Structural Rationality in Dynamic Games: ONLINE APPENDIX

Marciano Siniscalchi*

October 19, 2018

A Introduction

This Online Appendix contains supplemental material and elaborations upon the main results in the paper.

Section B provides an additional characterization of extensible CPSs via a strengthening of the chain rule.

Section C is devoted to the extensive form, and to results that depend upon its specifics. Subsection C.1 provides a the formal definition of game trees . Subsection C.2 proves Theorem 2 on the generic equivalence between structural and sequential rationality. Subsection

*Economics Department, Northwestern University, Evanston, IL 60208; marciano@northwestern.edu.
Earlier drafts were circulated with the titles 'Behavioral counterfactuals,' 'A revealed-preference theory of strategic counterfactuals,' 'A revealed-preference theory of sequential rationality,' and 'Sequential preferences and sequential rationality.' I thank Amanda Friedenberg, as well as Pierpaolo Battigalli, Gabriel Carroll, Drew Fudenberg, Alessandro Pavan, Phil Reny, and participants at RUD 2011, D-TEA 2013, and many seminar presentations for helpful comments on earlier drafts.

C.3 defines the extensive form of the elicitation game, of which Definition 9 is a reduced representation.

Section D analyzes two examples. Subsection D.1 provides a detailed analysis of the strategy-method elicitation example in Figure 6 of the paper. Subsection D.2 exemplifies one feature of Theorem 4 in the main text.

Section E explores alternative characterizations of structural preferences that use lexicographic preferences (§E.1 and E.2) or different representations of conditional beliefs (§E.3).

Finally, Section F collects a number of unsatisfactory definitions of preferences over strategies that, while apparently capturing certain intuitions about sequential rationality, they actually fail to formally imply it. Analyzing examples in which these unsatisfactory definitions fail provides another way to illustrate the features of structural preferences that link them to sequential rationality.

B Extensibility, Congruence, and Belief Trembles

Definition 1 *Fix a dynamic game* $\big(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot)\big)$*, a player* $i \in N$*, and a CPS* $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$*. The CPS* μ *is* ***congruent*** *if, for every ordered list* $I_1, \ldots, I_L \in \mathscr{I}_i$ *such that* $\mu([I_{\ell+1}|S_{-i}(I_\ell)) > 0$ *for all* $\ell = 1, \ldots, L-1$*, and all* $E \subseteq S_{-i}(I_1) \cap S_{-i}(I_L)$*,*

$$\mu(E|S_{-i}(I_1)) \cdot \prod_{\ell=1}^{L-1} \frac{\mu(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1})|S_{-i}(I_{\ell+1}])}{\mu([I_\ell) \cap S_{-i}(I_{\ell+1})|S_{-i}(I_\ell))} = \mu(E|S_{-i}(I_L))$$

Congruence is a strengthening of the chain rule of conditioning.[1] Furthermore, it sheds light on the pathological nature of the beliefs in Example 4. Take $I_1 = I$ and $I_2 = J$ in Definition 1, and note that $S_{-i}(I) \cap S_{-i}(J) = \{b, c\}$, $\mu(S_{-i}(I) \cap S_{-i}(J)||S_{-i}(I)) > 0$, and $\mu(S_{-i}(I) \cap S_{-i}(J)|S_{-i}(J)) > 0$. Then the equation in Definition 1 implies that, in particular,

$$\frac{\mu(\{b\}|S_{-i}(I))}{\mu(\{b, c\}|S_{-i}(I))} = \frac{\mu(\{b\}|S_{-i}(J))}{\mu(\{b, c\}|S_{-i}(J))}.$$

[1] To see that it implies the chain rule, take $L = 2$ and consider the case $E \subseteq S_{-i}(I_1) \subseteq S_{-i}(I_2)$ in Definition 1.

Intuitively, the probability of b given $\{b,c\}$ shoud be the same, whether it is calculated from the perspective of I or J. This is a reasonable requirement, given that Ann's information about the relative likelihood of b vs. c is the same at I and J—neither has yet been ruled out. Yet this condition is violated in Example 4: $\mu(\{b\}|S_{-i}(I))=1$, but $\mu(\{b\}|S_{-i}(J))=0$.

The following result complements Theorem 3 by showing that a CPS is extensible if and only if it is congruent.

Theorem 1 *Fix a dynamic game $\big(N,(S_i,\mathscr{I}_i,U_i)_{i\in N},S(\cdot)\big)$, a player $i\in N$, and a CPS $\mu\in\Delta(S_{-i},S_{-i}(\mathscr{I}_i))$. Then μ admits a unique extension $\nu\in\Delta(S_{-i},\mathscr{S}_{-i}(\mathscr{I}_i;\mu))$ if and only if it is congruent.*

The following Proposition contains the key measure-theoretic step.

Proposition 1 *Fix a non-empty collection $\mathscr{C}_i\subseteq 2^{S_{-i}}\setminus\{\emptyset\}$ and a CPS $\mu\in\Delta(S_{-i},\mathscr{C}_i)$ for player $i\in N$. The following are equivalent:*

1. *μ is congruent;*
2. *for every μ-sequence $F_1,\dots,F_K\in\mathscr{C}_i$, there exists $p\in\Delta(S_{-i})$ with $p(\cup_k F_k)=1$, such that, for every $\ell=1,\dots,K$ and $E\subseteq F_\ell$,*

$$p(E)=\mu(E|F_\ell)p(F_\ell). \tag{1}$$

If a probability p that satisfies the property in (2) exists, it is unique; furthermore, $p(F_K)>0$, and for all $\ell=1,\dots,K-1$, $p(F_\ell)>0$ iff $\mu(F_k|F_{k+1})>0$ for all $k=\ell+1,\dots,K$.

Note that, in part 2, the μ-sequence $F_1,\dots,F_K$ μ-supports p.

Proof: (1)$\Rightarrow$(2): assume that μ is congruent. Let $F_1,\dots,F_K\in\mathscr{C}_i$ be a μ-sequence.

Define $G_1=F_1$ and, inductively, $G_k=F_k\setminus(F_1\cup\ldots\cup F_{k-1})$ for $k=2,\dots,K$. Note that $F_1\cup\ldots\cup F_k=G_1\cup\ldots G_k$ for all $k=1,\dots,K$, [for $k=1$ this is by definition. By induction, $G_1\cup\ldots\cup G_{k+1}=(G_1\cup\ldots G_k)\cup G_{k+1}=(F_1\cup\ldots\cup F_k)\cup G_{k+1}=(F_1\cup\ldots\cup F_k)\cup[F_{k+1}\setminus(F_1\cup\ldots\cup F_k)]=F_1\cup\ldots\cup F_{k+1}]$ and $G_k\cap G_\ell=\emptyset$ for all $k\neq\ell$. [Let $\ell>k$: then $G_\ell=F_\ell\setminus(F_1\cup\ldots F_{\ell-1})=F_\ell\setminus(G_1\cup\ldots G_{\ell-1})$, and $k\in\{1,\dots,\ell-1\}$.] Also, $G_k\subseteq F_k$ for all $k=1,\dots,K$.

I now define a set function $\rho : 2^{S_{-i}} \to \mathbb{R}$. For every $\ell = 1, \ldots, K$ and $E \subseteq S_{-i}$ with $E \subseteq G_\ell$, let

$$\rho(E) \equiv \mu(E|F_\ell) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)},$$

with the usual convention that the product over an empty set of indices equals 1. By assumption, the denominators of the above fractions are all strictly positive. Also, since the sets $G_1, \ldots, G_k$ are disjoint by construction, if $\emptyset \neq E \subseteq G_\ell$ for some ℓ then $E \nsubseteq G_k$ for $k \neq \ell$, so $\rho(E)$ is uniquely defined; furthermore, $\emptyset \subseteq G_k$ for all k, but $\rho(\emptyset)$ is stil well-defined and equal to 0.

To complete the definition of $\rho(\cdot)$, for all events $E \subseteq S_{-i}$ such that $E \nsubseteq G_k$ for $k = 1, \ldots, K$ [i.e., E intersects two or more events G_k, or none], let

$$\rho(E) = \sum_{k=1}^{K} \rho(E \cap G_k).$$

The function $\rho(\cdot)$ thus defined takes non-negative values. I claim that $\rho(\cdot)$ is additive. Consider an ordered list $E_1, \ldots, F_M \subseteq S_{-i}$ such that $E_m \cap E_{\bar{m}} = \emptyset$ for $m \neq \bar{m}$. If there is $\ell \in \{1, \ldots, K\}$ such that $E_m \subseteq G_\ell$ for all m, then by additivity of $\mu(\cdot|F_\ell)$,

$$\rho\left(\bigcup_m E_m\right) = \mu\left(\bigcup_m E_m \middle| F_k\right) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \left(\sum_m \mu(E_m|F_k)\right) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} =$$

$$= \sum_m \left(\mu(E_m|F_\ell) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)}\right) = \sum_m \rho(E_m).$$

Thus, for a general ordered list $E_1, \ldots, F_M \subseteq S_{-i}$ of pairwise disjoint events,[2]

$$\rho\left(\bigcup_m E_m\right) = \sum_k \rho\left(\left[\bigcup_m E_m\right] \cap G_k\right) = \sum_k \rho\left(\bigcup_m [E_m \cap G_k]\right) =$$

$$= \sum_k \sum_m \rho(E_m \cap G_k) = \sum_m \sum_k \rho(E_m \cap G_k) = \sum_m \rho(E_m).$$

[2]For future reference, if the set S_{-i} is an arbitrary measurable space, and probabilities in i's CPS are countably additive, this step of the proof still holds, and shows that ρ is countably additive. Specifically, the derivation holds as written for a countable collection $F_1, F_2, \ldots$ (I purpoosely omitted limits from the summations). In particular, interchanging the order of the summation in the second line is allowed because all summands are non-negative and the derivation shows that $\sum_k \sum_m \rho(E_m \cap G_k) = \sum_k \rho([\cup_m E_m] \cap G_k)$, a sum of finitely many finite terms.

Now consider $E \subseteq S_{-i}$ with $E \subseteq F_m$ and $E \subseteq G_\ell$ for some $\ell, m \in \{1, \ldots, K\}$ with $\ell \neq m$. Since $F_m \subseteq F_1 \cup \ldots \cup F_m = G_1 \cup \ldots \cup G_m$, $\ell < m$. Consider the ordered list $F_\ell, \ldots, F_m \in \mathscr{C}_i$: since $F_1, \ldots, F_K$ is a μ-sequence, so is $F_\ell, \ldots, F_m$, so by congruence, since by assumption $E \subseteq F_m \cap G_\ell \subseteq F_m \cap F_\ell$,

$$\mu(E|F_\ell) \prod_{k=\ell}^{m-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \mu(E|F_m).$$

Multiply both sides by the positive quantity $\prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_n)}$ to get

$$\rho(E) = \mu(E|F_\ell) \prod_{k=\ell}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \mu(E|F_m) \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)}.$$

Therefore, for all $E \subseteq S_{-i}$ with $E \subseteq F_m$ for some $m \in \{1, \ldots, K\}$,

$$\mu(E|F_m) \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \sum_{\ell=1}^{K} \mu(E \cap G_\ell|F_m) \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} =$$
$$= \sum_{\ell=1}^{K} \rho(E \cap G_\ell) = \rho(E).$$

It follows that, for all $m \in \{1, \ldots, K\}$ and $E \subseteq S_{-i}$ with $E \subseteq F_m$,

$$\rho(F_m) = \mu(F_m|F_m) \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)}, \tag{2}$$

and therefore

$$\rho(E) = \mu(E|F_m)\rho(F_m). \tag{3}$$

Finally, notice that $\rho(\cup_k G_k) = \rho(\cup_k F_k) \geq \rho(F_K) = 1$; thus, one can define a probability measure $p \in \Delta(S_{-i})$ by letting

$$\forall E \subseteq S_{-i}, \qquad p(E) = \frac{\rho(E)}{\rho(\cup_k G_k)} = \frac{\rho(E)}{\rho(\cup_k F_k)}.$$

For every $\ell \in \{1, \ldots, K\}$ and every event $E \subseteq F_\ell$, p satisfies Eq. (1), as asserted.

To show that p is uniquely defined, let $q \in \Delta(S_{-i})$ be a measure that satisfies Eq.(1). I first claim that, for every $m = 1, \ldots, K$,

$$q(F_m) = \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} \cdot q(F_N) = \rho(F_m)q(F_K).$$

The claim is trivially true for $m = K$, so consider $m \in \{1, \ldots, K-1\}$ and assume that the claim holds for $m+1$. By Eq.(1),

$$\mu(F_m \cap F_{m+1}|F_{m+1})q(F_{m+1}) = q(F_m \cap F_{m+1}) = \mu(F_m \cap F_{m+1}|F_m)q(F_m);$$

since $\mu(F_m \cap F_{m+1}|F_m) > 0$ by assumption, solving for $q(F_m)$ and invoking the inductive hypothesis yields

$$q(F_m) = \frac{\mu(F_m \cap F_{m+1}|F_{m+1})}{\mu(F_m \cap F_{m+1}|F_m)} q(F_{m+1}) = \frac{\mu(F_m \cap F_{m+1}|F_{m+1})}{\mu(F_m \cap F_{m+1}|F_m)} \cdot \prod_{k=m+1}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} \cdot q(F_K) = \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} \cdot q(F_K).$$

Since $G_m \subseteq F_m$, Eq. (1) implies that

$$q(G_m) = \mu(G_m|F_m)q(F_m) = \mu(G_m|F_m) \cdot \rho(F_m) \cdot q(F_K) = \rho(G_m) \cdot q(F_K),$$

where the last equality follows from Eq.(3). Since $\sum_k q(G_k) = q(\cup_k G_k) = q(\cup_k F_k)$, if in addition q satisfies $q(\cup_k F_k) = 1$, then

$$1 = \sum_m \rho(G_m) \cdot q(F_K) = q(F_K)\rho(\cup_m G_m)$$

which implies that $q(F_K) > 0$, and indeed that

$$q(F_K) = \frac{1}{\rho(\cup_m G_m)} = \frac{\rho(F_K)}{\rho(\cup_m G_m)} = p(F_K).$$

so also $p(F_K) > 0$, as claimed. Furthermore, for $m = 1, \ldots, K-1$,

$$q(F_m) = \rho(F_m)q(F_N) = \rho(F_m)\frac{1}{\rho(\cup_m G_m)} = p(F_m).$$

Furthermore, let $k_0 \in \{1, \ldots, K-1\}$ be such that $\mu(F_k \cap F_{k+1}|F_{k+1}) > 0$ for all $k > k_0$, and $\mu(F_{k_0} \cap F_{k_0+1}|F_{k_0+1}) = 0$. By inspecting Eq. (2), it is clear that $\rho(F_k) = 0$ for $k = 1, \ldots, k_0$, and $\rho(F_k) > 0$ for $k = k_0+1, \ldots, K$. Then, $p(F_k) = 0$ for $k = 1, \ldots, k_0$, and $p(F_k) > 0$ for $k = k_0+1, \ldots, K$. From the above argument, it follows that the same is true for any $q \in \Delta(S_{-i})$ that satisfies Eq. (1) and $q(\cup_k F_k) = 1$. Thus, the last claim of the Proposition follows.

Finally, if $q \in \Delta(S_{-i})$ satisfies Eq.(1) and $q(\cup_k F_k) = 1$, for every $k = k_0+1, \ldots, K$ and $E \subseteq S_{-i}$ such that $E \subset F_k$,

$$q(E) = \mu(E|F_k)q(F_k) = \mu(E|F_k)p(F_k) = p(E)$$

and therefore, for every $E \subseteq S_{-i}$,

$$q(E) = \sum_k q(E \cap G_k) = \sum_{k=k_0+1}^{K} q(E \cap G_k) = \sum_{k=k_0+1}^{K} p(E \cap G_k) = \sum_k p(E \cap G_k) = p(E).$$

In other words, p is the unique probability measure that satisfies Eq. (1) and $p(\cup_k F_k) = 1$.

(2)⇒ (1): assume that (2) holds. Consider a μ-sequence $F_1, \ldots, F_K$. Fix an event $E \subseteq F_1 \cap F_K$. By assumption, there exists $p \in \Delta(S_{-i})$ that satisfies Eq. (1) for $k = 1, \ldots, K$, with $p(\cup_{k=1}^K F_k) = 1$.

Since $p(F_K) > 0$, $\mu(E|F_K) = \frac{p(E)}{p(F_K)}$. If $p(F_1) = 0$, then a fortiori $p(E) = 0$, so $\mu(E|F_K) = 0$; on the other hand, $p(F_1) = 0$ implies that there is $k = 1, \ldots, K-1$ such that $\mu(F_k \cap F_{k+1}|F_{k+1}) = 0$, so

$$\mu(E|F_1) \cdot \prod_{k=1}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \mu(E|F_1) \cdot 0 = 0 = \mu(E|F_K).$$

If instead $p(F_1) > 0$, then $\mu(E|F_1) = \frac{p(E)}{p(F_1)}$; furthermore, by the above argument $p(F_k) > 0$ for all $k = 2, \ldots, K-1$ as well, so

$$\mu(E|F_1) \cdot \prod_{k=1}^{K-1} \frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \frac{p(E)}{p(F_1)} \cdot \prod_{k=1}^{K-1} \frac{p(F_k \cap F_{k+1})}{p(F_{k+1})} \cdot \frac{p(F_k)}{p(F_k \cap F_{k+1})} = \frac{p(E)}{p(F_1)} \cdot \frac{p(F_1)}{p(F_K)} = \frac{p(E)}{p(F_K)} = \mu(E|F_K).$$

■

Corollary 1 *If μ is congruent, then for every μ-sequence $F_1, \ldots, F_K$ such that $\mu(F_1|F_K) > 0$, the reverse-ordered list $F_K, F_{K-1}, \ldots, F_1$ is also a μ-sequence: that is, $\mu(F_k|F_{k+1}) > 0$ for all $k = 1, \ldots, K-1$.*

In particular, this Corollary applies if $F_1 = F_K$.

Proof: Let $F_1, \ldots, F_K$ be as in the statement, and consider the ordered list $F_1, \ldots, F_K, F_{K+1}$ with $F_{k+1} = F_1$. Then $F_1, \ldots, F_{K+1}$ is also a μ-sequence. Let p be the unique measure in (2) of Proposition 1. The last claim of that Proposition shows that necessarily $p(F_{K+1}) > 0$, but since $F_{K+1} =$

F_1, also $p(F_1) > 0$. Again, the last claim in Proposition 1 implies that then $p(F_k) > 0$ for all $k = 1, \ldots, K$.

Then, for all $k = 1, \ldots, K-1$, $\mu(F_k \cap F_{k+1}|F_k) > 0$ implies that $p(F_k \cap F_{k+1}) > 0$, and so

$$\mu(F_k|F_{k+1}) = \mu(F_k \cap F_{k+1}|F_{k+1}) = \frac{p(F_k \cap F_{k+1})}{p(F_{k+1})} > 0.$$

■

Corollary 2 *Let $G_1, \ldots, G_N$ be a μ-sequence and p the measure in (2) of Proposition 1; consider $F \in S_{-i}(\mathscr{I}_i)$ such that $F \subset \cup_{k=1}^K G_k$. Then, for every $E \subseteq F$, $p(E) = \mu(E|F)p(F)$.*

Proof: It is enough to consider the case $p(F) > 0$.

Let $k \in \{1, \ldots, K\}$ be such that $p(G_k) > 0$ and $\mu(F|G_k) = \mu(F \cap G_k|G_k) > 0$. One such k must exist, because $p(F) > 0$ implies $p(F \cap G_m) > 0$ for some $m \in \{1, \ldots, K\}$, and by construction $p(F \cap G_m) = p(G_m)\mu(F \cap G_m|G_m)$.

I claim that, for any such k, $\mu(G_k|F) > 0$. Since $F \subseteq \cup_m G_m$ and $\mu(F|F) = 1$, $\mu(G_m|F) > 0$ for at least one $m \in \{1, \ldots, K\}$. If $m = k$, the claim is true. If $m < k$, then the ordered list $F, G_m, G_{m+1}, \ldots, G_k, F$ is a μ-sequence that satisfies the conditions of Corollary 1, so that in particular $\mu(G_k|F) > 0$, as claimed. Finally, suppose $m > k$. Since $p(G_k) > 0$, by the last claim of Proposition 1, $\mu(G_\ell|G_{\ell+1}) > 0$ for $\ell = k, \ldots, K-1$. Hence, since $\mu(G_m|F) > 0$, the ordered list $F, G_m, G_{m-1}, \ldots, G_{k+1}, G_k, F$ is a μ-sequence that satisfies the conditions in Corollary 1, so in particular $\mu(G_k|F) > 0$, as claimed.

This implies that the ordered list $G_1, \ldots, G_k, F, G_k, \ldots, G_K$ is a μ-sequence. Let p' be the measure delivered by Proposition 1 for this μ-sequence. Notice that $p(F \cup \bigcup G_k) = p'(F \cup \bigcup_k G_k) = 1$, and for all $\ell \in \{1, \ldots, K\}$ and $E \subseteq S_{-i}$ with $E \subset G_\ell$, $p'(E) = p'(G_\ell)\mu(E|G_\ell)$. Since p is the unique probability with these properties, $p = p'$. But then, for $E \subseteq S_{-i}$ with $E \subseteq F$,

$$p(E) = p'(E) = p'(F)\mu(E|F) = p(F)\mu(E|F),$$

as claimed. ■

Corollary 3 *Let $G_1,\dots,G_K$ and $F_1,\dots,F_M$ be μ-sequences with $\cup_m F_m \subseteq \cup_k G_k$. Let p and q be the probabilities associated with $G_1,\dots,G_K$ and $F_1,\dots,F_M$ respectively. Consider $E \subseteq \cup_m F_m$. Then $p(E) = p(\cup_m F_m)q(E)$.*

Proof: It is enough to consider the case $p(\cup_m F_m) > 0$.

Since, for every m, $F_m \subseteq \cup_k G_k$, Corollary 2 implies that, for every $E' \subseteq S_{-i}$ with $E' \subseteq F_m$,

$$p(E') = \mu(E'|F_m)p(F_m).$$

Hence, the measure $p' \in \Delta(S_{-i})$ defined by $p'(E) = p(E \cap \cup_m F_m)/p(\cup_m F_m)$ satisfies

$$\forall E' \subseteq S_i, E' \subseteq F_m, \quad p(E') = \mu(E'|F_m)p'(F_m) \quad \text{and} \quad p'(\cup_m F_m) = 1.$$

Therefore, $p' = q$, or $p(E') = p(\cup_m F_m)q(E')$ for every m and $E' \subseteq S_{-i}$ with $E' \subseteq F_m$. In particular, let $\bar{F}_1 = F_1$ and, for $m = 2,\dots,M$, let $\bar{F}_m = F_m \setminus (F_1 \cup \ldots \cup F_{m-1})$. Then, for every m,

$$p(E \cap \bar{F}_m) = p(\cup_\ell F_\ell)q(E \cap \bar{F}_m)$$

and so, since $\bar{F}_1,\dots,\bar{F}_M$ is a partition of $\cup_m F_m$ and $E \subseteq \cup_m F_m$, summing over all m yields $p(E) = p(\cup_m F_m)q(E)$, as required. ■

Finally, I prove Theorem 1.

Proof: Assume that μ is congruent. Let $\{F_1,\dots,F_L\}$ be a $\geq^\mu$-equivalence class. By Lemma 1, there is a μ-sequence $G_1,\dots,G_M$, with $G_1 = G_M = F_1$. By Proposition 1, there is a unique $p \in \Delta(S_{-i})$ such that $G_1,\dots,G_M$ *μ-supports* p—that is, $p(\cup_m G_m) = 1$ and $p(E) = \mu(E|G_m)p(G_m)$ for all m and all $E \subseteq G_m$. Furthermore, there is $\bar{m} \in \{1,\dots,M\}$ such that $m \geq \bar{m}$ if and only if $p(G_m) > 0$; since surely $p(G_M) > 0$ and $G_1 = G_M = F_1$, $\bar{m} = 1$. Therefore $p(G_m) > 0$ for all m, so $p(F_\ell) > 0$ for all ℓ. Finally, suppose that $\bar{F}_1,\dots,\bar{F}_{\bar{L}}$ is another $\geq^\mu$–equivalence class, with $[F_1]_\mu =$

$[\bar{F}_1]_\mu$. Again, by Lemma 1, there is a μ-sequence $\bar{G}_1,\ldots,\bar{G}_{\bar{M}}$ with $\{\bar{G}_1,\ldots,\bar{G}_{\bar{M}}\}=\{\bar{F}_1,\ldots,\bar{F}_{\bar{L}}\}$, and Corollary 3 implies that $\bar{F}_1,\ldots,\bar{F}_{\bar{L}}$ μ-supports the same probability p. Therefore, one can define an array $\nu\in\Delta(S_{-i})^{S_{-i}(\mathscr{I}_i)\cup B_\mu(i)}$ by letting $\nu(\cdot|F)=\mu(\cdot|F)$ for $F\in S_{-i}(\mathscr{I}_i)$, and $\nu(\cdot|[F]_\mu)=p$, where p is the unique probability that μ-support the $\geq^\mu$-equivalence class containing $F\in S_{-i}(\mathscr{I}_i)$.

It remains to be shown that ν is a CPS. Its construction implies that $\nu(G|G)=1$ for all $G\in S_{-i}(\mathscr{I}_i)\cup B_\mu(i)$. To show that the chain rule holds, consider $F,G\in S_{-i}(\mathscr{I}_i)\cup B_\mu(i)$ with $F\supseteq G$, and $E\subseteq F$. If $F,G\in S_{-i}(\mathscr{I}_i)$, then the conclusion follows from the fact that $\nu(\cdot|F)=\mu(\cdot|F)$ and $\nu(\cdot|G)=\mu(\cdot|G)$, because μ is a CPS. If $F=B_\mu(I)$ and $G=B_\mu(J)$, and $F\neq G$ (otherwise the conclusion is immediate), then I claim that $\nu(B_\mu(I)|B_\mu(J))=0$, which again that the chain rule holds. To see this, let $F_1,\ldots,F_L$ and $G_1,\ldots,G_M$ be the $\geq^\mu$-equivalence classes containing $S_{-i}(I)$ and $S_{-i}(J)$ respectively. If $\nu(\cup_\ell F_\ell|\cup_m G_m)>0$, then Lemma 2 implies that $\mu(F_{\bar{\ell}}|G_{\bar{m}})>0$ for some $\bar{\ell},\bar{m}$, so $F_{\bar{\ell}}\geq^\mu G_{\bar{m}}$ and thus, by transitivity, $S_{-i}(I)\geq^\mu S_{-i}(J)$. Since $\nu(\cup_m G_m|\cup_\ell F_\ell)\geq \nu(\cup_\ell F_\ell|\cup_\ell F_\ell)=1$, a symmetric argument shows that $S_{-i}(J)\geq^\mu S_{-i}(I)$, so $S_{-i}(I)=^\mu S_{-i}(J)$ and thus $F=B_\mu(I)=B_\mu(J)=G$, contradiction. This proves the claim.

Finally, consider the case of $F=S_{-i}(I)$ and $G=B_\mu(J)$. Lemma 2 implies that there is $\bar{m}$ with $\mu(F|G_{\bar{m}})>0$, so $F\geq^\mu G_{\bar{m}}$ and thus $F\geq^\mu S_{-i}(J)$. Since $\nu(G|F)\geq\nu(F|F)=1$, the same Lemma implies that there is G' with $G"=^\mu G$ and $\mu(G'|F)>0$, so $G'\geq^\mu F$ and $S_{-i}(J)\geq^\mu F$. Therefore, F is an element of the μ-equivalence class for $S_{-i}(J)$. But then, since this collection of events supports $\nu(\cdot|G)$, $\nu(E|G)=\mu(E|F)\nu(F|G)=\nu(E|F)\nu(F|G)$, as required.

Conversely, assume that μ admits an extension. Then, by Theorem 3, μ admits a structural perturbation $(p^n)\subset\Delta(S_{-i})$. Consider a μ-sequence $F_1,\ldots,F_L$ and an event $E\subseteq F_1\cap F_L$. Since $\mu(F_{\ell+1}|F_\ell)>0$ for all $\ell=1,\ldots,L-1$, there is $\bar{n}$ such that $n\geq\bar{n}$ implies $p^n(F_{\ell+1}\cap F_\ell)/p(F_\ell)>0$. For every such n and event $E\subseteq F_1\cap F_L$,

$$\frac{p^n(E)}{p^n(F_1)}\cdot\prod_{\ell=1}^{L-1}\frac{\frac{p^n(F_\ell\cap F_{\ell+1})}{p^n(F_{\ell+1})}}{\frac{p^n(F_\ell\cap F_{\ell+1})}{p^n(F_\ell)}}=\frac{p^n(E)}{p^n(F_1)}\cdot\prod_{\ell=1}^{L-1}\frac{p^n(F_\ell)}{p^n(F_{\ell+1})}=\frac{p^n(E)}{p^n(F_L)}$$

Since $p^n(E)/p^n(F_1)\to\mu(E|F_1)$, $p^n(F_\ell\cap F_{\ell+1})/p^n(F_{\ell+1})\to\mu(F_\ell\cap F_{\ell+1}|F_{\ell+1})$, $p^n(F_\ell\cap F_{\ell+1})/p^n(F_\ell)\to\mu(F_\ell\cap F_{\ell+1}|F_\ell)>0$, and $p^n(E)/p^n(F_L)\to\mu(E|F_L)$, it follows that Congruence holds. ■

C Game Trees and Generic Equivalence Theorem

I first provide a full, but concise description of game trees and extensive-form games (without chance nodes), using the notation in Osborne and Rubinstein (1994, Def. 200.1, pp-200-201). Additional notation and results can be found in Online Appendix C.1. I then formally state the generic equivalence result described in Section 5.1.

A **game tree** is a tuple $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i\in N})$; N is the set of *players*, A is a set of *actions*, and H is a finite collection of *histories*, i.e., finite sequences $(a_1, \ldots, a_n)$ of actions, which contains the empty sequence ϕ. For every history $h = (a_1, \ldots, a_L) \in H$, $A(h) \equiv \{a \in A : (a_1, \ldots, a_L, a) \in H\}$ is the set of actions available at h. A history $h \in H$ is *terminal* if $A(h) = \emptyset$; denote the set of terminal histories by Z.

$P : H \setminus Z \to N$ is the *player function*, which associates with each non-terminal history $h \in H \setminus Z$ the player on the move at h. Each $\mathscr{I}_i$ consists of a partition of $P^{-1}(i)$, plus the symbol ϕ, which corresponds to the beginning of the game (as explained in Section 2, this ensures that every player's CPS includes his prior beliefs). The elements of $\mathscr{I}_i$ are player i's *information sets*. For every $i \in N$, $I \in \mathscr{I}_i \setminus \{\phi\}$, and $h, h' \in I$, player i must have the same moves available at both h and h': that is, $A(h) = A(h')$.

The game form is assumed to have **perfect recall**, as per Def. 203.3 in OR. Briefly, for every $h \in P^{-1}(i)$, let $X_i(h)$ denote i's *experience* along the history h: that is, the ordered list of all information sets owned by i that i encountered along the history h, and the actions she played there.[3] Perfect recall is the requirement that, if $h, h' \in I \in \mathscr{I}_i \setminus \{\phi\}$, then $X_i(h) = X_i(h')$.

An **extensive-form game** is an game tree together with **payoff assignments** $u_i : Z \to \mathbb{R}$ for every player $i \in N$.

[3] Formally, if $h = (a_1, \ldots, a_L)$, let $\ell_1, \ldots, \ell_K$ be the set of indices $\ell \in \{1, \ldots, L-1\}$ such that $P((a_1, \ldots, a_{\ell-1})) = i$; let $I_1, \ldots, I_K$ be such that $(a_1, \ldots, a_{\ell_k - 1}) \in I_k$ for $k = 1, \ldots, K$. Then $X_i(h) = (I_1, a_{\ell_1}, \ldots, I_k, a_{\ell_k})$.

The strategic-form objects in Section 2 can be derived from the game form and the payoff assignments, as follows. For every player $i \in N$, a *strategy* is a map $s_i : H \setminus Z \to A$ such that $s_i(h) \in A(h)$ for all $h \in H \setminus Z$, and $s_i(h) = s_i(h')$ for all $h, h' \in I \in \mathscr{I}_i \setminus \{\phi\}$. S_i is the set of strategies for player $i \in N$, and as in the main text, the usual conventions for product sets apply. For every $s \in S$, $\zeta(s)$ is the terminal history induced by s.[4] The set of *strategy profiles reaching* $I \in \mathscr{I}_i \setminus \{\phi\}$ is $S(I) = \{s \in S : \zeta(s) = (a_1, \ldots, a_L), \exists \ell < L : (a_1, \ldots, a_\ell) \in I\}$; that is, $s \in S(I)$ if some initial segment of $\zeta(s)$ belongs to I. By convention, $S(\phi) = S$. Finally, for every $i \in N$, the *strategic-form payoff function* U_i is defined by letting $U_i(s) = u_i(\zeta(s))$ for every $s \in S$.

Under the above assumptions, $S(\cdot)$ and $U_i(\cdot)$ satisfy the properties in Section 2: see Online Appendix C.1.

For a fixed game tree $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i \in N})$ and player $i \in N$, a payoff assignment can be identified with a point in $\mathbb{R}^Z$. The generic equivalence theorem states that, except for a lower-dimensional set of payoff assignments, sequential rationality coincides with optimality with respect to structural preferences. Note that this is a stronger claim than is stated in the main text: it implies that, generically, there is no difference between optimality and maximality with respect to structural preferences.

For a given $u_i \in \mathbb{R}^Z$, the notions of sequential rationality and structural preferences are defined in the obvious way—by replacing $U_i(s_i, s_{-i})$ and $U_i(t_i, s_{-i})$ in Definitions 3 and 8 with $u_i(\zeta(s_i, s_{-i}))$ and $u_i(\zeta(t_i, s_{-i}))$ respectively. The structural preference determined by μ and u_i is denoted by $\succcurlyeq^{\mu, u_i}$. Finally, for a given CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i)$ for player i, let $V_i(\mu)$ denote the set of $u_i \in \mathbb{R}^Z$ for which there exists at least one strategy $s_i \in S_i$ that is sequentially rational for μ, but such that, for some $t_i \in S_i$, not $s_i \succcurlyeq^{\mu, u_i} t_i$ (thus, s_i is not optimal for $\succcurlyeq^{\mu, u_i}$).

Theorem 2 *Fix a game tree* $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i \in N})$*, a player* $i \in N$*, and an extensible CPS* $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$*. Then* $V_i(\mu)$ *has dimension strictly less than* $|Z|$*.*

[4] Formally, $\zeta(s) = (a_1, \ldots, a_L)$, where $a_1 = s_{P(\phi)}(\phi)$ and, inductively, $a_{\ell+1} = s_{P((a_1, \ldots, a_\ell))}((a_1, \ldots, a_\ell))$.

C.1 Preliminaries on extensive-form games

The proof of Theorem 2 requires certain additional definitions and facts about extensive games. Fix a game form $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i \in N})$.

It is convenient to define the concatenation of histories, and of histories and actions. If $h = (a_1, \ldots, a_L) \in A^L$ and $(b_1, \ldots, b_M) \in A^M$, then $(h, h') = (h, b_1, \ldots, b_M) = (a_1, \ldots, a_L, h') = (a_1, \ldots, a_L, b_1, \ldots, b_M)$.

Histories are ordered by the "initial segment" relation: $h < h'$ means that $h' = (h, b_1, \ldots, b_M)$ for some $b_1, \ldots, b_M \in A$; $h = \phi$ is a subhistory of all histories, and $h \leq h'$ means that either h and h' are the same sequence, or $h < h'$.

Information sets are also ordered by precedence: $I < I'$ iff for every $h' \in I'$ there is $h \in I$ with $h < h'$. The notation $I \leq I'$ means that either $I = I'$ or $I < I'$. For players i for which $\phi = \{\phi\}$ is not a partition cell, $\phi < I$ for all $I \in \mathscr{I}_i$.

Fix $I \in \mathscr{I}_i \setminus \{\phi\}$. Since $A(h) = A(h')$ for all $h \in I$, we can abuse notation slightly and write $A(I)$ to indicate $A(h)$ for any $h \in I$. Similarly, write $P(I)$ to indicate $P(h)$ for any $h \in I$.

Since a strategy $s_i : H \setminus Z \to A$ for $i \in N$ must satisfy $s_i(h) = s_i(h')$ for all $h, h' \in I \in \mathscr{I}_i \setminus \{\phi\}$, s_i can also be viewed as a map from $\mathscr{I}_i \setminus \{\phi\}$ to A.

It is convenient to define the set of strategy profiles reaching a history. For every $h \in H$ (terminal or non-terminal), $S(h) = \{s \in S : h \leq \zeta(s)\}$. In particular, if z is terminal, then $S(z) = \{s \in S : z = \zeta(s)\}$, because by definition a terminal history is not a subhistory of any other history. Notice that $s \in S(h)$ if there exists $z \in Z$ such that $h < z$ and $s \in S(h)$; furthermore, for every player $i \in N$ and $I \in \mathscr{I}_i \setminus \{\phi\}$, $S(I) = \bigcup_{h \in I} S(h)$.

It is also useful to define *player i's information sets $\mathscr{I}_i(s_i)$ allowed by strategy s_i*: that is, for every $I \in \mathscr{I}_i$, $I \in \mathscr{I}_i(s_i)$ if and only if $s_i \in S_i(I)$.

I now verify that the properties of $S(\cdot)$ assumed in Section 2 do hold under perfect recall. In addition, Properties (ii) and (iv) are used in the proof of Theorem 2.

Remark 1 *If $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i \in N})$ has perfect recall, then* *(i) for every $I, J \in \mathscr{I}_i$, $S(I) \cap S(J) \neq \emptyset$ implies that $S(I)$ and $S(J)$ are nested;*

(ii) for every $I, J \in \mathscr{I}_i \setminus \{\phi\}$ and $s_i, t_i \in S_i(I)$, if $J < I$ then $s_i(J) = t_i(J)$;

(iii) for every $I \in \mathscr{I}_i$, $S(I) = S_i(I) \times S_{-i}(I)$.

(iv) for all $z \in Z$, $S(z) = \prod_{j \in N} S_j(z)$.

Proof: (i) Suppose there are $r \in S(I) \cap S(J)$, $s \in S(I) \setminus S(J)$, and $t \in S(J) \setminus S(I)$. In particular, this implies that $I \neq J$. By definition, there are $h_r \in I$ and $h'_r \in J$ such that $h_r < \zeta(r)$ and $h'_r < \zeta(r)$. Since $I \neq J$, $h_r \neq h'_r$, so either $h_r < h'_r$ or $h'_r < h_r$. Suppose $h_r < h'_r$; then, by definition, $X_i(h'_r)$ contains I. Now let h' be such that $h' < \zeta(t)$ and $h' \in J$, which exists because $t \in S(J)$. Since $t \notin S(I)$, $X_i(h')$ does not contain I. But then $X_i(h'_r) \neq X_i(h')$, which contradicts perfect recall. Suppose instead $h'_r < h_r$; then $X_i(h_r)$ contains J. Let h be such that $h < \zeta(s)$ and $h \in I$, which exists because $s \in S(I)$. Since $s \notin S(J)$, $X_i(h)$ does not contain J. But then $X_i(h_r) \neq X_i(h)$, which again contradicts perfect recall.

(ii) Let $s_{-i}, t_{-i} \in S_{-i}$ be such that $s \equiv (s_i, s_{-i}), t \equiv (t_i, t_{-i}) \in S(I)$. By definition, there are $h, h' \in I$ such that $h < \zeta(s)$ and $h' < \zeta(t)$. Suppose that $J < I$, so by definition there are $\tilde{h}, \tilde{h}' \in J$ with $\tilde{h} < h$ and $\tilde{h}' < h'$. This implies that J and $s_i(J)$, and J and $t_i(J)$ respectively, are elements of $X_i(h)$ and $X_i(h')$ respectively. But then, by perfect recall, $s_i(J) = t_i(J)$.

(iii) Clearly, $S(I) \subseteq S_i(I) \times S_{-i}(I)$. For the converse inclusion, fix $s_{-i} \in S_{-i}(I)$ and $t_i \in S_i(I)$. Let $s_i \in S_i$ be such that $s = (s_i, s_{-i}) \in S(I)$. Let $h = (a_1, \ldots, a_L) \in I$ be such that $h < \zeta(s)$, and let $\ell_1, \ldots, \ell_K$ be such that $P((a_1, \ldots, a_{\ell-1})) = i$ if and only if $\ell = \ell_k$ for some k; also let I_k be such that $h_k \equiv (a_1, \ldots, a_{\ell_k - 1}) \in I_k$.

I claim that $I_k < I$ for all k. By contradiction, assume that there is $h' \in I$ such that $\tilde{h}' \notin I_k$ for every $\tilde{h}' < h'$. This implies that I_k is not an element of $X_i(h')$; however, since $h_k \in I_k$ and $h_k < h$, I_k *is* an element of $X_i(h)$, so $X_i(h) \neq X_i(h')$, which contradicts perfect recall.

Since $I_k < I$ for every k, part (ii) implies that $a_{\ell_k} = s_i(I_k) = t_i(I_k)$ for every k. Therefore, $h < \zeta(t_i, s_{-i})$, and so $(t_i, s_{-i}) \in S(I)$.

(iv) As in (iii), it is enough to show that $\prod_j S_j(z) \subseteq S(z)$. Write $z = (a_1, \ldots, a_L)$ and $h_\ell =$

$(a_1,\ldots,a_{\ell-1})$ for $\ell = 2,\ldots,L$; let $h_1 = \phi$. For every $j \in N$, fix $s^j \in S(z)$ arbitrarily. Then, by definition, $z = \zeta(s^j)$ for all j, so $s^j_{P(h_\ell)}(h_\ell) = a_\ell$ for all $\ell = 1,\ldots,L$ and all j. Now define $s = (s^j_j)_{j\in N}$. Then $s_{P(h_\ell)}(h_\ell) = s^{P(h_\ell)}_{P(h_\ell)}(h_\ell) = a_\ell$ for all $\ell = 1,\ldots,L$. Therefore, $\zeta(s) = z$, i.e., $s \in S(z)$. ■

Finally, recall that the strategic-form payoff function U_i is defined by $U_i(s) = u_i(\zeta(s))$ for all $s \in S$, where $u_i : Z \to \mathbb{R}$. I verify the *strategic independence* property in Section 2 of the paper.

Remark 2 *If* $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i\in N})$ *has perfect recall, then for all* $i \in N$, $I \in \mathscr{I}_i$, *and* $s_i, t_i \in S_i(I)$, *there is* $r_i \in S_i(I)$ *such that* $U_i(r_i, s_{-i}) = U_i(t_i, s_{-i})$ *for all* $s_{-i} \in S_{-i}(I) = S_{-i}(I)$, *and* $U_i(r_i, s_{-i}) = U_i(s_i, s_{-i})$ *for all* $s_{-i} \notin S_{-i}(I)$.

[The argument is based on Mailath, Samuelson, and Swinkels (1993), but it is included here for ease of reference.]

Proof: Let $r_i \in S_i$ be a strategy that agrees with s_i everywhere except at information sets that weakly follow I, where it agrees with t_i. Formally, for every $J \in \mathscr{I}_i$, $r_i(J) = t_i(J)$ if $I \le J$, and $r_i(J) = s_i(J)$ otherwise. By Remark 1, since $s_i, t_i \in S_i(I)$, $s_i(J) = t_i(J)$ for all $J \in \mathscr{I}_i$ with $J < I$; by construction, $r_j(J) = s_i(J)$ for such J. Therefore, $r_i \in S_i(I)$, and in addition, for every $s_{-i} \in S_{-i}(I)$, there is a unique $h \in I$ such that $(s_i, s_{-i}), (t_i, s_{-i}), (r_i, s_{-i}) \in S(h)$. At all $J \in \mathscr{I}_i$ with $I \le J$, by construction $r_i(J) = t_i(J)$, so $\zeta(r_i, s_{-i}) = (h, a_1, \ldots, a_M) = \zeta(t_i, s_{-i})$ for suitable $a_1, \ldots, a_M \in A$. Hence $U_i(r_i, s_{-i}) = u_i(\zeta(r_i, s_{-i})) = u_i(\zeta(t_i, s_{-i})) = U_i(t_i, s_{-i})$.

On the other hand, for $s_{-i} \notin S_{-i}(I)$, by perfect recall (again, see Remark 1) $(s_i, s_{-i}) \notin S(I)$, and hence also $(s_i, s_{-i}) \notin S(J)$ for any $J \in \mathscr{I}_i$ with $I \le J$. Then $(s_i, s_{-i}) \in S(J)$ implies that not $I \le J$, and therefore $r_i(J) = s_i(J)$ at all such J. Hence $\zeta(r_i, s_{-i}) = \zeta(s_i, s_{-i})$, and so $U_i(r_i, s_{-i}) = u_i(\zeta(r_i, s_{-i})) = u_i(\zeta(s_i, s_{-i})) = U_i(s_i, s_{-i})$. ■

C.2 Proof of Theorem 2

For every s_i, let $V_i(\mu, s_i)$ be the set of payoff assignments such that s_i is sequentially rational for μ, but not structurally optimal for μ. Then $V_i(\mu) = \cup_{s_i} V_i(\mu, s_i)$. Furthermore, let $V_i(\mu, s_i, t_i)$ be the set of payoff assignments u for i such that s_i is sequentially rational for μ, but not $s_i \succcurlyeq^{\mu,u} t_i$. Then $V_i(\mu, s_i) = \cup_{t_i} V_i(\mu, s_i, t_i)$. Since S_i is finite, it is sufficient to show that $V_i(\mu, s_i, t_i)$ is a lower-dimensional subset of $\mathbb{R}^Z$ for all s_i, t_i.

Similarly, fix s_i, t_i and, for every $I \in \mathscr{I}_i$, let $V_i(\mu, s_i, t_i, I)$ be the set of payoff assignments $u \in \mathbb{R}^Z$ for i such that s_i is sequentially rational for μ, but

$$\sum_{s_{-i}} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, t_{-i}))]\nu((s_{-i})|B_\mu(I)) < 0 \quad \text{and}$$
$$\sum_{s_{-i}} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, t_{-i}))]\nu((s_{-i})|B_\mu(J)) \leq 0 \quad \forall J : S_{-i}(J) >^\mu S_{-i}(I)$$

By Definition 8, $V_i(\mu, s_i, t_i) = \cup_{I \in \mathscr{I}_i} V_i(\mu, s_i, t_i, I)$. Since $\mathscr{I}_i$ is finite, it is sufficient to prove that each $V_i(\mu, s_i, t_i, I)$ is a lower-dimensional subset of $\mathbb{R}^Z$.

The problem can be further simplified. For all S_i, t_i, I let $V_i^=(\mu, s_i, t_i, I)$ be the set of payoff assignments $u \in \mathbb{R}^Z$ for i such that s_i is sequentially rational for μ,

$$\sum_{s_{-i}} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, t_{-i}))]\nu((s_{-i})|B_\mu(I)) < 0, \tag{4}$$
$$\sum_{s_{-i}} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, t_{-i}))]\nu((s_{-i})|B_\mu(J)) = 0 \quad \forall J : S_{-i}(J) >^\mu S_{-i}(I). \tag{5}$$

Lemma 1 *Fix $s_i, t_i \in S_i$, $\bar{I} \in \mathscr{I}_i$, and $u \in V_i(\mu, s_i, t_i, \bar{I})$. Then there is $I \in \mathscr{I}_i$ with $S_{-i}(I) \geq^\mu S_{-i}(\bar{I})$ and $B_\mu(I) \neq S_{-i}$, such that $V_i(\mu, s_i, t_i, \bar{I}) \subseteq V_i^=(\mu, s_i, t_i, I)$.*

Proof: Consider the following algorithm. At step $k = 1$, let $I_1 = \bar{I}$. Inductively, at each step $k > 1$, assume that I_{k-1} has been defined, and $S_{-i}(I_{k-1}) \geq^\mu S_{-i}(\bar{I})$. If $u \in V^=(s_i, t_i, I_{k-1})$ then STOP and let $I = I_{k-1}$. Otherwise, since $u \in V(s_i, t_i, \bar{I})$ and $S_{-i}(I_{k-1}) \geq^\mu S_{-i}(\bar{I})$, there is $\tilde{I}$ with $S_{-i}(\tilde{I}) >^\mu S_{-i}(I_{k-1}) \geq^\mu S_{-i}(\bar{I})$ and

$$\sum_{s_{-i}} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, t_{-i}))]\nu((s_{-i})|B_\mu(I)) < 0.$$

Let $I_k = \tilde{I}$. By transitivity, $S_{-i}(\tilde{I}) >^{\mu} S_{-i}(I_{k-1}) \geq^{\mu} S_{-i}(\bar{I})$. This completes the inductive step.

Since $\mathscr{I}_i$ is finite and, for $k > 1$ and all $\ell = 1, \ldots, k-1$, $S_{-i}(I_k) >^{\mu} S_{-i}(I_\ell)$, hence $I_l \neq I_\ell$, the process must stop at some step $K > 1$. This means that $u \in V^{=}(s_i, t_i, I_{K-1}) = V^{=}(s_i, t_i, I)$, with $S_{-i}(I) \geq^{\mu} S_{-i}(\bar{I})$. Finally, since s_i is sequentially rational for μ under the payoff assignment u, and $s_i, t_i \in S_i(\phi)$, it cannot be the case that $B_\mu(I) = S_{-i} = B_\mu(i)$. ■

By Lemma 1, it is enough to show that each $V_i^{=}(\mu, s_i, t_i, I)$ with $B_\mu(I) \neq S_{-i}$ is a lower-dimensional subset of $\mathbb{R}^Z$.

The basic intuition is that, by Eq. (5), if $u \in V_i^{=}(\mu, s_i, t_i, I)$ for some suitable I, then u must be subject to at least one linear restriction; in particuar, note that Eq. (5) must hold for $J = \phi$. However, there is a possible complication: it may be that, due to the structure of the game and/or the choice of μ, all the left-hand sides in Eq. (5) are identically zero, regardless of the choice of u. In this case, there obviously is no linear restriction on u.

In particular, the following situation may in principle arise: for all J with $S_{-i}(J) >^{\mu} S_{-i}(I)$, and all s_{-i} with $\nu(s_{-i}|B_\mu(J)) > 0$, s_i and t_i lead to the same terminal history when i's coplayers play s_{-i}: that is, $\zeta(s_i, s_{-i}) = \zeta(t_i, s_{-i})$. In this case, each summand in Eq. (5) is zero, either because the probability of the corresponding s_{-i} is zero, or because the payoff difference in square brackets is zero.

The remainder of the proof shows that (i) this is indeed the only pathology one needs to worry about [that is, the only case in which a left-hand side in Eq. (5) can be identically zero], and (ii) this pathology actually cannot arise, due to the assumptions that s_i is sequentially rational for μ under u and that Eq. (4) holds as well.

Fix $s_i, t_i \in S_i$, $I \in \mathscr{I}_i$, and $u \in V_i^{=}(\mu, s_i, t_i, I)$. Consider $J \in \mathscr{I}_i$ with $S_{-i}(J) >^{\mu} S_{-i}(I)$. The corresponding restriction in Eq. (5) can be rewritten as

$$\sum_z u(z)\nu(\{s_{-i} : z = \zeta(s_i, s_{-i})\}|B_\mu(J)) - \sum_z u(z)\nu(\{s_{-i} : z = \zeta(t_i, s_{-i})\}|B_\mu(J)) = 0.$$

Recall that $S(z) = \{s \in S : z = \zeta(s)\}$. By Remark 1, $S(z) = S_i(z) \times S_{-i}(z)$, where $S_i(z)$ and $S_{-i}(z)$ are the projections of $S(z)$ on S_i and S_{-i} respectively. Therefore, one can further rewrite Eq (5) as

$$\sum_{z: s_i \in S_i(z)} u(z)\nu(S_{-i}(z)|B_\mu(J)) - \sum_{z: t_i \in S_i(z)} u(z)\nu(S_{-i}(z)|B_\mu(J)) = 0.$$

Finally, by cancelling the terms corresponding to z's such that $s_i, t_i \in S_i(z)$, one obtains

$$\sum_{z: s_i \in S_i(z), t_i \notin S_i(z)} u(z)\nu(S_{-i}(z)|B_\mu(J)) - \sum_{z: t_i \in S_i(z), s_i \notin S_i(z)} u(z)\nu(S_{-i}(z)|B_\mu(J)) = 0. \tag{6}$$

Note that the left-hand side of Eq. (6) can only be identically zero, regardless of u, if $\nu(\cdot|B_\mu(J))$ assigns positive probability *only* to profiles $s_{-i} \in S_{-i}$ such that $\zeta(s_i, s_{-i}) = \zeta(t_i, s_{-i})$. Thus, as claimed, this is the only pathology one needs to rule out.

Let $\mathscr{I}_i(I) = \{J \in \mathscr{I}_i : s_i, t_i \in S_i(J), S_{-i}(J) \supseteq B_\mu(I)\}$. This is non-empty because it contains at least ϕ (as $S(\phi) = S$). Furthermore, suppose that $J, J' \in \mathscr{I}_i(I)$. If $J = \phi$, then $J \geq J'$; if $J' = \phi$, then $J' \geq J$. Otherwise, fix $s_{-i} \in B_\mu(I)$: then perfect recall implies that $s = (s_i, s_{-i}) \in S(J) \cap S(J')$. By definition, this means that there are $h \in J$, $h' \in J'$ such that $h < \zeta(s)$ and $h' < \zeta(s)$. If $h = h'$ then $J = J'$ because $\mathscr{I}_i \setminus \{\phi\}$ is a partition of $P^{(-1)}(i)$. If $h < h'$, then I claim that $J < J'$. Suppose not, so there is $\tilde{h}' \in J'$ such that $\tilde{h} \notin J$ for all $\tilde{h} < \tilde{h}'$. Then $X_i(\tilde{h}')$ does not include J, whereas $X_i(h')$ does: this contradicts perfect recall. Similarly, if $h' < h$,, then $J' < J$. Since h and h' are initial segments of the same terminal history $\zeta(s)$, there is no other possibility. Since J, J' were arbitrary elements of $\mathscr{I}_i(I)$, this set admits a $<$-maximal element, henceforth denoted I_0.

Define the set

$$D = \{s_{-i} \in S_{-i} \,:\, \nu(s_{-i}|B_\mu(I)) > 0,\ \zeta(s_i, s_{-i}) \neq \zeta(t_i, s_{-i})\}.$$

Eq. (4) implies that $D \neq \emptyset$. Also, for every $s_{-i} \notin D$, $[u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, s_{-i}))]\nu(s_{-i}|B_\mu(I)) = 0$, because either the term in square brackets, or the probability of s_{-i} (or both) are zero.

For every $s_{-i} \in D$, perfect recall implies that $(s_i, s_{-i}), (t_i, s_{-i}) \in S(I_0)$, because $s_i, t_i \in S_i(I_0)$ and $D \subseteq B_\mu(I) \subseteq S_{-i}(I_0)$. If $J \in \mathscr{I}_i$ and $J < I_0$, perfect recall (Remark 1) implies that $s_i(J) = t_i(J)$.

Finally, if $J, J' \in \mathscr{I}_i$ and $(s_i, s_{-i}) \in S(J) \cap S(J')$, then, as above, perfect recall implies that either $J < J'$, or $J' < J$, or $J = J'$. Hence, for every $s_{-i} \in D$, one can define $J(s_{-i})$ to be the $<$-maximal $J \in \mathscr{I}_i$ such that $(s_i, s_{-i}), (t_i, s_{-i}) \in S(J)$; note that $I_0 \leq J(s_{-i})$.

For any two $s_{-i}, s'_{-i} \in D$, the sets $S_{-i}(J(s_{-i}))$ and $S_{-i}(J(s'_{-i}))$ are either disjoint or nested (in particular, the two sets may coincide). To see this, suppose that there is $t_{-i} \in S_{-i}(J(s_{-i})) \cap S_{-i}(J(s'_{-i}))$. Then $(s_i, t_{-i}) \in S(J(s_{-i})) \cap S(J(s'_{-i}))$ by perfect recall. But perfect recall also implies that $S(J(s_{-i}))$ and $S(J(s'_{-i}))$ are nested. To see this, suppose for definiteness that $S(J(s_{-i})) \supseteq S(J(s'_{-i}))$, and pick an arbitrary $r_{-i} \in S_{-i}(J(s'_{-i}))$; by perfect recall, $(s_i, r_{-i}) \in S(J(s'_{-i}))$, so $(s_i, r_{-i}) \in S(J(s_{-i}))$ as well, which implies that $r_{-i} \in S_{-i}(J(s_{-i}))$, which proves the claim.

Now suppose that, for every $s_{-i} \in D$, $S_{-i}(J(s_{-i})) \subseteq B_\mu(I)$. Since the sets $S_{-i}(J(s_{-i}))$, $s_{-i} \in D$, are either disjoint or nested, there is a subset $\{s^1_{-i}, \ldots, s^M_{-i}\} \subseteq D$ such that (1) for every $s_{-i} \in D$, there is $m = 1, \ldots, M$ with $S_{-i}(J(s_{-i})) \subseteq S_{-i}(J(s^m_{-i}))$; and (2) for distinct $\ell, m = 1, \ldots, M$, $S_{-i}(J(s^\ell_{-i})) \cap S_{-i}(J(s^m_{-i})) = \emptyset$. Furthermore, for each $m = 1, \ldots, M$, $\nu(S_{-i}(J(s^m_{-i}))|B_\mu(I)) \geq \nu(s_{-i}|B_\mu(I)) > 0$. Finally, $D \subseteq \bigcup_{s_{-i} \in D} S_{-i}(J(s_{-i})) \subseteq \bigcup_m S_{-i}(J(s^m_{-i})) \subseteq B_\mu(I)$, so $s_{-i} \in B_\mu(I) \setminus \bigcup_m S_{-i}(J(s^m_{-i}))$ implies that $s_{-i} \notin D$ and so $[u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, s_{-i}))]\nu(s_{-i}|B_\mu(I)) = 0$. Therefore, defining $U^u_i : S \to \mathbb{R}$ by $U_i(s) = u(\zeta(s))$,

$$
\begin{aligned}
&\sum_{s_{-i}} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, s_{-i}))]\nu(s_{-i}|B_\mu(I)) = \\
&= \sum_m \sum_{s_{-i} \in S_{-i}(J(s^m_{-i}))} [u(\zeta(s_i, s_{-i})) - u(\zeta(t_i, s_{-i}))]\nu(s_{-i}|B_\mu(I)) = \\
&= \sum_m \nu(S_{-i}(J(s^m_{-i}))|B_\mu(I)) \left[U^u_i\left(s_i, \mu(\cdot|S_{-i}(J(s^m_{-i})))\right) - U^u_i\left(t_i, \mu(\cdot|S_{-i}(J(s^m_{-i})))\right)\right] \geq 0.
\end{aligned}
$$

The last equality follows from the fact that ν extends μ and satisfies the chain rule. The inequality follows from the assumption that s_i is sequentially rational for μ given u. But this conclusion contradicts Eq. (4).

Therefore, there is $s_{-i} \in D$ such that $S_{-i}(J(s_{-i})) \not\subseteq B_\mu(I)$. Since $s_{-i} \in S_{-i}(J(s_{-i}))$ and $s_{-i} \in D$, $\nu(S_{-i}(J(s_{-i}))|B_\mu(I)) > 0$; thus, a fortiori $\nu(B_\mu(J(s_{-i}))|B_\mu(I)) > 0$, so by Corollary 4, $S_{-i}(J(s_{-i})) \geq^\mu S_{-i}(I)$. Suppose that also $S_{-i}(I) \geq^\mu S_{-i}(J(s_{-i}))$, so $S_{-i}(J(s_{-i})) =^\mu S_{-i}(I)$: then $B_\mu(J(s_{-i})) = B_\mu(I)$ and so $S_{-i}(J(s_{-i})) \subseteq B_\mu(I)$, contradiction: thus, not $S_{-i}(I) \geq^\mu S_{-i}(J(s_{-i}))$, and so $S_{-i}(J(s_{-i})) >^\mu S_{-i}(I)$.

I claim that $s_i(J(s_{-i})) \neq t_i(J(s_{-i}))$. By contradiction, suppose that $s_i(J(s_{-i})) = t_i(J(s_{-i}))$. Write $\zeta(s_i, s_{-i}) = (a_1, \ldots, a_L)$ and $\zeta(t_i, s_{-i}) = (b_1, \ldots, m_M)$. Let $h_0 = \phi$. Then $h_0 < \zeta(s_i, s_{-i})$ and $h_0 < \zeta(t_i, s_{-i})$. If $P(h_0) \neq i$, then $a_1 = s_{P(h_0)}(h_0) = b_1$. If instead $P(h_0) = i$, then $\{h_0\} \in \mathscr{I}_i$ satisfies $\{h_0\} \leq J(s_{-i})$ and so, by Remark 1 or (in case $J(s_{-i}) = \phi$) the assumption that $s_i(J(s_{-i})) = t_i(J(s_{-i}))$, $a_1 = s_i(h_0) = t_i(h_0) = b_1$. Inductively, assume that, for some $\ell < \min(L, M)$, $a_k = b_k$ for all $k = 1, \ldots, \ell$, and consider $\ell + 1$. Let $h_\ell = (a_1, \ldots, a_\ell) = (b_1, \ldots, b_\ell)$, so $h_\ell < \zeta(s_i, s_{-i})$ and $h_\ell < \zeta(t_i, s_{-i})$. Again, if $P(h_\ell) \neq i$, then $a_{\ell+1} = s_{P(h_\ell)}(h_\ell) = b_{\ell+1}$. If instead $P(h_\ell) = i$, then $h_\ell \in J$ for some $J \in \mathscr{I}_i$. I claim that, in this case, $J \leq J(s_{-i})$, so that Remark 1 or the assumption that $s_i(J(s_{-i})) = t_i(J(s_{-i}))$ imply that $a_{\ell+1} = s_i(h_\ell) = t_i(h_\ell) = b_{\ell+1}$. To see this, observe that, since $(s_i, s_{-i}) \in S(J(s_{-i}))$, by definition there is $h < \zeta(s_i, s_{-i})$ such that $h \in J(s_{-i})$. Since both h and h_ℓ are subhistories of $\zeta(s_i, s_{-i})$, either $h = h_\ell$, or $h_\ell < h$, or $h < h_\ell$. If $h = h_\ell$, then $h_\ell \in J(s_{-i})$ and so $J = J(s_{-i})$. If $h_\ell < h$, then $X_i(h)$ contains J, and hence so does $X_i(h')$ for every $h' \in J(s_{-i})$: thus, $J < J(s_{-i})$. Finally, $h < h_\ell$ cannot actually hold: if $h < h_\ell$, then $X_i(h_\ell)$ contains $J(s_{-i})$; by perfect recall, every other $h' \in J$ must be such that $X_i(h')$ contains $J(s_{-i})$, so h' must have a subhistory in $J(s_{-i})$: that is, $J(s_{-i}) < J$. Since $h_\ell < \zeta(s_i, s_{-i})$ and $h_\ell < \zeta(t_i, s_{-i})$, $(s_i, s_{-i}), (t_i, s_{-i}) \in S(J)$, which contradicts the definition of $J(s_{-i})$. It follows that $L = M$ and $\zeta(s_i, s_{-i}) = \zeta(t_i, s_{-i})$, which contradicts the fact that $s_{-i} \in D$.

To complete the proof, fix an arbitrary $t_{-i} \in \operatorname{supp} \mu(\cdot|S_{-i}(J(s_{-i})))$; note that, since ν satisfies the chain rule and extends μ, and since $\nu(S_{-i}(J(s_{-i}))|B_\mu(J(s_{-i}))) > 0$ by the last claim of Lemma 2, also $\nu(t_{-i}|B_\mu(J(s_{-i}))) > 0$. Since s_i and t_i take different actions at $J(s_{-i})$, it follows that $z \equiv \zeta(s_i, t_{-i}) \neq \zeta(t_i, t_{-i}) \equiv z'$. Since, by Remark 1, $S(z) = S_i(z) \times S_{-i}(z)$ and similarly for $S(z')$, conclude that $s_i \in S_i(z)$ but $t_i \notin S_i(z)$, and $t_i \in S_i(z')$ but $s_i \notin S_i(z')$. Finally, $\nu(S_{-i}(z)|B_\mu(J(s_{-i}))) \geq \nu(t_{-i}|B_\mu(J(s_{-i}))) > 0$, and similarly $\nu(S_{-i}(z')|B_\mu(J(s_{-i}))) > 0$. Therefore, for $J = J(s_{-i})$, the l.h.s. of Eq. (6) is not identically zero. ■

C.3 Extensive form of the elicitation game

Fix the game tree and payoffs of the original game, namely $\Gamma = (N, A, H, P, (\mathscr{I}_i)_{i\in N})$ and $(u_i)_{i\in N}$, and a questionnaire $Q = (Q_i)_{i\in N}$. I now describe the game tree and payoff assignments of the elicitation game. The objective is to ensure that the corresponding strategy sets and other derived objects satisfy the properties in Definition 9.

Begin with a description of the elicitation game tree. The player set is $N^* = N \cup \{c\}$; the action set is $A^* = A \cup \{\varnothing, b\} \cup \{p_i : i \in N, Q_i = (I, b, p_i)\} \cup N$. It is useful to distinguish between first-stage and second-stage histories. In the *first stage*, Chance moves first, at the empty history ϕ^*, and chooses an element of $A_c^1 \equiv \{\varnothing\} \cup \{i : Q_i \neq \varnothing\}$. Then, players move according to their index; player i chooses from $A_i \equiv S_i \times W_i$, where $W_i = \{\varnothing\}$ if $Q_i = \varnothing$ and $W_i = \{b, p_i\}$ otherwise. Hence, stage-1 histories are of the form

$$\phi^* \qquad \text{or} \qquad \big(a_c, (s_1, w_1), \ldots, (s_{i-1}, w_{i-1})\big): \quad a_c^1 \in A_c,\ (s_j, w_j) \in A_j^1\ j = 1, \ldots, i-1. \tag{7}$$

Second-stage histories reflect the play of the strategies players have committed to in the first stage. Hence, they take the form

$$\big(a_c, (s_1, w_1), \ldots, (s_N, w_n), h\big): \qquad (s_1, \ldots, s_N) \in S(h). \tag{8}$$

It will be convenient to represent these histories by emphasizing strategy profiles, as in

$$(a_c, s, w, h) \quad \text{or} \quad (a_c, s_i, w_i, s_{-i}, w_{-i}, h).$$

For $h = \phi$, write (a_c, s, w, ϕ) simply as (a_c, s, w). The set of all histories will be denoted by H^*.

A history (a_c, s, w, z) is terminal if and only if z is terminal in the original game.

Turn now to information sets. The Chance player has a single one, the root $\{\phi^*\}$; with some notational abuse, denote this as ϕ^*. In the *first stage*, each player $i \in N$ has an information set

$$I_i^1 = \left\{(a_c, \big(s_1, w_1), \ldots, (s_{i-1}, w_{i-1})\big) \in H^* : (s_j, w_j) \in S_j \times W_j,\ j = 1, \ldots, i-1\right\}. \tag{9}$$

This formalizes the assumption that players do not observe each other's choices (nor Chance's move) in the first stage.

In the *second stage*, for each $i \in N$, $(s_i, w_i) \in S_i \times W_i$, and $I \in \mathscr{I}_i$ such that $s_i \in S_i$), keeping the notation of Definition 9,

$$(s_i, w_i, I) = \Big\{\big(a_c, (\bar{s}_i, \bar{w}_i), (s_{-i}, w_{-i}), h\big) \in H^* : \bar{s}_i = s_i, \bar{w}_i = w_i, \ s_{-i} \in S_{-i}(h), \ h \in I\Big\}. \tag{10}$$

Thus, player i does not observe Chance's move a_c and other players' choice of bet w_{-i}; however, she does recall her own first-stage choices, and does learn that her opponents chose a strategy that allows I in the original game.

Notice that, consistently with Definition 9, I do not assume that $\mathscr{I}_i^*$ includes the symbol ϕ^*. This is because I_i^1 serves the same purpose—it ensures that $S^*(I_i^1) = S^*$ is a conditioning event, and hence that a CPS for i includes i's unconditional beliefs.

Turn now to the payoff assignments u_j^*, for $j \in N^*$. For Chance, $u_c^* \equiv 0$. For each player $i \in N$, we let

$$u_i^*\big((a_c, s, w, z)\big) = \begin{cases} u_i(z) & a_c \neq i \\ 1 & a_c = i, Q_i = (I, E, p), w_i = b, s_{-i} \in E \\ 0 & a_c = i, w_i = b, s_{-i} \notin E \\ p & a_c = i, Q_i = (I, E, p), w_i = p_i. \end{cases} \tag{11}$$

I now verify that the induced strategy sets S_i^*, strategy correspondence $S^*(\cdot)$, and payoff functions $U_i^*(\cdot)$, satisfy the properties in Definition 9.

Chance has a unique information set ϕ^*, with action set A_c^1, so $S_c^* = A_c^1 = \{\varnothing\} \cup \{i \in N : Q_i \neq \varnothing\}$.

Now consider player $i \in N$. Eq. (7) and Eq. 8 for $h = \phi$ show that, for any first-period history $h^* \in I_i^1$ and action $(s_i, w_i) \in S_i \times W_i$, $(h^*, (s_i, w_i)) \in H^*$. Therefore, $A^*(I_i^1) = S_i \times W_i$. Given a second-period information set (s_i, w_i, I), Eq. (10) implies that, if $h^* \in (s_i, w_i, I)$, then

$h^* = (a_c, s, w, h)$ for some $a_c \in A_c$, $s \in S$, $w \in W$ and $h \in H$; Eq. (8) then implies that $(h^*, a) = (a_c, s, w, (h, a)) \in H^*$ iff $s \in S((h, a))$; and since $P(h) = i$ and $h \in I$, $a = s_i(I)$. Therefore, $A^*((s_i, w_i, I)) = \{s_i(I)\}$. This formalizes the statement that player i is committed to action $s_i(I)$ at (s_i, w_i, I).

It follows that, for every player $i \in N$, there is a bijection between S_i^* and $A^*(I_i^1) = S_i \times W_i$. Definition 9 abuses notation and sets $S_i^* = S_i \times W_i$.

Turn now to the strategy map $S^*(\cdot)$. First, every strategy profile reaches the initial history ϕ^*, so $S^*(\phi^*) = S^*$. For every other first-stage information set I_i^1, Eq. 7 implies that, for any profile $s^* \in S^*$, the induced partial history $(a_c, (s_1, w_1), \ldots, (s_{i-1}, w_{-i}))$ lies in I_i^1. Thus, $S^*(I_i^1) = S^*$.

Now consider a second-stage information set (s_i, w_i, I) and a strategy $\bar{s}^*$. Eq. (10) implies that, first of all, there is no restriction on Chance's move, but $\bar{s}_i^*(I_i^1) = (s_i, w_i)$. Additionally, let $\bar{s}_j^*(I_j^1) = (\bar{s}_j, \bar{w}_j)$ for all $j \neq i$: there is no restriction on w_{-i}, but $s_{-i} \in S_{-i}(I)$. Therefore, $S^*((s_i, w_i, I)) = \{(s_i, w_i)\} \times S_{-i}(I) \times W_{-i} \times S_c^*$.

Finally, turn to strategic-form payoffs. The definition of u_c^* implies that $U_c^* \equiv 0$. For players $i \neq c$, fix a profile $s^* = ((s_i, w_i), (s_{-i}, w_{-i}), s_c^*)$. The induced terminal history is then $(s_c^*, s, w, \zeta(s))$ [at $(s_c^*, s, w) = (s_c^*, s, w)$, the player on the move is $P(\phi)$; by Eq. (8), there is only one history featuring a single additional action, namely $(s_c^*, s, w, (s_{P(\phi)}(\phi)))$; inductively, if s^* induces (s_c^*, s, w, h), the only continuation history featuring a single additional action is $(s_c^*, s, w, (h, s_{P(h)}(h)))$]. Eq. (11) implies that

$$U_i^*(s^*) = u_i^*(\zeta^*(s^*)) = u_i^*((s_c^*, s, w, \zeta(s)) = \begin{cases} u_i(\zeta(s)) = U_i(s) & s_c^* \neq i \\ 1 & s_c^* = 1, Q_i = (I, e, p), w_i = b, s_{-i} \in E \\ 0 & s_c^* = 1, Q_i = (I, e, p), w_i = b, s_{-i} \notin E \\ p & s_c^* = i, Q_i = (I, E, p), w_i = p. \end{cases}$$

This completes the proof.

D Additional examples

D.1 Calculations for the game in Figure 6 (Section 5.2)

I first analyze Bob's preferences. We have (collapsing realization-equivalent strategies, as in the paper) $S_a = \{(Out,b),(Out,p),(InB,b),(InB,p),(InS,b),(InS,p)\}$, $S_b = \{\bar{B}B, \bar{S}S\}$, $\mathscr{I}_a = \{\phi, K\}$ with $S_a(\phi) = S_a(K) = S_a$, and $\mathscr{I}_b = \{\phi, I, I'\}$ with $S_a(I) = S_a(I') = \{(InB,b),(InB,p),(InS,b),(InS,p)\}$.

Assume that Bob's beliefs μ satisfy $\mu(\{(Out,b),(Out,p)\}|S_a) = 1$ and $\mu(\{(InS,b),(InS,p)\}|S_a(I)) = \mu(\{(InS,b),(InS,p)\}|S_a(I')) = \pi$. These assignments are enough to calculate all conditional expected payoffs for Bob, which are given in Table I.

s_b	$\mathrm{E}_{S_a} U_b(s_b,\cdot)$	$\mathrm{E}_{S_a(I)} U_b(s_b,\cdot)$
$\bar{B}B$	0	$1-\pi$
$\bar{S}S$	0	3π

Table I: Expected payoffs for Bob in Figure 6.

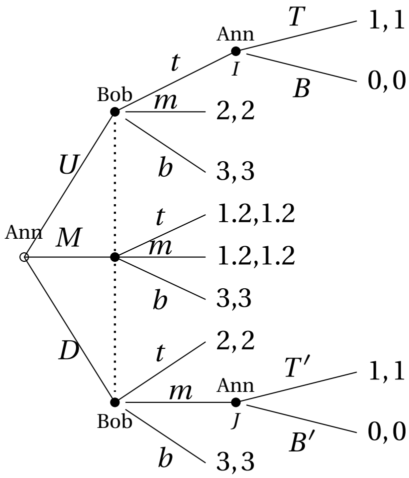
Applying Remark 1, one sees that, for instance, $\bar{S}S$ is structurally rational iff $\pi \geq \frac{1}{4}$. This is, of course, exactly the condition under which S is structurally and sequentially rational in the original game of Figure 1, if he expects Ann to choose S with probability π conditional upon having played *In*. Hence, as claimed, Bob's strategic incentives are preserved.

Now turn to Ann. Since the only conditioning event for her is S_b, her structural preferences are actually ex-ante EU. Hence, she will choose b at K if and only if she assigns probability at least p to Bob choosing $\bar{S}$ (and hence committing to S) at ϕ.

D.2 Structural rationality and trembles

As noted in the text, the characterization of structural preferences in terms of trembles implies that s_i is structurally rational if, for every $t_i \neq s_i$, there is *some* structural perturbation for which s_i is weakly better than t_i. This formulation indicates that this perturbation may well depend upon t_i. The following example shows that this dependence is necessary—there may not be a single perturbation that "works" for all strategies.

Example 1 Figure 1 differs from Figure 4 in that Ann has an additional action M at ϕ.



s_a	$\mathrm{E}_{\mu(\cdot\|S_b)} U_a(s_a,\cdot)$	$\mathrm{E}_{\mu(\cdot\|S_b(I))} U_a(s_a,\cdot)$	$\mathrm{E}_{\mu(\cdot\|S_b(J))} U_a(s_a,\cdot)$
UT	3	1	2
UB	3	0	2
M	3	1.2	1.2
DT'	3	2	1
DB'	3	2	0

Figure 1: No single rationalizing perturbation; $\mu(\{b\}|S_b(\phi)) = 1$

Ann's CPS μ assigns prior probability one to b; UT, DT' and M are the structurally rational strategies given μ. A perturbation (p^n) of μ is structural if and only if every p^n has full support. For any full-support $p \in \Delta(S)$, $U_a(M, p) \geq U_a(UT, p)$ if and only if $\frac{p(\{m\})}{p(\{t\})} \leq \frac{1}{4}$, and $U_a(M, p) \geq U_a(DT', p)$ if and only if $\frac{p(\{m\})}{p(\{t\})} \geq 4$. Thus, there is no single structural perturbation for which M is better than both UT and DT'.

E Equivalent definitions of structural preferences

E.1 Lexicographic preferences on every "path"

Section 3 provided a suggestive description of structural preferences as "lexicographic preferences on every path." Throughout, fix a game with nested strategic information (the arguments can be generalized to arbitrary games), a player i, and a CPS μ for player i. In the notation for game trees in Online Appendix C, a "path" is just a terminal history. For every terminal history $z \in Z$ there is an ordered list $(h_1 = \phi, h_2, \ldots, h_L)$ of partial histories such that, for every $\ell = 1, \ldots, L$, $h_\ell < z$ and $h_\ell \in I_\ell$ for some $I_\ell \in \mathscr{I}_i$ such that $S_{-i}(I_\ell)$ is μ-basic. One can then require that strategy s_i be lexicographically weakly better than strategy t_i given the lexicographic probability system (LPS) $\big(\mu(\cdot|S_{-i}(I_1)), \ldots, \mu(\cdot|S_{-i}(I_L))\big)$. The discussion in Section 3 suggests that, "loosely," s_i is structurally preferred to t_i when this requirement is satisfied for *every* path, or terminal history, $z \in Z$. (This is indeed correct for the example games considered in the Introduction and in Section 3.)

This pathwise lexicographic criterion is not quite a characterization of structural rationality. Consider two μ-basic information sets $I, J \in \mathscr{I}_i$. While it is the case that, if I is encountered earler on a path than J, then $S_{-i}(I) \supset S_{-i}(J)$, it may also be the case that $S_{-i}(I) \supset S_{-i}(J)$ even though there is *no* path that crosses both I and J. When this occurs, the pathwise-lexicographic criterion leads to questionable conclusions that differ from those of structural rationality. The following example illustrates.

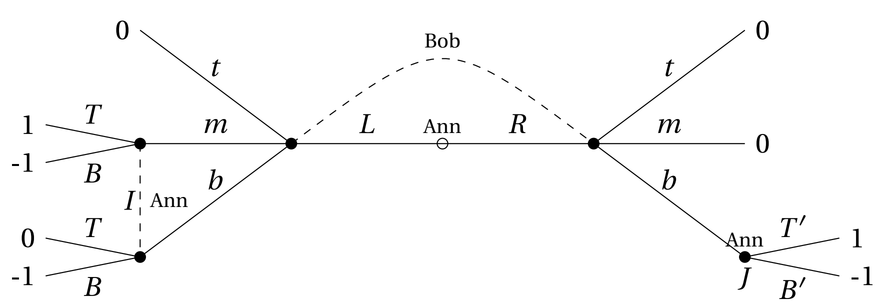


Figure 2: Inclusion-ordered information sets not on the same path (Ann's payoffs shown)

Assume that Ann's CPS μ satisfies $\mu(\{t\}|S_b) = 1$ and $\mu(\{m\}|S_b(I)) = 1$; by definition, $\mu(\{b\}|S_b(J)) = 1$. In this game, $S_b \supset S_b(I) \supset S_b(J)$. However, there is no path that crosses both I and J. Definition 5 implies that LT is the only structural best reply to μ; in particular, $LT \succ^\mu RT'$. Instead, the pathwise-lexicographic criterion leads to the conclusion that LT and RT' are incomparable, and both maximal. The reason is that LT is (lexicographically) better than RT' with respect to the LPS $\big(\mu(\cdot|S_b),\mu(\cdot|S_b(I))\big)$, but worse than RT' given the LPS $\big(\mu(\cdot|S_b),\mu(\cdot|S_b(J))\big)$.

Ranking LT above RT' is appropriate, as the CPS μ indicates that Bob is infinitely more likely to play m than b. The same conclusion arises from an argument based on belief perturbations. Note that, if (p^n) is a perturbation of μ, then eventually LT *must* yield a strictly greater expected payoff than RT' given p^n. The reason is that, in order to generate μ, the sequence (p^n) must be such that $\frac{p^n(S_b(J))}{p^n(S_b(I))}=\frac{p^n(\{b\})}{p^n(\{m,b\})}\to 0=\mu(\{b\}|S_b(I))$. Intuitively, (p^n) induces a likelihood ranking of $S_b(I)$ and $S_b(J)$, necessarily consistent with set inclusion, even though I and J are not on any common path. Definition 5 takes this ranking into account, whereas pathwise-lexicographic optimality does not.

I now show how to modify the pathwise-lexicographic criterion described in Section 3 so as to obtain a full characterization of structural rationality. To minimize notation, I do so here for games with nested strategic information, but a similar characterization holds for general games. I do, however, use the notation $B_\mu(I)$ in Section 4; in particular, recall that, by Remark 3, $B_\mu(\mathscr{I}_i)$ is the set of conditioning events corresponding to μ-basic information sets for player i. Finally, let $\geq_L$ denote the lexicographic order (that is, given vectors $v,w\in\mathbb{R}^K$, $v\geq_L w$ iff $v_\ell<w_\ell$ for some $\ell\in\{1,\dots,K\}$ implies that there is $k\in\{1,\dots,\ell-1\}$ with $v_k>w_k$).

Proposition 2 *Assume the dynamic game has nested strategic infornation. For all $s_i,t_i\in S_i$, $s_i\succcurlyeq^\mu t_i$ if and only if, for every maximal chain $(F_1,\dots,F_K)$ in the partially ordered set $(B_\mu(\mathscr{I}_i),\supseteq)$,*

$$\left(\mathrm{E}_{\mu(\cdot|F_k)}U_i(s_i,\cdot)\right)_{k=1}^K\geq_L\left(\mathrm{E}_{\mu(\cdot|F_k)}U_i(t_i,\cdot)\right)_{k=1}^K.$$

To relate this remark to the informal discussion in the text, fix a terminal history z and list the μ-basic information sets crossed by z, in the order they are encountered, as $I_1=\phi,I_2,\dots,I_K$. Then $(S_{-i}(I_1),\dots,S_{-i}(I_K))$ is a chain in $(B_\mu(\mathscr{I}_i),\supseteq)$, but it need not be a *maximal* chain. In the above example, $(S_b,S_{-i}(J))$ is a chain in $(B_\mu(\mathscr{I}_a),\supseteq)$, but it is not a maximal chain, because it is contained in the chain $(S_b,S_b(I),S_b(J))$.

Proof: If $(F_1,\dots,F_K)$ is a chain in $(B_\mu(\mathscr{I}_i),\supseteq)$, then there is a maximal chain that contains it. (This is a general property of posets, but it follows from an elementary construction here.)

Proof: consider $G \in B_\mu(\mathscr{I}_i)$ such that $F_k \neq G$ for all k. If $G \cap F_K \neq \emptyset$, then by nested strategic information either $F_K \supset G$ or $G \supset F_K$. In the former case, $(F_1, \ldots, F_K, G)$ is also a chain. In the latter, since $F_k \supseteq F_K$ for all k, $F_k \cap G \neq \emptyset$ for all k. Hence, by nested strategic information and the assumption that $G \notin \{F_1, \ldots, F_K\}$, the set $\{F_1, \ldots, F_K, G\}$ is totally ordered by $\supset$; hence, there is a chain that contains $(F_1, \ldots, F_K)$ and G Now let $F_k^0 = F_k$ for all k, and $K_0 = K$. Inductively, if F_k^n and K_0 have been defined for some $n \geq 0$ and $k = 1, \ldots, K_0$, and there exists $G \in B_\mu(\mathscr{I}_i) \setminus \{F_1^n, \ldots, F_{K_n}^n\}$ such that $G \cap F_{K_n}^n \neq \emptyset$, let $K_{n+1} = K_n + 1$ and define $F_1^{n+1}, \ldots, F_{K_{n+1}}^{n+1}$ to be the chain that consists of $\{F_1^n, \ldots, F_{K_n}^n, G\}$. Since $B_\mu(\mathscr{I}_i)$ is finite, this process must stop at some finite $\bar{n}$. For all $G \in B_\mu(\mathscr{I}_I)$, either $G = F_k^{\bar{n}}$ for some k, or $G \cap F_{K_{\bar{n}}}^{\bar{n}} = \emptyset$. In the latter case, G and $F_{K_{\bar{n}}}^{\bar{n}}$ are not by inclusion, so there is no chain that contains $\{F_1^n, \ldots, F_{K_n}^n, G\}$. In other words, the chain $(F_1^{\bar{n}}, \ldots, F_{K_{\bar{n}}}^{\bar{n}})$ cannot be extended, and is thus maximal. By construction, it contains $(F_1, \ldots, F_K)$.

Now suppose that the condition in the Proposition holds. Let $I \in \mathscr{I}_i$ be μ-basic and such that $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}[U_i(s_i, \cdot) - U_i(t_i, \cdot)] < 0$. The (trivial) chain $(S_{-i}(I))$ belongs to some maximal chain $(F_1, \ldots, F_K)$; suppose that $S_{-i}(I) = F_\ell$. Then by assumption there is $k < \ell$ with $\mathrm{E}_{\mu(\cdot|F_k)}[U_i(s_i, \cdot) - U_i(t_i, \cdot)] > 0$. Since $k < \ell$, $F_\ell \supset S_{-i}(I)$. Finally, by construction there exists a μ-basic $J \in \mathscr{I}_i$ with $F_k = S_{-i}(J)$. Since I as above was arbitrary, $s_I \succcurlyeq^\mu t_i$.

Conversely, suppose that $s_i \succcurlyeq^\mu t_i$ and consider a maximal chain $(F_1, \ldots, F_K)$ in $(B_\mu(\mathscr{I}_i), \supseteq)$. Suppose further that $\mathrm{E}_{\mu(\cdot|F_\ell)}[U_i(s_i, \cdot) - U_i(t_i, \cdot)] < 0$ for some ℓ. By construction, there is a μ-basic $I \in \mathscr{I}_i$ with $S_{-i}(I) = F_\ell$. Then, since $s_i \succcurlyeq^\mu t_i$, there is a μ-basic $J \in \mathscr{I}_i$ with $S_{-i}(J) \supset S_{-i}(I)$ and $\mathrm{E}_{\mu(\cdot|S_{-i}(J))}[U_i(s_i, \cdot) - U_i(t_i, \cdot)] > 0$. Furthermore, there is a maximal chain $(G_1, \ldots, G_L)$ in $(B_\mu(\mathscr{I}_i), \supseteq)$ that contains the chain $(S_{-i}(J), S_{-i}(I))$. Let $\bar{k}, \bar{\ell}$ be such that $G_{\bar{k}} = S_{-i}(J)$ and $G_{\bar{\ell}} = S_{-i}(I)$. Then $G_{\bar{k}} \supset G_{\bar{\ell}} = S_{-i}(I) = F_\ell$. Hence, for all $\tilde{k} \in \{1, \ldots, \ell\}$, $G_{\bar{k}} \cap F_{\tilde{k}} \supseteq F_\ell \neq \emptyset$. Furthermore, for all $\tilde{k} \in \{\ell + 1, \ldots, K\}$, $G_{\bar{k}} \supset F_\ell \supset F_{\tilde{k}}$. In other words, the set $\{F_1, \ldots, F_K, G_{\bar{k}}\}$ is totally ordered, and therefore must be contained in a maximal chain. Since $(F_1, \ldots, F_K\}$ is itself a maximal chain, there must be k with $F_k = F_{\bar{k}}$; and since $G_{\bar{k}} \supset F_\ell$, $k < \ell$. Thus, the displayed equation in the Proposition must hold. ■

E.2 Lexicographic rationality for completions of the plausibility ordering

This section shows that the property exhibited in Example 2 holds generally for all dynamic games: a strategy is structurally rational given a CPS μ if and only if it is lexicogrpahically rational with respect to some completion of the plausibility ordering of μ-basic events. This statement is reminiscent of the characterization of structural rationality via belief perturbations in Theorem 4; indeed, there is a connection between these two characterizations.

Throughout, fix a dynamic game $(N,(S_i,\mathscr{I}_i,U_i)_{i\in N},S(\cdot))$, a player $i\in N$, and a CPS $\mu\in\Delta(S_{-i},S_{-i}(\mathscr{I}_i))$, with extension $\nu\in\Delta(S_{-i},S_{-i}(\mathscr{I}_i)\cup B_\mu(\mathscr{I}_i))$.

Recall that a lexicographic *conditional* probability system (LCPS) is an LPS $(p_1,\ldots,p_L)$ in which $\operatorname{supp} p_\ell\cap\operatorname{supp} p_k=\emptyset$ (Blume, Brandenburger, and Dekel, 1991a). I now define the set of LCPSs that are obtained by completing the partial order induced by $\geq^\mu$ over the measures $\{\nu(\cdot|B_\mu(I)) : I\in\mathscr{I}_i\}$. Conceptually, given two measures $\nu(\cdot|B_\mu(I))$ and $\nu(\cdot|B_\mu(J))$ such that the sets $S_{-i}(I)$ and $S_{-i}(J)$ are not ranked by $\geq^\mu$, a LCPS may rank one measure as more plausible (i.e., lower-order) than the other, or it may rank then as equally plausible. In the latter case, there is some measure p_ℓ in the LCPS whose support contains the (disjoint) supports of both measures.

To formalize this, as in the proof of Theorems 3 and 4, it is convenient to fix a collection $\{I_1,\ldots,I_M\}\subset\mathscr{I}_i$ such that, for every $I\in\mathscr{I}_i$, there is a unique $m\in\{1,\ldots,M\}$ such that $S_{-i}(I)=^\mu S_{-i}(I_m)$. Thus, structural preferences $\succcurlyeq^\mu$ are defined in terms of the probabilities $\nu(\cdot|B_\mu(I_1)),\ldots,\nu(\cdot|B_\mu(I_M))$, ordered by the plausibility relation $\geq^\mu$.

Definition 2 *An LCPS $\sigma=(p_1,\ldots,p_L)$ is a **completion** of μ, written $\sigma\in\mathscr{C}(\mu)$, if there is an onto function $\ell:\{1,\ldots,M\}\to\{1,\ldots,L\}$ such that,*

1. *for every $\ell=1,\ldots,L$, p_ℓ is a convex combination of $\{\nu(\cdot|B_\mu(I_m)) : \ell(m)=\ell\}$ with strictly positive weights;*

2. *for every* $m, n \in \{1, \dots, M\}$, $S_{-i}(I_m) >^{\mu} S_{-i}(I_n)$ *implies* $\ell(m) < \ell(n)$.

Recall that the probabilities $\nu(\cdot|B_\mu(I_1)), \dots, \nu(\cdot|B_\mu(I_M))$ have disjoint supports (Corollary 4 in the Appendix of the paper). Therefore, (1) the support of each probability p_ℓ in the LCPS σ is the union of supports of the probabilities $\nu(\cdot|B_\mu(I_m))$, for $m \in \ell^{-1}(\ell)$; consequently (2) $\cup_\ell \text{supp } p_\ell = \cup_m \text{supp } \nu(\cdot|B_\mu(I_m)) = \cup_{I \in \mathscr{I}_i} \text{supp } \mu(\cdot|I_m)$.

To streamline notation, for $p \in \Delta(S_{-i})$, let $U_i(s_i, p) = \mathrm{E}_p U_i(s_i, \cdot)$. Given an L(C)PS $\sigma = (p_1, \dots, p_L)$, let $\succsim^\sigma$ denote the lexicographic preference over strategies induced by σ: that is, $s_i \succsim^\sigma t_i$ iff $\left(U_i(s_i, p_\ell)\right)_{\ell=1}^L \geq_L \left(U_i(t_i, p_\ell)\right)_{\ell=1}^L$. We then have:

Theorem 3 *For every* $s_i, t_i \in S_i$*:*

(1) $s_i \succsim^\mu t_i$ *if and only if* $s_i \succsim^\sigma t_i$ *for every* $\sigma \in \mathscr{C}(\mu)$*;*

(2) $s_i \succ^\mu t_i$ *if and only if* $s_i \succ^\sigma t_i$ *for every* $\sigma \in \mathscr{C}(\mu)$*.*

Therefore, a strategy $s_i \in S_i$ *is structurally rational for* μ *if, for every* $t_i \in S_i$*, there is* $\sigma \in \mathscr{C}(\mu)$ *such that* $s_i \succsim^\sigma t_i$*.*

The game in Example 1 demonstrates that, as is the case for perturbations in Theorem 4, it may be necessary to choose a different LCPS for every alternative strategy t_i.

Proof: As in the proof of Theorem 4, I prove sufficiency and necessity for (1) and (2) jointly; the last statement follows immediatley from (2).

(Necessity): fix $\bar{s}_i, \bar{t}_i \in S_i$ and suppose that $\bar{s}_i \succsim^\mu \bar{t}_i$. If $\bar{s}_i \sim^\mu \bar{t}_i$, then Definition 8 implies that $U_i(\bar{s}_i, \nu(\cdot|B_\mu(I_m))) = U_i(\bar{t}_i, \nu(\cdot|B_\mu(I_m)))$ for all m, and hence $U_i(\bar{s}_i, p) = U_i(\bar{t}_i, p)$ for every $p \in \Delta(S_{-i})$ that is a convex combination of $\nu(\cdot|B_\mu(I_m))$, $m \in \{1, \dots, M\}$. Thus, in this case $\bar{s}_i \sim^\sigma \bar{t}_i$ for all $\sigma \in \mathscr{C}(\mu)$.

If instead $\bar{s}_i \succ^\mu \bar{t}_i$, then by Theorem 4 $U_i(\bar{s}_i, q^n) > U_i(\bar{t}_i, q^n)$ eventually for all structural perturbations (q^n) of μ.

Fix $\sigma = (p_1, \dots, p_L) \in \mathscr{C}(\mu)$. Proposition 1 in Blume, Brandenburger, and Dekel (1991b) states that there exist $(\epsilon_\ell^k)_{k \geq 1} \in (0,1]^L$, with $\epsilon_\ell^k \to 0$ for all $\ell = 1, \dots, L-1$ and $\epsilon_L^k = 1$ for all k, such that the sequence $(q^k)_{k\geq 1}$ defined by $q^k = \sum_{\ell=1}^{L}\left(\prod_{r=1}^{\ell-1}\epsilon_r^k\right)(1-\epsilon_\ell^k)p_\ell$ satisfies $U_i(\bar{s}_i, q^k) > U_i(\bar{t}_i, q^k)$ for all k if and only if $\bar{s}_i \succ^\sigma \bar{t}_i$. I claim that $(q^k)_{k\geq 1}$ is a structural perturbation of μ.

By the construction of the measures q^k and the fact that $\sigma \in \mathscr{C}(\mu)$, supp $q_k = \cup_\ell$supp $p_\ell = \cup_{I\in\mathscr{I}_i}$supp $\mu(\cdot|S_{-i}(I))$. Now fix $I \in \mathscr{I}_i$. Since $\mu(S_{-i}(I)|S_{-i}(I)) = 1$, the equality just established implies $q^k(S_{-i}(I)) > 0$. Moreover, consider $s_i, t_i \in$ supp $\mu(\cdot|S_{-i}(I))$. Let $m \in \{1,\ldots,M\}$ be such that $S_{-i}(I) =^\mu S_{-i}(I_m)$. Then $s_i, t_i \in$ supp $\nu(\cdot|B_\mu(I_m))$, and therefore $s_i, t_i \in$ supp $p_{\ell(m)}$, where $\ell(\cdot)$ is the function in Definition 2. Then

$$\frac{q^k(\{s_i\})}{q^k(\{t_i\})} = \frac{p_{\ell(m)}(\{s_i\})}{p_{\ell(m)}(\{t_i\})} = \frac{\nu(\{s_i\}|B_\mu(I_m))}{\nu(\{t_i\}|B_\mu(I_m))} = \frac{\mu(\{s_i\}|S_{-i}(I))}{\mu(\{t_i\}|S_{-i}(I))};$$

the first equality follows by cancelling the common weight $\prod_{r=1}^{\ell(m)-1}\epsilon_r^k(1-\epsilon_{\ell(m)}^k)$ that q^k attaches to $p_{\ell(m)}$; the second follows from the fact that the supports of the measures $\nu(\cdot|B_\mu(I_n))$, with $\ell(n) = \ell(m)$, are disjoint, and $p_{\ell(m)}$ is a convex combination of these measures with strictly positive weights; the last one follows from the fact that $\nu(S_{-i}(I)|B_\mu(I_m)) > 0$ and $\mu(\cdot|S_{-i}(I))$ is the update of $\nu(\cdot|B_\mu(I_m))$.

It remains to be shown that $q^k(\cdot|S_{-i}(I)) \to \mu(\cdot|S_{-i}(I))$. Let $m \in \{1,\ldots,M\}$ be such that $I =^\mu I_m$. Fix $t_{-i} \in$ supp $\mu(\cdot|S_{-i}(I))$, so also $t_{-i} \in$ supp $\nu(\cdot|B_\mu(I_m))$. By construction, $t_{-i} \in$ supp $p_{\ell(m)}$, and therefore $q^k(\{t_{-i}\}|S_{-i}(I)) > 0$. Now consider an arbitrary s_{-i} such that $q^k(\{s_{-i}\}|S_{-i}(I)) > 0$. Then there is a unique $\ell \in \{1,\ldots,L\}$ such that $p_\ell(\{s_{-i}\}) > 0$. By construction, this means that there is $n \in \{1,\ldots,M\}$ such that $\nu(\{s_{-i}\}|B_\mu(I_n)) > 0$. Since $S_{-i}(I) \subseteq B_\mu(I) = B_\mu(I_m)$, Corollary 4 implies that $S_{-i}(I_m) \geq^\mu S_{-i}(I_n)$. If $S_{-i}(I_m) =^\mu S_{-i}(I_n)$, then $B_\mu(I_n) = B_\mu(I_m)$, so $\nu(\{s_{-i}\}|B_\mu(I_m)) > 0$ and, by the chain rule, $s_{-i} \in$ supp $\mu(\cdot|S_{-i}(I))$; in this case,

$$\frac{q^k(\{s_{-i}\}|S_{-i}(I))}{q^k(\{t_{-i}\}|S_{-i}(I))} = \frac{q^k(\{s_{-i}\})}{q^k(\{t_{-i}\})} = \frac{\mu(\{s_{-i}\}|S_{-i}(I))}{\mu(\{t_{-i}\}|S_{-i}(I))}.$$

Otherwise, $S_{-i}(I_m) >^\mu S_{-i}(I_n)$, which implies that $\ell(m) < \ell(n)$ and so, by the construction of q^k,

$$\frac{q^k(\{s_{-i}\}|S_{-i}(I))}{q^k(\{t_{-i}\}|S_{-i}(I))} = \frac{q^k(\{s_{-i}\})}{q^k(\{t_{-i}\})} \to 0.$$

It follows that $q^k(\cdot|S_{-i}(I)) \to \mu(\cdot|S_{-i}(I))$, as claimed.

Thus, (q^k) is a structural perturbation of μ, soTheorem 4 implies that $U_i(\bar{s}_i, q^k) > U_i(\bar{t}_i, q^k)$ eventually, and Propsition 1 in Blume et al. (1991b) yields $\bar{s}_i \succ^\sigma \bar{t}_i$. This establishes necessity in both (1) and (2).

(Sufficiency): Define a function $f : \{1,\ldots,M\} \to \mathbb{R}$ by letting

$$f(m) = \#\Big\{n \in \{1,\ldots,M\} : S_{-i}(I_n) >^\mu S_{-i}(I_m)\Big\} + \frac{m}{1+M}. \tag{12}$$

The function f is one-to-one, because, for $m,n \in \{1,\ldots,M\}$ $|\frac{m}{1+M} - \frac{n}{1+M}|$ is at most $\frac{M-1}{1+M} < 1$ and the first term in the rhs of the above equation is integer-valued. Furthermore, if $S_{-i}(I_m) >^\mu S_{-i}(I_n)$, then $f(m) < f(n)$, because $S_{-i}(I_\ell) >^\mu S_{-i}(I_m)$ implies $S_{-i}(I_\ell) >^\mu S_{-i}(I_n)$ by transitivity, but in addition $S_{-i}(I_m) >^\mu S_{-i}(I_n)$ and not $S_{-i}(I_n) >^\mu S_{-i}(I_n)$.

Now suppose that $s_i \succcurlyeq^\sigma t_i$ for all $\sigma \in \mathscr{C}(\mu)$. Let $m^* \in \{1,\ldots,M\}$ be such that $U_i(s_i, \nu(\cdot|B_\mu(I_{m^*})) < U_i(\cdot, \nu(\cdot|B_\mu(I_{m^*}))$. It must be shown that there is $m \in \{1,\ldots,M\}$ with $S_{-i}(I_m) >^\mu S_{-i}(I_{m^*})$ and $U_i(s_i, \nu(\cdot|B_\mu(I_m)) > U_i(\cdot, \nu(\cdot|B_\mu(I_m))$.

Define $g : \{1,\ldots,M\} \to \mathbb{R}$ by

$$g(m) = \begin{cases} f(m) & S_{-i}(I_m) \geq^\mu S_{-i}(I_{m^*}) \\ f(m) + M + 1 & \text{otherwise.} \end{cases}$$

Consider $m,n \in \{1,\ldots,M\}$ such that $S_{-i}(I_m) >^\mu S_{-i}(I_n)$. I claim that $g(m) < g(n)$. If $S_{-i}(I_n) \geq^\mu S_{-i}(I_{m^*}))$, then $g(m) = f(m) < f(n) = g(n)$. If instead $S_{-i}(I_n) \not\geq^\mu S_{-i}(I_{m^*})$, then $g(n) = f(n) + M + 1 \geq M + 1$, and there are two cases to consider. If $S_{-i}(I_m) \geq^\mu S_{-i}(I_{m^*})$, then $g(m) = f(m) < M + 1 \leq g(n)$. Finally, if also $S_{-i}(I_m) \not\geq^\mu S_{-i}(I_{m^*})$, then $g(m) = f(m) + M + 1 < f(n) + M + 1 = g(n)$. This proves the claim.

The function g is one-to-one, because f is one-to-one and strictly less than $M+1$, and $g(m) \geq M+1$ if $S_{-i}(I_m) \not\geq^\mu S_{-i}(I_{m^*})$. Furthermore, if $g(m) \leq g(m^*) = f(m^*)$, then $g(m) = f(m)$ and so $S_{-i}(I_m) \geq^\mu S_{-i}(I_{m^*})$; in particular, either $m = m^*$ or $S_{-i}(I_m) >^\mu S_{-i}(I_{m^*})$

Finally, define $\ell : \{1,\ldots,M\} \to \{1,\ldots,M\}$ by $\ell(m) = \#\{n : g(n) \leq g(m)\}$and let $\sigma = (p_1,\ldots,p_M)$, with $p_m = \nu(\cdot|B_\mu(I_{\ell^{-1}(m)}))$ for all $m \in \{1,\ldots,M\}$. Then $\sigma \in \mathscr{C}(\mu)$; in addition, for $\ell \in \{1,\ldots,M\}$,

$\ell < \ell(m^*)$ implies that $S_{-i}(I_{\ell^{-1}(\ell)}) > S_{-i}(I_{m^*})$.

By assumption, $s_i \not\succsim^{\sigma} t_i$, and by construction $U_i(s_i, p_{\ell(m^*)}) < U_i(t_i, p_{\ell(m^*)})$. Then, by definition, there is $\ell < \ell(m^*)$ with $U_i(s_i, p_\ell) > U_i(t_i, p_\ell)$. Letting $m = \ell^{-1}(\ell)$, $U_i(s_i, \nu(\cdot|B_\mu(I_m))) > U_i(t_i, \nu(\cdot|B_\mu(I_m)))$ and $S_{-i}(I_m) >^{\mu} S_{-i}(I_{m^*})$. Since m^* was arbitrary, $s_i \not\succsim^{\mu} t_i$. This completes the proof of sufficiency in (1).

In addition, if $s_i \succ^{\sigma} t_i$ for some $\sigma = (p_1, \ldots, p_L) \in \mathscr{C}(\mu)$, then there must be $\ell \in \{1, \ldots, L\}$ with $U_i(s_i, p_\ell) \neq U_i(t_i, p_\ell)$, and hence $m \in \{1, \ldots, M\}$ with $U_i(s_i, \nu(\cdot|B_\mu(I_m))) \neq U_i(t_i, \nu(\cdot|B_\mu(I_m)))$. Definition 8 implies that $s_i \sim^{\mu} t_i$ if and only if $U_i(s_i, \nu(\cdot|B_\mu(I_m))) = U_i(t_i, \nu(\cdot|B_\mu(I_m)))$ for all m. Therefore, $s_i \succ^{\mu} t_i$ for some (a fortiori, all) $\sigma \in \mathscr{C}(\mu)$ implies $s_i \succ^{\mu} t_i$. Thus, sufficiency holds in (2) as well. ■

E.3 Partially ordered probability systems and structural preferences

Definition 8 employs the extension of player i's CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$ to a CPS $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i))$. However, only the conditioning events of the form $B_\mu(I)$ for some $I \in \mathscr{I}_i$ are actually used. In addition, Definition 8 uses two CPSs: μ is used to define the likelihood relation $\geq^{\mu}$, but expected payoffs are computed using its extension ν.

One can restate the definition of structural preferences in terms of an alternative representation of beliefs that avoids both kinds of (formal) redundancy. Consider the following definition.

Definition 3 *A **partially ordered probability system (POPS)** for player $i \in N$ is a collection $(p_I)_{I \in \mathscr{I}_i} \in \Delta(S_{-i})^{\mathscr{I}_i}$ that satisfies*

1. *for every $I, J \in \mathscr{I}_i$, $p_I = p_J$ if and only if there exist $M > 1$ and $I_1, \ldots, I_M \in \mathscr{I}_i$ such that $I_1 = I_M = I$, $I_L = J$ for some $L \in \{1, \ldots, M\}$, and $p_{I_\ell}(S_i(I_\ell) \cap S_{-i}(I_{\ell+1})) > 0$ for $\ell = 1, \ldots, M-1$;*
2. *for every $I \in \mathscr{I}_i$, $p_I(\cup\{S_{-i}(J) : J \in \mathscr{I}_i, p_J = p_I\}) = 1$.*

If the CPS μ admits an extension ν, then $\left(\nu(\cdot|B_\mu(I))\right)_{I\in\mathscr{I}_i}$ is a POPS. Conversely, if $\mathbf{p}=(p_I)_{I\in\mathscr{I}_i}$ is a POPS, then one can define a CPS μ by letting $\mu(E|S_{-i}(I)) = p_I(E\cap S_{-i}(I))/p_I(S_{-i}(I))$ for all $I\in\mathscr{I}_i$ and $E\subseteq S_{-i}(I)$. (Proofs are available upon request.)

Remark 3 *Fix strategies $s_i, t_i \in S_i$. Let $\mathbf{p} = (p_I)_{I\in\mathscr{I}_i}$ be a POPS for $i\in N$, and let μ be the CPS generated by $\mathbf{p}$ as above. Then $s_i \succ^\mu t_i$ iff, for every $J\in\mathscr{I}_i$ such that $\mathrm{E}_{p_J}U_i(s_i,\cdot) < \mathrm{E}_{p_J}U_i(t_i,\cdot)$, there is $I\in\mathscr{I}_i$ such that $\mathrm{E}_{p_I}U_i(s_i,\cdot) > \mathrm{E}_{p_I}U_i(t_i,\cdot)$ and $S_{-i}(I) >^{\mathbf{p}} S_{-i}(J)$.*

Thus, as claimed above, the definition of structural preferences can be given entirely in terms of a player's POPS.

[The name "partially ordered probability system" reflects the fact that the relation $\geq^{\mathbf{p}}$ induces a partial order on the probabilities $\{p_I : I\in\mathscr{I}_i\}$: with some abuse of notation, this order is defined by $p_I \geq^{\mathbf{p}} p_J$ iff $S_{-i}(I)\geq^{\mathbf{p}} S_{-i}(J)$.]

F Unsatisfactory definitions of structural preferences

This subsection collects alternatives to Definition 8 (or, for games with nested strategic information, Definition 5) that, while apparently sensible, do not achieve the principal objective of this project—they do not imply sequential rationality.

F.1 Requiring greater payoffs at every information set

The following "eventwise dominance" definition may seem particularly close in spirit to sequential rationality:

Unsatisfactory definition DOM: $s_i \succ^\mu_{DOM} t_i$ iff, for every $I\in\mathscr{I}_i$, $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(s_i,\cdot) \geq \mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(t_i,\cdot)$.

Somewhat surprisingly, this definition actually fails to imply sequential rationality, even in simple, perfect-information games with nested strategic information. Consider for instance the Centipede game of Figure 3 in the paper, and assume that Ann's CPS μ is consistent with backward-induction reasoning. As the table in Figure 3 shows, Ann's strategy A_1D_2 does strictly better than D_1D_2 given $\mu(\cdot|S_b(I))$—that is, in case Bob chooses a at the second node. Even though D_1D_2 does strictly better than A_1D_2 given Ann's prior beliefs, Unsatisfactory Definition DOM still deems D_1D_2 and A_1A_2 incomparable. As a result, A_1D_2 is maximal in the order $\succ^{\mu}_{DOM}$, even though it is not even optimal ex-ante—let alone sequentially rational.

Notice that Unsatisfactory Definition DOM considers all information sets, rather than just the ones that are basic. However, in the example just shown, both ϕ and I are μ-basic, so modifying Unsatisfactory Definition DOM by restricting attention to basic information sets would still not resolve the issue.

This example demonstrates that, in order to deliver sequential rationality, it is crucial to take into account the likelihood ordering of (basic) information sets. Structural rationality recognizes that Ann's prior beliefs should take priority over $\mu(\cdot|S_b(I))$, and for this reason it discards A_1D_2.

F.2 A definition that considers all conditional beliefs

Definitions 5 and 8 restrict attention to basic information sets. Sequential rationality instead requires optimality at every information set. One might then be led to consider a notion that takes all information sets into account, but still ranks them in terms of likelihood:

Unsatisfactory definition ACB: $s_i \succ^{\mu}_{ACB} t_i$ iff, for every $I \in \mathscr{I}_i$ with $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(s_i,\cdot) < \mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(t_i,\cdot)$, there is $J \in \mathscr{I}_i$ such that $S_{-i}(J) >^{\mu} S_{-i}(I)$ and $\mathrm{E}_{\mu(\cdot|S_{-i}(J))}U_i(s_i,\cdot) > \mathrm{E}_{\mu(\cdot|S_{-i}(J))}U_i(t_i,\cdot)$.

To see why this definition is inadequate, consider the game in Figure 3.

Strategy L is strictly dominated for Ann. In addition, if Ann's CPS μ assigns equal probability ex-ante to a, b and c, strategy D yields strictly higher unconditional expected payoff than R, because $\epsilon > 0$. Thus, D is the unique sequentially rational strategy given μ. Furthermore, the same payoff inequality implies that it is not the case that $R \succ^{\mu}_{ACB} D$. However, consider the non-basic information set I. Given the associated belief $\mu(\cdot|S_b(I))$, R yields an expected

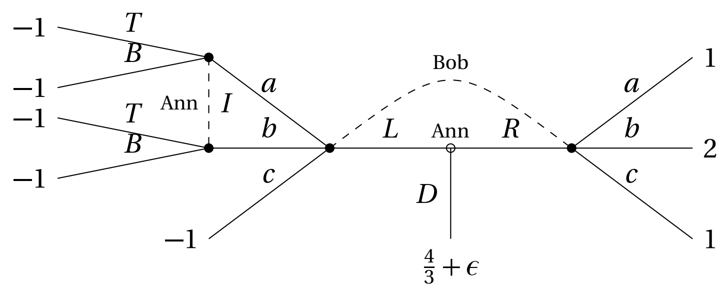


Figure 3: Ann's CPS: $\mu(a|S_b) = \mu(b|S_b) = \mu(c|S_b) = \frac{1}{3}$; $0 < \epsilon < \frac{1}{6}$.

payoff of $\frac{3}{2}$; since $\epsilon < \frac{1}{6}$, D yields a strictly lower expected payoff. As was just noted, D *does* do strictly better than R given the prior belief $\mu(\cdot|S_b)$; however, it is *not* the case that $S_b >^\mu S_b(I)$, because $\mu(S_b(I)|S_b) = \frac{2}{3}$. Hence, it is not the case that $D \succcurlyeq^\mu_{ACB} R$. So, R and D are incomparable according to Unsatisfactory Definition ACB; in particular, R is maximal, even though it is not sequentially rational.

Definition 5 avoids this issue because it restricts attention to the sole basic information set in this example, namely ϕ.

One might consider "fixing" Unsatisfactory Definition ACB by replacing the condition that $S_{-i}(J) >^\mu S_{-i}(I)$ with set inclusion:

Unsatisfactory definition ACB': $s_i \succcurlyeq^\mu_{ACB'} t_i$ iff, for every $I \in \mathscr{I}_i$ with $\mathrm{E}_{\mu(\cdot|S_{-i}(I))} U_i(s_i,\cdot) < \mathrm{E}_{\mu(\cdot|S_{-i}(I))} U_i(t_i,\cdot)$, there is $J \in \mathscr{I}_i$ such that $S_{-i}(J) \supset S_{-i}(I)$ and $\mathrm{E}_{\mu(\cdot|S_{-i}(J))} U_i(s_i,\cdot) > \mathrm{E}_{\mu(\cdot|S_{-i}(J))} U_i(t_i,\cdot)$.

One can no longer interpret the resulting preference relation as stating that s_i is "infinitely more likely" to be better than t_i, than to be worse than t_i. However, this modification does address the issue that arises in the example of Figure 3: since $S_b \supset S_b(I)$, one has $D \succ^\mu_{ACB'} R$.

Yet, Unsatisfactory definition ACB' also fails to deliver sequential rationality in general games. Consider the game in Fig. 5 of Example 3 in the paper, but now assume that Ann's beliefs μ are given by $\mu(o|S_b) = \mu(t|S_b(I)) = \mu(m|S_b(J)) = 1$. Notice that then $S_b >^\mu S_b(I) >^\mu S_b(J)$, so $S_b(\mathscr{I}_a;\mu) = S_b(\mathscr{I}_a)$: thus, the extension of μ is μ itself, and Definition 8 reduces to Definition

5, even though that game does not have nested strategic information.

Observe theat RT yields strictly higher expected payoff than RB given $\mu(\cdot|S_b(I))$, whereas the opposite is true given $\mu(\cdot|S_b(J))$. Both strategies have the same expected payoff given the prior belief $\mu(\cdot|S_b)$. Given the beliefs μ, $S_b(I) >^\mu S_b(J)$; according to Definition 8, this implies that $RT \succ^\mu RB$. However, $S_b(I)$ and $S_b(J)$ are not nested, so Unsatisfactory definition ACB' implies that RT and RB are incomparable, and hence that RB is maximal for $\succcurlyeq^\mu_{ACB'}$. Yet, RB is not sequentially rational given μ.

Thus, comparing the relative likelihood of information sets, or even basic information sets, by set inclusion alone is *not* appropriate in general games. Remark 2 in the paper *only* applies to basic information sets in games with nested strategic information.

F.3 Comparing payoffs conditional on events allowed by both strategies

The definition of structural preferences compares the expected payoff of strategies s_i, t_i given beliefs conditional upon events that may not be allowed by s_i, t_i, or even both. As noted in the main text, this is motivated by the ex-ante nature of structural preferences. However, one may consider the following alternative, which restricts attention to "common conditioning events." These are events $F \in S_{-i}(\mathscr{I}_i)$ for which there exist $I, I' \in \mathscr{I}_i$ with $s_i \in S_i(I)$, $t_i \in S_i(I')$, and $S_{-i}(I) = S_{-i}(I') = F$. (Of course, a special case is $I = I'$).

Unsatisfactory definition COM: $s_i \succcurlyeq^\mu_{COM} t_i$ iff, for all $I, I' \in \mathscr{I}_i$ such that $s_i \in S_i(I)$, $t_i \in S_i(I')$, $S_{-i}(I) = S_{-i}(I')$, and $\mathrm{E}_{\mu(\cdot|S_{-i}(I))} U_i(s_i,\cdot) < \mathrm{E}_{\mu(\cdot|S_{-i}(I))} U_i(t_i,\cdot)$, there are $J, J' \in \mathscr{I}_i$ such that $s_i \in S_i(J)$, $t_i \in S_i(J')$, $S_{-i}(J) = S_{-i}(J') \supset S_{-i}(I) = S_{-i}(I')$, and $\mathrm{E}_{\mu(\cdot|S_{-i}(J))} U_i(s_i,\cdot) > \mathrm{E}_{\mu(\cdot|S_{-i}(J))} U_i(t_i,\cdot)$.

Note: one may also consider further modifications whereby the information sets I and J are required to be basic for μ, and/or set inclusion is replaced with $>^\mu$. However, I am going to provide a counterexample in which the game satisfies nested strategic information, and in addition every conditioning event is basic. By Remark 2 in the paper, these possible modifications are thus immaterial to the argument.

One can show that, if a strategy s_i is *optimal* with respect to $\succcurlyeq^{\mu}_{COM}$ (that is, $s_i \succcurlyeq^{\mu}_{COM} t_i$ for all $t_i \in S_i$), then s_i is sequentially rational given μ. However, since $\succcurlyeq^{\mu}_{COM}$ is incomplete, optimal strategies may fail to exist. I have been unable to show that, if s_i is *maximal* with respect to $\succcurlyeq^{\mu}_{COM}$ (that is, $t_i \succ^{\mu}_{COM} s_i$ for no $t_i \in S_i$), then s_i is sequentially rational (whereas Theorem 1 establishes this implication for structural preferences). However, even if such a result were true, *it would only hold vacuously in some games.* The relation $\succcurlyeq^{\mu}_{COM}$ is not acyclic, and consequently even $\succcurlyeq^{\mu}_{COM}$–maximal strategies may fail to exist. (Structural preferences are transitive, so that maximal strategies exist for all finite games.)

To illustrate, consider the game in Figure 4. Notice that this game has nested strategic information, and a relatively simple multistage structure: Ann and Bob first move simultaneously, and then Ann makes a further choice after observing Bob's action.

Assume that Ann's CPS satisfies $\mu(\{o\}|S_b) = 1$. All information sets in $\mathscr{I}_a$ are basic for μ. Thus, as noted above, modifying Unsatisfactory Definition COM by requiring that the relevant events be basic, or replacing set inclusion with $>^{\mu}$, would not change the analysis.

To simplify the presentation, I denote Ann's strategies by indicating only the actions specified at information sets not precluded by Ann's initial choices. Thus, I write $UT\bar{T}$, without specifying whether Ann chooses T' or B' at I', etc.

First, note that $UT\bar{T} \succ^{\mu}_{COM} UT\bar{B}$. The common conditioning events for these strategies are S_b, $S_b(I) = \{t\}$ and $S_b(\bar{I}) = \{m\}$, and $DT\bar{B}$ does strictly worse than $UT\bar{T}$ conditional on $S_b(\bar{I})$—indeed, it makes a sequentially irrational choice at $\bar{I}$.

Second, $DT''\bar{T}'' \succ^{\mu}_{COM} UT\bar{T}$. The reason is that the only common conditioning events are S_b and $S_b(\bar{I}) = S_b(I'') = \{m\}$, and $DT''\bar{T}''$ yields 5 given $\mu(\cdot|\{m\})$, whereas $UT\bar{T}$ only yields 3 given $\mu(\cdot|\{m\})$.

Third, $MT'\bar{T}' \succ^{\mu}_{COM} DT''\bar{T}''$. The common conditioning events are now S_b and $S_b(\bar{I}') = S_b(\bar{I}'') = \{b\}$, and given $\mu(\cdot|\{b\})$, $MT'\bar{T}'$ does strictly better.

Finally, $UT\bar{B} \succ^{\mu}_{COM} MT'\bar{T}'$. The reason is that the only common conditioning events are S_b and $S_b(I) = S_b(I') = \{t\}$, and $UT\bar{B}$ yields 5, rather than 3, given $\mu(\cdot|\{t\})$. In particular, the fact that $UT\bar{B}$ makes the wrong choice at $\bar{I}$ is not relevant to the comparison, because $S_b(\bar{I})=\{m\}$ is *not* a commn conditioning event for $UT\bar{B}$ and $MT"\bar{T}'$.

This example demonstrates three points. First, the relation $\succsim^{\mu}_{COM}$ admits a strict cycle. Second, there is *no* maximal strategy for the relation $\succsim^{\mu}_{COM}$. In particular, the three strategies that are sequentially rational given μ, namely $UT\bar{T}$, $MT'\bar{T}'$ and $DT''\bar{T}''$, are all deemed strictly worse than some other strategy by $\succsim^{\mu}_{COM}$. Finally, a cycle can include strategies that are *not* sequentially rational.

All difficulties in this example arise because the relation $\succsim^{\mu}_{COM}$ is not transitive. In turn, this is a consequence of the fact that the set of conditioning events that determine the ranking of two given strategies depends upon the strategies themselves.[5] Structural preferences are instead defined via a *fixed* collection of conditioning events (those corresponding to the basic information sets for the player's CPS); this delivers transitivity.

References

L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica: Journal of the Econometric Society*, 59(1):61–79, 1991a.

L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and equilibrium refinements. *Econometrica: Journal of the Econometric Society*, pages 81–98, 1991b.

George J Mailath, Larry Samuelson, and Jeroen M Swinkels. Extensive form reasoning in normal form games. *Econometrica*, 61:273–302, 1993.

Martin J. Osborne and A. Rubinstein. *A Course on Game Theory.* MIT Press, Cambridge, MA, 1994.

[5]The fact that $r_i \succsim^{\mu}_{COM} s_i$ and $s_i \succsim^{\mu}_{COM} t_i$ does not necessarily yield restrictions on the payoffs conditional upon reaching information sets that are allowed by boht r_i and t_i.

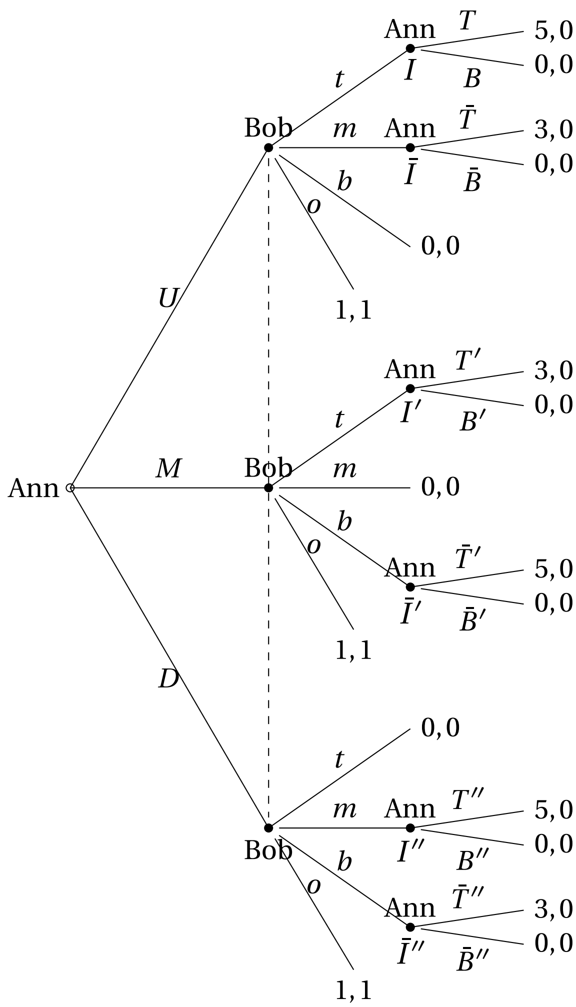


Figure 4: A strict cycle including a sequentially irrational strategy. Ann's CPS: $\mu(\{o\}|S_b) = 1$.