τ_j to $\tau_j' : \bar{s}_j \in \mathrm{Proj}_{S_j} F_j \mapsto$

---

[21] Note that the maps are injective.

[22] The event E can be empty, just like $CSB^{\infty}(R) \cap [\Delta]$ in Theorem 1.

[23] $\subseteq$ is guaranteed by the completeness property of E; $\supseteq$ is guaranteed by the fact that $E \subseteq R$.

[24] For $n > 1$, $\mathrm{Proj}_{T_i} CSB^{n-1}(E)$ must be substituted by $f_i(\mathrm{Proj}_{T_i} CSB_i(CSB^{n-2}(E)))$ because for some $j \neq i$, $CSB_j(CSB^{n-2}(E))$ may be empty (and thus $CSB^{n-1}(E)$ too) while $\Delta_i^{n,G}$ is not.

[25] Finiteness implies that for every $s_i \in S_{i,G}^{\infty}$, $s_i \in \rho(\mu_i)$ for some $\mu_i \in \Delta_i^{\infty,G}$, but it can be substituted by mild regularity conditions (see, for instance, Battigalli [5]).

$(\bar{s}_j, t_j) \in F_j$ in such a way that for every $s_j \in \mathrm{Proj}_{S_j} G_j$, $\tau'_j(s_j) = \tau_j(s_j)$. By the Induction Hypothesis, F has the completeness property. So, there exists $t'_i \in T_i$ such that $(s_i, t'_i) \in F_i$, $f_i(t'_i) = f_i(t_i)$, and for every $h \in H_i$, $j \neq i$, and $s_j \in \mathrm{Proj}_{S_j} F_j \supseteq \mathrm{Proj}_{S_j} G_j$, $g_{i,h}(t'_i)[\tau'_j(s_j) \times \Omega_{-i,j}] = f_{i,h}(t_i)[s_j \times S_{-i,j}]$. Since $t_i \in \mathrm{Proj}_{T_i} SB_i(F_{-i})$, $f_i(t_i)$ strongly believes $\mathrm{Proj}_{S_j} F_j$ for all $j \neq i$. Then, by construction, $g_i(t'_i)$ strongly believes F_j. So $(s_i, t'_i) \in SB_i(F_{-i})$. Thus, $(s_i, t'_i) \in G_i$. Note that for every $h \in H_i$, $j \neq i$, and $s_j \in \mathrm{Proj}_{S_j} G_j$, $g_{i,h}(t'_i)[\tau_j(s_j) \times \Omega_{-i,j}] = f_{i,h}(t_i)[s_j \times S_{-i,j}]$. $\square$

Now I prove that the lemma holds for $n = \infty$. By finiteness, there is $M \in \mathbb{N}$ such that $S^{\infty}_G = S^M_G$. Let $F := CSB^M(E)$ and $G := CSB^{\infty}(E) = \cap_{n \geq 0} CSB^n(E)$.

Fix $i \in I$ and $\mu_i \in \Delta^{\infty,G}_i = \Delta^{M+1,G}_i$.

By finiteness, the existence of such μ_i implies that $\Delta^{\infty,G}_j \neq \emptyset$ for all $j \neq i$, so $S^{\infty}_G \neq \emptyset$. As shown above, for every $s \in S^{\infty}_G$ and $q \geq 0$, $(\{s\} \times T) \cap CSB^q(E) \neq \emptyset$. Since each $CSB^q(E)$ and $(\{s\} \times T)$ are closed (see Section 4), $((\{s\} \times T) \cap CSB^q(E))_{q \geq 0}$ is a sequence of nested, nonempty closed sets, so it has the finite intersection property. Since Ω is compact, $(\{s\} \times T) \cap G \neq \emptyset$. Then, for each $j \in I$, there exists $\tau_j : \bar{s}_j \in \mathrm{Proj}_{S_j} F \mapsto (\bar{s}_j, t_j) \in \mathrm{Proj}_{\Omega_j} G$.

As shown above, there exists $t_i \in \mathrm{Proj}_{T_i} F$ such that $f_i(t_i) = \mu_i$, and F has the completeness property. So, there exists $(s'_i, t'_i) \in \mathrm{Proj}_{\Omega_i} F$ such that for every $i \in I$, $f_i(t'_i) = f_i(t_i)$, and for every $h \in H_i$, $j \neq i$, and $s_j \in \mathrm{Proj}_{S_j} F$, $g_{i,h}(t'_i)[\tau_j(s_j) \times \Omega_{-i,j}] = f_{i,h}(t_i)[s_j \times S_{-i,j}]$. Then, since $f_i(t_i) = \mu_i \in \Delta^{\infty,G}_i$ strongly believes $S^{\infty}_{j,G} = \mathrm{Proj}_{S_j} F$ (shown above), $g_i(t'_i)$ strongly believes $\mathrm{Proj}_{\Omega_j} G$. Hence, for each $q \geq M$, since $S^q_G = S^{\infty}_G$ and $CSB^q(E) \supset G$, $g_i(t'_i)$ strongly believes $\mathrm{Proj}_{\Omega_j} CSB^q(E)$. So, $(s'_i, t'_i) \in SB_i(CSB^q(E))$. Repeating for each $i \in I$, $CSB^{q+1}(E) \neq \emptyset$. Then $(s'_i, t'_i) \in \mathrm{Proj}_{\Omega_i} CSB^{q+1}(E)$ for all $q \geq M$. Thus $(s'_i, t'_i) \in \mathrm{Proj}_{\Omega_i} G \neq \emptyset$.

Fix $i \in I$ and $t_i \in \mathrm{Proj}_{T_i} G$. For every $q \geq 1$, $t_i \in \mathrm{Proj}_{T_i} CSB^{q-1}(E)$, thus, as shown above, $f_i(t_i) \in \Delta^{q,G}_i$. Then, $f_i(t_i) \in \Delta^{\infty,G}_i$.

Now I show that G has the completeness property. Fix $i \in I$, $t_i \in \mathrm{Proj}_{T_i} G \subseteq \mathrm{Proj}_{T_i} F$, $s_i \in \rho(f_i(t_i))$, and maps $(\tau_j)_{j \neq i}$ with $\tau_j : \bar{s}_j \in \mathrm{Proj}_{S_j} G \mapsto (\bar{s}_j, t_j) \in \mathrm{Proj}_{\Omega_j} G \subseteq \mathrm{Proj}_{\Omega_j} F$ for all $j \neq i$. As shown above, $\mathrm{Proj}_S G = S^{\infty}_G = S^M_G = \mathrm{Proj}_S F$, and F has the completeness property. So, there exists $t'_i \in T_i$ such that $(s_i, t'_i) \in \mathrm{Proj}_{\Omega_i} F$, $f_i(t'_i) = f_i(t_i)$, and for every $h \in H_i$, $j \neq i$, and $s_j \in \mathrm{Proj}_{S_j} F = \mathrm{Proj}_{S_j} G$, $g_{i,h}(t'_i)[\tau_j(s_j) \times \Omega_{-i,j}] = f_{i,h}(t_i)[s_j \times S_{-i,j}]$. Since $t_i \in \mathrm{Proj}_{T_i} SB_i(F)$, $f_i(t_i)$ strongly believes $\mathrm{Proj}_{S_j} F = \mathrm{Proj}_{S_j} G$ for all $j \neq i$. Then, $g_i(t'_i)$ strongly believes $\mathrm{Proj}_{\Omega_j} G$. Hence, for each $q \geq M$, since $\mathrm{Proj}_S F = \mathrm{Proj}_S CSB^q(E) = \mathrm{Proj}_S G$ and $CSB^q(E) \supset G$, $g_i(t'_i)$ strongly believes $\mathrm{Proj}_{\Omega_j} CSB^q(E)$. So, $(s_i, t'_i) \in SB_i(CSB^q(E))$. Then $(s_i, t'_i) \in \mathrm{Proj}_{\Omega_i} CSB^{q+1}(E)$ for all $q \geq M$. Thus $(s_i, t'_i) \in \mathrm{Proj}_{\Omega_i} G$. $\blacksquare$

Proof of Theorem 1.

For each $i \in I$, let $\Delta^{\infty,G}_i$ be the set of CPS's that satisfy S3. Theorem 1 is given by Lemma 2 with $E = [\Delta] \cap CSB^{\infty}(R)$ and $\Delta^G_i = \Delta^{\infty,G}_i$ for all $i \in I$. I show that $[\Delta] \cap CSB^{\infty}(R)$ satisfies the hypotheses of Lemma 2.

The event $R = \times_{i \in I} R_i$ is closed (see Section 4). Now I show that it has the completeness property. Fix $i \in I$, $t_i \in \mathrm{Proj}_{T_i} R$, $s_i \in \rho(f_i(t_i))$, and, for each $j \neq i$, $\tau_j : \bar{s}_j \in \mathrm{Proj}_{S_j} R \mapsto (\bar{s}_j, t_j) \in R_j$. Extend each τ_j to $\tau'_j : \bar{s}_j \in S_j \mapsto (\bar{s}_j, t_j) \in \Omega_j$ in such a way that for every $s_j \in \mathrm{Proj}_{S_j} R$, $\tau'_j(s_j) = \tau_j(s_j)$. Define $\nu_i \in (\Delta(S_{-i} \times T_{-i}))^{H_i}$ as $\nu_i(\times_{j \neq i} \tau'_j(s_j) | h) = f_{i,h}(t_i)[s_{-i}]$ for all $h \in H_i$ and $s_{-i} = (s_j)_{j \neq i} \in S_{-i}$ (it is well defined because each τ'_j is injective). It is easy to verify that ν_i is a CPS given that $f_i(t_i)$ is a CPS.[26] By ontoness

[26] A detailed argument for this under finiteness can be found in [7], in the proof of Lemma 1.

of g_i, there exists $t_i' \in T_i$ such that $g_i(t_i') = \nu_i$. Clearly, $f_i(t_i') = f_i(t_i)$, which implies $(s_i, t_i') \in R_i$, and $g_{i,h}(t_i')\,[\tau_j(s_j) \times \Omega_{-i,j}] = f_{i,h}(t_i)\,[s_j \times S_{-i,j}]$ for all $h \in H_i$, $j \neq i$, and $s_j \in \mathrm{Proj}_{S_j} R$.

Note that (trivially) $f_i(\mathrm{Proj}_{T_i} R_i) = \Delta^{H_i}(S_{-i})$.

So, I can apply Lemma 2 with $E = R$ and $\Delta_i^G = \Delta^{H_i}(S_{-i})$, so that $((S_{i,G}^n)_{i\in I})_{n=0}^{\infty}$ is Rationalizability. Thus, $CSB^{\infty}(R)$ is a closed Cartesian event with the completeness property where $f_i(\mathrm{Proj}_{T_i} CSB^{\infty}(R)) = \Delta_i^{\infty,G}$ for all $i \in I$. Then, it is easy to check that $E = [\Delta] \cap CSB^{\infty}(R)$ is a closed Cartesian event with the completeness property where $f_i(\mathrm{Proj}_{T_i} E) = \Delta_i \cap \Delta_i^{\infty,G}$ for all $i \in I$. ■

References

[1] R. Aumann (1987): *Correlated Equilibrium as an Expression of Bayesian Rationality*. Econometrica 55, doi:10.2307/1911154.

[2] P. Battigalli (1996): *Strategic Rationality Orderings and the Best Rationalization Principle*. Games and Economic Behavior 13(2), doi:10.1006/game.1996.0033.

[3] P. Battigalli (1997): *Dynamic Consistency and Imperfect Recall*. Games and Economic Behavior 20(1), doi:10.1006/game.1997.0535.

[4] P. Battigalli (1997): *On rationalizability in extensive games*. Journal of Economic Theory 13, doi:10.1006/jeth.1996.2252.

[5] P. Battigalli (2003): *Rationalizability in Infinite, Dynamic Games of Incomplete Information*. Research in Economics 57, doi:10.1016/S1090-9443(02)00054-6.

[6] P. Battigalli & A. Friedenberg (2012): *Forward induction reasoning revisited*. Theoretical Economics 7, doi:10.3982/TE598.

[7] P. Battigalli & A. Prestipino (2013): *Transparent Restrictions on Beliefs and Forward Induction Reasoning in Games with Asymmetric Information*. The B.E. Journal of Theoretical Economics 13(1), doi:10.1515/bejte-2012-0005.

[8] P. Battigalli & M. Siniscalchi (1999): *Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games*. Journal of Economic Theory 88, doi:10.1006/jeth.1999.2555.

[9] P. Battigalli & M. Siniscalchi (2002): *Strong Belief and Forward Induction Reasoning*. Journal of Economic Theory 106, doi:10.1006/jeth.2001.2942.

[10] P. Battigalli & M. Siniscalchi (2003): *Rationalization and Incomplete Information*. The B.E. Journal of Theoretical Economics 3(1), doi:10.2202/1534-5963.1073.

[11] A. Brandenburger & A. Friedenberg (2008): *Intrinsic correlation in games*. Journal of Economic Theory 141, doi:10.1016/j.jet.2007.09.012.

[12] E. Catonini (2017): *On non-monotonic strategic reasoning*. Working paper.

[13] E. Catonini (2017): *The overlap between extensive-form rationalizability and subgame perfection*. Working paper.

[14] E. Catonini (2017): *Self-enforcing agreements and forward induction reasoning*. Working paper.

[15] I.K. Cho & D. Kreps (1987): *Signaling Games and Stable Equilibria*. Quarterly Journal of Economics 102, doi:10.2307/1885060.

[16] A. Friedenberg (2010): *When Do Type Structures Contain All Hierarchies of Beliefs?* Games and Economic Behavior 68, doi:10.1016/j.geb.2009.05.005.

[17] D. Fudenberg & D. Levine (1993): *Self-confirming equilibrium*. Econometrica 61, doi:10.2307/2951716.

[18] S. Govindan & R. Wilson (2009): *On Forward Induction*. Econometrica 77, doi:10.3982/ECTA6956.

[19] A. Heifetz & A. Perea (2015): *On the Outcome Equivalence of Backward Induction and Extensive Form Rationalizability*. International Journal of Game Theory 44(1), doi:10.1007/s00182-014-0418-x.

[20] E. Kohlberg & J.F. Mertens (1986): *On the Strategic Stability of Equilibria*. Econometrica 54, doi:10.2307/1912320.

[21] P. Man (2012): *Forward Induction Equilibrium*. Games and Economic Behavior 75(1), doi:10.1016/j.geb.2011.12.004.

[22] M. Osborne (1990): *Signaling, Forward Induction, and Stability in Finitely Repeated Games*. Journal of Economic Theory 50, doi:10.1016/0022-0531(90)90083-V.

[23] D. Pearce (1984): *Rational Strategic Behavior and the Problem of Perfection*. Econometrica 52, doi:10.2307/1911197.

[24] A. Renyi (1955): *On a New Axiomatic Theory of Probability*. Acta Mathematica Academiae Scientiarum Hungaricae 6, doi:10.1007/BF02024393.

[25] M. Shimoji (2004): *On the Equivalence of Weak Dominance and Sequential Best Response*. Games and Economic Behavior 48, doi:10.1016/j.geb.2003.09.003.