Structural Rationality in Dynamic Games: ONLINE APPENDIX

Marciano Siniscalchi

May 2, 2020

A Introduction

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

Section B is devoted to the extensive form, and to results that depend upon its specifics. Subsection B.1 provides a the formal definition of game trees . Subsection B.2 proves Theorem 3. Subsection B.3 defines the extensive form of the elicitation game, of which Definition 8 is a reduced representation.

Section C proves Proposition 4 and discusses structural assumptions that are sufficient to ensure that the nested-supports condition holds.

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.

Section D analyzes the elicitation example in Figure 6 of the paper.

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

Finally, Section F reviews alternative definitions of structural preferences, including one that was proposed in previous versions of this paper, and explains how the present version improves upon them. This section also collects a number of unsatisfactory definitions of preferences over strategies that, while apparently capturing certain intuitions about (weak) sequential rationality, they actually fail to formally imply it.

B 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 B.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 CCPS 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.[1] 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 $i \in N$.

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.[2] 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: since this requires notation that is also used in the proof of Theorem 3, I prove it in the next subsection.

B.1 Further details and properties of extensive-form games

This subsection contains more detailed definitions and properties of game trees. All results are essentially known and included here only for ease of reference and notational consistency. 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 $h' = (h, b_1, \ldots, b_M)$

[1] 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 $X_i(h) = (I_1, a_{\ell_1}, \ldots, I_k, a_{\ell_k})$.

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

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 $s_i \in S_i(I)$.

The following properties are immediate consequences of the definitions.

Remark 1 *(i) For every $z, z' \in Z$ there is $h \in H \setminus Z$ such that (a) $h < z$ and $h < z'$, and (b) for all $h' \in H$ with $h' < z$ and $h' < z'$, $h' \leq h$.*

(ii) For all $I, J \in \mathscr{I}_i$, if $I < J$ then $S(I) \supseteq S(J)$.

Proof: (i) Write $z = (a_1, \ldots, a_L)$ and $z' = (b_1, \ldots, b_M)$. If $z = z'$, then take $h = (a_1, \ldots, a_{L-1}) = (b_1, \ldots, b_{M-1})$; this may be the empty history if $L = M = 1$. Otherwise, there is $m \in \{1, \ldots, \min(L, M)\}$ such that $a_\ell = b_\ell$ for $1 \leq \ell \leq m-1$, and $a_m \neq b_m$. Then take $h = (a_1, \ldots, a_{m-1})$; again, this may be the empty history if $m = 1$.

(ii) Fix $s \in S(J)$. By definition, there is $h \in J$ with $h < \zeta(s)$. But $J < J$ implies that $h' < h$ for some $h' \in I$. Then $h' < \zeta(s)$, so $s \in S(I)$. ■

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 3.

Remark 2 *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 3 *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* $s_{-i} \notin S_{-i}(I)$.

[This argument is due to Mailath, Samuelson, and Swinkels (1993).]

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 $I \leq J$, and $r_i(J) = s_i(J)$ otherwise. By Remark 2, 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 $I \leq J$, 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 2) $(s_i, s_{-i}) \notin S(I)$, and hence also $(s_i, s_{-i}) \notin S(J)$ for any $J \in \mathscr{I}_i$ with $I \leq J$. Then $(s_i, s_{-i}) \in S(J)$ implies that not $I \leq J$, 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 $U_i(r_i, s_{-i}) = u_i(\zeta(r_i, s_{-i})) = u_i(\zeta(s_i, s_{-i})) = U_i(s_i, s_{-i})$. ■

B.2 Proof of Theorem 3

Assume that s_i is weakly sequentially rational given μ, but there is $t_i \in S_i$ such that $t_i \succ^\mu s_i$. For every $I \in \mathscr{I}_i$, let $B_\mu(I) = \cup\{S_{-i}(J) : J =^\mu I\}$.

I first analyze the case in which $U_i(s_i, s_{-i}) \leq U_i(t_i, s_{-i})$ for all s_{-i}. In this case, $U_i(t_i, P_\mu(I)) \geq U_i(s_i, P_\mu(I))$ for all $I \in \mathscr{I}_i$, and Theorem 1 implies that $U_i(t_i, P_\mu(I^*)) > U_i(s_i, P_\mu(I^*))$ for some $I^* \in \mathscr{I}_i$. Hence, there must be $t_{-i} \in B_\mu(I^*)$ such that $U_i(t_i, t_{-i}) > U_i(s_i, t_{-i})$. In particular, this means that $z \equiv \zeta(s_i, t_{-i}) \neq \zeta(t_i, t_{-i}) \equiv z'$. Now let h be the longest non-terminal history such that $h < z$ and $h < z'$, per Remark 1 part (i). Then $P(h) = i$: otherwise, the move at h is determined by t_{-i}, so $h < (h, t_{P(h)}(h)) < z, z'$, contradiction. Let $J^* \in \mathscr{I}_i$ be such that $h \in J^*$. Then $s_i(J^*) \neq t_i(J^*)$: otherwise, $h < (h, a) < z, z'$, where $a = s_i(J^*) = t_i(J^*)$, contradiction.

By weak sequential rationality, $U_i(s_i, \mu(\cdot|J^*)) \geq U_i(t_i, \mu(\cdot|J^*))$. Since $U_i(s_i, s_{-i}) \leq U_i(t_i, s_{-i})$ for all s_{-i}, $U_i(s_i, \mu(\cdot|J^*)) = U_i(t_i, \mu(\cdot|J^*))$. Hence there is $t^*_{-i} \in S_{-i}(J^*)$ such that $U_i(s_i, t^*_{-i}) = U_i(t_i, t^*_{-i})$. But since $s_i(J^*) \neq t_i(J^*)$, $\zeta(s_i, t^*_{-i}) \neq \zeta(t_i, t^*_{-i})$. Therefore, there is a relevant tie for i at J^*.

Now suppose that there is $s^*_{-i} \in S_{-i}$ such that $U_i(s_i, s^*_{-i}) > U_i(t_i, s^*_{-i})$. Since $t_i \succ^\mu s_i$, by Theorem 1 there are $\tilde{I}, \tilde{J} \in \mathscr{I}_i$ with $\tilde{I} \geq^\mu \tilde{J}$, $s^*_{-i} \in S_{-i}(\tilde{J})$, and $U_i(t_i, P_\mu(\tilde{I})) > U_i(s_i, P_\mu(\tilde{I}))$.

Let $I \in \mathscr{I}_i$ be a $\geq^\mu$-maximal element of

$$\tilde{\mathscr{I}}_i = \{I' \in \mathscr{I}_i : I' \geq^\mu \tilde{J}, U_i(t_i, P_\mu(I')) > U_i(s_i, P_\mu(I'))\}. \tag{1}$$

Such an element exists because $\tilde{I} \in \tilde{\mathscr{I}}_i$, $\geq^\mu$ is transitive, and $\mathscr{I}_i$ is finite.

Also, define the set

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

Since $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$, $D \neq \emptyset$. Also, for every $s_{-i} \notin D$, $[U_i(s_i, s_{-i}) - U_i(t_i, s_{-i})] \cdot P_\mu(I)(s_{-i}) = 0$, because either $\zeta(s_i, s_{-i}) = \zeta(t_i, s_{-i})$, so that the term in square brackets is zero, or the probability of s_{-i} is zero (or both).

For every $s_{-i} \in D$, let $\mathscr{J}_i(s_{-i}) = \{J' \in \mathscr{I}_i : (s_i, s_{-i}), (t_i, s_{-i}) \in S(J')\}$, which is non-empty because it contains ϕ. Then let $J(s_{-i})$ a $<$-maximal element of $\mathscr{J}_i(s_{-i})$.

I claim that, 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. By Remark 2 part (i), $S(J(s_{-i}))$ and $S(J(s'_{-i}))$ are nested. 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 $r_{-i} \in S_{-i}(J(s_{-i}))$, as claimed.

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 $m = 1, \ldots, M$, $P_\mu(I)(S_{-i}(J(s^m_{-i}))) \geq P_\mu(I)(\{s_{-i}\}) > 0$. In particular, by Corollary 4 in the text, $J(s^m_{-i}) \geq^\mu I$; since $S_{-i}(J(s^m_{-i})) \subseteq B_\mu(I)$, also $\mu(B_\mu(I))|J(s^m_{-i}) > 0$, so $P_\mu(J(s^m_{-i}))(B_\mu(I)) > 0$ and therefore, by the same Corollary, $I \geq^\mu J(s^m_{-i})$: thus, $I =^\mu J(s^m_{-i})$. 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 $[U_i(s_i, s_{-i}) - U_i(t_i, s_{-i})] \cdot P_\mu(I)(\{s_{-i}\}) = 0$. Therefore,

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

The last equality follows from the properties of $P_\mu(I)$ and the fact that $I =^\mu J(s^m_{-i})$. The inequality follows from the assumption that s_i is sequentially rational for μ given u. But this conclusion contradicts the assumption that $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$. 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 $s_{-i} \in D$, $P_\mu(I)(S_{-i}(J(s_{-i}))) > 0$, so by Corollary 4, $J(s_{-i}) \geq^\mu I$. Suppose that also $I \geq^\mu J(s_{-i})$, so $J(s_{-i}) =^\mu I$: then $B_\mu(J(s_{-i})) = B_\mu(I)$ and so $S_{-i}(J(s_{-i})) \subseteq B_\mu(I)$, contradiction: thus, not $I \geq^\mu J(s_{-i})$, and so $J(s_{-i}) >^\mu I$.

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 $\zeta(t_i, s_{-i}) = (b_1, \ldots, b_M)$. 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 2 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 2 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 $h_\ell < \zeta(t_i, s_{-i})$, $(s_i, s_{-i}), (t_i, s_{-i}) \in S(J)$: but then, $J(s_{-i})$ is not the $<$-maximal element of $\mathscr{J}_i(s_{-i})$, contradiction. 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, sequential rationality implies that $U_i(s_i, \mu(\cdot|J(s_{-i})) \geq U_i(t_i, \mu(\cdot|J(s_{-i}))$, so there is $t_{-i} \in \text{supp}\ \mu(\cdot|J(s_{-i}))$ such that $U_i(s_i, t_{-i}) \geq U_i(t_i, t_{-i})$. By contradiction, suppose that $U_i(s_i, t_{-i}) > U_i(t_i, t_{-i})$. Then $t_i \succ^\mu s_i$ implies that there must be $\bar{I}, \bar{J} \in \mathscr{I}_i$ with $t_{-i} \in S_{-i}(\bar{J})$, $\bar{I} \geq^\mu \bar{J}$, and $U_i(t_i, P_\mu(\bar{I})) > U_i(s_i, P_\nu(\bar{I}))$. Then $\mu(S_{-i}(\bar{J})|J(s_{-i})) \geq \mu(\{t_{-i}\}|J(s_{-i})) > 0$, so $\bar{J} \geq^\mu J(s_{-i})$ by Corollary 4. By transitivity of $\geq^\mu$, $\bar{I} \geq^\mu J(s_{-i})$. Furthermore, as shown above, $J(s_{-i}) >^\mu I$, so transitivity also implies that $\bar{I} >^\mu I \geq^\mu \tilde{J}$. But this contradicts the choice of I as a $\geq^\mu$-maximal element of the set $\tilde{\mathscr{I}}_i$ in Eq. (1). Therefore, $U_i(s_i, t_{-i}) = U_i(t_i, t_{-i})$. But since $s_i(J(s_{-i})) \neq t_i(J(s_{-i}))$, $\zeta(s_i, t_{-i}) \neq \zeta(t_i, t_{-i})$. Thus, there is a relevant tie at $J(s_{-i})$. ■

B.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 8.

Begin with a description of the elicitation game tree. The player set is $N^* = N \cup \{c\}$; the action set is $A^* = A \cup \{h, t\}$. 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 $A_c^1 \equiv \{h, t\}$. Then, players move according to their index; player i chooses from $A_i^1 \equiv S_i \times W_i$. Hence, stage-1 histories are of the form

$$\phi^* \qquad \text{or} \qquad \big(a_c, (s_1, w_1), \ldots, (s_{i-1}, w_{i-1})\big): \quad a_c^1 \in A_c,\ (s_j, w_j) \in A_j^1\ j = 1, \ldots, i-1. \tag{2}$$

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

$$\big(a_c, (s_1, w_1), \ldots, (s_N, w_n), h\big): \qquad (s_1, \ldots, s_N) \in S(h). \tag{3}$$

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

$$I_i^1 = \Big\{(a_c, \big(s_1, w_1), \ldots, (s_{i-1}, w_{i-1})\big) \in H^* : a_c \in A_c^1, (s_j, w_j) \in A_j^1,\ j = 1, \ldots, i-1\Big\}. \tag{4}$$

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 8,

$$(s_i, w_i, I) = \left\{ \left(a_c, (\bar{s}_i, \bar{w}_i), (s_{-i}, w_{-i}), h\right) \in H^* : \bar{s}_i = s_i, \bar{w}_i = w_i,\ s_{-i} \in S_{-i}(h),\ h \in I \right\}. \tag{5}$$

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 8, 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 CCPS 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

$$u_i^*\big((a_c, s, w, z)\big) = \begin{cases} u_i(z) & a_c = h \\ 1 & a_c = t, Q_i = (I, (E, p)), w_i = E, s_{-i} \in E \\ 0 & a_c = t, w_i = E, s_{-i} \notin E \\ p & a_c = t, Q_i = (I, (E, p)), w_i = p, s_{-i} \in S_{-i}(I). \end{cases} \tag{6}$$

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 8.

Chance has a unique information set ϕ^*, with action set A_c^1, so $S_c^* = A_c^1 = \{h, t\}$.

Now consider player $i \in N$. Eq. (2) and Eq. (3) 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. (5) 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 $h \in I$; Eq. (3) 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 8 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. 2 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. (5) 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. (3), 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. (6) implies that

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

This completes the proof.

C Computational considerations

C.1 Proof of Proposition 4

Claim 1: for all μ-sequences $I_1,\dots,I_L$ such that $I_L,\dots,I_1$ is also a μ-sequence, there is $\ell \in \{1,\dots,L\}$ such that $\operatorname{supp}\mu(\cdot|I_\ell) \supseteq \operatorname{supp}\mu(\cdot|I_m)$ for all $m \in \{1,\dots,L\}$.

If $L = 1$, the claim is trivially true. Thus, suppose it is true for some $L \geq 1$, and consider a μ-sequence $I_1,\dots,I_L,I_{L+1}$. Then the inductive hypothesis yields $\ell \in \{1,\dots,L\}$ such that $\operatorname{supp}\mu(\cdot|I_\ell) \supseteq \operatorname{supp}\mu(\cdot|I_m)$ for all $m \in \{1,\dots,L\}$. In particular, $\operatorname{supp}\mu(\cdot|I_\ell) \supseteq \operatorname{supp}\mu(\cdot|I_L)$.

I claim that $\mu(S_{-i}(I_{L+1})|I_\ell) > 0$. By assumption $\mu(S_{-i}(I_{L+1}|I_L) > 0$, so there is $s_{-i} \in S_{-i}(I_{L+1}) \cap \operatorname{supp}\mu(\cdot|I_L)$. By the inductive hypothesis, $s_{-i} \in \operatorname{supp}\mu(\cdot|I_\ell)$, so $\mu(S_{-i}(I_{L+1})|I_\ell) \geq \mu(\{s_{-i}\}|I_\ell) > 0$.

Next, I claim that $\mu(S_{-i}(I_\ell)|I_{L+1}) > 0$. Again, by assumption $\mu(S_{-i}(I_L L)|I_{L+1}) > 0$, and since $\mu(S_{-i}(I_{L+1})|I_L) > 0$ as well, by nested supports either $\operatorname{supp}\mu(\cdot|I_L) \supseteq \operatorname{supp}\mu(\cdot|I_{L+1})$ or $\operatorname{supp}\mu(\cdot|I_{L+1}) \supseteq \operatorname{supp}\mu(\cdot|I_L)$. In the first case, $\mu(S_{-i}(I_\ell)|I_{L+1}) \geq \mu(\operatorname{supp}\mu(\cdot|I_\ell)|I_{L+1}) \geq \mu(\operatorname{supp}\mu(\cdot|I_L)|I_{L+1}) \geq \mu(\operatorname{supp}\mu(\cdot|I_{L+1})|I_{L+1}) = 1$. In the second case, since $\operatorname{supp}\mu(\cdot|I_\ell) \supseteq \mu(\cdot|I_L)$, there is $s_{-i} \in \operatorname{supp}\mu(\cdot|I_L)$ with $s_{-i} \in \operatorname{supp}\mu(\cdot|I_\ell) \subseteq S_{-i}(I_\ell)$; since $\operatorname{supp}\mu(\cdot|I_{L+1}) \supseteq \operatorname{supp}\mu(\cdot|I_L)$, also $s_{-i} \in \operatorname{supp}\mu(\cdot|I_{L+1})$; but then $\mu(S_{-i}(I_\ell)|I_{L+1}) \geq \mu(\{s_{-i}\}|I_{L+1}) > 0$, as claimed.

Then, by nested supports, either $\operatorname{supp}\mu(\cdot|I_\ell) \supseteq \operatorname{supp}\mu(\cdot|I_{L+1})$ or $\operatorname{supp}\mu(\cdot|I_{L+1}) \supseteq \operatorname{supp}\mu(\cdot|I_\ell)$. In the first case, ℓ has the required property for the μ-sequence $I_1,\dots,I_{L+1}$; in the second, $L+1$ does. In either case, the inductive step is complete.

By Lemma 1 in the paper, the elements of the $\geq^\mu$-equiv class of any $I \in \mathscr{I}_i$ can be arranged (potentially with duplicates) in a μ-sequence that begins and ends with I. Let $I_1,\dots,I_L$ be such a μ-sequence; by Corollary 3, since $I_1 = I_L = I$, $I_L,\dots,I_1$ is also a μ-sequence. Hence, Claim 1 applies; denote by ℓ the index with the properties therein.

Since $I_\ell =^\mu I$, $P_\mu(I)(S_{-i}(I_\ell)) > 0$ and $\mu(\cdot|I_\ell) = P_\mu(I)(\cdot|S_{-i}(I_\ell))$ by Proposition 3. Moreover, consider $s_{-i} \in \operatorname{supp} P_\mu(I)$. Since $P_\mu(I)(\cup\{S_{-i}(I_m) : m = 1,\dots,L\}) = 1$ by the same Proposition and the definition of the μ-sequence $I_1,\dots,I_L$, $s_{-i} \in S_{-i}(I_m)$ for some m. Furthermore,

since $P_\mu(I)(S_{-i}(I_m)) > 0$ and $\mu(\cdot|I_m) = P_\mu(I)(\cdot|S_{-i}(I_m))$, $\mu(\{s_{-i}\}|I_m) > 0$. Then $s_{-i} \in \text{supp}\,\mu(\cdot|I_m) \subseteq \text{supp}\,\mu(\cdot|I_\ell)$. Therefore, $\text{supp}\,P_\mu(I) \subseteq \text{supp}\,\mu(\cdot|I_\ell)$, so $P_\mu(I)(S_{-i}(I_\ell)) \geq P_\mu(I)(\text{supp}\,\mu(\cdot|I_\ell)) \geq P_\mu(I)(\text{supp}\,P_\mu(I)) = 1$, and so $P_\mu(I) = \mu(\cdot|I_\ell)$, as claimed.

C.2 A sufficient condition for nested supports

The following condition can be checked directly on the game form:

Definition 1 *A dynamic game has **nested information** if, for all $i \in N$ and $I, J \in \mathscr{I}_i$, either $S_{-i}(I) \cap S_{-i}(J) = \emptyset$ or $S_{-i}(I) \subseteq S_{-i}(J)$ or $S_{-i}(J) \subseteq S_{-i}(I)$*

It is immediate to see that nested information implies nested supports for every player i and CCPS $\mu \in \Delta(S_{-i}, \mathscr{I}_i)$: if $\mu(S_{-i}(I)|J) > 0$ and $\mu(S_{-i}(J)|I) > 0$, then $S_{-i}(I) \cap S_{-i}(J) \neq \emptyset$; if $S_{-i}(I) \subseteq S_{-i}(J)$, then for all $E \subseteq S_{-i}(I)$, $\mu(E|I) = \frac{\mu(E|J)}{\mu(S_{-i}(I)|J)}$ by Eq. (1) in the paper, so $\text{supp}\,\mu(\cdot|I) \subseteq \text{supp}\,\mu(\cdot|J)$; similarly, if $S_{-i}(J) \subseteq S_{-i}(I)$, then $\text{supp}\,\mu(\cdot|J) \subseteq \text{supp}\,\mu(\cdot|I)$.

Any game in which every player moves at most once on every path has nested information. Fix one such game, a player i, and $I, J \in \mathscr{I}_i$. If (wlog) $I = \phi$, then trivially $S_{-i}(I) = S_{-i} \supseteq S_{-i}(J)$. Otherwise, I and J belong to distinct paths: that is, if $h \in I$, there is no $h' < h$ such that $h' \in J$, and conversely. Now fix $s_{-i} \in S_{-i}(I)$. By contradiction, assume $s_{-i} \in S_{-i}(J)$ as well. Let $s_i \in S_i(I)$ and $t_i \in S_i(J)$. By perfect recall, $(s_i, s_{-i}) \in S(I)$ and $(s_i, s_{-i}) \in S(J)$: that is, there are $h, h' \in H \setminus Z$ such that $h < \zeta(s_i, s_{-I})$, $h' < \zeta(t_i, s_{-i})$, $h \in I$, and $h' \in J$. Let h'' be the longest history such that $h'' \leq h$ and $h'' \leq h'$. Since i moves only once on each path, $P(h'') \neq i$; in particular, $h'' \notin \{h, h'\}$. Moreover, by definition, there must be actions $a, a' \in A$ such that $(h'', a) \leq h$ and $(h'', a') \leq h'$, with $a \neq a'$. But since $P(h'') \neq i$, $s_{P(h'')}(h'') = a \neq a' = s_{P(h'')}(h'')$, contradiction. Thus, $s_{-i} \notin S_{-i}(J)$. Similarly, if $s_{-i} \in S_{-i}(J)$, then $s_{-i} \notin S_{-i}(I)$, so the game has nested information.

One can model a signalling game as a three-player game, in which the first player is Nature, the second is the sender, and the third is the receiver. Each player moves only once on each

path, so the game has nested information, and hence every CCPS for each of the personal players has nested supports. Selten's Horse is also a game in which every player moves once on each path.

Finally, the Battle of the Sexes with an outside option, Burning Money, and centipede games all feature nested information. On the other hand, the game in Figure 4 does not have nested information, and the CCPS in Example 2 does not have nested supports; however, the CCPS $\nu \in \Delta(S_{-i}, \mathscr{I}_a)$ defined by $\nu(\{d\}|\phi) = \nu(\{a\}|I) = \nu(\{c\}|J)$ does have nested supports.

D Calculations for the game in Figure 6 (Section 6.2)

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 $\mu(\{(Out, b), (Out, p)\}|\phi) = 1$ and $\mu(\{(InS, b), (InS, p)\}|I) = \mu(\{(InS, b), (InS, p)\}|I') = \pi$. One readily verifies that $P_\mu(I) = P_\mu(I') = \mu(\cdot|I) = \mu(\cdot|I')$. Table I summarizes Bob's payoffs as a function of Ann's strategy, as well as his conditional expected payoffs.

s_b	$(Out, b), (Out, p)$	$(InB, b), (InB, p)$	$(InS, b), (InS, p)$	ϕ	I, I'
$\bar{B}B$	2	1	0	2	$1-\pi$
$\bar{S}S$	2	0	3	2	3π

Table I: Payoffs and expected payoffs for Bob in Figure 6.

By Theorem 1, since Bob's payoff is 2 if Ann plays (Out, p) or (Out, b), and $S_b(I) = S_b(I') = S_a \setminus \{(Out, b), (Out, p)\}$, the ranking of $\bar{B}B$ vs. $\bar{S}S$ is pinned down by expected payoffs given $\mu(\cdot|I) = \mu(\cdot|I')$. 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 weakly 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.

E Equivalent definitions of structural preferences

E.1 LPSs and Proof of Theorem 5

Following Blume, Brandenburger, and Dekel (1991b), for an LPS $\lambda = (p_1, \ldots, p_n)$ and a vector $r = (r_1, \ldots, r_{n-1}) \in (0,1)^{n-1}$, let $r \square \lambda = (1-r_1)p_1 + r_1((1-r_2)p_2 + r_2 \ldots + r_{n-1}p_n)\ldots)$.

(Necessity): Assume that $t_i \succ^{\mu} s_i$ and consider an LPS $\lambda = (p_1, \ldots, p_n)$ that generates μ.

Claim: for any sequence $(r^k)_{k\geq 1} \subset (0,1)^{n-1}$ with $r^k \to 0$, $(r^k \square \lambda)_{k\geq 1}$ is a perturbation of μ.

Proof: since λ generates μ, for every $I \in \mathscr{I}_i$ there is m with $p_m(S_{-i}(I)) > 0$ and $p_\ell(S_{-i}(I)) = 0$ for $\ell = 1, \ldots, m-1$. Thus $(r^k \square \lambda)(S_{-i}(I)) > 0$ for all k. Moreover, for every $E \subseteq S_{-i}(I)$,

$$
\begin{aligned}
(r^k \square \lambda)(E|S_{-i}(I)) &= \frac{\prod_{m'=1}^{m-1} r^k_{m'} \cdot (1-r^k_m) p_m(E) + \sum_{\ell=m+1}^{n} \prod_{m'=1}^{\ell-1} r^k_{m'} \cdot (1-r^k_\ell) p_\ell(E)}{\prod_{m'=1}^{m-1} r^k_{m'} \cdot (1-r^k_m) p_m(S_{-i}(I)) + \sum_{\ell=m+1}^{n} \prod_{m'=1}^{\ell-1} r^k_{m'} \cdot (1-r^k_\ell) p_\ell(S_{-i}(I))} = \\
&= \frac{(1-r^k_m) p_m(E) + \sum_{\ell=m+1}^{n} \prod_{m'=m}^{\ell-1} r^k_{m'} \cdot (1-r^k_\ell) p_\ell(E)}{(1-r^k_m) p_m(S_{-i}(I)) + \sum_{\ell=m+1}^{n} \prod_{m'=m}^{\ell-1} r^k_{m'} \cdot (1-r^k_\ell) p_\ell(S_{-i}(I))} = \\
&= \frac{p_m(E|S_{-i}(I)) + \sum_{\ell=m+1}^{n} \prod_{m'=m}^{\ell-1} r^k_{m'} \cdot \frac{(1-r^k_\ell) p_\ell(E)}{(1-r^k_m) p_m(S_{-i}(I))}}{1 + \sum_{\ell=m+1}^{n} \prod_{m'=m}^{\ell-1} r^k_{m'} \cdot \frac{(1-r^k_\ell) p_\ell(S_{-i}(I))}{(1-r^k_m) p_m(S_{-i}(I))}} \to \mu(E|I):
\end{aligned}
$$

the second equality follows by dividing the numerator and the denominator by the common factor $\prod_{m'=1}^{m-1} r^k_{m'}$, the third by dividing the numerator and denominator by $(1-r^k_m) p_m(S_{-i}(I)) > 0$; and the limit follows because $(p^k)_{k\geq 1}$ is a perturbation of μ, and each summation is either empty (if $m = n$), or contains terms multiplied by at least one factor $r^k_{m'} \to 0$. This proves the claim.

Since $(r^k \square \lambda)_{k\geq 1}$ is a perturbation of μ, by assumption $U_i(t_i, r^k \square \lambda) > U_i(s_i, r^k \square \lambda)$ eventually. By Proposition 1 in Blume et al. (1991b) (see the Remark at the end of the proof in the Appendix of that paper), $t_i \succ^\lambda s_i$. This completes the proof of necessity.

(Sufficiency): Suppose that $t_i \succ^\sigma s_i$ for all LPSs σ that generate μ. To show that $t_i \succ^\mu s_i$, I leverage Theorem 1 in the paper.

As in Section B, for $I \in \mathscr{I}_i$, let $B_\mu(I) = \cup\{S_{-i}(J) : J =^\mu I\}$. Also write $I \geq^\mu s_{-i}$ for $I \in \mathscr{I}_i$ and $s_{-i} \in S_{-i}(I)$ to mean that there is $J \in \mathscr{I}_i$ with $I \geq^\mu J$ and $s_{-i} \in S_{-i}(J)$. Thus in particular $I \geq^\mu s_{-i}$ for all $s_{-i} \in S_{-i}(I)$.

I invoke definitions and results from Appendix A in the paper. Let $I_1, \ldots, I_M$ be a representative collection for μ, and f the canonical map for $I_1, \ldots, I_M$.

Claim: Consider a one-to-one function $g : \{1, \ldots, M\} \to \mathbb{R}$ that agrees with $>^\mu$, and define the LPS $\lambda = (P_\mu(I_{n_1}), \ldots, P_\mu(I_{n_M}))$, where $n_1, \ldots, n_M$ is the unique permutation of $1, \ldots, M$ such that $\ell < m$ iff $g(I_{n_\ell}) < g(I_{n_m})$. Then λ generates μ.

Proof: for every I there is m such that $I =^\mu I_{n_m}$, and thus $P_\mu(I_{n_m})(S_{-i}(I)) > 0$ and $P_\mu(I_{n_m})(\cdot|S_{-i}(I)) = \mu(\cdot|I)$ by the properties of $P_\mu(\cdot)$. Furthermore, suppose that $P_\mu(I_{n_\ell})(S_{-i}(I)) > 0$, so that $P_\mu(I_{n_\ell})(B_\mu(I_{n_m})) > 0$. By Corollary 4, $I_{n_m} \geq^\mu I_{n_\ell}$. In particular, if $m \neq \ell$, then $I_{n_m} >^\mu I_{n_\ell}$, so $g(I_{n_m}) < g(I_{n_\ell})$ and therefore $m < \ell$. Thus, for $\ell < m$, $P_\mu(I_{n_\ell})(S_{-i}(I)) = 0$. Thus λ generates μ.

To invoke Theorem 1, it is sufficient to show that (i) there is at least one $I \in \mathscr{I}_i$ such that $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$, and (ii) for all $s_{-i} \in S_{-i}$, if $U_i(t_i, s_{-i}) < U_i(s_i, s_{-i})$ then there is $I \in \mathscr{I}_i$ such that $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$ and $I \geq^\mu s_{-i}$. [This is because then one can take the information sets in the statement of Theorem 1 to be all $I \in \mathscr{I}_i$ for which $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$.]

For (i), consider the LPS λ obtained by taking $g = f$ in the above Claim. Then λ generates μ and, by assumption, $t_i \succ^\lambda s_i$, so (in the notation of the Claim) there must be some m such that $U_i(t_i, P_\mu(I_{n_m})) > U_i(s_i, P_\mu(I_{n_m}))$.

Now consider (ii) and let $s_{-i} \in S_{-i}$ be such that $U_i(s_i, s_{-i}) < U_i(t_i, s_{-i})$. Let $\mathscr{I}_i^+ = \{I \in \mathscr{I}_i : I \geq^\mu s_{-i}\}$. By contradiction, assume that $U_i(t_i, P_\mu(I)) \leq U_i(s_i, P_\mu(I))$ for all $I \in \mathscr{I}_i^+$.

Define $g:\{1,\dots,M\}\to\mathbb{R}$ by letting $g(m)=f(m)$ if $I_m\in\mathscr{I}_i^+$, $f_{\max}=\max_{m:I_m\in\mathscr{I}_i^+}f(m)$, and $g(m)=f(m)+f_{\max}+1$ otherwise. Then g is one-to-one. Furthermore, suppose $I_\ell >^\mu I_m$. If $I_m\geq^\mu s_{-i}$, then also $I_\ell\geq^\mu s_{-i}$, and so $I_\ell, I_m\in\mathscr{I}_i^+$, which implies that $g(\ell)=f(\ell)<f(m)=g(m)$. If not $I_m\geq^\mu s_{-i}$ but $I_\ell\geq^\mu s_{-i}$, then $g(\ell)=f(\ell)<f_{\max}+1<f(m)+f_{\max}+1=g(m)$. Finally, if neither $I_m\geq^\mu s_{-i}$ nor $I_\ell\geq^\mu s_{-i}$, then $g(\ell)=f(\ell)+f_{\max}+1)<f(m)+f_{\max}+1=g(m)$. Thus, g agrees with $>^\mu$, so by the Claim, the LPS $\lambda=(P_\mu(I_{n_1}),\dots,P_\mu(I_{n_M}))$ induced by g generates μ (again, $n_1,\dots,n_M$ are as in the Claim).

By construction, there is m such that $I_{n_\ell}\in\mathscr{I}_i^+$ iff $\ell\leq m$. Let

$$\rho=(P_\mu(I_{n_1}),\dots,P_\mu(I_{n_m}),\delta_{s_{-i}},P_\mu(I_{n_{m+1}}),\dots P_\mu(I_{n_M})),$$

where as usual $\delta_{s_{-i}}$ is the Dirac measure concentrated on s_{-i}. Then ρ also generates μ: if $\delta_{s_{-i}}(S_{-i}(K))>0$ for some $K\in\mathscr{I}_i$, then $s_{-i}\in S_{-i}(K)$, so $S_{-i}(K)\geq^\mu\{s_{-i}\}$ and therefore $K=^\mu I_{n_\ell}$ for some n_ℓ with $I_{n_\ell}\in\mathscr{I}_i^+$; but then, $\ell\leq m$ and $P_\mu(I_{n_\ell}))(S_{-i}(K))=P_\mu(K)(S_{-i}(K))>0$. Thus, $\delta_{s_{-i}}$ is never the first measure to assign positive probability to any information set K. On the other hand, λ does generate μ, and therefore so does ρ. By construction, for all $\ell\leq m$, $I_{n_\ell}\in\mathscr{I}_i^+$, and by assumption this implies $U_i(s_i,P_\mu(I_{n_\ell}))\leq U_i(t_i,P_\mu(I_{n_\ell}))$. Furtherore, again by assumption $U_i(s_i,\delta_{s_{-i}})<U_i(t_i,\delta_{s_{-i}})$. Hence $t_i\prec^\rho s_i$: contradiction.Thus, (ii) must hold.

E.2 Partially ordered probability systems and structural preferences

Theorem 1 employs the belief $P_\mu(I)$ associated with each $I\in\mathscr{I}_i$ via Definition 6. The CCPS μ only appears indirectly, via the definition of the preorder $\geq^\mu$ per Definition 5.

One can restate the definition of structural preferences in terms of an alternative representation of beliefs that, in a sense, involves only the objects of interest—the beliefs $P_\mu(I)$, $I\in\mathscr{I}_i$. Consider the following definition.

Definition 2 *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* $I \in \mathscr{I}_i$, $p_I(\cup\{S_{-i}(J) : J \in \mathscr{I}_i, p_J = p_I\}) = 1$.

Notice that, in part 2 of the definition, one can take $I = J$, $M = 2$, and $I_1 = I_2$ (so e.g. $J = 1$) to obtain $p_I(S_{-i}(I)) > 0$.

Given a POPS $\mathbf{p}$, one can define a preorder $\geq^{\mathbf{p}}$ on $\mathscr{I}_i$ by letting $I \geq_0^{\mathbf{I}} J$ iff $p_J(S_{-i}(I)) > 0$, and letting $\geq^{\mathbf{p}}$ be the transitive closure of $\geq_0^{\mathbf{p}}$, as in Definition 5 in the paper. Then part 2 of the definition states that $p_I = p_J$ iff $I =^{\mathbf{p}} J$. Furthermore, when restricted to its equivalence classes, $\geq^{\mathbf{p}}$ is a partial order, and hence induces a partial order over $\{p_I : I \in \mathscr{I}_i\}$.

For any CCPS μ, the collection $\left(P_\mu(I)\right)_{I \in \mathscr{I}_i}$ is a POPS. Conversely, if $\mathbf{p} = (p_I)_{I \in \mathscr{I}_i}$ is a POPS, then one can define an array $\mu = (\mu(\cdot|I))_{I \in \mathscr{I}_i}$ by letting $\mu(E|I) = p_I(E \cap S_{-i}(I))/p_I(S_{-i}(I))$ for all $I \in \mathscr{I}_i$ and $E \subseteq S_{-i}(I)$; the properties in Definition 2 imply that μ is a CCPS. Furthermore, in either case $\geq^{\mathbf{p}} = \geq^{\mu}$. (Proofs are available upon request.)

Remark 4 *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* μ *the CCPS it generates. Then* $t_i \succ^{\mu} s_i$ *iff there exist* $M > 1$ *and* $I_1, \ldots, I_M \in \mathscr{I}_i$ *such that* $U_i(t_i, p_{I_m}) > U_i(s_i, p_{I_m})$ *for* $m = 1, \ldots, M$, *and* $U_i(t_i, s_{-i}) \geq U_i(t_i, s_{-i})$ *for all* $s_{-i} \notin \bigcup_m \bigcup_{J : I_m \geq^{\mathbf{p}} J} S_{-i}(J)$.

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

F Unsatisfactory definitions of structural preferences

This subsection first briefly comments on the definition of structural preferences proposed in previous versions of this paper, and on an alternative definition based on perturbations. It emphasizes in what respects the definition in the present version is preferable—in particular, as regards minimality.

It then examines alternatives to Definition 4 (or, more precisely, the characterization thereof in Theorem 1) that, while apparently sensible, do not imply weak sequential rationality.

F.1 Non-minimal definitions of structural preferences

In previous versions of this paper, structural preferences were defined as a weak, rather than a strict order.

Definition 3 *For $s_i, t_i \in S_i$, $t_i \succcurlyeq^{\mu}_{OLD} s_i$ if, for every $J \in \mathscr{I}_i$ with $U_i(t_i, P_\mu(J)) < U_i(s_i, P_\mu(J))$, there is $I \in \mathscr{I}_i$ with $I >^{\mu} J$ and $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$.*

This definition emphasizes the similarity with lexicographic weak preference. One can also provide a characterization closer to that in Theorem 1: The starting point is the observation that, if one defines $t_i \succ^{\mu}_{OLD} s_i$ as $t_i \succcurlyeq^{\mu}_{OLD} s_i$ and not $s_i \succcurlyeq^{\mu}_{OLD} t_i$, then $t_i \succ^{\mu}_{OLD} s_i$ is equivalent to: $t_i \succcurlyeq^{\mu}_{OLD} s_i$ and $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$ for some $I \in \mathscr{I}_i$.

Remark 5 *$t_i \succ^{\mu}_{OLD} s_i$ iff there are $M \geq 1$ and $I_1, \ldots, I_M \in \mathscr{I}_i$ such that $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$ for all m, and $U_i(t_i, P_\mu(J)) \geq U_i(s_i, P_\mu(J))$ for all $J \notin \bigcup_m \{K \in \mathscr{I}_i : I_m >^{\mu} K\}$.*

The key difference between this definition and the one now adopted in the paper (or its characterization in Theorem 1) arises if there is a profile s_{-i} such that $U_i(t_i, s_{-i}) < U_i(s_i, s_{-i})$ at an information set J such that $U_i(t_i, P_\mu(J)) = U_i(s_i, P_\mu(J))$. The characterization of $\succ^{\mu}_{OLD}$ just given implies that the existence of such a profile does *not* rule out the possibility that $t_i \succ^{\mu}_{OLD} s_i$. By way of contrast, Theorem 1 shows that, if such a profile exists, there *must* be m and J with $s_{-i} \in S_{-i}(J)$ and $I_m \geq^{\mu} J$. Intuitively, the definition in the present paper adds a robustness requirement with respect to the exact spcification of the probabilities $P_\mu(J)$, or, more broadly, of the CCPS μ. Equalities in expected payoffs are regarded as knife-edge cases that cannot support a strict preference for t_i over s_i.

Example 1 in the paper illustrates this point. In this game, $P_\mu(K) = \mu(\cdot|K)$ for all information sets K of Ann. Recall that $\mu(\{t\}|\phi) = \mu(\{t\}|\phi) = \frac{1}{2}$, so $U_a(HL, \mu(\cdot|\phi)) = U_a(T, \mu(\cdot|\phi))$.

Moreover, $\mu(\{o\}|I) = 1$, and $U_a(HL, o) = 2 > 0 = U_a(T, o)$. Thus, according to Definition 3, $HL \succ^{\mu}_{OLD} T$. On the other hand, according to the definition in the present version, HL and T are incomparable. As explained in the paper, this accounts for the possibility of vanishing perturbations of Ann's prior beliefs that, in particular, place slighlty more weight on t than on h.

In this particular example, ruling out T as a rational strategy has a further unpleasant consequence. The game under consideration is a modification of "Matching Pennies," and has a unique Nash equilibrium in which Ann plays HL and T with equal probability. If one rules out T as a rational strategy, then one must conclude that this game has no Nash equilibrium in structurally rational strategies (i.e., a Nash equilibrium in which, additionally, each strategy is a structural best reply to a CCPS for which the prior coincides with the equilibrium conjecture). The definition adopted in this paper does not suffer from this limitation.

Incidentally, the unique Nash equilibrium of this game is, of course, also the unique perfect equilibrium. To justify the play of T, one perturbs the Nash equilibrium more towards t than h—which is exactly the sort of perturbations that are used to argue that T is structurally rational in the paper.

An alternative to Definition 4 in the present version of the paper uses weak rather than strict preferences:

Definition 4 *$s_i \succcurlyeq^{\mu}_{*} t_i$ if, for all perturbations $(p^k)_{k\geq 1}$ of μ, $U_i(s_i, p^k) \geq U_i(t_i, p^k)$ eventually. s_i is structurally rational given μ if there is no $t_i \in S_i$ with $t_i \succ^{\mu}_{*} s_i$ (i.e., $t_i \succcurlyeq^{\mu}_{*} s_i$ and not $s_i \succcurlyeq^{\mu}_{*} t_i$.)*

This definition of strutural rationality is stronger than the one in the paper. If s_i is not structurally rational in the sense of this paper, there is t_i with $U_i(t_i, p^k) > U_i(s_i, p^k)$ eventually for all perturbations (p^k) of μ. Thus, a fortiori $t_i \succcurlyeq^{\mu}_{*} s_i$, and it is not the case that $s_i \succcurlyeq^{\mu}_{*} t_i$. But then $t_i \succ^{\mu}_{*} s_i$, so s_i is not structurally rational according to Definition 4.

In addition, Definition 4 rules out strategies that are weakly dominated by other pure strategies. If $U_i(t_i, s_{-i}) \geq U_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i}$, with a strict inequality for at least one $t_{-i} \in S_{-i}$,

then $U_i(t_i, p^k) \geq U_i(s_i, p^k)$ for all k and all perturbations (p^k), and $U_i(t_i, q^k) > U_i(s_i, q^k)$ for any perturbation $*q^k)_{k\geq 1}$ such that $q^k(\{t_{-i}\}) > 0$ for all k. Therefore, $t_i \succcurlyeq_*^\mu s_i$ and not $s_i \succcurlyeq_*^\mu t_i$, so $t_i \succ_*^\mu s_i$ and s_i is not structurally rational according to this definition.

One may wish to rule out strategies weakly dominated by pure strategies for reasons independent from elicitation. However, weak sequential rationality does not eliminate such strategies in general, and elicitation does not require that they be eliminated. One aspect of the minimality that characterizes the preferences defined in this paper is that it, too, does not rule out such strategies.

F.2 Requiring greater payoffs at every information set

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

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

Somewhat surprisingly, this definition actually fails to imply weak sequential rationality, even in simple, perfect-information games with nested strategic information.

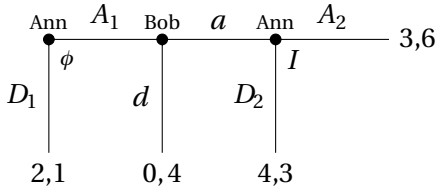

Figure 1: A centipede game. Ann's CCPS: $\mu(\{d\}|\phi) = \mu(\{a\}|I) = 1$

Consider for instance the Centipede game of Figure 1, and assume that Ann's CCPS μ is consistent with backward-induction reasoning. Ann's strategy $A_1 D_2$ does strictly better than $D_1 D_2$ given $\mu(\cdot|I) = P_\mu(I)$—that is, in case Bob chooses a at the second node. Even though $D_1 D_2$ does strictly better than $A_1 D_2$ 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 $\succcurlyeq^{\mu}_{DOM}$, even though it is not even optimal ex-ante—let alone weakly sequentially rational.

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

F.3 A definition that considers all conditional beliefs

The characterization in Theorem 1 restricts attention to the probabilities $P_\mu(I)$. Even in games in which this coincides with an element of the player's CCPS, this still implies that not all elements of μ play a direct role. For instance, if $S_{-i}(I) \supset S_{-i}(J)$ and $\mu(S_{-i}(J)|I) > 0$, then $\mu(\cdot|J)$ is not used directly. Weak sequential rationality instead requires optimality at every information set. One might then be led to consider a notion that takes the original elements of information sets into account, but still ranks them in terms of likelihood:

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

To see why this definition is inadequate, consider the game in Figure 2.

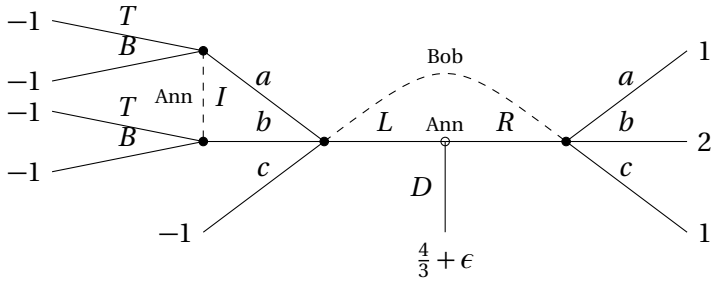

Figure 2: Ann's CCPS: $\mu(a|) = \mu(b|) = \mu(c|) = \frac{1}{3}$; $0 < \epsilon < \frac{1}{6}$.

Strategy L is strictly dominated for Ann. In addition, if Ann's CCPS 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 weakly sequentially rational strategy given μ. Furthermore, the same payoff inequality implies that it is not the case that $R \succcurlyeq^{\mu}_{ACB} D$. However, consider information set I. Given the associated belief $\mu(\cdot|S_b(I))$, R yields an expected 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 $\mu(\cdot|)$; however, it is *not* the case that $S_b >^{\mu} S_b(I)$, because $\mu(S_b(I)|) = \frac{2}{3}$. 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 weakly sequentially rational.

Theorem 1 avoids this issue because it employs only the probabilities $P_{\mu}(\cdot)$—in this example, $P_{\mu}(\phi) = \mu(\cdot|\phi)$.

F.4 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 $\mathrm{E}_{\mu(\cdot|I)} U_i(s_i, \cdot) < \mathrm{E}_{\mu(\cdot|I)} U_i(t_i, \cdot)$, 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 $\mathrm{E}_{\mu(\cdot|J)} U_i(s_i, \cdot) > \mathrm{E}_{\mu(\cdot|J)} U_i(t_i, \cdot)$.

Note: one may also consider further modifications whereby $P_{\mu}(\cdot)$ is used in lieu of μ, and/or set inclusion is replaced with $>^{\mu}$. However, I am going to provide a counterexample in which the game satisfies nested information (cf. Section C), and in addition $P_{\mu}(I) = \mu(\cdot|I)$ for all I.

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 weakly 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 weakly sequentially rational (whereas Theorem 2 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 3. 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 CCPS satisfies $\mu(\{o\}|) = 1$. As asserted, $P_{\mu}(K) = \mu(\cdot|K)$ for all $K \in \mathscr{I}_a$, and the game has nested information. 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 $\mu(\cdot|I'') = \mu(\cdot|\bar{I})$, whereas $UT\bar{T}$ only yields 3 given $\mu(\cdot|\bar{I}) = \mu(\cdot|I'')$.

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 $\mu(\cdot|\bar{I}') = \mu(\cdot|\bar{I}''$, $MT'\bar{T}'$ 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 $\mu(\cdot|I) = \mu(\cdot|I')$. 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 $\succcurlyeq^{\mu}_{COM}$ admits a strict cycle. Sec-

ond, there is *no* maximal strategy for the relation $\succcurlyeq^{\mu}_{COM}$. In particular, the three strategies that are weakly 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 $\succcurlyeq^{\mu}_{COM}$. Finally, a cycle can include strategies that are *not* weakly sequentially rational.

All difficulties in this example arise because the relation $\succcurlyeq^{\mu}_{COM}$ 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.[3] Structural preferences are instead defined via all measures $P_{\mu}(I)$, $I \in \mathcal{I}_i$; this delivers transitivity.

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

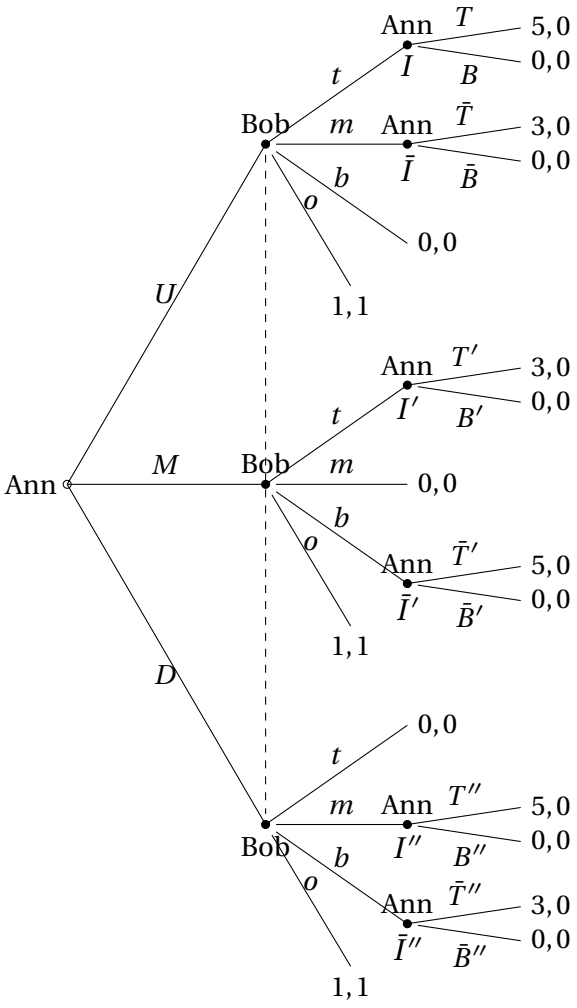

Figure 3: A strict cycle including a sequentially irrational strategy. Ann's CCPS: $\mu(\{o\}|) = 1$.