Structural Rationality in Dynamic Games

Marciano Siniscalchi*

November 30, 2018

Abstract

The analysis of dynamic games hinges on assumptions about players' actions and beliefs at information sets that are not expected to be reached during game play. However, under the standard assumption that players are sequentially rational, these assumptions cannot be tested on the basis of observed, on-path behavior. This paper introduces a novel optimality criterion, *structural rationality*, which addresses this concern. In any dynamic game, structural rationality implies sequential rationality. If players are structurally rational, assumptions about on-path beliefs concerning off-path actions, as well as off-path beliefs, can be tested via suitable "side bets." Structural rationality can also be characterized via trembles, or belief perturbations. Finally, structural rationality is consistent with experimental evidence about play in the extensive and strategic form, and justifies the use of the strategy method (Selten, 1967) in experiments.

Keywords: conditional probability systems, sequential rationality, strategy method.

*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.

1 Introduction

Solution concepts for dynamic games, such as subgame-perfect, sequential, or perfect Bayesian equilibrium, often predict that certain information sets will not be reached during game play. At the same time, these concepts aim to ensure that on-path play is sustained by "credible threats:" players believe that the (optimal) continuation play following any deviation from the predicted path would lead to a lower payoff.

A credible threat involves two types of assumptions about beliefs. The first pertains to on-path beliefs about off-path play: what is the threat? The second pertains to beliefs at off-path information sets about subsequent play: why is the threatened course of action credible? What is it a best reply to? The assumptions placed on such beliefs are an important dimension in which solution concepts differ. A key conceptual aspect of Savage (1954)'s foundational analysis of expected utility (EU) is to argue that the psychological notion of "belief" can and should be related to observable behavior. The objective of this paper is to characterize the behavioral content of assumptions on players' beliefs at both on-path and off-path information sets. The motivation is both methodological and practical: the results in this paper strengthen the behavioral foundations of dynamic game theory, but also broaden the range of game-theoretic predictions that can be tested experimentally.

In a single-person decision problem, the individual's beliefs can be elicited by offering her "side bets" on the relevant uncertain events, with the stipulation that both the choice in the original problem and the side bets contribute to the overall payoff. Similarly, in a game with simultaneous moves, a player's beliefs can be elicited by offering side bets on her opponents' actions (Luce and Raiffa, 1957, §13.6); for game-theoretic experiments implementing side bets, see e.g. Nyarko and Schotter (2002), Costa-Gomes and Weizsäcker (2008), Rey-Biel (2009), and Blanco, Engelmann, Koch, and Normann (2010).[1]

However, in a dynamic game, the fact that certain information sets may be off the predicted

[1] For related approaches, see Aumann and Dreze, 2009 and Gilboa and Schmeidler, 2003.

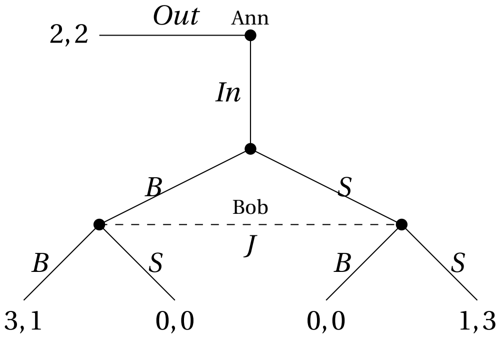


Figure 1: The Battle of the Sexes with an Outside Option

path of play poses additional challenges. For instance, the game of Figure 1 (cf. Ben-Porath and Dekel, 1992) has a subgame-perfect equilibrium in which Ann chooses *Out* at the initial node, under the threat that the Nash profile (S,S) would prevail in the subgame following *In*. Suppose an experimenter wishes to verify that, if Ann played *In*, Bob would indeed expect her to continue with S. (It turns out that testing Ann's initial beliefs is also problematic; since the discussion is more subtle, I defer it to Section 5.2.) If the simultaneous-move subgame was reached, the experimenter could offer Bob side bets on Ann's actions B vs. S. However, Ann is expected to play *Out* at the initial node, so the subgame is never actually reached. Alternatively, the experimenter could attempt to elicit Bob's conditional beliefs (i.e., the beliefs he would hold following *In*) from suitable betting choices observed at the beginning of the game. I now argue that, under textbook rationality assumptions, this approach, too, is not feasible; however, the discussion motivates the approach taken in the present paper.

In the game of Figure 2, before Ann chooses between *In* and *Out*, Bob can either secure a betting payoff of p close to but smaller than 1, or $b$et on Ann choosing S in the subgame, in which case his betting payoff is 1 for a correct guess and 0 otherwise. (All payoffs are denominated in "utils.") If Ann chooses *Out*, the bet is "called off," and Bob's betting payoff is 0. At every terminal node, a coin toss determines whether Bob receives his game payoff (which is as in Figure 1) or his betting payoff; these are displayed as an ordered pair in Figure 2. Ann's pay-

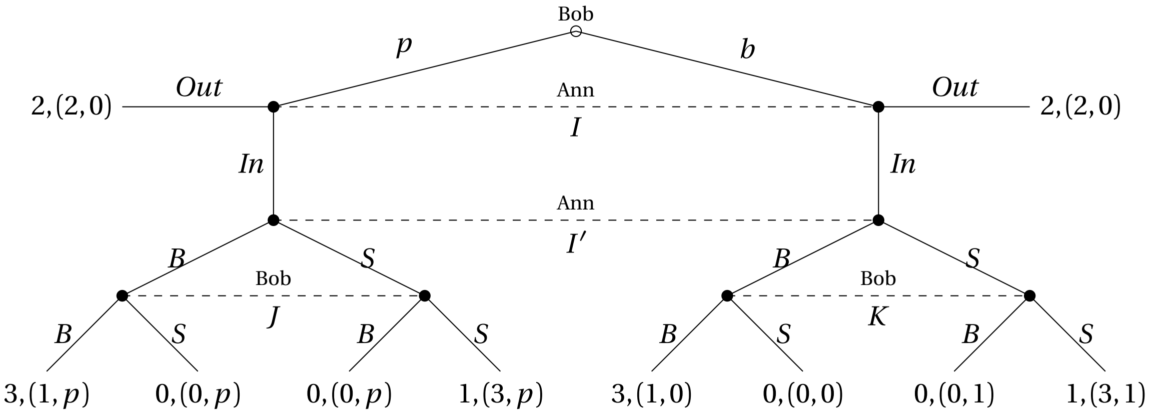


Figure 2: Eliciting Bob's conditional on *In* with ex-ante side bets.

off is as in Figure 1, independently of Bob's betting choice.[2] If Bob assigns positive probability to *In*, then it is optimal for him to bet on S if and only if he assigns probability greater than p to Ann's move S conditional on her playing *In*. However, if Bob is certain that Ann will choose *Out*, the standard assumption of sequential rationality (Kreps and Wilson, 1982) places no restriction on his initial betting choice.[3] In fact, there is a sequential equilibrium in which Bob chooses p at the initial node, Ann plays *Out* at I, and both would play S following *In*.

Whether in a game or in a single-person choice problem, assumptions about beliefs cannot be tested without also assuming a specific form of rationality, which relates beliefs to observable choices. The example suggests that the joint assumption that a player is sequentially rational *and* holds a given belief at an off-path information set may be intrinsically untestable—even in a simple game played "in the lab." The reason is that sequential rationality only requires that the action taken at an information set (such as b vs. p at the initial node of Figure 2) be optimal given the beliefs the player holds at *that* point in the game; it does *not* require that

[2] This is a simplified version of the elicitation mechanism formally analyzed in Section 5.2.

[3] This choice-based argument corresponds to the observation that, if Bob assigns positive probability to the event that Ann chooses *In*, his beliefs in the subgame can be derived by first eliciting his *prior* beliefs, and then *conditioning* on this event; however, this is not possible if Bob is certain of *Out*.

the player also take into account the beliefs she would hold at subsequent, zero-probability information sets (such as J and K in Figure 2). Hence, on-path choices cannot convey any information about off-path beliefs.

With this motivation, the present paper proposes a novel optimality criterion, *structural rationality*. Loosely speaking, this criterion requires that every action choice take into account both a player's current beliefs, and the beliefs she would hold following unexpected moves by coplayers (Definitions 5 and 8). In Figure 2, structural rationality implies that Bob *must* play b at the initial node if he is certain of S at J and K (Section 5.2.1). A structurally rational player is sequentially rational (Theorem 1), but in addition, her on-path betting choices fully pin down her beliefs, when a suitable implementation of side bets is employed (Theorem 2). Yet, for generic payoff assignments at terminal nodes, structural and sequential rationality coincide (Theorem 2 in Online Appendix C). Indeed, the objective of structural rationality is to provide only a "minimal" refinement of sequential rationality that enables the elicitation of conditional beliefs: see Example 2, and Sections 6.F and 6.G.

Structural rationality can also be seen as arising from "trembles," or belief perturbations. For instance, in Figure 2, Bob should play b if he assigns arbitrarily small but positive probability to *In*. Theorem 4 shows that structural rationality is fully characterized by a novel class of belief perturbations that, loosely speaking, assign positive probability to every information set but preserve the relative likelihood of strategies whenever they are positive and finite.

Finally, structural preferences can account for experimental evidence on the *strategy method* (Selten, 1967). This evidence suggests that, when confronted with a dynamic game, subjects make qualitatively similar choices when they play the game directly and when they are instead required to commit to extensive-form strategies, which the experimenter then implements (Brandts and Charness, 2011; Fischbacher, Gächter, and Quercia, 2012; Schotter, Weigelt, and Wilson, 1994). Unlike sequential rationality, structural rationality predicts that, under a suitable implementation of the strategy method, subjects should indeed exhibit the same behavior as in the original game: see Corollary 1 in Section 5.2. This provides a theoretical rationale for the strategy method. (Conversely, the cited evidence provides a degree of indirect support in favor of structural rationality.) At the same time, structural rationality reduces to EU maximization in games with simultaneous moves; hence, in general, it has different behavioral predictions for dynamic games and for their strategic form. This, too, is in accordance with experimental evidence: see (Cooper, DeJong, Forsythe, and Ross, 1993; Schotter et al., 1994; Cooper and Van Huyck, 2003; Huck and Müller, 2005).[4]

The companion paper Siniscalchi (2016) provides an axiomatization of structural rationality, taking as primitive a preference relation over acts à la Anscombe and Aumann (1963). Thus, the present paper and Siniscalchi (2016) jointly establish proper choice-theoretic foundations for dynamic game theory, comparable to the foundations that exist for the theory of games with simultaneous moves.

The present paper does not introduce specific restrictions on beliefs, or analyze particular solution concepts. Rather, it studies structural rationality as a notion of best response to given beliefs. In on-going work (available upon request), I incorporate structural rationality into equilibrium and non-equilibrium solution concepts.

Organization. Section 2 introduces notation. Section 3 defines structural preferences for dynamic games in which information sets satisfy a regularity condition. This class includes several games of interest in applications and experiments, and permits a simpler definition of structural preferences. Section 4 generalizes this definition to arbitrary dynamic games. Section 5 contains the main results. Section 6 discusses the related literature and extensions. All proofs are in the Appendix. The Online Appendix contains additional results, some omitted proofs, equivalent characterizations of structural preferences, as well as several alternative definitions that *do not* achieve the intended objectives.

[4] To the best of my knowledge, no known theory of play can accommodate both findings. Sequential rationality predicts different behavior in the strategic and extensive form, but—as noted above—also in the strategy method and under direct play. The invariance hypothesis (Kohlberg and Mertens, 1986) predicts that behavior should be the same in all presentations of the game.

2 Basic notation and definitions

This paper considers dynamic games with imperfect information. The analysis only requires that certain familiar reduced-form objects be defined. Online Appendix C describes how these objects are derived from a complete description of the underlying game, as e.g. in Osborne and Rubinstein (1994, Def. 200.1, pp. 200-201; OR henceforth) Section 6 indicates how to extend the notation to allow for incomplete information as well.

A dynamic game will be represented by a tuple $\big(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot)\big)$, where:

- N is the set of **players**.
- S_i is the set of **strategies** of player i.
- $\mathscr{I}_i$ is the collection of **information sets** of player i; it is convenient to assume that the **root**, ϕ, is an information set for all players.
- $U_i : \prod_{j\in N} S_j \to \mathbb{R}$ is the reduced-form **payoff function** for player i (see Section 6).
- For every $i \in N$ and $I \in \mathscr{I}_i$, $S(I)$ is the set of strategy profiles $(s_j)_{j\in N} \in \prod_j S_j$ that **reach** I. In particular, for every $i \in N$, $S(\phi) = S$.

I adopt the usual conventions for product sets: $S_{-i} = \prod_{j\neq i} S_j$ and $S = S_i \times S_{-i}$. I assume that the game has **perfect recall**, as per Def. 203.3 in OR. In particular, this implies that, for every $i \in N$ and $I \in \mathscr{I}_i$, $S(I) = S_i(I) \times S_{-i}(I)$, where $S_i(I) = \mathrm{proj}_{S_i} S(I)$ and $S_{-i}(I) = \mathrm{proj}_{S_{-i}} S(I)$. If $s_{-i} \in S_{-i}(I)$, say that s_{-i} **allows** I.[5] The range of the map $\mathrm{proj}_{S_{-i}} S : \mathscr{I}_i \to 2^{S_{-i}}$ plays an important role:

$$S_{-i}(\mathscr{I}_i) = \{S_{-i}(I) : I \in \mathscr{I}_i\}. \tag{1}$$

Finally, for every player $i \in N$ and information set $I \in \mathscr{I}_i$, the set $S(I)$ is required to satisfy **strategic independence** (Mailath, Samuelson, and Swinkels, 1993, Def. 2): for every $s_i, t_i \in S_i(I)$ there is $r_i \in S_i(I)$ such that $U_i(r_i, s_{-i}) = U_i(t_i, s_{-i})$ for all $s_{-i} \in S_{-i}(I)$, and $U_i(r_i, s_{-i}) = U_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i} \setminus S_{-i}(I)$. Intuitively, r_i is the strategy that coincides with s_i everywhere except

[5] That is: if i's coplayers follow the profile s_{-i}, I *can* be reached; whether it *is* reached depends upon whether or not i plays a strategy in $S_i(I)$.

at I and all subsequent information sets, where it coincides with t_i: see Remark 2 in Online Appendix C.1, or Theorem 1 in Mailath et al. (1993).

Section 3 restricts attention to games that satisfy an additional regularity condition. While this requirement rules out many games of interest (see Section 4), it does allow for a simple definition of structural preferences and structural rationality.

Definition 1 *A dynamic game* $\left(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot)\right)$ *has* ***nested strategic information*** *if*

$$\forall i \in N, I, J \in \mathscr{I}_i: \qquad \textit{either } S_{-i}(I) \cap S_{-i}(J) = \emptyset \textit{ or } S_{-i}(I) \subseteq S_{-i}(J) \textit{ or } S_{-i}(J) \subseteq S_{-i}(I). \tag{2}$$

In a game with nested strategic information, either *every* strategy profile that allows I (resp. J) set also allows J (resp. I), or *no* strategy profile allows both I and J. The games in Figure 1 and 2 satisfy this condition. All signaling games, and, more broadly, all games in which each player moves only once on each path of play, have nested strategic information. So do centipede game forms, and ascending-clock auctions.

At any information set $I \in \mathscr{I}_i$, player i's beliefs about the past and future moves of her coplayers are represented by a probability distribution over S_{-i}. These beliefs are conditional upon the (possibly partial) information she has at I about coplayers' previous moves; this information is represented by the event $S_{-i}(I)$. Collectively, player i's beliefs are required to satisfy the chain rule of conditioning: starting with the prior, player i updates her beliefs in the usual way "whenever possible"—that is, until a zero-probability event occurs. Then, player i formulates a new belief, but from that point on, she again applies the updating formula, until a new, zero-probability event is observed; and so on.

Definition 2 below takes as given an arbitrary collection $\mathscr{C}_i$ of conditioning events. The preceding paragraph suggests the specification $\mathscr{C}_i = S_{-i}(\mathscr{I}_i)$; this is sufficient to define sequential rationality, and is also enough to define structural rationality in the class of games considered in Section 3. The general definition of structural rationality requires a richer set of conditioning events; see Section 4. Myerson (1986) considers the case $\mathscr{C}_i = 2^{S_{-i}} \setminus \{\emptyset\}$.

Definition 2 *(Rényi, 1955; Myerson, 1986; Ben-Porath, 1997; Battigalli and Siniscalchi, 1999, 2002) Fix a dynamic game $\left(N,(S_i,\mathscr{I}_i,U_i)_{i\in N},S(\cdot)\right)$, a player $i\in N$, and a non-empty collection $\mathscr{C}_i$ of non-empty subsets of S_{-i}. A* ***conditional probability system (CPS) on*** *$(S_{-i},\mathscr{C}_i)$ is a collection $\mu_i \equiv \left(\mu_i(\cdot|F)\right)_{F\in\mathscr{C}_i}$ such that:*

(1) for every $F\in\mathscr{C}_i$, $\mu_i(\cdot|F)\in\Delta(S_{-i})$ and $\mu_i(F|F)=1$;

(2) for every $E\subseteq S_{-i}$ and $F,G\in\mathscr{C}_i$ such that $E\subseteq F\subseteq G$,

$$\mu_i(E|G)=\mu_i(E|F)\cdot\mu_i(F|G). \tag{3}$$

The set of CPS on $(S_{-i},\mathscr{C}_i)$ is denoted by $\Delta(S_{-i},\mathscr{C}_i)$. For any probability distribution $\pi\in\Delta(S_{-i})$ and function $a:S_{-i}\to\mathbb{R}$, let $\mathrm{E}_\pi a=\sum_{s_{-i}\in S_{-i}} a(s_{-i})\pi(s_{-i})$.

Following Reny (1992), Rubinstein (1991), and Battigalli and Siniscalchi (2002), I define sequential rationality to mean that a strategy is optimal at every information set that it does not preclude. This imposes no restrictions on actions at information sets that the strategy itself precludes. One implication is that this definition does not distinguish between realization-equivalent strategies.[6] Structural rationality (Definitions 5 and 8) has the same feature.

Definition 3 (Sequential rationality) *Fix a dynamic game $\left(N,(S_i,\mathscr{I}_i,U_i)_{i\in N},S(\cdot)\right)$, a player $i\in N$, and a CPS $\mu\in\Delta(S_{-i},S_{-i}(\mathscr{I}_i))$ for player i. Strategy $s_i\in S_i$ is* ***sequentially rational given*** *μ if, for every $I\in\mathscr{I}_i$ such that $s_i\in S_i(I)$, and all $t_i\in S_i(I)$, $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(s_i,\cdot)\geq \mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(t_i,\cdot)$.*

3 Structural preferences under nested strategic information

To streamline definitions and results, throughout this section, fix a dynamic game with nested strategic information $\left(N,(S_i,\mathscr{I}_i,U_i)_{i\in N},S(\cdot)\right)$, a player $i\in N$, and a CPS $\mu\in\Delta(S_{-i},S_{-i}(\mathscr{I}_i))$.

[6] More precisely, it does not distinguish between *payoff*-equivalent strategies: if $U_i(s_i,s_{-i})=U_i(t_i,s_{-i})$ for all $s_{-i}\in S_{-i}$, then s_i is sequentially rational given a CPS μ if and only if t_i is.

Recall the notions of "structural" and "lexicographic consistency" put forth by Kreps and Wilson (1982, p. 873) to motivate their definition of consistent assessments[7]:

> Fix a player i. His "primary hypothesis" as to how the game will be played is $P = \mu(\cdot|S_{-i})$, and if his beliefs obey Eq. (3), then he applies P to compute $\mu(\cdot|S_{-i}(I))$ whenever possible. We might assume that when P does not apply—when he comes to an information set I with $P(S_{-i}(I)) = 0$—then he has a "second most likely hypothesis" P' that he attempts to apply. If that fails, he tries his "third most likely hypothesis" P'', and so on.

Structural rationality, as put forth in the present paper, *identifies* and *partially orders* a player's probabilistic "hypotheses" on the basis of her conditional beliefs at every information set—i.e. her CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$. It then defines best replies using a generalized lexicographic criterion formulated in terms of these alternative probabilistic hypotheses.

In games with nested strategic information, one can identify the player's alternative hypotheses leveraging Equation (3). Fix an information set $I \in \mathscr{I}_i$; by Equation (3), if there is $J \in \mathscr{I}_i$ such that $S_{-i}(J) \supset S_{-i}(I)$ and $\mu(S_{-i}(I)|S_{-i}(J)) > 0$, then $\mu(\cdot|S_{-i}(I))$ is derived from $\mu(\cdot|S_{-i}(I))$ by updating. If instead $\mu(S_{-i}(I)|S_{-i}(J)) = 0$ for every information set $J \in \mathscr{I}_i$ such that $S_{-i}(J) \supset S_{-i}(I)$, then $\mu(\cdot|S_{-i}(I))$ cannot be derived from any other element of the CPS μ by updating. Hence, $\mu(\cdot|S_{-i}(I))$ can be regarded as one of player i's alternative hypotheses:

Definition 4 *An information set $I \in \mathscr{I}_i$ is **basic for** μ (or μ-**basic**) if, for all $J \in \mathscr{I}_i$, $S_{-i}(J) \supset S_{-i}(I)$ implies $\mu(S_{-i}(I)|S_{-i}(J)) = 0$. If $I \in \mathscr{I}_i$ is basic for μ, then $\mu(\cdot|S_{-i}(I))$ is a **(μ-) basic belief** of player i, and $S_{-i}(I)$ is a μ-**basic (conditioning) event**.*

[7] In this quotation, I have adapted the notation to that of the present paper. The actual definition of structural and lexicographic consistency in Kreps and Wilson (1982) uses a different formalism and is tailored to equilibrium analysis; it also incorporates additional independence assumptions that play no role in the present paper. See Kreps and Ramey (1987) and Battigalli (1994) for further details.

The prior $\mu(\cdot|S_{-i})$ is always basic: it is the player's primary hypothesis. Moreover, under nested strategic information, for every information set I, there is a *unique* μ-basic belief $\mu(\cdot|S_{-i}(J))$ that can generate $\mu(\cdot|S_{-i}(I))$ by updating (see Corollary 3 and Observation 2 in the Appendix).

Player i's CPS μ also reveals the relative plausibility of specific pairs of μ-basic beliefs. Consider two μ-basic information sets $I, J \in \mathscr{I}_i$ such that $S_{-i}(I) \supset S_{-i}(J)$. Then $\mu(S_{-i}(I)|S_{-i}(J)) = 1 > 0$, so the belief $\mu(\cdot|S_{-i}(J))$ is a *possible* hypothesis at I, in the sense that it is consistent with reaching I. Appealing to the logic of lexicographic consistency, the fact that player i's beliefs at I are $\mu(\cdot|S_{-i}(I))$ and not $\mu(\cdot|S_{-i}(J))$[8] suggests that i deems $\mu(\cdot|S_{-i}(I))$ a *more plausible* hypothesis than $\mu(\cdot|S_{-i}(J))$. While this logic yields a partial, rather than a complete ordering of μ-basic beliefs, this is sufficient for the present purposes: see Example 2 and Section 6.F.

Finally, to define best replies, I use a generalized lexicographic criterion[9] wherein μ-basic beliefs are partially ordered by set inclusion of the corresponding conditioning events.

Definition 5 *For any two strategies $s_i, t_i \in S_i$, s_i is (weakly)* ***structurally preferred*** *to t_i given μ ($s_i \succsim^\mu t_i$) iff, for any μ-basic $I \in \mathscr{I}_i$ such that $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(s_i,\cdot) < \mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(t_i,\cdot)$, there is a μ-basic $J \in \mathscr{I}_i$, with $S_{-i}(J) \supset S_{-i}(I)$, such that $\mathrm{E}_{\mu(\cdot|S_{-i}(J))}U_i(s_i,\cdot) > \mathrm{E}_{\mu(\cdot|S_{-i}(J))}U_i(t_i,\cdot)$.*

A strategy $s_i \in S_i$ is ***structurally rational for*** *μ if there is no $t_i \in S_i$ such that $t_i \succ^\mu s_i$.*[10]

In words, $s_i \succsim^\mu t_i$ means that, if t_i yields a strictly higher payoff given a hypothesis (i.e., μ-basic belief) $\mu(\cdot|S_{-i}(I))$, then s_i must yield a strictly higher payoff given an alternative hypothesis $\mu(\cdot|S_{-i}(J))$ that the player deems more plausible.

More informally, to compare s_i and t_i, player i considers all possible paths through the tree. For each such path, she keeps track of the expected payoffs of s_i and t_i at each μ-basic

[8] Note that $\mu(\cdot|S_{-i}(I)) \neq \mu(\cdot|S_{-i}(J))$: in particular, since J is μ-basic, $S_{-i}(I) \supset S_{-i}(J)$ implies $\mu(S_{-i}(J)|S_{-i}(I)) = 0$.

[9] The usual lexicographic order $\geq_L$ on $\mathbb{R}^L$ can be defined as follows: given $a, b \in \mathbb{R}^L$, $a \geq_L b$ if, for every $k \in \{1,\dots,L\}$ such that $a_k < b_k$, there is $\ell \in \{1,\dots,k-1\}$ with $a_\ell > b_\ell$. Definition 5 generalizes this formulation.

[10] As usual, strict preference, denoted "$s_i \succ^\mu t_i$," is defined as "$s_i \succsim^\mu t_i$ and not $t_i \succsim^\mu s_i$;" indifference, denoted "$s_i \sim^\mu t_i$," is defined as "both $s_i \succsim^\mu t_i$ and $t_i \succsim^\mu s_i$."

information set, in the order these information sets are crossed. Then, loosely, $s_i \succsim^\mu t_i$ requires that, *on each path*, the resulting ordered list of expected payoffs for s_i be lexicographically weakly greater than the corresponding list for t_i. (I provide a precise formalization of this intuition in Online Appendix E.1.) Section 6.G discusses similarities and differences between structural and lexicographic preferences (Blume, Brandenburger, and Dekel, 1991a).

Structural rationality for a CPS μ is defined as *maximality* with respect to $\succsim^\mu$. Structural preferences are transitive (see Siniscalchi, 2016, Appendix B); hence, every finite game admits a structurally rational strategy for every CPS. With complete preferences, maximality coincides with optimality (s_i is at least as good as any other strategy). However, as Example 2 shows, structural preferences may be incomplete.

Structural preferences are tied to the *extensive form* of the game, and specifically to the conditioning events $S_{-i}(\mathscr{I}_i)$. These play a key role in the definition of player i's CPS, of basic beliefs, and of their plausibility ordering.

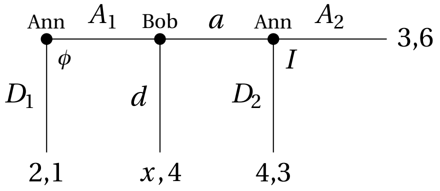
Structural rationality reduces to ex-ante EU maximization in two cases: when the game has simultaneous moves, and when every information set of player i has positive prior probability (because in this case only the prior $\mu(\cdot|S_{-i})$ is a basic belief). Outside of these special cases, structural preferences refine ex-ante EU maximization. This immediately delivers the first prediction anticipated in the Introduction: if a player has structural preferences, her behavior will, in general, differ in a dynamic game and in the associated strategic form.

The following characterization is immediate from Definition 5:

Remark 1 *Strategy s_i is structurally rational for μ if, for every $t_i \in S_i$, either (i) $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(s_i,\cdot) = \mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(t_i,\cdot)$ for every μ-basic $I \in \mathscr{I}_i$, or (ii) there is a μ-basic $I \in \mathscr{I}_i$ such that $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(s_i,\cdot) > \mathrm{E}_{\mu(\cdot|S_{-i}(I))}U_i(t_i,\cdot)$ and $\mathrm{E}_{\mu(\cdot|S_{-i}(J))}U_i(s_i,\cdot) = \mathrm{E}_{\mu(\cdot|S_{-i}(J))}U_i(t_i,\cdot)$ for all μ-basic $J \in \mathscr{I}_i$ with $S_{-i}(J) \supset S_{-i}(I)$.*

Finally, it is immediate from Definition 5 that structural preferences (and hence structural rationality) do not distinguish between realization-equivalent strategies. In particular, if s_i and t_i are realization-equivalent, then $s_i \sim^\mu t_i$ for every CPS μ (cf. footnote 6).

Example 1 The game in Figure 3 is parameterized by $x \in [0,2]$; for $x < 2$, it is a Centipede game. Ann's beliefs μ satisfy $\mu(\{d\}|S_b) = \mu(\{a\}|S_b(I)) = 1$. The events S_b and $S_b(I)$ are μ-basic, and the table in Figure 3 shows Ann's expected payoff conditional upon these events.



s_a	$\mathrm{E}_{\mu(\cdot\|S_b(\phi))} U_a(s_a,\cdot)$	$\mathrm{E}_{\mu(\cdot\|S_b(I))} U_a(s_a,\cdot)$
D_1D_2, D_1A_2	2	2
A_1D_2	x	4
A_1A_2	x	3

Figure 3: A centipede-like game. Ann's CPS: $\mu(\{d\}|S_b(\phi)) = \mu(\{a\}|S_b(I)) = 1$. $x \in [0,2]$

Denote by D_1 either one of the realization-equivalent (hence, indifferent) strategies D_1D_2, D_1A_2. First, I apply Remark 1. If $x < 2$, then D_1 is the unique structurally rational strategy given μ (again, up to realization equivalence), as it yields strictly higher ex-ante expected payoff than A_1D_2 and A_1A_2. This is also the unique sequential best reply to μ. For $x = 2$, A_1D_2 is the unique structurally rational strategy given μ: all other strategies also yield an ex-ante payoff of 2, but A_1D_2 yields a strictly higher payoff conditional upon $S_b(I) = \{a\}$.[11] By comparison, both D_1 and A_1D_2 are sequentially rational given μ.

Next, I apply Definition 5 directly. Assume first that $x < 2$: then $D_1 \succ^\mu A_1D_2 \succ^\mu A_1A_2$. To see this, consider D_1 and A_1D_2. Although $\mathrm{E}_{\mu(\cdot|S_b(I))} U_a(D_1,\cdot) = 2 < 4 = \mathrm{E}_{\mu(\cdot|S_b(I))} U_a(A_1D_2,\cdot)$, we have $S_b(\phi) = S_b \supset S_b(I)$, and $\mathrm{E}_{\mu(\cdot|S_b)} U_a(D_1,\cdot) = 2 > x = \mathrm{E}_{\mu(\cdot|S_b)} U_a(A_1D_2,\cdot)$: thus, $D_1 \succcurlyeq^\mu A_1D_2$. On the other hand, $\mathrm{E}_{\mu(\cdot|S_b)} U_a(D_1,\cdot) > \mathrm{E}_{\mu(\cdot|S_b)} U_a(A_1D_2,\cdot)$ and the fact that no (μ-basic) $J \in \mathscr{I}_i$ can satisfy $S_b(J) \supset S_b$ imply that not $A_1D_2 \succcurlyeq^\mu D_1$. Thus, $D_1 \succ^\mu A_1D_2$. Now consider A_1D_2 and A_1A_2. There is no μ-basic $J \in \mathscr{I}_a$ with $\mathrm{E}_{\mu(\cdot|S_{-i}(J))} U_a(A_1D_2,\cdot) < \mathrm{E}_{\mu(\cdot|S_{-i}(J))} U_a(A_1A_2,\cdot)$, so trivially $A_1D_1 \succcurlyeq^\mu A_1A_2$. Moreover, $\mathrm{E}_{\mu(\cdot|S_b(I))} U_a(A_1D_2,\cdot) > \mathrm{E}_{\mu(\cdot|S_b(I))} U_a(A_1A_2,\cdot)$ and $\mathrm{E}_{\mu(\cdot|S_b)} U_a(A_1D_2,\cdot) = \mathrm{E}_{\mu(\cdot|S_b)} U_a(A_1A_2,\cdot)$, so not $A_1A_2 \succcurlyeq^\mu A_1D_2$. Hence, $A_1D_2 \succ^\mu A_1A_2$.

Finally, consider $x = 2$. All strategies yield the same ex-ante payoff, so Definition 5 implies

[11] Indeed, A_1D_2 is the unique structurally rational strategy for *any* CPS of Ann.

that preferences are determined by expectations given $S_b(I)$. Thus, $A_1D_2 \succ^\mu A_1A_2 \succ^\mu D_1$. □

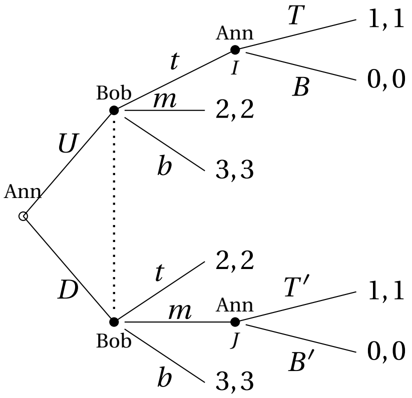
In Example 1, the fact that D_1 is a sequential, but not structural best reply to Ann's CPS μ when $x = 2$ illustrates that, while sequential rationality is defined in terms of interim or "local" optimality, structural rationality is a "global" optimality criterion. This corresponds to the intuition provided in the Introduction: at every point in the game, a structurally rational player takes into account possible surprise moves, in a disciplined way.[12] In the example, this leads to more refined predictions about play, given the same conditional beliefs.

Indeed, Theorem 2 and Remark 4 below imply that, for beliefs to be elicitable from on-path betting choices, structural rationality *must* refine sequential rationality. Yet, the next example shows that Definition 5 is a *parsimonious*, or limited, refinement, in two respects. First, structural preferences need not rank *all* strategies in the game: they only guarantee that, if a strategy is not sequentially rational, there is always another strategy that is strictly preferred to it. Second, structural rationality employs the "minimal" amount of information about a player's conditional beliefs required to ensure sequential rationality.

Example 2 (Incompleteness and Minimality) Consider the game in Fig. 4, and assume that Ann's CPS μ satisfies $\mu(\{b\}|S_b(\phi)) = 1$ (the other conditional probabilities are pinned down by Definition 2). All information sets for Ann are μ-basic. Moreover, $S_b(I) = \{t\}$ and $S_b(J) = \{m\}$ are disjoint, hence not ordered by set inclusion. Denote by UT either one of the realization-equivalent strategies UTT', UTB'; interpret UB, DT', DB' similarly.

Incompleteness: Definition 5 implies that UT (resp. DT') is strictly preferred to the sequentially irrational strategy UB (resp. DB'). The strategies UT and DT' are incomparable according to Definition 5: UT yields a strictly higher expected payoff given $\mu(\cdot|S_b(I))$ than DT', but DT does strictly better given $\mu(\cdot|S_b(J))$, and these μ-basic beliefs are not ordered by

[12] Online Appendix F discusses different definitions of preferences that also take into account possible surprise moves, but do so in ways that do not deliver the properties of interest in this paper (for instance, they do not ensure sequential rationality).



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

Figure 4: Incompleteness and minimality. Ann's CPS: $\mu(\{b\}|S_b(\phi))=1$

set inclusion of the corresponding conditioning events.[13] However, UT and DT' are also the sequentially rational best replies to μ. Since the objective of Definition 5 is to (i) guarantee sequential rationality, and (ii) allow the elicitation of conditional beliefs, the fact that UT and DT' are not ranked by $\succcurlyeq^{\mu}$ is immaterial. See Section 6.F for further discussion.

Minimality: Consider a modification of Definition 5 that only takes into account Ann's basic beliefs $\mu(\cdot|S_b)$ and $\mu(\cdot|S_b(I))$, but not $\mu(\cdot|S_b(J))$. Then DB' is undominated in the resulting ordering; however, it is *not* sequentially rational. Definition 5 rules out DB' precisely because it takes the basic belief $\mu(\cdot|S_b(J))$ into account as well. □

4 Structural preferences for general dynamic games

While convenient, the assumption of nested strategic information rules out several games of economic interest; notable examples include, but are not limited to, English (rather than ascending-clock) auctions, alternating-offer bargaining, and the chain-store game.[14] Unfor-

[13] UT and DT' are *maximal* but not *optimal*. Indeed, this game has no optimal strategies.

[14] In a chain-store game with players i, e_1 and e_2, in the obvious notation, let I (J) be i's node reached when e_1 enters, i fights (acquiesces), and e_2 enters. The profile s_{-i} such that e_1 enters and e_2 enters regardless of i's first-period choice allows both I and J. However, the profile s'_{-i} (s''_{-i}) such that e_1 enters and e_2 only enters if i

tunately, without nested strategic information, a strategy may be maximal in the order formalized by Definition 5, and yet fail to be sequentially rational.

Example 3 (Non-nested information) Consider the "signal-choice" game in Figure 5. Ann and Bob choose an action simultaneously. If Bob chooses o, the game ends. Otherwise, Ann's action determines what she learns about Bob's action. This game does not have nested strategic information: $S_b(I)=\{t,m\}$ and $S_b(J)=\{m,b\}$, so $S_b(I)\cap S_b(J)\neq\emptyset$ but $S_b(I)$ and $S_b(J)$ are not nested. Bob's payoffs are omitted in Fig. 5 as they are not relevant to the discussion.

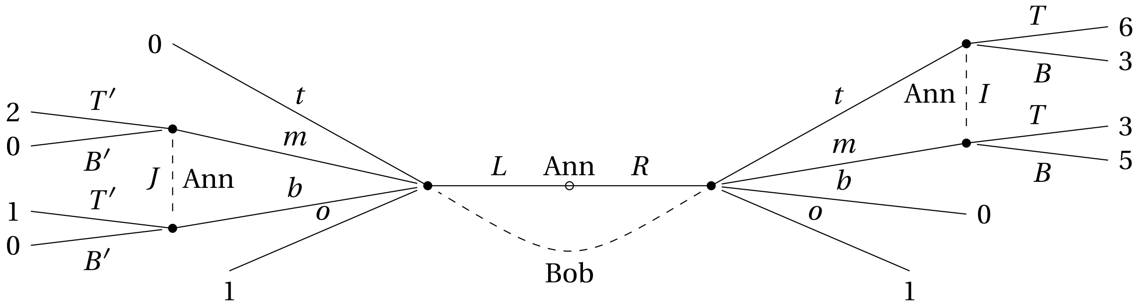


Figure 5: A signal-choice game.

Define Ann's CPS μ by $\mu(\{o\}|S_b(\phi))=1$, $\mu(\{t\}|S_b(I))=\mu(\{m\}|S_b(I))=\frac{1}{2}$, and $\mu(\{m\}|S_b(J))=\mu(\{b\}|S_b(J))=\frac{1}{2}$. All three information sets for Ann are basic for μ; expected payoffs are displayed in Table I.

Strategy RB is not sequentially rational given μ: it chooses the inferior action B, rather than T, at I. Yet, According to Definition 5, RB is structurally rational given μ. In particular, RT is *not* strictly preferred to RB. While $\mathrm{E}_{\mu(\cdot|S_b(I))}U_a(RT,\cdot)>\mathrm{E}_{\mu(\cdot|S_b(I))}U_a(RB,\cdot)$, it is also the case that $\mathrm{E}_{\mu(\cdot|S_b(J))}U_a(RB,\cdot)>\mathrm{E}_{\mu(\cdot|S_b(J))}U_a(RT,\cdot)$, and the conditioning events $S_{-}(I)$ and $S_{-i}(J)$ are not ordered by inclusion. Thus, RT and RB are not comparable according to Definition 5. □

The failure of sequential rationality in Example 3 reflects a deeper, conceptual issue. In

fights (acquiesces) in the first period allows I but not J (J but not I). Thus, $S_{-i}(I)$ and $S_{-i}(J)$ are not nested.

s_a	$\mathrm{E}_{\mu(\cdot\mid S_b(\phi))}U_a(s_a,\cdot)$	$\mathrm{E}_{\mu(\cdot\mid S_b(I))}U_a(s_a,\cdot)$	$\mathrm{E}_{\mu(\cdot\mid S_b(J))}U_a(s_a,\cdot)$
RT	1	4.5	1.5
RB	1	4	2.5
LT'	1	1	1.5
LB'	1	0	0

Table I: $\mu(\{o\}|S_b(\phi))=1$, $\mu(\{t\}|S_b(I))=\mu(\{m\}|S_b(I))=\mu(\{m\}|S_b(J))=\mu(\{b\}|S_b(J))=\frac{1}{2}$

games without nested strategic information, the simple approach used in Section 3 to identify and rank a player's alternative hypotheses from her CPS is not satisfactory. In Example 3, $\mu(\cdot|S_b(I)|S_b(J))>0$, even though $S_b(I)\not\supset S_b(J)$: thus, the alternative hypothesis Ann uses at J—which Section 3 identifies with $\mu(\cdot|S_b(J))$—is also consistent with reaching I. Yet $\mu(\cdot|S_b(I))$ is *not* derived from $\mu(\cdot|S_b(J))$ by conditioning on $S_b(I)$. The logic of lexicographic consistency then suggests that, at I, Ann uses a more plausible hypothesis than the one she employs at J. But, switching the roles of I and J, a symmetric argument leads to the conclusion that she uses a more plausible hypothesis at J than at I.

The key insight is that, to avoid this contradiction, $\mu(\cdot|S_b(I))$ and $\mu(\cdot|S_b(J))$ should be regarded as being derived from *the same* alternative theory. Indeed, there is a *unique* probability distribution $p\in\Delta(S_b)$ that satisfies $p(S_b(I))>0$, $p(S_b(J))>0$, and $p(S_b(I)\cup S_b(J))=1$, and such that $\mu(\cdot|S_{-i}(I))$ and $\mu(\cdot|S_{-i}(J))$ are its updates given $S_{-i}(I)$ and $S_{-i}(J)$ respectively. This suggests that, in addition to the prior $\mu(\cdot|S_b)$, Ann entertains only one additional hypothesis, p. Furthermore, she deems this hypothesis less plausible than the prior: p assigns positive probability to $S_b=S_b(\phi)$, but it is different from Ann's prior belief. Finally, this approach restores the connection with sequential rationality: if expected payoffs given p are used to break ties in ex-ante expectations, *RT* is ranked strictly above *RB*.

One conclusion from the preceding discussion is that, in general dynamic games, alternative probabilistic hypotheses need not coincide with μ-basic beliefs; they may instead generate two or more μ-basic beliefs by updating. Despite this, a player's plausibility ranking among such theories is still partially revealed by her CPS. As was just argued, in Example 3, the fact that $\mu(S_b(I)|S_b(J)) > 0$ suggests that the theory that Ann updates to obtain her beliefs at I is at least as plausible as the theory she uses at J. In that example, it was argued that the *same* theory must be used at I and J; more generally, the theories used may be different, but if so the theory used at J is revealed to be less plausible than the theory used at I. Definition 6 below generalizes this intuition by adding the requirement that plausibility be transitive.

It is conceptually appropriate to formalize plausibility as a ranking over *conditioning events*. After all, if a player associates a more plausible hypothesis to reaching I than J, then arguably she should deem it more plausible to reach I than J (and conversely). As will be clear momentarily, defining plausibility as a ranking over events is also notationally convenient.

Throughout the remainder of this section, fix an arbitrary dynamic game $(N, (S_i, \mathscr{I}_i, U_i)_{i\in N})$, a player $i \in N$, and a CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$.

Definition 6 *Consider two information sets $I, J \in \mathscr{I}_i$. Then $S_{-i}(I)$ is **at least as plausible as** $S_{-i}(J)$ **given** μ ($S_{-i}(I) \geq^{\mu} S_{-i}(J)$) if there is an ordered list $I_1, \ldots, I_L \in \mathscr{I}_i$ such that $I_1 = J$, $I_L = I$, and $\mu(S_{-i}(I_{\ell+1})|S_{-i}(I_\ell)) > 0$ for all $\ell = 1, \ldots, L$.*

The relation $\geq^{\mu}$ is a preorder (i.e., reflexive and transitive). Its strict (i.e., asymmetric) part $>^{\mu}$ is defined as usual by letting $F >^{\mu} G$ iff $F \geq^{\mu} G$ and not $G \geq^{\mu} F$.

With nested strategic information, the plausibility relation of Definition 6 reduces to set inclusion (see Corollary 3 in the Appendix); this ensures that the approach of this section is a proper generalization of the approach in Section 3:

Remark 2 *If the game has nested strategic information, then for all μ-basic information sets $I, J \in \mathscr{I}_i$: $S_{-i}(J) >^{\mu} S_{-i}(I)$ if and only if $S_{-i}(J) \supset S_{-i}(I)$.*

The relation $\geq^{\mu}$ also helps identify them from the player's CPS. Observe that, in Example 3, $S_b(I) =^{\mu} S_b(J)$, and the support of the probability p, Ann's secondary hypothesis, is $S_b(I) \cup$

$S_b(J) = \{t, m, b\}$. Note also that one can define a CPS $\nu \in \Delta(S_b, S_b(\mathscr{I}_a) \cup \{t, m, b\})$ by letting $\nu(\cdot|F) = \mu(\cdot|F)$ for $F \in S_b(\mathscr{I}_i)$ and $\nu(\cdot|\{t, m, b\}) = p$. Rather than saying that the CPS μ permits one to uniquely identify Ann's secondary hypothesis p, one could equivalently state that Ann's CPS μ admits a unique extension to a CPS ν defined over a broader family of conditioning events. The following definition formalizes and generalizes these observations.

Definition 7 *For every $I \in \mathscr{I}_i$, let*

$$B_\mu(I) = \bigcup \Big\{ S_{-i}(J) : S_{-i}(J) =^\mu S_{-i}(I) \Big\}.$$

*A CPS $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i))$ is an **extension** of μ if $\nu(\cdot|S_{-i}(I)) = \mu(\cdot|S_{-i}(I))$ for all $I \in \mathscr{I}_i$.*

Analogously to $S_{-i}(\mathscr{I}_i)$, $B_\mu(\mathscr{I}_i)$ denotes the range of the function $B_\mu : \mathscr{I}_i \to 2^{S_{-i}}$. If μ has an extension, it is called **extensible**. Theorem 3 in Section 5.3 characterizes extensible CPSs, and establishes that, if an extension exists, it is unique. The probabilities $\nu(\cdot|B_\mu(I))$ are the counterpart of μ-basic beliefs for general dynamic games; indeed, in games with nested strategic information, these notions coincide:

Remark 3 *If the game has nested strategic information, then for every $I \in \mathscr{I}_i$, I is μ-basic if and only if $B_\mu(I) = S_{-i}(I)$; therefore, the extension of μ is μ itself.*

To extend Definition 5 to general dynamic games, I replace μ-basic beliefs with the probabilities $\nu(\cdot|B_\mu(I))$ (with $I \in \mathscr{I}_i$ μ-basic), and set inclusion with the ordering of Definition 6.

Definition 8 *Let $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i))$ be an extension of μ. For every pair of strategies $s_i, t_i \in S_i$, s_i is (weakly) **structurally preferred** to t_i given μ ($s_i \succsim^\mu t_i$) iff, for every μ-basic $I \in \mathscr{I}_i$ such that $\mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(s_i, \cdot) < \mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(t_i, \cdot)$, there is a μ-basic $J \in \mathscr{I}_i$, with $S_{-i}(J) >^\mu S_{-i}(I)$, such that $\mathrm{E}_{\nu(\cdot|B_\mu(J))} U_i(s_i, \cdot) > \mathrm{E}_{\nu(\cdot|B_\mu(J))} U_i(t_i, \cdot)$.*

Remark 3 implies that Definition 8 reduces to Definition 5 for games with nested strategic information. The intuition given in Section 3 for structural preference applies here as well:

it is more plausible that s_i is better than t_i. The obvious counterpart to Remark 1 holds for general games. Online Appendix E explores equivalent definitions of structural preferences.

5 Main Results

Throughout subsections 5.1–5.3, fix an arbitrary dynamic game $\big(N,(S_i,\mathscr{I}_i,U_i)_{i\in N},S(\cdot)\big)$.

5.1 Structural and Sequential Rationality

The first main result of this paper can now be stated. For readers who skipped Section 4: in a game with nested strategic information, every CPS is extensible.

Theorem 1 *Fix a player $i\in N$ and an extensible CPS $\mu\in\Delta(S_{-i},S_{-i}(\mathscr{I}_i))$ for i. If strategy $s_i\in S_i$ is structurally rational for μ, then it is sequentially rational for μ.*

The intuition[15] behind Theorem 1 is reminiscent of the argument establishing that, with EU preferences, an ex-ante optimal strategy s_i of player i must prescribe an optimal continuation at every positive-probability information set $I\in\mathscr{I}_i$: if s_i of player i is not conditionally optimal at I, and $S_{-i}(I)$ has positive prior probability, then there is a strategy s_i^* that differs from s_i only at I and subsequent information sets, and which yields strictly higher ex-ante expected payoff than s_i. The additional power of structural preferences allows one to extend this argument to the case in which $S_{-i}(I)$ has zero prior probability.

The logic of the proof is easiest to convey if the game has nested strategic information, and I is a μ-basic information set. In this case, the noted strategy s_i^* has strictly higher expected payoff than s_i conditional on $S_{-i}(I)$. Furthermore, since I is basic, and s_i^* coincides with s_i at information sets that do not (weakly) follow I, s_i and s_i^* have the same conditional expectation at every information set J with $S_{-i}(J)\supset S_{-i}(I)$. Then, by Definition 5, s_i is *not* weakly preferred

[15] The actual proof relies on the characterization of structural rationality via belief perturbations (Theorem 4).

to s_i^*. In addition, consider an arbitrary μ-basic information set J. If the expected payoff of s_i at J is strictly greater than that of s_i^*, one can show that, with nested strategic information, it *must* be the case that $S_{-i}(I) \supset S_{-i}(J)$. Since s_i^* has a greater conditional expectation than s_i at I, Definition 5 implies that s_i^* is weakly preferred to s_i. Since we also argued that the converse does not hold, s_i^* is *strictly* preferred to s_i. Hence, as in the argument for EU preferences, if s_i is structurally rational, it must maximize the conditional expected payoff at I.

Example 1 (in the case $x = 2$) shows that the converse to Theorem 1 does not hold. However, structural and sequential rationality are "generically" equivalent. A precise statement of this result requires an explicit description of extensive-form games that goes beyond the notation introduced in Section 2.[16] In order to focus on the important issue of elicitation, I relegate the formal statement and proof of this result (Theorem 2) to Online Appendix C, and instead provide an informal description. Fix an extensive game *tree*, having z terminal nodes, and a player i. A *payoff assignment* for i is a specification of i's payoff at each terminal node. Each payoff assignment is thus a z-dimensional vector.

> **Theorem.** Fix a CPS μ and a strategy s_i for player i. Except for a set of payoff assignments for i of dimension strictly less than z, s_i is sequentially rational given μ *if and only if* it is structurally rational given μ.

5.2 Eliciting Conditional Beliefs

5.2.1 Structural preferences in the elicitation game of Figure 2

I first show that, if Bob has structural preferences in the game of Figure 2, then his initial choice conveys information about his beliefs conditional upon observing Ann's move *In*. The strategy sets are $S_a = \{Out, InB, InS\}$ and $S_b = \{pB, pS, bB, bS\}$, where, as in previous examples, *Out* denotes either one of the realization-equivalent strategies *OutB*, *OutS*, etc.. Furthermore,

[16] In addition, the proof of this result relies on the properties of extensive game forms with perfect recall.

$\mathscr{I}_b = \{\phi, J, K\}$ with $S_a(J) = S_a(K) = \{InB, InS\}$. The game has nested strategic information. Assume that Bob's CPS μ satisfies $\mu(\{Out\}|S_a) = 1$ and $\mu(\{S\}|S_a(J)) = \mu(\{S\}|S_a(K)) = \pi \in [0,1]$ (the Introduction focused on the case $\pi = 1$). Bob's expected payoffs are depicted in Table II. Recall that Figure 2 displays both a "game" and a "betting" payoff for Bob at each terminal node, and a fair coin toss determines which one Bob receives. Each entry in Table II is thus the expectation with respect to the relevant belief on S_a as well as the lottery probabilities.

s_b	S_a	$S_a(J) = S_a(K)$
pB	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot (1-\pi) + \frac{1}{2} \cdot p$
pS	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 3\pi + \frac{1}{2} \cdot p$
bB	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot (1-\pi) + \frac{1}{2} \cdot \pi$
bS	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 3\pi + \frac{1}{2} \cdot \pi$

Table II: Bob's expected payoffs for the game in Figre 2.

This randomization ensures that Bob has strict incentives to choose the best "game" action (B vs. S) and the best "betting" action (b vs. p).[17] In particular, the best game action is S if and only if $\pi > \frac{1}{4}$ and, crucially, the best betting action is b if and only if $\pi > p$. Remark 1 then readily imply that Bob's choice at the initial node reveals whether or not he assigns probability greater than p to S conditional upon Ann choosing *In*. (For instance, if $\pi = 1$, the unique structurally rational strategy for Bob is bS; if $\pi = 0$, it is pB.)

This construction only allows one to conclude whether $\pi > p$ or $\pi \leq p$. To obtain tigher bounds on beliefs, one can employ richer betting choices, such as price lists, scoring rules,

[17] In several experimental papers (e.g., Van Huyck, Battalio, and Beil, 1990; Nyarko and Schotter, 2002; Costa-Gomes and Weizsäcker, 2008; Rey-Biel, 2009), payoffs are monetary, and game and betting payoffs are simply added. Under risk neutrality, this provides correct incentives. Blanco et al. (2010) argue that, if players are risk-averse, randomization addresses the concern that betting choices may be used to "hedge" against uncertainty in the game. I use randomization primarily because, throughout this paper, outcomes are expressed in *utils*, so randomization is the appropriate way to combine game and betting payoffs.

or the Becker, DeGroot, and Marschak (1964) mechanism. Incorporating these mechanisms into the elicitation game does not change the basic insight, but requires additional notation. To streamline the exposition, this section focuses on simple bets as in this example.

5.2.2 On-path beliefs about off-path moves: the strategy method

Assume again that the subgame-perfect equilibrium in which Ann plays *Out* prevails. To elicit Ann's initial beliefs, an experimenter could in principle offer her side bets on Bob's choice of B vs S. These would be offered at the initial node, so Ann's betting behavior would be observable.

However, a new issue arises. If Ann's *game* choice at the initial node is *Out*, Bob's move is not observed. Thus, whatever Ann's *betting* choice may be, it is not in response to real incentives: after all, Ann understands that the simultaneous-move subgame will not be reached.[18]

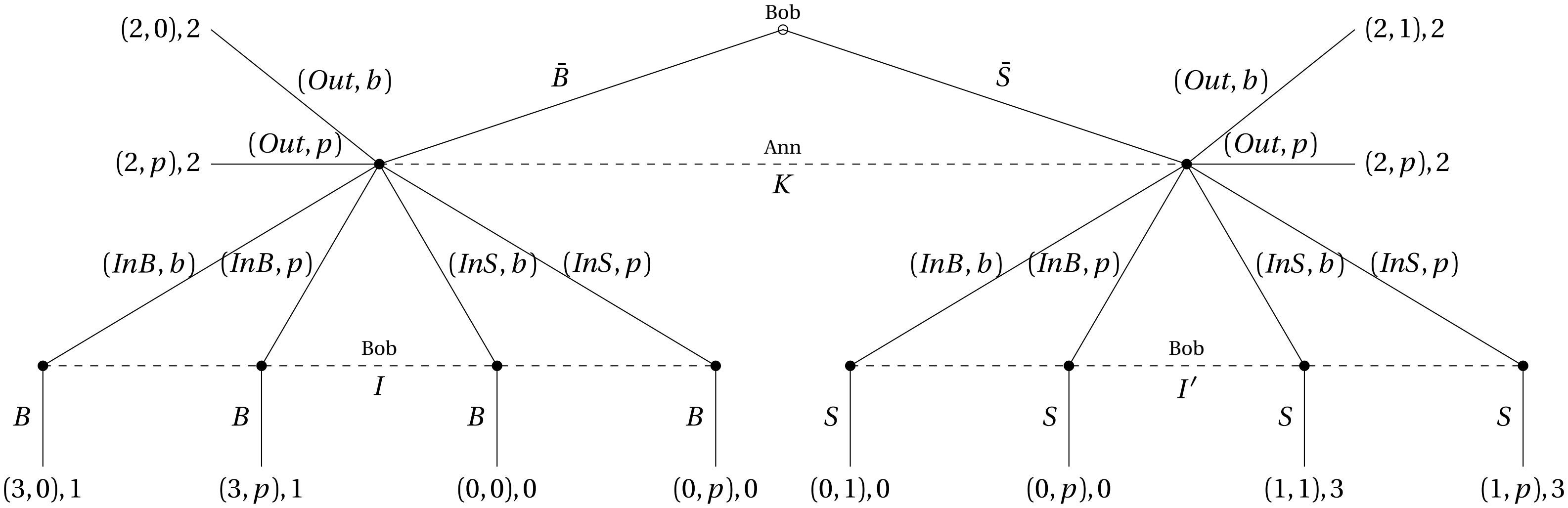


Figure 6: Eliciting Ann's initial beliefs in the game of Figure 1.

The approach I propose implements the game (and bets) using the *strategy method* of Selten (1967). Recall that, in this protocol, players simultaneously commit to extensive-form

[18] Modifying the game so that the subgame *is* reached, perhaps with small probability, may change the nature of the strategic interaction and so invalidate the elicitation exercise: see §6.E.

strategies; the experimenter then implements them. Figure 6 depicts a simplified[19] strategy-method elicitation game in which Ann bets on Bob's choice of S. In this game, Ann's choice of b vs. p is observed *and* has actual payoff consequences: betting incentives are real.[20]

A feature of Figure 6 is that, at information sets I and I', Bob learns that Ann chose *In*. This is exactly what he learns at J in Figure 1.[21] Thus, the conditioning events for Bob in Figure 6 "correspond to" his conditioning events in Figure 1. This implies that any CPS for Bob in Figure 1 can be used to define a CPS in Figure 6 that preserves Bob's beliefs about Ann's choices of *In* vs. *Out* and B vs. S. (Definitions 9 and 10 formalize this.) The same is true for Ann's conditioning events and beliefs. The key insight of this subsection is that, under structural rationality, if Bob's conditional beliefs about Ann are the same in Figures 1 and 6, then *by definition, so will his preferences*, and hence his behavior. In particular, if Bob assigns probability one to S following Ann's unexpected choice of *In*, he must play $\bar{S}$ in Figure 6. Consequently, if Ann anticipates this, (Out, b) is her unique (ex-ante and structural) best reply for any $p < 1$. (The calculations are in Online Appendix D.1.) Thus, analogously to Figure 2, Ann's initial betting choice conveys information about the beliefs she holds in both Figures 6 and 1.

Assuming structural, rather than sequential rationality, is crucial to this conclusion. For instance, with $p = \frac{2}{3}$ in Figure 6, there is a sequential equilibrium in which Bob plays $\bar{B}$ and $\bar{S}$ with equal probability, Ann plays (Out, p), and at both I and I' Bob assigns probability one to *InS*. Bob's beliefs about Ann's game actions are as in the $(Out, (S, S))$ equilibrium of Figure 1. However, sequential rationality allows Bob's behavior to differ in the two games. Consequently, while Ann's betting behavior correctly reveals her (prior) beliefs in Figure 6, these do not correspond to her beliefs in the posited equilibrium of the game in Figure 1.

[19] Figure 6 does not distinguish between Ann's commitment choice of a strategy and its implementation. This is inessential for structural rationality, because the only conditioning event for Ann is S_b in both cases.

[20] Strictly speaking, Ann bets on $\bar{S}$ in Figure 6; however, $\bar{S}$ commits Bob to choosing S at I'.

[21] The information sets I and I' in Figure 6 are distinct only because they also encode Bob's own past choice of $\bar{B}$ vs. $\bar{S}$. Note also that, at I and I', Bob is committed to playing the action he has chosen at the initial node.

5.2.3 The general elicitation game

I now formalize the construction of the elicitation game associated with an arbitrary dynamic game. As in Figure 6, I employ a specific implementation of the strategy method in which, as the experimenter executes the strategies chosen in the commitment stage, players receive the same information about opponents' actions as in the original game. A coin toss, modeled as the choice of a dummy chance player, and not observed until a terminal node is reached, determines whether subjects receive their game or betting payoff.[22]

As in Figures 2 and 6, I restrict attention to bets that only reveal whether the probability a player assigns to a given event at a given information set is above or below a certain value; further extensions are only a matter of additional notation. I allow for belief bounds to be simultaneously elicited from zero, one, or more of players.[23]

Definition 9 [24] *A **questionnaire** is a collection* $Q = (I_i, W_i)_{i \in N}$ *such that, for every* $i \in N$*,* $I_i \in \mathscr{I}_i$ *and either* $W_i = \{*\}$ *or* $W_i = (E, p)$ *for some* $E \subseteq S_{-i}(I)$ *and* $p \in [0,1]$*. The **elicitation game for the questionnaire*** $Q = (I_i, W_i)_{i \in N}$ *is the tuple* $\big(N \cup \{c\}, (S_i^*, \mathscr{I}_i^*, U_i^*)_{i \in N \cup \{c\}}, S^*(\cdot)\big)$*, where* $S_c^* = \{h, t\}$*,* $\mathscr{I}_c^* = \{\phi^*\}$*,* $U_c^* \equiv 0$*, and for all* $i \in N$*:*

1. *(Strategies)* $S_i^* = S_i \times W_i$*;*
2. *(Information)* $\mathscr{I}_i^* = \{\phi^*, I_i^1\} \cup \{(s_i, w_i, I) : (s_i, w_i) \in S_i^*, I \in \mathscr{I}_i, s_i \in S_i(I)\}$*;*
3. *(First stage)* $S^*(I_i^1) = S^*$
4. *(Second stage) for all* $(s_i, w_i, I) \in \mathscr{I}_i^*$*,* $S^*\big((s_i, w_i, I)\big) = \{(s_i, w_i)\} \times S_{-i}(I) \times W_{-i} \times S_c^*$*;*[25]
5. *(Payoffs) for all* $\big((s_i, w_i), (s_{-i}, w_{-i}), s_c^*\big) \in S^*$*: if* $s_c^* = h$ *or* $W_i = \{*\}$*, then* $U_i^*\big((s_i, w_i), (s_{-i}, w_{-i}), s_c^*\big) =$

[22] For notational simplicity, in Definition 9 the same coin toss selects game or betting payoffs for all players. One can alternatively assume i.i.d. coin tosses for each player, and/or i.i.d. coin tosses at each terminal node, provided one makes the appropriate assumptions on players' beliefs about the chance player (cf. Definition 10).

[23] Thus, a justification for the use of the strategy method without belief elicitation follows as a corollary.

[24] I use the formalism of Section 2. Online Appendix C.3 formalizes the extensive form of the elicitation game.

[25] Here and in part 5, it is convenient to decompose $S^* = (S_i \times W_i) \times (S_{-i} \times W_{-i}) \times S_c^*$.

$U_i(s_i, s_{-i})$; and if $s_c^* = t$ and $W_i = (E, p)$, then

$$U_i^*\big((s_i, E), (s_{-i}, w_{-i}), t\big) = \begin{cases} 1 & s_{-i} \in E \\ 0 & \textit{otherwise} \end{cases} \quad \textit{and} \quad U_i^*\big((s_i, p), (s_{-i}, w_{-i}), t\big) = \begin{cases} p & s_{-i} \in S_{-i}(I_i) \\ 0 & \textit{otherwise.} \end{cases}$$

Thus, chance can select either h, in which case payoffs are as in the original game, or t, in which case payoffs are given by betting choices for every player whose beliefs are being elicited. Each player i chooses a strategy s_i and betting action w_i at her first-stage information set I_i^1, without any knowledge of coplayers' moves. At every second-stage information set (s_i, w_i, I), player i recalls her first-stage choice (s_i, w_i); furthermore, what i learns about her (real) coplayers at (s_i, w_i, I) is precisely what she learns about them at I in the original game.[26]

Next, I formalize the assumptions that players hold (a) the same beliefs about coplayers in the original game and in the elicitation game, and (b) view chance moves as independent of coplayers' strategies. This ensures that conditional expected payoffs are $\frac{1}{2} : \frac{1}{2}$ mixtures of game and betting payoffs, as in Table II (cf. Lemma 6).

Definition 10 *Fix a questionnaire $(I_i, W_i)_{i\in N}$. Let $\big(N^*, (S_i^*, \mathscr{I}_i^*, U_i^*)_{i\in N^*}, S^*(\cdot)\big)$ be the associated elicitation game. For any $i \in N$ and CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$, the CPS $\mu^* \in \Delta(S_{-i}^*, S_{-i}^*(\mathscr{I}_i^*))$* ***agrees with*** *$\mu$ if, for every $I^* \in \mathscr{I}_i^*$,*

$$\text{marg}_{S_{-i}\times S_c^*}\mu^*\big(\cdot\,|S_{-i}^*(I^*)\big) = \frac{1}{2}\mu\big(\cdot\,|\text{proj}_{S_{-i}}S_{-i}^*(I^*)\big).^{27} \tag{4}$$

More than one CPS for player i in the elicitation game may agree with her CPS in the original game. This is because i may assign different probabilities to her coplayers' choices of side bets in the elicitation game. However, these differences are irrelevant for her strategic reasoning, because her payoff does not depend on these choices. On the other hand, independence

[26] Part 4 of the definition also indicates that i has a single action available at (s_i, w_i, I); see Appendix C.3.

[27] By Definition 9, $\text{proj}_{S_i}S_{-i}^*(\phi^*) = \text{proj}_{S_{-i}}S_{-i}^*(I_i^1) = S_{-i}$ and, for all $(s_i, w_i, I) \in \mathscr{I}_i^*$, $\text{proj}_{S_{-i}}S_{-i}^*((s_i, w_i, I)) = S_{-i}(I)$.

of Chance's move is important: if i believes that her coplayers correlate their choices with Chance, this may impact her expected payoffs, and hence her strategic incentives.

The main result of this section can now be stated: if the strategy method is implemented as described above, and players' beliefs about others' moves are the same as in the original game, then (1) players' preferences are also unchanged, and (2) as a result, belief bounds can be elicited from initial, observable betting choices.

Theorem 2 *Fix a questionnaire $(I_i, W_i)_{i\in N}$. Let $\big(N^*, (S_i^*, \mathscr{I}_i^*, U_i^*)_{i\in N^*}, S^*(\cdot)\big)$ be the associated elicitation game. For any player $i \in N$, fix an extensible CPS $\mu_i \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$. Then there exists an extensible CPS $\mu_i^* \in \Delta(S_{-i}^*, S_{-i}(\mathscr{I}_i^*))$ that agrees with μ_i. For any such CPS,*

(1) for all $(s_i, w_i), (t_i, w_i) \in S_i^*$, $(s_i, w_i) \succsim^{\mu_i^*} (t_i, w_i)$ if and only if $s_i \succsim_i^{\mu} t_i$;

(2) if $W_i = (E, p)$, then for all $s_i \in S_i$, $p > \mu_i(E|S_{-i}(I_i))$ implies $(s_i, p) \succ^{\mu_i^*} (s_i, b)$ and $p < \mu_i(E|S_{-i}(I_i))$ implies $(s_i, b) \succ^{\mu_i^*} (s_i, p)$.

Hence, if $W_i = (E, p)$ and (s_i, b) (resp. (s_i, p)) is structurally rational in the elicitation game, then s_i is structurally rational in the original game, and $\mu_i(E|S_{-i}(I_i)) \geq p$ (resp. $\mu_i(E|S_{-i}(I_i)) \leq p$).[28]

This result also provides a positive rationale for the use of the strategy method:

Corollary 1 *Under the assumptions of Theorem 2, suppose that $W_i = \{*\}$ for all $i \in N$. Then, for all $i \in N$ and all $s_i, t_i \in S_i$, $s_i \succsim^{\mu_i} t_i$ if and only if $(s_i, *) \succsim^{\mu_i^*} (t_i, *)$. In particular, s_i is structurally rational in the original game if and only if $(s_i, *)$ is structurally rational in the elicitation game.*

Theorem 2 and Corollary 1 depend crucially on the assumption that players are structurally rational. Sequential rationality is not sufficient to deliver these results, *even if players' conditional beliefs are the same as in the original game*:

Remark 4 *Under the assumptions of Theorem 2, for every player $i \in N$, $(s_i, w_i) \in S_i^*$ is sequentially rational in the elicitation game if and only if (i) $s_i \in \arg\max_{t_i \in S_i} \mathrm{E}_{\mu_i(\cdot|S_{-i})} U_i(t_i, \cdot)$, and (ii) if*

[28] A weak inequality is needed because, if $p = \mu_i(E|S_{-i})$, the strategies (s_i, b) and (s_i, p) may be incomparable.

$W_i = (E, p)$ and $w_i = b$ (resp. $w_i = p$), then $\mu_i(E|S_{-i}) \geq p \cdot \mu_i(S_{-i}(I)|S_{-i})$) (resp. $\mu_i(E|S_{-i}) \leq p \cdot \mu_i(S_{-i}(I)|S_{-i})$)).

This is an immediate consequence of the fact that, for each player i, the only information set in the elicitation game where more than one action is available is I_i^1.

To reconcile Theorem 2 and Remark 4 with the generic equivalence result described in Section 5.1, notice that elicitation games feature numerous ties at terminal nodes: see Fig. 6. In other words, by construction elicitation games are non-generic, and such that structural rationality is strictly stronger than sequential rationality.

5.3 Extensibility, Structural Rationality, and Trembles

The final set of results relates the approach in this paper with the traditional notion of "trembles," or belief perturbations. First, I show that a CPS is extensible if and only if it can be obtained by taking limits of perturbed beliefs. Second, I provide a characterization of structural rationality as optimality given a novel class of belief perturbations.

Example 4 (Non-extensible beliefs) Consider the game in Figure 7.

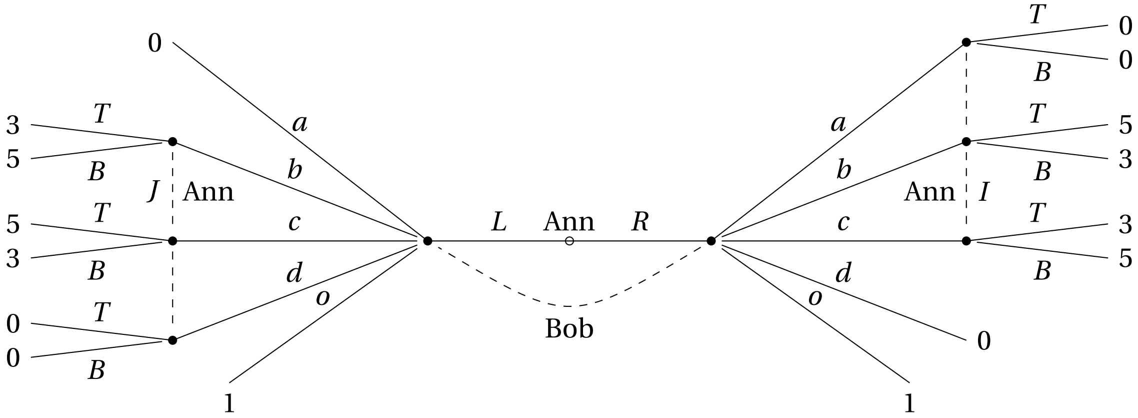


Figure 7: A conditional Newcomb's paradox (only Ann's payoffs are shown)

Suppose that Ann's CPS μ satisfies $\mu(\{o\}|S_b(\phi)) = \mu(\{b\}|S_b(I)) = \mu(\{c\}|S_b(J)) = 1$. Observe that $S_b(\phi) >^{\mu} S_b(I) =^{\mu} S_b(J)$. Hence, $B_\mu(I) = B_\mu(J) = S_b(I) \cup S_b(J) = \{a, b, c, d\}$, and any extension $\nu \in \Delta(S_b, S_b(\mathscr{I}_a) \cup B_\mu(\mathscr{I}_a))$ must satisfy

$$\mu(\{b\}|S_b(I)) \cdot \nu(S_b(I)|B_\mu(I)) \;=\; \nu(\{b\}|B_\mu(I)) \;=\; \nu(\{b\}|B_\mu(J)) = \mu(\{b\}|S_b(J)) \cdot \nu(S_b(J)|B_\mu(J)), \text{ and}$$
$$\mu(\{c\}|S_b(I)) \cdot \nu(S_b(I)|B_\mu(I)) \;=\; \nu(\{c\}|B_\mu(I)) \;=\; \nu(\{c\}|B_\mu(J)) = \mu(\{c\}|S_b(J)) \cdot \nu(S_b(J)|B_\mu(J)).$$

Since $\mu(\{b\}|S_b(I)) > 0$ and $\mu(\{b\}|S_b(J)) = 0$, the first equation implies $\nu(S_b(I)|B_\mu(I)) = 0$; similarly, since $\mu(\{c\}|S_b(I)) = 0$ and $\mu(\{c\}|S_b(J)) > 0$, the second equation implies $\nu(S_b(J)|B_\mu(J)) = 0$. But $B_\mu(I) = B_\mu(J) = S_b(I) \cup S_b(J)$, contradiction. Hence μ is not extensible. □

A peculiar feature of the CPS μ in Example 4 is that Ann's own initial choice of R vs. L determines her conditional beliefs on the relative likelihood of b and c, despite the fact that Bob does not observe Ann's initial choice. (In fact, Ann's first action and Bob's move may well be simultaneous.) This phenomenon is reminiscent of Newcomb's paradox (Weirich, 2016).

If $S_b(I)$ and $S_b(J)$ both had positive prior probability, the definition of conditional probability would imply that the relative likelihood of b and c must be the same at both information sets. The same conclusion holds in any consistent assessment in the sense of Kreps and Wilson (1982), and for CPSs for which *all* non-empty subsets of S_{-i} are conditioning events, as in Myerson (1986). The reason is that, in both cases, Ann's conditional beliefs at I and J are obtained by fixing a sequence (p^n) of strictly positive probability distributions on S_b, and taking the limit of the conditional probabilities $p^n(\cdot|S_b(I))$ and $p^n(\cdot|S_b(J))$.[29]

To sum up, if a CPS is extensible, or if it is the limit of belief perturbations, the pathologies in Example 4 do not arise. The next result shows that, in fact, belief perturbation *characterize* extensible CPSs. It also establishes uniqueness of the extension of a CPS. Throughout this subsection, in addition to the game $\big(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot)\big)$, also fix a player $i \in N$ and a CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$.

[29] Modulo notational differences, this is true by definition for consistent assessments; for complete CPSs, it follows from a result in Myerson (1986).

Definition 11 *A sequence* $(p^n)_{n\geq 1} \subset \Delta(S_{-i})$ *is a* ***perturbation*** *of* μ *if* $p^n(S_{-i}(I)) > 0$ *for all* $n \geq 1$ *and* $I \in \mathscr{I}_i$*, and* $p^n(\cdot|S_{-i}(I)) \to \mu(\cdot|S_{-i}(I))$ *for all* $I \in \mathscr{I}_i$*.*

A perturbation need not consist of full-support probabilities, so long as every conditioning event has positive probability. (The reason for this will be clear momentarily.) A particular class of perturbations plays a key role in the characterization of structural preferences:

Definition 12 *A perturbation* $(p^n)_{n\geq 1}$ *of* μ *is* ***structural*** *if* $\operatorname{supp} p^n = \bigcup_{I\in\mathscr{I}_i} \mu(\cdot|S_{-i}(I))$ *for every* $n \geq 1$*, and* $\frac{p^n(\{s_{-i}\})}{p^n(\{t_{-i}\})} = \frac{\mu(\{s_{-i}\}|S_{-i}(I))}{\mu(\{t_{-i}\}|S_{-i}(I))}$ *for all* $n \geq 1$*,* $I \in \mathscr{I}_i$*, and all* $s_{-i}, t_{-i} \in \operatorname{supp} \mu(\cdot|S_{-i}(I))$*.*

The rationale for this definition is given below. The first main result of this section is:

Theorem 3 *The following are equivalent:*

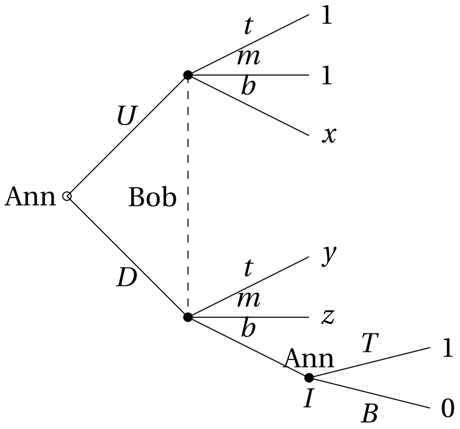
1. μ *admits an extension* $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i))$*;*
2. *there is a structural perturbation* $(p^n) \subset \Delta(S_{-i})$ *of* μ*;*
3. *there is a perturbation* $(p^n) \subset \Delta(S_{-i})$ *of* μ*.*

If μ *is extensible, then its extension is unique.*

Online Appendix B provides a further characterization of extensibility via an intrinsic property of CPSs that strengthens the chain rule of conditioning.

The next result shows that structural rationality can itself be characterized via structural perturbations of the player's beliefs. Trembles have been used in the literature to refine equilibria (e.g. Myerson, 1978; Van Damme, 1984; Kohlberg and Mertens, 1986). However, the result presented below is closer in spirit to the use of trembles in Selten (1975), who introduced perturbations to ensure "perfection" (optimality at unreached information sets). The objective of Theorem 4 below is to identify perturbations that do *not* refine structural rationality, and indeed characterize it exactly. The following example illustrates.

Example 5 Consider the game in Fig. 8. I analyze two parameterizations of Ann's payoffs and prior belief, with $p \equiv \mu(\{t\}|S_b) = 1 - \mu(\{m\}|S_b)$ and $\mu(\{b\}|S_b) = 0$.



s_a	$\mathrm{E}_{\mu(\cdot\vert S_b)} U_a(s_a,\cdot)$	$\mathrm{E}_{\mu(\cdot\vert S_b(I))} U_a(s_a,\cdot)$
U	1	x
DT	$py+(1-p)z$	1
DB	$py+(1-p)z$	0

Figure 8: Structural perturbations. Ann's payoffs shown. $p=\mu(\{t\}|S_b)=1-\mu(\{m\}|S_b)$.

Supports: let $x=1$, $y=1$, $z=0$, and $p=1$. Then U weakly dominates DT, and so $\mathrm{E}_{p^n} U_a(U,\cdot) > \mathrm{E}_{p^n} U_a(DT,\cdot)$ whenever (p^n) is a full-support perturbation of μ. However, $UT \sim^{\mu} DT$. Indeed, for DT to be a best reply to a perturbation of Ann's CPS μ, it must be the case that $p^n(\{m\})=0$. Note that the CPS μ assigns probability 0 to $\{m\}$ both ex-ante and given $S_b(I)$.

Relative Likelihoods: now let $x=0$, $y=0$, $z=2$, and $p=\frac{1}{2}$. Then $DT \succ^{\mu} U$. Define a sequence $(p^n) \in \Delta(S_b)$ by letting $p^n(\{t\})=\frac{1}{2}$, $p^n(\{m\})=\frac{1}{2}-\frac{1}{2n}$, and $p^n(\{b\})=\frac{1}{2n}$ for every $n \geq 1$. Then (p^n) is a perturbation of μ; however, $\mathrm{E}_{p^n} U_a(U,\cdot) > \mathrm{E}_{p^n} U_a(DT,\cdot)$ for all $n>1$. Here, the fact that $\mu(\cdot|S_b)$ assigns equal weight to t and m ensures that U and DT have the same ex-ante expected payoff, and hence is crucial to conclude that $DT \succ^{\mu} U$. However, p^n does not preserve the relative likelihood of t and m.

Example 5 motivates the definition of structural perturbations given above. These trembles preserve two key features of the limiting CPS: supports and relative likelihoods. Loosely, the objective is to modify the player's beliefs only insofar as it is necessary to ensure that all her information sets are reached.

Note that, if the game has simultaneous moves, or if all information sets of i have positive prior probability under μ, then the *unique* structural perturbation (p^n) of μ is defined by $p^n = \mu(\cdot|S_{-i})$ for all n. This underscores the fact that structural perturbations only modify a player's (prior) beliefs in a "minimal" way so as to ensure that there are no unexpected information sets: if no information set is unexpected to begin with, there is no need for *any* perturbation.

Theorem 4 *For every $s_i, t_i \in S_i$:*

(1) $s_i \succsim^{\mu} t_i$ if and only if, for every structural perturbation $(p^n)_{n\geq 1}$ of μ, there is $\bar{n} \geq 1$ such that $\mathrm{E}_{p^n} U(s_i, \cdot) \geq \mathrm{E}_{p^n} U_i(t_i, \cdot)$ for all $n \geq \bar{n}$;

(2) $s_i \succ^{\mu} t_i$ if and only if, for every structural perturbation $(p^n)_{n\geq 1}$ of μ, there is $\bar{n} \geq 1$ such that $\mathrm{E}_{p^n} U(s_i, \cdot) > \mathrm{E}_{p^n} U_i(t_i, \cdot)$ for all $n \geq \bar{n}$.

Therefore, a strategy $s_i \in S_i$ is structurally rational for μ if, for every $t_i \in S_i$, there is a structural perturbation (p^n) of μ such that $\mathrm{E}_{p^n} U(s_i, \cdot) \geq \mathrm{E}_{p^n} U_i(t_i, \cdot)$ for all $n \geq 1$.

Theorem 4 yields a tight characterization of structural preferences and rationality. As shown in Example 5, non-structural perturbations may fail to rationalize a given structurally rational strategy. In addition, the second parameterization in the example shows that a best reply to a non-structural perturbation may fail to be structurally rational given the limiting CPS.

In the last statement of Theorem 4, the structural perturbation for which s_i yields a weakly higher expected payoff than an alternative strategy t_i may well vary with t_i. This is a direct implication of part (2) in the Theorem; Example 1 in Online Appendix D.2 illustrates this point.

6 Discussion

6.A Material payoffs. The partial representation of a dynamic game given in Section 2 is sufficient to state the main definitions and results in this paper. One can enrich this representation by replacing player i's reduced-form payoff functions $U_i : S \to \mathbb{R}$ with (i) a set of *material consequences* X_i, (ii) a *consequence function* $C_i : S \to X_i$, and (iii) a (von Neumann-Morgenstern) *utility function* $u_i : X_i \to \mathbb{R}$: thus, $U_i = u_i \circ C_i$. In this case, (i) and (ii) are part of the description of the game; (iii) is part of the representation of players' preferences. If Definitions 3, 5, 8, 11 and 12 are modified in the obvious way, Theorems 1, 3 and 4 continue to hold.

Furthermore, if the sets X_i are sufficiently rich (e.g., the set of lotteries on some prize space X_0), Theorem 2 can be adapted so that both beliefs and utilities can be elicited in the game.

6.B Incomplete-information games The analysis may also be adapted to accommodate incomplete information. Fix a dynamic game with N players, strategy sets S_i and information sets $\mathscr{I}_i$ for each $i \in N$, and a strategy profile correspondence $S(\cdot)$. Consider sets Θ_i of possible "types" for each $i \in N$, and a set Θ_0 that captures residual uncertainty not reflected in players' types. Player i's payoff function is a map $U_i : S \times \Theta \to \mathbb{R}$, where $\Theta = \Theta_0 \times \prod_{j \in N} \Theta_j$. The set of conditioning events for player i is $\mathscr{F}_i = \{S_{-i}(I) \times \Theta_{-i} : I \in \mathscr{I}_i\}$, where $\Theta_{-i} = \Theta_0 \times \prod_{j \in N \setminus \{i\}} \Theta_j$. The conditional beliefs of player i's type θ_i can then be represented via a CPS μ_{θ_i} on $S_{-i} \times \Theta_{-i}$, with conditioning events $\mathscr{F}_i$. Definitions 3, 5 and 8 can be readily adapted to characterize notions of sequential and structural rationality for any given type $\theta_i \in \Theta_i$). Theorems 1 and 2 have straightforward extensions, even if some of the sets Θ_j are uncountably infinite. For finite Θ, Definitions 11 and 12, as well as Theorems 3 and 4, also extend readily.

6.C Higher-order beliefs The proposed approach can also be adapted to elicit higher-order beliefs. Consider a two-player game for simplicity. The analyst begins by eliciting Ann's first-order beliefs about Bob's strategies, as in Section 5.2. She can then elicit Bob's second-order beliefs by offering him side bets on both Ann's strategies *and* on her first-order beliefs. The required formalism is analogous to that for incomplete information, taking $\Theta_i = \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$ for each player i. The incomplete-information extension of Theorem 2 ensures that they can be elicited in an incentive-compatible way. The argument extends to beliefs of higher orders.

6.D Elicitation: ex-ante analysis. As argued in the Introduction, Bob's beliefs at J in the subgame-perfect equilibrium$(Out, (S, S))$ of the game of Figure 1 cannot be elicited under sequential rationality. A possible response is to note that the strategic reasoning that supports this equilibrium can be restated entirely in terms of Ann's *ex-ante, second-order* beliefs, without reference to Bob's *actual* (first-order) beliefs at J.[30] However, the issue is how to elicit Ann's initial second-order beliefs in an incentive-compatible way. As discussed above, this involves asking Ann to bet on Bob's actual, elicited beliefs at J. Thus, from a behavioral perspective, the elicitation of off-path beliefs *is* relevant in an ex-ante view of strategic reasoning as well.

6.E Elicitation: modified or perturbed games. In the equilibrium $(Out, (S, S))$ of the game of Figure 1, Ann's initial move prevents J from being reached. One might consider modifying the game so that J *is* actually reached, perhaps with small probability, regardless of Ann's initial move. However, such modifications may have a significant impact on players' strategic reasoning and behavior—and therefore on elicited beliefs. For instance, in the game of Figure 1, *forward-induction* reasoning selects the equilibrium $(In, (B, B))$ (cf. e.g. Ben-Porath and Dekel, 1992). Thus, if Ann follows the logic of forward induction, she should expect Bob to play B in the subgame. However, consider the extreme case in which action *Out* is removed. The game of Figure 1 reduces to the simultaneous-move Battle of the Sexes, in which forward induction has no bite. Ann may well expect Bob to play B in the game of Figure 1, and S in the game with *Out* removed. Thus, Ann's beliefs elicited in the latter game may differ from her actual beliefs in the former. Similar conclusions hold if one causes Ann to play *In* with positive probability when she chooses *Out*. Analogous arguments apply to backward-induction reasoning: see e.g. Ben-Porath (1997), Example 3.2 and p. 36.

By way of contrast, the elicitation approach in Section 5.2 only modifies the game in ways that, as per Statement (1) of Theorem 2, are inessential for each player's structural preferences.

6.F Partial and complete ordering of beliefs Recall from Section 3 that the notion of lexicographic consistency in (Kreps and Wilson, 1982) involves a complete ordering of a player's alternative probabilistic hypotheses. The ordering of beliefs used in Definitions 5 and 8 is

[30] That is: whether Bob would *actually* assign high probability to S at J is irrelevant; what matters is that Ann *initially believe* that he would, and that this would induce him to play S. I thank Phil Reny for this observation.

instead only partial. I offer two complementary interpretation.

In the first interpretation, structural rationality reflects a generalization of lexicographic consistency whereby a player *actually* only entertains a partial order over alternative hypotheses, which the analysis in Sections 3 and 4 elicits from conditional beliefs at each information set. To preserve the spirit of the Kreps and Wilson (1982) definition, one does need to ensure that, for each information set I, there is a unique "most plausible" hypothesis that is consistent with it. Observation 2 in the Appendix establishes this uniqueness.

In the second interpretation, the player's actual ordering of alternative hypotheses may be richer than the one elicited in Sections 3 and 4, and possibly complete. However, this richer ranking *cannot be elicited* from the player's collection of beliefs at each information set. Structural rationality is defined in a way that is *robust* to the specification of such unelicitable rankings. Example 2 illustrates this: if Ann actually deems $\mu(\cdot|S_b(I))$ more (resp. less) plausible than $\mu(\cdot|S_b(J))$, then UT (resp. DT') is the only (lexicographically) rational strategy. Since Ann's CPS does not identify the ranking of $\mu(\cdot|S_b(I))$ vs. $\mu(\cdot|S_b(J))$, structural rationality allows both UT and DT' as best replies. (Recall that these are also the sequential best replies to Ann's CPS.) Online Appendix E.2 shows that this holds generally: indeed, as I now discuss, there is a tight connection between structural preferences and lexicographic preferences defined by completions of the elicited ranking of basic beliefs.

6.G Lexicographic expected utility. As noted in Section 3, structural preferences are formally a generalization of lexicographic preferences (Blume et al., 1991a). Whereas lexicographic preferences were introduced into game theory in order to study *strategic-form* refinements (Blume, Brandenburger, and Dekel, 1991b), the definition of structural preferences is clearly tied to a specific extensive form. Also, recall that lexicographic maximization with respect to a full-support lexicographic probability system implies invariance in the sense of Kohlberg and Mertens (1986): the same strategies will be optimal regardless of the extensive form of the game. Structural rationality is conceptually closer to sequential rationality, in that the given extensive form is essential.

That said, there are useful connections between structural and lexicographic rationality. First, as informally described in Section 3, structural rationality can be thought of as "lexicographic rationality along each path:" see Online Appendix E.1. Second, strategy s_i is structurally preferred to strategy t_i *if and only if* s_i is lexicographically preferred to t_i for every completion of the elicited plausibility ordering of basic events: see Online Appendix E.2.

6.H Preferences for the timing of uncertainty resolution The fact that structural preferences depend upon the extensive form of the dynamic game can be seen as loosely analogous to the issue of sensitivity to the timing of uncertainty resolution: see e.g. Kreps and Porteus (1978); Epstein and Zin (1989), and in particular Dillenberger (2010). In the latter paper, preferences are allowed to depend upon whether information is revealed gradually rather than in a single period, even if no action can be taken upon the arrival of partial information. This is close in spirit to the observation that subjects behave differently in the strategic form of a dynamic game (where all uncertainty is resolved in one shot), and when the game is played with commitment as in the strategy method (where information arrives gradually). However, for structural preference, this dependence on the timing of uncertainty resolution is only allowed when some piece of partial information has zero prior probability—that is, when there is *unexpected* partial information.

A Appendix: Preliminary results on extensible CPSs

Throughout, fix a dynamic game $(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot))$.

For every player i, collection $\mathscr{C}_i \subseteq 2^{S_{-i}} \setminus \{\emptyset\}$,and and CPS $\mu \in \Delta(S_{-i}, \mathscr{C}_i)$, a μ-**sequence** is an ordered list $F_1, \dots, F_K \in \mathscr{C}_i$ such that $\mu(F_{k+1}|F_k) > 0$ for all $k = 1, \dots, K-1$. Thus, for all $F, G \in \mathscr{C}_i$, $F \geq^\mu G$ iff there is a μ-sequence $F_1, \dots, F_K$ with $F_1 = G$ and $F_K = F$.

The following result states that every equivalence class of $\geq^\mu$ can be arranged in a μ-sequence.

Observe that the elements of the μ-sequence constructed in the proof are not all distinct.

Lemma 1 *For every player i, collection $\mathscr{C}_i \subseteq 2^{S_{-i}} \setminus \{\emptyset\}$, CPS $\mu \in \Delta(S_{-i}, \mathscr{C}_i)$, and event $F \in \mathscr{C}_i$, there is a μ-sequence $F_1, \dots, F_M \in \mathscr{C}_i$ such that $F_1 = F_M = F$ and, for all $G \in \mathscr{C}_i$, $G =^\mu F$ if and only if $G = F_m$ for some $m = 1, \dots, M$.*

Proof: let $\{F_1, \dots, F_L\}$ be an enumeration of the equivalence class of $\geq^\mu$ containing F; in particular, assume without loss that $F_1 = F$. Then in particular $F_1 \geq^\mu F_2 \geq^\mu \dots \geq^\mu F_L$ and $F_L \geq^\mu F_1$. By definition, for every $\ell = 1, \dots, L-1$, there is a μ-sequence $F^\ell_1, \dots, F^\ell_{M(\ell)}$ such that $F^\ell_1 = F_{\ell+1}$ and $F^\ell_{M(\ell)} = F_\ell$; furthermore, there is a μ-sequence $F^L_1, \dots, F^L_{M(L)}$ such that $F^L_1 = F_1$ and $F^L_{M(L)} = F^L$. Then the ordered list

$$F^L_1, F^L_2, \dots, F^L_{M(L)} = F^{L-1}_1, \dots, F^{L-1}_{M(L-1)} = F^{L-2}_1, \dots, F^1_{M(1)}.$$

is a μ-sequence, with $F^L_1 = F_1 = F$ and $F^1_{M(1)} = F_1 = F$.

By construction, $F_\ell = F^\ell_{M(\ell)}$ for every $\ell = 1, \dots, L$, so this μ-sequence contains the equivalence class $\{F_1, \dots, F_L\}$ for F. Finally, notice that, for every $\ell = 1, \dots, L$ and $m = 1, \dots, M(\ell)$, the ordered sublist beginning with F^L_1 and ending with F^ℓ_m, and the ordered sublist beginning with F^ℓ_m and ending with $F^1_{M(1)}$, are both μ-sequences, so $F^\ell_m \geq^\mu F^L_1$ and $F^L_{M(1)} \geq F^\ell_m$. Furthermore, $F^L_1 = F^1_{M(1)} = F_1 = F$, so in fact $F^\ell_m \geq^\mu F$ and $F \geq^\mu F^\ell_m$, so $F^\ell_m = F_\ell$ for some $\bar{\ell} = 1, \dots, L$. ■

Corollary 2 *Fix a player $i \in N$ and a CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$. For every $I \in \mathscr{I}_i$, there is a μ-basic $I' \in \mathscr{I}_i$ such that $S_{-i}(I) =^\mu S_{-i}(I')$*

Observation 1 By transitivity of $\geq^\mu$, $S_{-i}(I) =^\mu S_{-i}(I')$ implies $B_\mu(I) = B_\mu(I')$. Hence, by this Corollary, one can drop the requirement that the information sets in Definition 8 be μ-basic.

Proof: By Lemma 1 there is a μ-sequence $F_1, \dots, F_M \in S_{-i}(\mathscr{I}_i)$ such that $F_1 = F_M = S_{-i}(I)$ and $G =^\mu S_{-i}(I)$ iff $G = F_m$ for some $m \in \{1, \dots, M\}$. To simplify the exposition, in this proof only, call an event $F \in S_{-i}(\mathscr{I}_i)$ "μ-basic" if $F = S_{-i}(J)$ for some μ-basic $J \in \mathscr{I}_i$.

Consider the following algorithm. First, set $m_1 = 1$ and $k = 1$. Then, for each $k \geq 1$, if F_{m_k} is μ-basic, STOP; otherwise, (1) find $G \in S_{-i}(\mathscr{I}_i)$ such that $G \supset F_{m_k}$ and $\mu(F_{m_k}|G) > 0$, and (2) find m_{k+1} such that $G = F_{m_{k+1}}$: this must exist, because $\mu(F_{m_k}|G) > 0$ and $\mu(G|F_{m_k}) \geq \mu(F_{m_k}|F_{m_k}) = 1$ imply $G =^{\mu} F_{m_k} =^{\mu} S_{-i}(I)$. Set $k := k+1$ and repeat.

This algorithm produces a sequence $F_{m_1}, F_{m_2}, \ldots$ such that $F_{m_k} \supset F_{m_{m+1}}$ for every $k \geq 1$: thus, the elements of this sequence are all distinct. Since each F_{m_k} is a member of the $\geq^{\mu}$-equivalence class of $S_{-i}(I)$, which is finite, the algorithm must stop. If the algorithm stops at step k, by construction $F_{m_k} =^{\mu} S_{-i}(I)$ and F_{m_k} is μ-basic. Since $F_{m_k} \in S_{-i}(\mathscr{I}_i)$, there is a ($\mu$-basic) $I' \in \mathscr{I}_i$ such that $S_{-i}(I') = F_{m_k}$. ■

Corollary 3 (Remarks 2 and 3) *Fix a player $i \in N$ and a CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$. If the game has nested strategic information, then for every μ-basic $I \in \mathscr{I}_i$, $B_\mu(I) = S_{-i}(I)$. Furthermore, if $I, J \in \mathscr{I}_i$ are μ-basic, then $S_{-i}(I) >^{\mu} S_{-i}(J)$ iff $S_{-i}(I) \supset S_{-i}(J)$.*

Proof: By Lemma 1 there is a μ-sequence $F_1, \ldots, F_M \in S_{-i}(\mathscr{I}_i)$ such that $F_1 = F_M = S_{-i}(I)$ and $G =^{\mu} S_{-i}(I)$ iff $G = F_m$ for some $m \in \{1, \ldots, M\}$, so that $B_\mu(J) = \cup_m F_m$.

I claim that $F_m \subseteq F_M = S_{-i}(I)$ for all m. The claim is trivially true for $m = M$. Assume it is true for some $m \in \{2, \ldots, M\}$. By the definition of a μ-sequence, $\mu(F_m|F_{m-1}) > 0$; by the induction hypothesis $F_M \supseteq F_m$, so $\mu(F_M|F_{m-1}) > 0$, and hence $F_M \cap F_{m-1} \neq \emptyset$. Since the game has nested strategic information, either $F_M \supseteq F_{m-1}$ or $F_{m-1} \supseteq F_M$. By assumption, I is μ-basic, and $\mu(F_M|F_{m-1}) > 0$; thus, it cannot be the case that $F_{m-1} \supseteq F_M$. This proves the claim for $m-1$. Therefore, $B_\mu(I) = \cup_{m=1}^{M} F_m = F_M = S_{-i}(I)$.

Finally, suppose $I, J \in \mathscr{I}_i$ are μ-basic and $S_{-i}(I) \supset S_{-i}(J)$. Then $S_{-i}(I) \geq^{\mu} S_{-i}(J)$; if also $S_{-i}(J) \geq^{\mu} S_{-i}(I)$, then $S_{-i}(I) =^{\mu} S_{-i}(J)$, and so, by the definition of μ-basis and transitivity of $\geq^{\mu}$, $S_{-i}(I) = B_\mu(I) = B_\mu(J) = S_{-i}(J)$, contradiction. Hence, $S_{-i}(I) >^{\mu} S_{-i}(J)$. Conversely, suppose that $S_{-i}(I) >^{\mu} S_{-i}(J)$. Then there is a μ-sequence $F_1, \ldots, F_M$ with $F_1 = S_{-i}(J)$ and $F_M = S_{-i}(J)$. I claim that $F_m \subseteq F_M$ for all m. For $m = M$ the claim is trivial. Suppose the claim is true for

$m \in \{2,\dots,M\}$. Since $\mu(F_m|F_{m-1}) > 0$, also $\mu(F_M|F_{m-1}) > 0$. By nested strategic information and the assumption that I is μ-basic, $F_{m-1} \subseteq F_M$, as claimed. Hence, $S_{-i}(J) = F_1 \subseteq F_M = S_{-i}(I)$. If $S_{-i}(I) = S_{-i}(J)$, then $S_{-i}(I) =^{\mu} S_{-i}(J)$, contradiction: thus, $S_{-i}(I) \supset S_{-i}(J)$. ■

The next result is useful to analyze extensions of CPSs.

Lemma 2 *Fix a player $i \in N$, a CPS $\mu \in \Delta(S_{-i}, S_i(\mathscr{I}_i))$ with extension $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_{\mu}(\mathscr{I}_i))$, and an information set $I \in \mathscr{I}_i$. Consider a collection $F_1,\dots,F_L \in S_{-i}(\mathscr{I}_i)$ such that $F_\ell =^{\mu} F_m$ for all $\ell, m \in \{1,\dots,L\}$, and $\nu(\cup_\ell F_\ell | B_\mu(I)) > 0$. Then there are $\hat{\ell} \in \{1,\dots,L\}$ and $\hat{I} \in \mathscr{I}_i$ such that $S_{-i}(\hat{I}) =^{\mu} S_{-i}(I)$ and $\mu(F_{\hat{\ell}}|S_{-i}(\hat{I})) > 0$. Therefore $F_\ell \geq^{\mu} S_{-i}(I)$ for all ℓ. In particular, $\nu(S_{-i}(J)|B_\mu(I)) > 0$ for all $J \in \mathscr{I}_i$ such that $S_{-i}(J) =^{\mu} S_{-i}(I)$.*

Proof: Denote by $G_1,\dots,G_M$ the $\geq^{\mu}$-equivalence class for $S_{-i}(I)$. Then $B_\mu(I) = \cup_m G_m$. Since $\nu(\cup_m G_m|B_\mu(I)) = \nu(B_\mu(I)|B_\mu(I)) = 1$,

$$0 < \nu(\cup_\ell F_\ell|B_\mu(I)) \leq \sum_{\bar{m}} \nu\big(G_{\bar{m}} \cap [\cup_\ell F_\ell]\big| B_\mu(I)\big),$$

so there must be $\hat{m}$ with $\nu\big(G_{\hat{m}} \cap [\cup_\ell F_\ell] | B_\mu(I)\big) > 0$. Furthermore,

$$0 < \nu\big(G_{\hat{m}} \cap [\cup_\ell F_\ell]\big| B_\mu(I)\big) \leq \sum_{\bar{\ell}} \nu(G_{\hat{m}} \cap F_{\bar{\ell}}|B_\mu(I)),$$

so there is $\hat{\ell}$ with $\nu(G_{\hat{m}} \cap F_{\hat{\ell}}|B_\mu(I)) > 0$. A fortiori, $\nu(F_{\hat{\ell}}|B_\mu(I)) > 0$ and $\nu(G_{\hat{m}}|B_\mu(I)) > 0$. Since ν extends μ, $0 < \nu(F_{\hat{\ell}} \cap G_{\hat{m}}|B_\mu(I)) = \mu(F_{\hat{\ell}} \cap G_{\hat{m}}|G_{\hat{m}}) \cdot \nu(G_{\hat{m}}|B_\mu(I))$, so $\mu(F_{\hat{\ell}}|G_{\hat{m}}) = \mu(F_{\hat{\ell}} \cap G_{\hat{m}}|G_{\hat{m}}) > 0$. Since $G_{\hat{m}} \in S_{-i}(\mathscr{I}_i)$, there is $\hat{I}$ such that $G_{\hat{m}} = S_{-i}(\hat{I})$. Thus, as claimed, $\mu(F_{\hat{\ell}}|S_{-i}(\hat{I})) > 0$. In turn, this implies that $F_\ell =^{\mu} F_{\hat{\ell}} \geq^{\mu} S_{-i}(\hat{I}) =^{\mu} S_{-i}(I)$ for all ℓ.

For the last statement, fix $m \in \{1,\dots,M\}$. By Lemma 1, there is a μ-sequence $F_1,\dots,F_L \in \mathscr{C}_i$ such that $F_1 = F_L = G_m$ and $\{F_1,\dots,F_L\}$ is the $\geq^{\mu}$–equivalence class of G_m—hence, it coincides with $\{G_1,\dots,G_M\}$. Therefore $\nu(\cup_\ell F_\ell|B_\mu(I)) = \nu(\cup_m G_m|B_\mu(I)) = 1$. As shown above, there is $\bar{\ell} \in \{1,\dots,L\}$ such that $\nu(F_{\bar{\ell}}|B_\mu(I)) > 0$.

I claim that this implies $\nu(F_\ell|B_\mu(I)) > 0$ for all $\ell = \bar{\ell}+1,\dots,L$. The claim is trivially true if $\bar{\ell} = L$; otherwise, suppose that $\nu(F_\ell|B_\mu(I)) > 0$ for some $\ell = \bar{\ell},\dots,L-1$, and consider $\ell+1$. By the chain rule, since by construction $F_\ell \in \{G_1,\dots,G_M\}$, $\nu(F_\ell \cap F_{\ell+1}|B_\mu(I)) = \mu(F_\ell \cap F_{\ell+1}|F_\ell)\nu(F_\ell|B_\mu(I))$. By the induction hypothesis, $\nu(F_\ell|B_\mu(I)) > 0$; and since $F_1,\dots,F_L$ is a μ-sequence, $\mu(F_\ell \cap F_{\ell+1}|F_\ell) = \mu(F_{\ell+1}|F_\ell) > 0$. Thus, $\nu(F_{\ell+1}|B_\mu(I)) \geq \nu(F_\ell \cap F_{\ell+1}|B_\mu(I)) > 0$, as claimed. Since $F_L = G_m$, this completes the proof. ■

Corollary 4 *For all $I, J \in \mathscr{I}_i$, $\nu(B_\mu(I)|B_\mu(J)) > 0$ implies $S_{-i}(I) \geq^\mu S_{-i}(J)$. Hence $S_{-i}(I) \neq^\mu S_{-i}(J)$ implies* supp $\nu(\cdot|B_\mu(I)) \cap$ supp $\nu(\cdot|B_\mu(J)) = \emptyset$.

Proof: If $\nu(B_\mu(I)|B_\mu(J)) > 0$, then there are $t_{-i} \in B_\mu(I_m)$ and $I' \in \mathscr{I}_i$ with $t_{-i} \in S_{-i}(I')$ and $S_{-i}(I') =^\mu S_{-i}(I)$ such that $\nu(\{t_{-i}\}|B_\mu(J)) > 0$. Then $\nu(S_{-i}(I')|B_\mu(J)) > 0$ and so, by Lemma 2 and transitivity, $S_{-i}(I) =^\mu S_{-i}(I') \geq^\mu S_{-i}(J))$, as claimed. Therefore, $S_{-i}(I) \neq^\mu S_{-i}(J)$ implies that $\nu(B_\mu(I)|B_\mu(J)) = \nu(B_\mu(J)|B_\mu(I)) = 0$. ■

Observation 2 Corolllary 4 also implies that, for every $I \in \mathscr{I}_i$, $S_{-i}(I)$ is $\geq^\mu$-maximal in the set $\{S_{-i}(J) : J \in \mathscr{I}_i, \nu(S_{-i}(I)|B_\mu(J)) > 0\}$. If $S_{-i}(J)$, $J \neq I$, is also $\geq^\mu$-maximal in this set, then $S_{-i}(I) =^\mu S_{-i}(J)$ and so $B_\mu(I) = B_\mu(J)$ and $\nu(\cdot|B_|\mu(I)) = \nu(\cdot|B_\mu(J))$.

B Appendix: Proofs of the main results

B.1 Trembles

Lemma 3 *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 $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$ that admits an extension $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i))$. Choose $I_1,\dots,I_M \in \mathscr{I}_i$ so that, for every $I \in \mathscr{I}_i$, there is a unique $m \in \{1,\dots,M\}$ such that $S_{-i}(I) =^\mu S_{-i}(I_m)$. Then a sequence*

$(p^n) \subset \Delta(S_{-i})$ *is a structural perturbation of* μ *if and only if, for all* $s_{-i} \in S_{-i}$,

$$p^n(\{s_{-i}\}) = \sum_{m=1}^{M} \alpha_m^n \, \nu(\{s_{-i}\}|B_\mu(I_m)), \tag{5}$$

where the collection of sequences $\big((\alpha_m^n)_{n\geq1}\big)_{m=1}^{N}$ *satify*

(i) for all n *and* m, $\alpha_m^n \in (0,1]$;

(ii) for all n, $\sum_m \alpha_m^n = 1$; *and*

(iii) for all $\ell, m \in \{1,\ldots,M\}$, $S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$ *implies* $\frac{\alpha_\ell^n}{\alpha_m^n} \to 0$.

Also, if (p^n) *is a structural perturbation of* μ*, then* $p^n(\cdot|B_\mu(I_m)) \to \nu(\cdot|B_\mu(I_m))$ *for all* $m = 1,\ldots,M$.

Proof: ($\Leftarrow$): let $((\alpha_m^n)_{n\geq1})_{m=1}^M$, and $(p^n)_{n\geq1}$ be as in the statement of the Lemma. By construction, $p^n(B_\mu(I_m)) > 0$ for all m. By Corollary 4, the supports of the probabilities $\nu(\cdot|B_\mu(I_m))$ for different indices m are disjoint, and moreover, for all $t_{-i} \in B_\mu(I_m)$ and all ℓ, $\nu(\{t_{-i}\}|B_\mu(I_\ell)) > 0$ implies $S_{-i}(I_m) \geq^\mu S_{-i}(I_\ell))$. Thus, fix $s_{-i} \in B_\mu(I_m)$. If $\nu(\{s_{-i}\}|B_\mu(I_\ell)) = 0$ for all ℓ, then $p^n(\{s_{-i}\}) = 0$ and so $\frac{p^n(\{s_{-i}\})}{p^n(B_\mu(I_m))} = 0 = \nu(\{s_{-i}\}|B_\mu(I_m))$. Otherwise, let $\bar\ell$ be the unique index ℓ such that $s_{-i} \in \operatorname{supp} \nu(\cdot|B_\mu(I_\ell))$. As just noted, $S_{-i}(I_m) \geq^\mu S_{-i}(I_{\bar\ell})$; indeed, by construction, either $\bar\ell = m$ or $S_{-i}(I_m) >^\mu S_{-i}(I_{\bar\ell})$. Therefore,

$$\begin{aligned}
&\frac{p^n(\{s_{-i}\})}{p^n(B_\mu(I_m))} = \\
&= \frac{\alpha_{\bar\ell}^n \, \nu(\{s_{-i}\}|B_\mu(I_{\bar\ell}))}{\sum_{t_{-i} \in \operatorname{supp} \nu(\cdot|B_\mu(I_m))} \alpha_m^n \, \nu(\{t_{-i}\}|B_\mu(I_m)) + \sum_{\ell: S_{-i}(I_m) >^\mu S_{-i}(I_\ell)} \sum_{t_{-i} \in B_\mu(I_m) \cap \operatorname{supp} \nu(\cdot|B_\mu(I_\ell))} \alpha_\ell^n \, \nu(\{t_{-i}\}|B_\mu(I_\ell))} = \\
&= \frac{\nu(\{s_{-i}\}|B_\mu(I_{\bar\ell}))}{\sum_{t_{-i} \in \operatorname{supp} \nu(\cdot|B_\mu(I_m))} \frac{\alpha_m^n}{\alpha_{\bar\ell}^n} \nu(\{t_{-i}\}|B_\mu(I_m)) + \sum_{\ell: S_{-i}(I_m) >^\mu S_{-i}(I_\ell)} \sum_{t_{-i} \in B_\mu(I_m) \cap \operatorname{supp} \nu(\cdot|B_\mu(I_\ell))} \frac{\alpha_\ell^n}{\alpha_{\bar\ell}^n} \nu(\{t_{-i}\}|B_\mu(I_\ell))} = \\
&= \frac{\nu(\{s_{-i}\}|B_\mu(I_m))}{\frac{\alpha_m^n}{\alpha_{\bar\ell}^n} + \sum_{\ell: S_{-i}(I_m) >^\mu S_{-i}(I_\ell)} \sum_{t_{-i} \in B_\mu(I_m) \cap \operatorname{supp} \nu(\cdot|B_\mu(I_\ell))} \frac{\alpha_\ell^n}{\alpha_{\bar\ell}^n} \nu(\{t_{-i}\}|B_\mu(I_\ell))}.
\end{aligned}$$

If $\bar\ell = m$, then $\frac{\alpha_m^n}{\alpha_{\bar\ell}^n} = 1$ for all n, and $\frac{\alpha_\ell^n}{\alpha_{\bar\ell}^n} = \frac{\alpha_\ell^n}{\alpha_m^n} \to 0$ for all ℓ with $S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$. Hence, the first term in the denominator equals 1, and all other terms vanish as $n \to \infty$, so $\frac{p^n(\{s_{-i}\})}{p^n(B_\mu(I_m))} \to \nu(\{s_{-i}\}|B_\mu(I_m))$. If instead $\bar\ell \neq m$, then $S_{-i}(I_m) >^\mu S_{-i}(I_{\bar\ell})$, so $\frac{\alpha_m^n}{\alpha_{\bar\ell}^n} \to \infty$; also, for ℓ with $S_{-i}(I_m) >^\mu S_{-i}(\ell)$, either $\frac{\alpha_\ell^n}{\alpha_{\bar\ell}^n} \to 0$ or $\frac{\alpha_\ell^n}{\alpha_{\bar\ell}^n} \to \infty$ as well. Hence, $\frac{p^n(\{s_{-i}\})}{p^n(B_\mu(I_m))} \to 0 = \nu(\{s_{-i}\}|B_\mu(I_m))$.

For every $I \in \mathscr{I}_i$, if $S_{-i}(I) =^{\mu} S_{-i}(I_m)$ then Lemma 2 implies that $\nu(S_{-i}(I)|B_\mu(I_m)) > 0$, and so by construction $p^n(S_{-i}(I)) > 0$; therefore, for all $s_{-i} \in S_{-i}(I)$, $\frac{p^n(\{s_{-i}\}}{p^n(S_{-i}(I))} = \frac{p^n(\{s_{-i}\}}{p^n(B_\mu(I))} \cdot \frac{p^n(B_\mu(I)}{p^n(S_{-i}(I))} \to \frac{\nu(\{s_{-i}\}|B_\mu(I))}{\nu(S_{-i}(I)|B_\mu(I))} = \mu(\{s_{-i}\}|S_{-i}(I))$, by the chain rule and the assumption that ν extends μ. Therefore (p^n) is a perturbation of μ. Furthermore it is a structural perturbation: it is immediate to see that $p^n(\{s_{-i}\}) > 0$ iff $\nu(\{s_{-}\}|B_\mu(I_m)) > 0$ for some m, and hence iff $\mu(\{s_{-i}\}|S_{-i}(I)) > 0$ for some $I \in \mathscr{I}_i$; and if $s_{-i}, t_{-i} \in \operatorname{supp}\mu(\cdot|S_{-i}(I))$ for some $I \in \mathscr{I}_i$, and $S_{-i}(I) =^{\mu} S_{-i}(I_m)$, then $\frac{p^n(\{s_{-i}\})}{p^n(\{t_{-i}\})} = \frac{\nu(\{s_{-i}\}|B_\mu(I_m))}{\nu(\{t_{-i}\}|B_\mu(I_m))} = \frac{\mu(\{s_{-i}\}|S_{-i}(I_m))}{\mu(\{t_{-i}\}|S_{-i}(I_m))}$ by the chain rule and the assumption that ν extends μ.

($\Rightarrow$): given a structural perturbation (p^n) of μ, define $((\alpha^n_m)_{n\ge1})_{m=1}^{M}$ by letting

$$\forall m = 1,\dots,M,\ n \ge 1: \quad \alpha^n_m = p^n(\operatorname{supp}\ \nu\big(\cdot\,\big|B_\mu(I_m))\big). \tag{6}$$

I claim that, for every m, $\operatorname{supp}\ \nu(\cdot|B_\mu(I_m)) = \cup\{\operatorname{supp}\mu(\cdot|S_{-i}(I) : S_{-i}(I) =^{\mu} S_{-i}(I_m)\}$. If $s_{-i} \in \operatorname{supp}\ \nu(\cdot|B_\mu(I_m))$, then $s_{-i} \in B_\mu(I_m) = \cup\{S_{-i}(I) : I \in \mathscr{I}_i, S_{-i}(I) =^{\mu} S_{-i}(I_m)\}$, so there is $J \in \mathscr{I}_i$ with $s_{-i} \in S_{-i}(J)$ and $S_{-i}(J) =^{\mu} S_{-i}(I)$. By the chain rule and the fact that ν extends μ, $\mu(\{s_{-i}\}|S_{-i}(J)) = \nu(\{s_{-i}\}|S_{-i}(J)) > 0$. Conversely, if $s_{-i} \in \operatorname{supp}\mu(\cdot|S_{-i}(I))$ for some $I \in \mathscr{I}_i$ with $S_{-i}(I) =^{\mu} S_{-i}(I_m)$, then, since $\nu(S_{-i}(I)|B_\mu(I_m)) > 0$ by Lemma 2, the chain rule, and the fact that ν extends μ imply that $\nu(\{s_{-i}\}|B_\mu(I_m)) = \mu(\{s_{-i}\}|S_{-i}(I))\,\nu(S_{-i}(I)|B_\mu(I_m)) > 0$.

Since p^n is a structural perturbation of μ, the claim implies that

$$\operatorname{supp} p^n = \bigcup_I \operatorname{supp}\mu(\cdot|S_{-i}(I)) = \bigcup_m \bigcup \{\operatorname{supp}\mu(\cdot|S_{-i}(I)) : S_{-i}(I) =^{\mu} S_{-i}(I_m)\} = \bigcup_m \operatorname{supp}\ \nu(\cdot|B_\mu(I_m)), \tag{7}$$

For every $m = 1,\dots,M$ and $n \ge 1$, Eq. (7) implies that $\operatorname{supp}\ \nu(\cdot|B_\mu(I_m)) \subset \operatorname{supp} p^n$, so $\alpha^n_m \in (0,1]$. By Corollary 4, the supports of $\nu(\cdot|B_\mu(I_m))$ and $\nu(\cdot|B_\mu(I_\ell))$ are disjoint for $\ell \ne m$; hence, Eq. (7) also implies that $\sum_m \alpha^n_m = \sum_m p^n(\operatorname{supp}\ \nu(\cdot|B_\mu(I_m)) = p^n\left(\cup_m \operatorname{supp}\ \nu(\cdot|B_\mu(I_m))\right) = p^n(\operatorname{supp} p^n) = 1$. Thus, (i) and (ii) hold.

To prove (iii), consider first arbitrary $I, J \in \mathscr{I}_i$ such that $S_{-i}(I) >^{\mu} S_{-i}(I)$. Then in particular there are $J_1,\dots,J_K \in \mathscr{I}_i$ such that $J_1 = J$, $J_K = I$, and $S_{-i}(J_1),\dots,S_{-i}(J_K)$ is a μ-sequence. Then

$$\frac{p^n(S_{-i}(J))}{p^n(S_{-i}(I))} = \prod_{k=2}^{K} \frac{p^n(S_{-i}(J_{k-1}))}{p^n(S_{-i}(J_{k-1}) \cap S_{-i}(J_k))} \cdot \frac{p^n(S_{-i}(J_{k-1}) \cap S_{-i}(J_k))}{p^n(S_{-i}(J_k))} \to \prod_{k=2}^{K} \frac{\mu(S_{-i}(J_{k-1}) \cap S_{-i}(J_k)|S_{-i}(J_k))}{\mu(S_{-i}(J_{k-1}) \cap S_{-i}(J_k)|S_{-i}(J_{k-1}))}.$$

By assumption, the denominators in the limit expression are all positive. However, the numerators cannot be all positive, because otherwise $S_{-i}(J_K), S_{i}(J_{K-1}), \dots, S_{-i}(J_1)$ would also be a μ-sequence and so $S_{-i}(J) = S_{-i}(J_1) \geq^\mu S_{-i}(J_K) = S_{-i}(I)$, contradiction. Therefore, $\frac{p^n(S_{-i}(J))}{p^n(S_{-i}(I))} \longrightarrow 0$. Now suppose that, for some $m, \ell \in \{1, \dots, M\}$, $S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$. Fix $s_{-i} \in \operatorname{supp} \mu(\cdot|S_{-i}(I_m))$, so as shown above $s_{-i} \in \operatorname{supp} \nu(\cdot|B_\mu(I_m))$ and, by Eq. (7), $p^n(\{s_{-i}\}) > 0$ as well. Then

$$\frac{\alpha^n_\ell}{\alpha^n_m} = \frac{p^n(\operatorname{supp} \nu(\cdot|B_\mu(I_\ell)))}{p^n(\operatorname{supp} \nu(\cdot|B_\mu(I_m)))} \leq \frac{p^n(B_\mu(I_\ell))}{p^n(\{s_{-i}\})} = \frac{p^n(S_{-i}(I_m))}{p^n(\{s_{-i}\})} \cdot \frac{p^n(B_\mu(I_\ell))}{p^n(S_{-i}(I_m))} \leq \frac{p^n(S_{-i}(I_m))}{p^n(\{s_{-i}\})} \cdot \sum_{I : S_{-i}(I) =^\mu S_{-i}(I_\ell} \frac{p^n(S_{-i}(I))}{p^n(S_{-i}(I_{}))} \longrightarrow 0,$$

because by assumption $\frac{p^n(\{s_{-i}\})}{p^n(S_{-i}(I_m))} \longrightarrow \mu(\{s_{-i}\}|S_{-i}(I_m)) > 0$ and $\frac{p^n(S_{-i}(I))}{p^n(S_{-i}(I_m))} \longrightarrow 0$ for all $I \in \mathscr{I}_i$ with $S_{-i}(I) =^\mu S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$. This completes the proof of (iii).

Now consider $m \in \{1, \dots, M\}$. I claim that

$$\forall s_{-i}, t_{-i} \in \operatorname{supp} \nu(\cdot|B_\mu(I_m)), \qquad \frac{p^n(\{s_{-i}\})}{p^n(\{t_{-i}\})} = \frac{\nu(\{s_{-i}\}|B_\mu(I_m))}{\nu(\{t_{-i}\}|B_\mu(I_m))}. \tag{8}$$

Fix two such s_{-i}, t_{-i} (recall that $p^n(\{t_{-i}\}) > 0$). By the definition of $B_\mu(I_m)$, there are $I, J \in \mathscr{I}_i$ such that $S_{-i}(I) =^\mu S_{-i}(J) =^\mu S_{-i}(I_m)$, $s_{-i} \in S_{-i}(I)$, and $t_{-i} \in S_{-i}(J)$. By Lemma 1, there is a μ-sequence $F_1, \dots, F_{\bar{L}} \in S_{-i}(\mathscr{I}_i)$ such that $F_1 = F_{\bar{L}} = S_{-i}(I)$ and $\{F_1, \dots, F_{\bar{L}}\} = \{S_{-i}(J) : J \in \mathscr{I}_i, S_{-i}(J) =^\mu S_{-i}(I_m)\}$. In particular, $S_{-i}(J) = F_L$ for some $L \in \{1, \dots, \bar{L}\}$. By Lemma 2, $\nu(F_\ell|B_\mu(I_m)) > 0$ for all ℓ. By definition, $\mu(F_{\ell+1}|F_\ell) = \mu(F_\ell \cup F_{\ell+1}|F_\ell) > 0$ for all $\ell = 1, \dots, L-1$, so by the chain rule and the extension property $\nu(F_\ell \cup F_{\ell+1}|B_\mu(I_m)) > 0$ as well, and one can find $s^\ell_{-i} \in F_\ell \cup F_{\ell+1}$ such that $\nu(\{s^\ell_{-i}\}|B_\mu(I_m)) > 0$ for every such ℓ. Then

$$\frac{\nu(\{s_{-i}\}|B_\mu(I_m))}{\nu(\{t_{-i}\}|B_\mu(I_m))} = \frac{\nu(\{s_{-i}\}|B_\mu(I_m))}{\nu(\{s^1_{-i}\}|B_\mu(I_m))} \cdot \prod_{\ell=1}^{L-2} \frac{\nu(\{s^\ell_{-i}\}|B_\mu(I_m))}{\nu(\{s^{\ell+1}_{-i}\}|B_\mu(I_m))} \cdot \frac{\nu(\{s^{L-1}_{-i}\}|B_\mu(I_m))}{\nu(\{t_{-i}\}|B_\mu(I_m))}$$

Let $s^0_{-i} = s_{-i}$, $s^L_{-i} = t_{-i}$, and $F_0 = S_{-i}(I) = F_1$. By construction, for every $\ell = 0, \dots, L-1$, $s^\ell_{-i}, s^{\ell+1}_{-i} \in F_\ell$, so $\frac{\nu(\{s^\ell_{-i}\}|B_\mu(I_m))}{\nu(\{s^{\ell+1}_{-i}\}|B_\mu(I_m))} = \frac{\nu(\{s^\ell_{-i}\}|F_\ell)\nu(F_\ell|B_\mu(I_m))}{\nu(\{s^{\ell+1}_{-i}\}|F_\ell)\nu(F_\ell|B_\mu(I_m))} = \frac{\mu(\{s^\ell_{-i}\}|F_\ell)}{\mu(\{s^{\ell+1}_{-i}\}|F_\ell)} = \frac{p^n(\{s^\ell_{-i}\})}{p^n(\{s^{\ell+1}_{-i}\})}$ by the chain rule, the extension property, and the fact that (p^n) is a structural perturbation of μ. Therefore,

$$\frac{\nu(\{s_{-i}\}|B_\mu(I_m))}{\nu(\{t_{-i}\}|B_\mu(I_m))} = \frac{p^n(\{s_{-i}\})}{p^n(\{s^1_{-i}\})} \cdot \prod_{\ell=1}^{L-2} \frac{p^n(\{s^\ell_{-i}\})}{p^n(\{s^{\ell+1}_{-i}\})} \cdot \frac{p^n(\{s^{L-1}_{-i}\})}{p^n(\{t_{-i}\})} = \frac{p^n(\{s_{-i}\})}{p^n(\{t_{-i}\})}.$$

By Lemma 2, $\nu(S_{-i}(I)|B_\mu(I_m)) > 0$ and $\nu(S_{-i}(J)|B_\mu(I_m)) > 0$. Since ν extends μ, $\nu(S_{-i}(I)\cap S_{-i}(J)|S_{-i}(J)) = \nu(S_{-i}(I)|S_{-i}(J)) = \mu(S_{-i}(I)|S_{-i}(J)) > 0$; by the chain rule, $\nu(S_{-i}(I)\cap S_{-i}(J)|B_\mu(I_m)) = \nu(S_{-i}(I)\cap S_{-i}(J)|S_{-i}(J))\,\nu(S_{-i}(J)|B_\mu(I_m)) > 0$, so applying the chain rule again yields $\mu(S_{-i}(J)|S_{-i}(I)) = \mu(S_{-i}(I)\cap S_{-i}(J)|S_{-i}(I)) = \nu(S_{-i}(I)\cap S_{-i}(J)|S_{-i}(I)) = \frac{\nu(S_{-i}(I)\cap S_{-i}(J)|B_\mu(I_m))}{\nu(S_{-i}(I)|B_\mu(I_m))} > 0$.

The claim implies that, for all $m = 1,\dots,M$ and $s_{-i} \in \operatorname{supp}\ \nu(\cdot|B_\mu(I_m))$,

$$\frac{p^n(\{s_{-i}\})}{p^n(\operatorname{supp}\ \nu(\cdot|B_\mu(I_m)))} = \frac{1}{1+\sum_{t_{-i}\in \operatorname{supp}\ \nu(\cdot|B_\mu(I_m))\setminus\{s_i\}} \frac{p^n(\{t_{-i}\})}{p^n(\{s_{-i}\})}} = \frac{1}{1+\sum_{t_{-i}\in \operatorname{supp}\ \nu(\cdot|B_\mu(I_m))\setminus\{s_i\}} \frac{\nu(\{t_{-i}\}|B_\mu(I_m))}{\nu(\{s_{-i}\}|B_\mu(I_m))}} =$$
$$= \frac{\nu(\{s_{-i}\}|B_\mu(I_m)}{\nu(\operatorname{supp}\ \nu(\cdot|B_\mu|B_\mu(I_m(I_m)))} = \nu(\{s_{-i}\}|B_\mu(I_m)).$$

Therefore, as required, $p^n = \sum_m \alpha_m^n\, \nu(\cdot|B_\mu(I_m))$. ■

Proof of Theorem 3: (2) ⇒ (3) is immediate. To prove (3) ⇒ (1), for every $n \geq 1$, define $w_n \in [0,1]$ and $q^n \in \Delta(S_{-i})$ by $w_n = \frac{1}{n}\min\{p^n(\{s_{-i}\}) : s_{-i} \in \operatorname{supp} p^n\}$ and $q^n(\{s_{-i}\}) = (1-w_n)p^n(s_{-i}) + w_n\frac{1}{|S_{-i}|}$ for every $s_{-i} \in S_{-i}$. Then $w_n \to 0$, and for every $I \in \mathscr{I}_i$ and $s_{-i} \in S_{-i}(I)$,

$$\frac{q^n(\{s_{-i}\})}{q^n(S_{-i}(I))} = \frac{(1-w_n)p^n(\{s_{-i}\}) + w_n\frac{1}{|S_{-i}|}}{(1-w_n)p^n(S_{-i}(I)) + w_n\frac{|S_{-i}(I)|}{|S_{-i}|}} = \frac{(1-w_n)\frac{p^n(\{s_{-i}\})}{p^n(S_{-i}(I))} + \frac{w_n}{p^n(S_{-i}(I))}\frac{1}{|S_{-i}|}}{(1-w_n) + \frac{w_n}{p^n(S_{-i}(I))}\frac{|S_{-i}(I)|}{|S_{-i}|}} \to \mu(\{s_{-i}\}|S_{-i}(I)),$$

because $\frac{w_m}{p^n(S_{-i}(I))} = \frac{1}{n}\cdot\frac{\min\{p^n(\{s_{-i}\}):s_{-i}\in\operatorname{supp} p^n\}}{p^n(S_{-i}(I))} \in (0,\frac{1}{n}]$ for every n, as by assumption $p^n(S_{-i}(I)) > 0$ and so $\min\{p^n(\{s_{-i}\}) : s_{-i} \in \operatorname{supp} p^n\} \leq p^n(S_{-i}(I))$. By Myerson (1986, Theorem 1), (q^n) generates a CPS $\rho \in \Delta(S_{-i}, 2^{S_{-i}} \setminus \{\emptyset\})$. The restriction of ρ to conditioning events in $S_{-i}(\mathscr{I}_i)$ is μ, and its restriction to $S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i)$ is thus an extension of μ.

Finally, to show (1) ⇒ (2), let $I_1,\dots,I_M \in \mathscr{I}_i$ be such that, for every $I \in \mathscr{I}_i$ there is a unique $m \in \{1,\dots,M\}$ with $S_{-i}(I) =^\mu S_{-i}(I_m)$. For every $m = 1,\dots,M$ and $n \geq 1$, let $\beta_m^n = n^{-|\{\ell:S_{-i}(I_\ell)>^\mu S_{-i}(I_m)\}|}$ and $\alpha_m^n = \frac{\beta_m^n}{\sum_\ell \beta_\ell^n}$. Finally, for every $n \geq 1$, let $p^n = \sum_m \alpha_m^n\, \nu(\cdot|B_\mu(I_m))$. By construction, $\alpha_m^n \in (0,1]$ for all m, n, and $\sum_m \alpha_m^n = 1$ for all $n \geq 1$. Furthermore, if $S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$, then

$$\frac{\alpha_\ell^n}{\alpha_m^n} = n^{-|\{k:S_{-i}(I_k)>^\mu S_{-i}(I_\ell)\}|+|\{k:S_{-i}(I_k)>^\mu S_{-i}(I_m)\}|} \leq n^{-1} \to 0,$$

because $S_{-i}(I_k) >^\mu S_{-i}(I_m)$ implies $S_{-i}(I_k) >^\mu S_{-i}(I_\ell)$, and in addition $S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$ but not $S_{-i}(I_m) >^\mu S_{-i}(I_m)$. Therefore, by Lemma 3, (p^n) is a structural perturbation of μ. ■

Observation 3 The proof of (3) ⇒ (1) actually establishes a slightly stronger fact (leveraging Theorem 1 in Myerson, 1986): if an array $\mu = (\mu(\cdot|S_{-i}(I))_{I\in\mathscr{I}_i} \in \Delta(S_{-i})^{\mathscr{I}_i}$ and a sequence $(p^n) \subset \Delta(S_{-i})$ satisfy $\mu(S_{-i}(I)|S_{-i}(I)) = 1$ and $\mu(\cdot|S_{-i}(I)) = \lim_{n\to\infty} p^n(\cdot|S_{-i}(I))$ for every $I \in \mathscr{I}_i$, where $p^n(S_{-i}(I)) > 0$ for all I and n, then $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$ and, in addition, μ is extendible. In other words, it is not necessary to *assume* that μ is a CPS.

Proof of Theorem 4: let $I_1, \dots, I_M$ be as in the statement of Lemma 3. It is convenient to prove sufficiency in (1) and (2) jointly, then do the same for necessity. In turn, (2) implies the last claim, possibly taking a subsequence for the ⇒ direction.

(⇒) assume that $s_i \succcurlyeq^\mu t_i$. To simplify notation, let $\mu_m \equiv \mathrm{E}_{\nu(\cdot|B_\mu(I_m))}[U_i(s_i,\cdot) - U_i(t_i,\cdot)]$ for $m = 1, \dots, M$; also let $\pi^n \equiv \mathrm{E}_{p^n}[U_i(s_i,\cdot) - U_i(t_i,\cdot)]$ for all $n \geq 1$. With this,

$$\pi^n = \sum_m \alpha_m^n \mu_m. \tag{9}$$

where $(\alpha_m^n)_{m=1,\dots,M;n\geq 1}$ are as in Lemma 3.

If $s_i \sim^\mu t_i$, then Definition 8 implies that $\mu_m = 0$ for all m; then Eq. (9) implies $\pi^n = 0$ for all n. If instead $s_i \succ^\mu t_i$, then there is some ℓ^* with $\mu_{\ell^*} \neq 0$. I now analyze this case.

For every $\ell \in \{1, \dots, M\}$ with $\mu_\ell \neq 0$, I claim that there is $m(\ell)$ such that $\mu_{m(\ell)} > 0$, $S_{-i}(I_{m(\ell)}) \geq^\mu S_{-i}(I_\ell)$, and $\mu_m = 0$ for all $m \in \{1, \dots, M\}$ with $S_{-i}(I_m) >^\mu S_{-i}(I_{m(\ell)})$. By contradiction, suppose no such $m(\ell)$ exists for some $\ell \in \{1, \dots, M\}$ with $\mu_\ell \neq 0$. Construct a sequence $m^1, m^2, \dots$ with $\mu_{m^k} \neq 0$ and $S_{-i}(I_{m^{k+1}}) >^\mu S_{-i}(I_{m^k})$ for all $k \geq 1$, as follows. Let $m^1 = \ell$. Inductively, assume m^k with $\mu_{m^k} \neq 0$ has been defined for some $k \geq 1$. If $\mu_{m^k} > 0$, then there exists $m \in \{1, \dots, M\}$ with $S_{-i}(I_m) >^\mu S_{-i}(I_{m^k})$ and $\mu_m \neq 0$ (otherwise one could take $m(\ell) = m^k$); let $m^{k+1} = m$. If instead $\mu_{m^k} < 0$, then $s_i \succcurlyeq^\mu t_i$ implies that there is $m \in \{1, \dots, M\}$ with $\mu_m > 0$ and $S_{-i}(I_m) >^\mu S_{-i}(I_{m^k})$;

let $m^{k+1} = m$. This completes the inductive step. Since $S_{-i}(I_{m^{k+1}}) >^{\mu} S_{-i}(I_{m^k})$ for all $k \geq 1$, the indices m^k, $k \geq 1$, are all distinct; but $m^k \in \{1,\dots,M\}$ and $M < \infty$, contradiction. This shows that $m(\ell)$ with the noted properties exists.

Now let $\mathscr{M} = m(\{\ell : \mu_\ell \neq 0\})$, which is non-empty because there is ℓ^* with $\mu_{\ell^*} \neq 0$. Then

$$\pi^n = \sum_m \alpha^n_m \mu_m = \sum_{m \in \mathscr{M}} \alpha^n_m \mu_m + \sum_{m \in \{1,\dots,M\} \setminus \mathscr{M} : \mu_m \neq 0} \alpha^n_m \mu_m.$$

Let $M(n) = \arg\max_{m \in \mathscr{M}} \alpha^n_m$ for each n (pick one arbitrarily if there are ties): then

$$\frac{\pi^n}{\alpha^n_{M(n)}} = \mu_{M(n)} + \sum_{m \in \mathscr{M} \setminus \{M(n)\}} \frac{\alpha^n_m}{\alpha^n_{M(n)}} \mu_m + \sum_{\ell \in \{1,\dots,M\} \setminus \mathscr{M} : \mu_\ell \neq 0} \frac{\alpha^n_\ell}{\alpha^n_{M(n)}} \mu_\ell.$$

The first term on the rhs is not smaller than $\min_{m \in \mathscr{M}} \mu_m > 0$. Each summand in the second term is also positive. Finally, summands in the third term may be negative; however, if $\mu_\ell < 0$, then $m(\ell) \neq \ell$; since $S_{-i}(I_{m(\ell)}) \geq^{\mu} S_{-i}(I_\ell)$ by definition, and it cannot be that $S_{-i}(I_{m(\ell)}) = S_{-i}(I_\ell)$ by the choice of indices $1,\dots,M$, conclude that $S_{-i}(I_{m(\ell)}) >^{\mu} S_{-i}(I_\ell)$. Then, by the definition of $M(n)$ and (iii) in Lemma 3, $\frac{\alpha^n_\ell}{\alpha^n_{M(n)}} \leq \frac{\alpha^n_\ell}{\alpha^n_{m(\ell)}} \to 0$. Hence, for n large, $\pi^n > 0$.

Summing up, if $s_i \sim^{\mu} t_i$ then $\pi^n = 0$ for all n, and if $s_i \succ^{\mu} t_i$,, then π^n eventually. Thus, necessity holds in both (1) and (2).

($\Leftarrow$): suppose that, for every structural perturbation (p^n) of μ, $\mathrm{E}_{p^n} U_i(s_i,\cdot) \geq \mathrm{E}_{p^n} U_i(t_i,\cdot)$ for all large n. Again let $\mu_m = \mathrm{E}_{\nu(\cdot|B_\mu(I_m))}[U_i(s_i,\cdot) - U_i(t_i,\cdot)]$ for $m = 1,\dots,M$. Suppose that $\mu_{\bar\ell} < 0$ for some $\bar\ell$. It must be shown that there is m with $S_{-i}(I_m) >^{\mu} S_{-i}(I_{\bar\ell})$ such that $\mu_m > 0$.

For every $m \in \{1,\dots,M\}$ and $n \geq 1$, let $\beta^n_m = n^{-|\{\ell : S_{-i}(I_\ell) >^{\mu} S_{-i}(I_m)\}|}$. If $S_{-i}(I_m) >^{\mu} S_{-i}(I_\ell)$, then

$$\frac{\beta^n_\ell}{\beta^n_m} = n^{-\left[|\{r : S_i(I_r) >^{\mu} S_{-i}(I_\ell)\}| - |\{r : S_i(I_r) >^{\mu} S_{-i}(I_m)\}|\right]} \leq n^{-1} \to 0.$$

Next, for $m \in \{1,\dots,M\}$ and $n \geq 1$, let $\gamma^n_m = \beta^n_m$ if $S_{-i}(I_m) \geq^{\mu} S_{-i}(I_{\bar\ell})$, and $\gamma^n_m = \beta^n_{\bar\ell} \cdot \beta^n_m$ otherwise. I claim that, for all $m, \ell \in \{1,\dots,M\}$, $S_{-i}(I_m) >^{\mu} S_{-i}(I_\ell)$ implies $\frac{\gamma^n_\ell}{\gamma^n_m} \to 0$. If $S_{-i}(I_\ell) \geq^{\mu} S_{-i}(I_{\bar\ell})$, then also $S_{-i}(I_m) >^{\mu} S_{-i}(I_{\bar\ell})$, so $\frac{\gamma^n_\ell}{\gamma^n_m} = \frac{\beta^n_\ell}{\beta^n_m} \to 0$. If not $S_{-i}(I_\ell) \geq^{\mu} S_{-i}(I_{\bar\ell})$ and also not $S_{-i}(I_m) \geq^{\mu} S_{-i}(I_{\bar\ell})$, then $\frac{\gamma^n_\ell}{\gamma^n_m} = \frac{\beta^n_{\bar\ell} \cdot \beta^n_\ell}{\beta^n_{\bar\ell} \cdot \beta^n_m} = \frac{\beta^n_\ell}{\beta^n_m} \to 0$. Finally, if not $S_{-i}(I_\ell) \geq^{\mu} S_{-i}(I_{\bar\ell})$ but $S_{-i}(I_m) \geq^{\mu} S_{-i}(I_{\bar\ell})$, then $\frac{\gamma^n_\ell}{\gamma^n_m} = \frac{\beta^n_{\bar\ell} \cdot \beta^n_\ell}{\beta^n_m} = \beta^n_{\bar\ell} \cdot \frac{\beta^n_\ell}{\beta^n_m} \to 0$, because $0 < \beta^n_{\bar\ell} \leq 1$ for all n and $\frac{\beta^n_\ell}{\beta^n_m} \to 0$.

Finally, for $m \in \{1,\dots,M\}$ and $n \geq 1$, let $\alpha^n_m = \frac{\gamma^n_m}{\sum_\ell \gamma^n_m}$, and note that, again, $S_{-i}(I_m) >^\mu S_{-i}(I_\ell)$ for $\ell, m \in \{1,\dots,M\}$ implies $\frac{\alpha^n_m}{\alpha^n_\ell} \to 0$. For every $n \geq 1$, define $p^n = \sum_m \alpha^n_m \nu(\cdot|B_\mu(I_m))$ and, as above, $\pi^n = \mathrm{E}_{p^n}[U_i(s_i,\cdot) - U_i(t_i,\cdot)]$.

By Lemma 3, (p^n) is a structural perturbation of μ. Therefore, by assumption $\pi^n \geq 0$ eventually. By contradiction, assume that $\mu_m \leq 0$ for all m such that $S_{-i}(I_m) >^\mu S_{-i}(I_{\bar\ell})$. Then, since $S_{-i}(I_\ell) =^\mu S_{-i}(I_{\bar\ell})$ for no $\ell \in \{1,\dots,M\} \setminus \{\bar\ell\}$,

$$\pi^n \leq \alpha^n_{\bar\ell}\mu_{\bar\ell} + \sum_{\ell:\ \text{not}\ S_{-i}(I_\ell) \geq^\mu S_{-i}(I_{\bar\ell})} \alpha^n_\ell \mu_\ell.$$

Dividing throughout by $\alpha^n_{\bar\ell}$,

$$\begin{aligned}\frac{\pi^n}{\alpha^n_{\bar\ell}} \leq& \mu_{\bar\ell} + \sum_{\ell:\ \text{not}\ S_{-i}(I_\ell) \geq^\mu S_{-i}(I_{\bar\ell})} \frac{\alpha^n_\ell}{\alpha^n_{\bar\ell}}\mu_\ell = \mu_{\bar\ell} + \sum_{\ell:\ \text{not}\ S_{-i}(I_\ell) \geq^\mu S_{-i}(I_{\bar\ell})} \frac{\gamma^n_\ell}{\gamma^n_{\bar\ell}}\mu_\ell = \\ =& \mu_{\bar\ell} + \sum_{\ell:\ \text{not}\ S_{-i}(I_\ell) \geq^\mu S_{-i}(I_{\bar\ell})} \frac{\beta^n_{\bar\ell} \cdot \beta^n_\ell}{\beta^n_{\bar\ell}}\mu_\ell = \mu_{\bar\ell} + \sum_{\ell:\ \text{not}\ S_{-i}(I_\ell) \geq^\mu S_{-i}(I_{\bar\ell})} \beta^n_\ell \mu_\ell.\end{aligned}$$

If not $S_{-i}(I_\ell) \geq^\mu S_{-i}(I_{\bar\ell})$, then in particular $I_\ell \neq \phi$. Hence, $S_{-i}(\phi) >^\mu S_{-i}(I_\ell)$, and there is $r \in \{1,\dots,M\}$ such that $S_{-i}(I_r) =^\mu S_{-i}(\phi)$. Therefore, $\beta^n_\ell \leq n^{-1} \to 0$. Since, by assumption, $\mu_{\bar\ell} < 0$, this implies that, for n large, $\pi^n < 0$: contradiction. Thus, $s_i \succsim^\mu t_i$, and the proof of (1) is complete. If in addition $\mathrm{E}_{p^n}[U_i(s_i,\cdot) - U_i(t_i,\cdot)] > 0$ eventually for every structural perturbation (p^n) of μ, then by (1) not $t_i \succsim^\mu s_i$; hence, $s_i \succ^\mu t_i$, and the proof of (2) is also compete. ■

B.2 Theorem 1 (structural and sequential rationality)

Suppose that s_i is structurally rational, and fix $I \in \mathscr{I}_i$ and t_i such that $s_i, t_i \in S_i(I)$. By the strategic independence property, there is $r_i \in S_i(I)$ such that $U_i(r_i, s_{-i}) = U_i(t_i, s_{-i})$ for all $s_{-i} \in S_{-i}(I)$, and $U_i(r_i, s_{-i}) = U_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i} \setminus S_{-i}(I)$. Since s_i is structurally rational, by Theorem 4 there is a structural perturbation (q^n) of μ such that $\mathrm{E}_{q^n} U(s_i,\cdot) \geq \mathrm{E}_{q^n} U_i(r_i,\cdot)$ for all

n. Since $q^n(S_{-i}(I)) > 0$, this implies that

$$\begin{aligned}\mathrm{E}_{q^n(\cdot|S_{-i}(I))}U_i(s_i,\cdot) = \frac{\mathrm{E}_{q^n}U_i(s_i,\cdot) - \mathrm{E}_{q^n}1_{S_{-i}\setminus S_{-i}(I)}U_i(s_i,\cdot)}{q^n(S_{-i}(I))} &= \frac{\mathrm{E}_{q^n}U_i(s_i,\cdot) - \mathrm{E}_{q^n}1_{S_{-i}\setminus S_{-i}(I)}U_i(r_i,\cdot)}{q^n(S_{-i}(I))} \geq \\ \geq \frac{\mathrm{E}_{q^n}U_i(r_i,\cdot) - \mathrm{E}_{q^n}1_{S_{-i}\setminus S_{-i}(I)}U_i(r_i,\cdot)}{q^n(S_{-i}(I))} &= \mathrm{E}_{q^n(\cdot|S_{-i}(I))}U_i(r_i,\cdot) = \mathrm{E}_{q^n(\cdot|S_{-i}(I))}U_i(t_i,\cdot).\end{aligned}$$

Taking limits as $n \to \infty$, $\mathrm{E}_{\mu(\cdot|S_{-i}(I))}[U_i(s_i,\cdot) - U_i(t_i,\cdot)] \geq 0$. Since I and $t_i \in S_i(I)$ were arbitrary, s_i is sequentially rational. ■

B.3 Elicitation

Throughout this section, fix a dynamic game $(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot))$, a questionnaire $Q = (Q_i)_{i\in N}$, and the corresponding elicitation game $\big(N \cup \{c\}, (S_i^*, \mathscr{I}_i^*, U_i^*)_{i\in N\cup\{c\}}, S^*(\cdot)\big)$, according to Definition 9. It is convenient to let $N^* = N \cup \{c\}$. Also, as in part 1 of Definition 9, for every $i \in N$, let $W_i = \{\varnothing\}$ if $Q_i = \varnothing$ and $W_i = \{b, p\}$ if $Q_i = (I, E, p)$.

B.3.1 Preliminaries

I first verify that the elicitation game satisfies two properties in Section 2. This is necessary to ensure that definitions and results on structural rationality in Section 4 apply.

It is immediate by inspecting Definition 9 that, for every $i \in N^*$ and $I^* \in \mathscr{I}_i^*$, $S^*(I) = S_i^*(I^*) \times S_{-i}^*(I^*)$. Second, fix $i \in N$ (so $i \neq c$) and $I^*, J^* \in \mathscr{I}_i^*$: it must be shown that either $S^*(I^*) \cap S^*(J^*) = \emptyset$, or $S^*(I^*)$ and $S^*(J^*)$ are nested. This is immediate if I^* or J^* equal I_i^1. Otherwise, $I^* = (s_i, w_i, I)$ and $J^* = (s_i', w_i', J)$, where $s_i \in S_i(I)$ and $s_i' \in S_i(J)$; then $S_i^*(I^*) = \{(s_i, w_i)\}$ and $S_i^*(J^*) = \{(s_i', w_i')\}$. If either $s_i \neq s_i'$ or $w_i \neq w_i'$, then $S^*(I^*) \cap S^*(J^*) = \emptyset$. Thus, suppose $s_i = s_i'$ and $w_i = w_i'$. By part 4 of Definition 9, $S^*(I^*) = \{(s_i, w_i)\} \times S_{-i}(I) \times W_{-i} \times S_c^*$ and $S^*(J^*) = \{(s_i, w_i)\} \times S_{-i}(J) \times W_{-i} \times S_c^*$. Therefore, $S^*(I^*) \cap S^*(J^*) \neq \emptyset$ implies $S_{-i}(I) \cap S_{-i}(J) \neq \emptyset$, and so $S(I) \cap S(J) \supseteq [\{s_i\} \times S_{-i}(I)] \cap [\{s_i\} \times S_{-i}(J)] \neq \emptyset$. Therefore $S(I)$ and $S(J)$ are nested: say $S(I) \supseteq S(J)$, and so $S_{-i}(I) \supseteq S_{-i}(J)$. But then, part 4 of Definition 9 implies that $S^*(I^*)$ and $S^*(J^*)$ are nested, as required.

Next, it must be verified that the elicitation game satisfies strategic independence. Again, it is enough to focus on $i \in N$ and $I^* = (s_i, w_i, I) \in \mathscr{I}_i^*$, with $s_i \in S_i(I)$, because $S^*_{-i}(I^*) = S^*_{-i}$ for all other I^* (including for $i = c$ and $I^* = \phi$). But part 4 of Definition 9 implies that $S_i^*(I^*) = \{(s_i, w_i)\}$, a singleton set, so strategic independence holds trivially.

Remark 5 *If $(N, (S_i, \mathscr{I}_i, U_i)_{i\in N}, S(\cdot))$ has nested strategic information, so does the associated elicitation game.*

Proof: Suppose the original game has nested strategic information, and fix a player $i \in N$. It is enough to consider information sets of the form $(s_i, w_i, I), (s_i', w_i', I') \in \mathscr{I}_i$. Suppose that $\big((s_{-i}, w_{-i}), s_c^*\big) \in S^*_{-i}\big((s_i, w_i, I)\big) \cap S^*_{-i}\big((s_i', w_i', I')\big)$; then, by Definition 9 part 4, $s_{-i} \in S_{-i}(I) \cap S_{-i}(I')$. Since the original game has nested strategic information, $S_{-i}(I)$ and $S_{-i}(I')$ are nested; assume that $S_{-i}(I) \subseteq S_{-i}(I')$. Then, Definition 9 part 4 implies that $S^*_{-i}\big((s_i, w_i, I)\big) \subseteq S^*_{-i}\big((s_i', w_i', I')\big)$. ■

B.3.2 Proof of Theorem 2

Throughout this subsection, fix a player $i \in N$ and a CPS $\mu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i))$ with extension $\nu \in \Delta(S_{-i}, S_{-i}(\mathscr{I}_i) \cup B_\mu(\mathscr{I}_i))$.

Lemma 4 *Let $\mu^* \in \Delta(S^*_{-i}, \mathscr{I}_i^*)$ agree with μ (Definition 10). Then:*

(0) $I^* \in \mathscr{I}_i^*$ if and only if $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$ for some $I \in \mathscr{I}_i$.

(1) for every $I^*, J^* \in \mathscr{I}_i^*$, $\mu^*(S^*_{-i}(I^*)|S^*_{-i}(J^*)) = \mu(\text{proj}_{S_{-i}} S^*_{-i}(I^*)|\text{proj}_{S_{-i}} S^*_{-i}(J^*))$.

(2) for every $I^*, J^* \in \mathscr{I}_i^*$, $S^*_{-i}(I^*) \geq^{\mu^*} S^*_{-i}(J^*)$ if and only if $\text{proj}_{S_{-i}} S^*_{-i}(I^*) \geq^{\mu} \text{proj}_{S_{-i}} S^*_{-i}(J^*)$.

(3) for every $I \in \mathscr{I}_i$ and $I^* \in \mathscr{I}_i^*$, if $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$, then $B^*_{\mu^*}(I^*) = B_\mu(I) \times W_{-i} \times S_c^*$.

Proof: (0): Fix $I^* \in \mathscr{I}_i^*$. If $I^* = \phi^*$ or $I^* = I_i^1$, then $S^*_{-i}(I^*) = S^*_{-i} = S_{-i} \times W_{-i} \times S_c^* = S_{-i}(\phi) \times W_{-i} \times S_c^*$. If instead $I^* = (s_i, w_i, I)$ for some $s_i \in S_i$, $w_i \in W_i$ and $I \in \mathscr{I}_i$, then $S_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$. Conversely, for every $I \in \mathscr{I}_i$, $s_i \in S_i(I)$ and $w_i \in W_i$, $I^* = (s_i, w_i, I) \in \mathscr{I}_i$ satisfies $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$.

(1): if $I^* = \phi^*$ or $I^* = I_i^1$, then $S^*_{-i}(I^*) = S^*_{-i}$, so both conditional probabilities equal 1. Otherwise, by part 4 of Definition 9, $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S^*_c$ for some $I \in \mathscr{I}_i$, so

$$\mu^*(S^*_{-i}(I^*)|S^*_{-i}(J^*)) = \mu^*(S_{-i}(I) \times W_{-i} \times S^*_c|S^*_{-i}(J^*)) = \left[\text{marg}_{S_{-i}\times S^*_c}\mu^*(\cdot|S^*_{-i}(J^*))\right](S_{-i}(I) \times S^*_c) =$$
$$= \mu(S_{-i}(I)|\text{proj}_{S_{-i}} S^*_{-i}(J^*)) = \mu(\text{proj}_{S_{-i}} S^*_{-i}(I^*)|\text{proj}_{S_{-i}} S^*_{-i}(J^*)),$$

where the second equality follows from marginalization and the third from Definition 10.

(2): suppose that $S^*_{-i}(I^*) \geq^{\mu^*} S^*_{-i}(J^*)$. Then there are $I^*_1, \ldots, I^*_L \in \mathscr{I}^*_i$ such that $I^*_1 = J^*$, $I^*_L = J^*$, and $\mu^*(S^*_{-i}(I^*_{\ell+1})|S^*_{-i}(I^*_\ell))$ for $\ell = 1, \ldots, L-1$. Hence (1) implies that $\mu(\text{proj}_{S_{-i}} S^*_{-i}(I^*_{\ell+1})|\text{proj}_{S_{-i}} S^*_{-i}(I^*_\ell)) > 0$ for $\ell = 1, \ldots, L-1$, which implies that $\text{proj}_{S_{-i}} S^*_{-i}(I^*) = \text{proj}_{S_{-i}} S^*_{-i}(I^*_L) \geq^\mu \text{proj}_{S_{-i}} S^*_{-i}(I^*_1) = \text{proj}_{S_{-i}} S^*_{-i}(J^*)$.

Conversely, suppose that $\text{proj}_{S_{-i}} S^*_{-i}(I^*) \geq^\mu \text{proj}_{S_{-i}} S^*_{-i}(J^*)$, so there are $I_1, \ldots, I_L \in \mathscr{I}_i$ such that $S_{-i}(I_1) = \text{proj}_{S_{-i}} S^*_{-i}(J^*)$, $S_{-i}(I_L) = \text{proj}_{S_{-i}} S^*_{-i}(I^*)$, and $\mu(S_{-i}(I_{\ell+1})|S_{-i}(I_\ell)) > 0$ for all $\ell = 1, \ldots, L-1$. Let $I^*_1 = J^*$, $I^*_L = I^*$, and $I^*_\ell = (s_i, w_i, I_\ell)$, with $s_i \in S_{-i}(I_\ell)$ and $w_i \in W_i$, for all $\ell = 2, \ldots, L-1$. Then, for all $\ell = 1, \ldots, L$, $\text{proj}_{S_{-i}} S^*_{-i}(I^*_\ell) = S_{-i}(I_\ell)$, so part (1) implies that $\mu^*(S_{-i}(I^*_{\ell+1})|S_{-i}(I^*_\ell)) > 0$ for all $\ell = 1, L-1$. Therefore $S^*_{-i}(I^*) \geq^{\mu^*} S_{-i}(J^*)$.

(3) Let $I^*_1, \ldots, I^*_L \in \mathscr{I}^*_i$ be an enumeration of $\{J^* : S^*_{-i}(J^*) =^{\mu^*} S^*_{-i}(I^*)\}$. By part (0), for every $\ell = 1, \ldots, L$, there is $I_\ell \in \mathscr{I}_i$ such that $S^*_{-i}(I^*_\ell) = S_{-i}(I_\ell) \times W_{-i} \times S^*_c$. By part (2), $S^*_{-i}(J^*) =^{\mu^*} S^*_{-i}(I^*)$ iff $\text{proj}_{S_{-i}} S^*_{-i}(J^*) =^\mu \text{proj}_{S_{-i}} S^*_{-i}(I^*) = S_{-i}(I)$; hence, $S_{-i}(I_1), \ldots, S_{-i}(I_L)$ is an enumeration of $\{J : S_{-i}(J) =^\mu S_{-i}(I)\}$, and therefore $B^*_{\mu^*}(I^*) = \cup_\ell S^*_{-i}(I^*_\ell) = [\cup_\ell S_{-i}(I_\ell)] \times W_{-i} \times S^*_c = B_\mu(I) \times W_{-i} \times S^*_c$. ■

Lemma 5 *There is an extensible CPS $\mu^* \in \Delta(S^*_{-i}, \mathscr{I}^*_i)$ that agrees with μ.*

Proof: Since μ is extensible, by Theorem 3 there is a perturbation (p^n) of μ. Fix an arbitrary element $w_{-i} \in W_{-i}$ (cf. part 1 of Definition 9) and define a sequence $(q^n) \subset \Delta(S^*_{-i})$ by letting

$$q^n(\{(s_{-i}, w_{-i}, s^*_c\}) = \frac{1}{2} p^n(\{s_{-i}\}) \qquad \forall n \geq 1,\ s_{-i} \in S_{-i},\ s^*_c \in S^*_c.$$

Fix $I^* \in \mathscr{I}^*_i$. By Lemma 4 part (0), there is $I \in \mathscr{I}_i$ such that $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S^*_c$, so $\text{proj}_{S_{-i}} S^*_{-i}(I^*) = S_{-i}(I)$. Therefore, $q^n(S^*_{-i}(I^*)) = q^n(S_{-i}(I) \times W_{-i} \times S^*_c) = q^n(S_{-i}(I) \times \{w_{-i}\} \times S^*_c) =$

$p^n(S_{-i}(I)) > 0$; furthermore, for every $s_{-i} \in S_{-i}$ and $s_c^* \in S_c^*$,

$$\frac{q^n(\{(s_{-i}, w_{-i}, s_c^*\}}{q^n(S_{-i}^*(I^*))} = \frac{\frac{1}{2}p^n(\{s_{-i}\})}{p^n(\operatorname{proj}_{S_{-i}} S_{-i}^*(I^*)))} \to \frac{1}{2}\mu(\{s_{-i}\}|\operatorname{proj}_{S_{-i}} S_{-i}^*(I^*)).$$

If instead $w'_{-i} \in W_{-i} \setminus \{w_{-i}\}$, then $\frac{q^n(\{(s_{-i}, w'_{-i}, s_c^*\}}{q^n(S_{-i}^*(I^*))} = 0$. Therefore, by Observation 3, the array $\mu^* = (\mu^*|S_{-i}^*(I^*)))_{I^* \in \mathscr{I}_i^*}$ defined by $\mu^*(\{(s_{-i}, w'_{-i}, s_c^*)\}|S_{-i}^*(I^*)) = 1_{w'_{-i}=w_{-i}}\frac{1}{2}\mu(\{s_{-i}\}|\operatorname{proj}_{S_{-i}} S_{-i}^*(I^*))$ for all $s_{-i} \in S_{-i}$, $w'_{-i} \in W_{-i}$ and $s_c^* \in S_c^*$, is a CPS, (q^n) is a perturbation of it, and by construction μ^* agrees with μ. By Theorem 3, μ^* is extensible. ■

Lemma 6 *Let $\mu^* \in \Delta(S_{-i}^*, \mathscr{I}_i^*)$ a CPS that agrees with μ and admits an extension $\nu^* \in \Delta(S_{-i}^*, S_{-i}^*(\mathscr{I}_i^*) \cup B_{\mu^*}^*(\mathscr{I}_i^*))$. Fix $I^* \in \mathscr{I}_i^*$ and let $I \in \mathscr{I}_i$ be such that $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$ (cf. Lemma 4 part 0). Then, for all $s_{-i} \in S_{-i}$ and $s_c^* \in S_c^*$,*

$$\nu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}|B_{\mu^*}^*(I^*)) = \frac{1}{2}\nu(\{s_{-i}\}|B_\mu(I)). \tag{10}$$

Furthermore, for every $s_i \in S_i$, if $Q_i = (\hat{I}, E, p)$ then

$$\begin{aligned} \mathrm{E}_{\nu^*(\cdot|B_{\mu^*}^*(I^*))} U_i^*\big((s_i, b), \cdot\big) &= \frac{1}{2}\mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(s_i, \cdot) + \frac{1}{2}\nu(S_{-i}(\hat{I})|B_\mu(I))\mu(E|S_{-i}(\hat{I})) \quad (11) \\ \mathrm{E}_{\nu^*(\cdot|B_{\mu^*}^*(I^*))} U_i^*\big((s_i, p), \cdot\big) &= \frac{1}{2}\mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(s_i, \cdot) + \frac{1}{2}\nu(S_{-i}(\hat{I})|B_\mu(I))p, \quad (12) \end{aligned}$$

whereas, if $Q_i = \emptyset$, then

$$\mathrm{E}_{\nu^*(\cdot|B_{\mu^*}^*(I^*))} U_i^*\big((s_i, *), \cdot\big) = \mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(s_i, \cdot). \tag{13}$$

Proof: By Lemma 4, $B_{\mu^*}^*(I^*) = B_\mu(I) \times W_{-i} \times S_c^*$. Now consider $s_{-i} \in S_{-i}$, and $s_c^* \in S_c^*$. If $s_{-i} \notin B_\mu(I)$, then $[\{s_{-i}\} \times W_{-i} \times \{s_c^*\}] \cap B_{\mu^*}^*(I^*) = \emptyset$, so $\nu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}|B_{\mu^*}^*(I^*)) = 0 = \nu(\{s_{-i}\}|B_\mu(I))$. Thus, assume $s_{-i} \in B_\mu(I)$, so there is $J \in \mathscr{I}_i$ such that $s_{-i} \in S_{-i}(J)$ and $S_{-i}(J) =^\mu S_{-i}(I)$, so $S_{-i}(J) \subseteq B_\mu(I)$. By the last claim of Lemma 2, $\nu(S_{-i}(J)|B_\mu(I)) > 0$. Finally, $S_{-i}(J) \times W_{-i} \times S_c^* \in S_{-i}^*(\mathscr{I}_i^*)$ and $S_{-i}(J) \times W_{-i} \times S_c^* \subseteq B_{\mu^*}^*(I^*)$. Then, by the chain rule, the assumptions that ν^* extends μ^* and μ^*

agrees with μ, and the fact that $\nu(S_{-i}(J)|B_\mu(I)) > 0$,

$$\begin{aligned}\nu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}|B^*_{\mu^*}(I^*)) =& \nu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}|S_{-i}(J) \times W_{-i} \times S_c^*) \cdot \nu^*(S_{-i}(J) \times W_{-i} \times S_c^*|B^*_{\mu^*}(I^*)) = \\ =& \mu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}|S_{-i}(J) \times W_{-i} \times S_c^*) \cdot \nu^*(S_{-i}(J) \times W_{-i} \times S_c^*|B^*_{\mu^*}(I^*)) = \\ =& \frac{1}{2}\mu(\{s_{-i}\}|S_{-i}(J)) \cdot \nu^*(S_{-i}(J) \times W_{-i} \times S_c^*|B^*_{\mu^*}(I^*)) = \\ =& \frac{1}{2}\nu(\{s_{-i}\}|B_\mu(I)) \cdot \frac{\nu^*(S_{-i}(J) \times W_{-i} \times S_c^*|B^*_{\mu^*}(I^*))}{\nu(S_{-i}(J)|B_\mu(I))} \equiv \frac{1}{2}\nu(\{s_{-i}\}|B_\mu(I)) \cdot \kappa_J.\end{aligned}$$

It must thus be shown that $\kappa_J = 1$ for all $J \in I_i$ such that $S_{-i}(I) =^\mu S_{-i}(J)$.

To do so, let $\{I_1, \dots, I_L\}$ be such that $\{S_{-i}(I_1), \dots, S_{-i}(I_L)\}$ is the $\geq^\mu$-equivalence class containing $S_{-i}(I)$. By Lemma 1, there is a μ-sequence $F_1, \dots, F_M$ such that $\{F_1, \dots, F_M\} = \{S_{-i}(I_1), \dots, S_{-i}(I_L)\}$. Note that, for every $m = 1, \dots, M$, there is $I_{(m)} \in \{I_1, \dots, I_L\}$ such that $F_m = S_{-i}(I_{(m)})$.[31] For every $m = 1, \dots, M-1$, $\mu(F_{m+1}|F_m) > 0$, so there is $s^m_{-i} \in F_m \cap F_{m+1}$. Then also $s_{-i} \in B_\mu(I)$ and $\{s_{-i}\} \times W_{-i} \times S_c^* \subseteq B^*_{\mu^*}(I^*)$, so (adding over all $s_c^* \in \{h, t\}$)

$$\nu(\{s^m_{-i}\}|B_\mu(I)) \cdot \kappa_{I_{(m)}} = \nu^*(\{s_{-i}\} \times W_{-i} \times S_c^*|B^*_{\mu^*}(I^*)) = \nu(\{s^m_{-i}\}|B_\mu(I)) \cdot \kappa_{I_{(m+1)}},$$

which implies that $\kappa_m = \kappa_{m+1}$. Therefore, there is $\kappa \in \mathbb{R}$ such that $\kappa_J = \kappa$ for all $J \in \mathscr{I}_i$ with $S_{-i}(I) =^\mu S_{-i}(J)$. But

$$1 = \sum_{s_{-i} \in B_\mu(I), s_c^* \in \{h,t\}} \nu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}|B^*_\mu(I^*)) = \sum_{s_{-i} \in B_\mu(I), s_c^* \in \{h,t\}} \frac{1}{2}\nu(\{s_{-i}|B_\mu(I)) \cdot \kappa = \kappa,$$

which completes the proof of Eq. (10).

[31] Recall that, to obtain the μ-sequence $F_1, \dots, F_M$, it may be necessary to rearrange and/or repeat the sets $S_{-i}(I_1), \dots, S_{-i}(I_L)$; hence the need for a separate indexing of the information sets $I_1, \dots, I_L$.

Finally, fix $s_i \in S_i^*$. If $Q_i = (\hat{I}, E, p)$, then:

$$
\begin{aligned}
&\mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(I^*))} U_i^*\big((s_i, b), \cdot\big) = \\
&= \sum_{s_{-i} \in S_{-i}} \sum_{w_{-i} \in W_{-i}} \sum_{s_c^* \in \{h,t\}} \nu^*(\{(s_{-i}, w_{-i}, s_c^*)\}|B^*_{\mu^*}(I^*)) U_i^*\big((s_i, b), (s_{-i}, w_{-i}, s_c^*)\big) = \\
&= \sum_{s_{-i} \in S_{-i}} \sum_{w_{-i} \in W_{-i}} \nu^*(\{(s_{-i}, w_{-i}, h)\}|B^*_{\mu^*}(I^*)) U_i\big(s_i, s_{-i}\big)\big) + \sum_{s_{-i} \in S_{-i}} \sum_{w_{-i} \in W_{-i}} \nu^*(\{(s_{-i}, w_{-i}, t)\}|B^*_{\mu^*}(I^*)) 1_E(s_{-i}) = \\
&= \sum_{s_{-i} \in S_{-i}} \nu^*(\{(s_{-i}\} \times W_{-i} \times \{h\}|B^*_{\mu^*}(I^*)) U_i\big(s_i, s_{-i}\big)\big) + \sum_{s_{-i} \in S_{-i}} \nu^*(\{(s_{-i}\} \times W_{-i} \times \{t\}|B^*_{\mu^*}(I^*)) 1_E(s_{-i}) = \\
&= \sum_{s_{-i} \in S_{-i}} \frac{1}{2} \nu(\{(s_{-i}\}|B_\mu(I)) U_i\big(s_i, s_{-i}\big)\big) + \sum_{s_{-i} \in S_{-i}} \frac{1}{2} \nu(\{(s_{-i}\}|B_\mu(I)) 1_E(s_{-i}) = \\
&= \frac{1}{2} \mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(s_i, \cdot) + \frac{1}{2} \nu(E|B_\mu(I)) = \frac{1}{2} \mathrm{E}_{\nu(\cdot|B_\mu(I))} U_i(s_i, \cdot) + \frac{1}{2} \nu(S_{-i}(\hat{I})|B_\mu(I)) \mu(E|S_{-i}(\hat{I})),
\end{aligned}
$$

i.e., Eq. (11) holds. The other equations are proved similarly. ■

The proof of Theorem 2 can now be completed. Lemma 5 shows that there exists an extensible CPS $\mu^* \in \Delta(S^*_{-i}, S^*_{-i}(\mathscr{I}_i^*))$ that agrees with μ; call ν^* its extension. By Lemma 4, for all $I^*, J^* \in \mathscr{I}_i^*$, $S^*_{-i}(I^*) \geq^{\mu^*} S^*_{-i}(J^*)$ if and only if $S_{-i}(I) \geq^{\mu} S_{-i}(J)$, where $I, J \in \mathscr{I}_i$ are such that $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$ and $S^*_{-i}(J^*) = S_{-i}(J) \times W_{-i} \times S_c^*$. Then part (1) of Theorem 2 follows from the definition of structural preferences (Definition 8), Observation 1, and Eqs. (11)-(13).

For part (2), let $Q_i = (I, E, p)$, and fix $s_i \in S_i$. Suppose that $p > \mu(E|S_{-i}(I))$. Let $I^* \in \mathscr{I}_i^*$ be such that $S^*_{-i}(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$, which exists by part (0) of Lemma 4. By Eqs. (11) and (12), and the fact that $\nu(S_{-i}(I)|B_\mu(I)) > 0$ by the last claim of Lemma 2, $\mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(I^*))} U_i^*((s_i, b), \cdot) <$ $\mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(I^*))} U_i^*((s_i, p), \cdot)$. Furthermore, consider $J^* \in \mathscr{I}_i^*$ such that $S^*_{-i}(J^*) >^{\mu^*} S^*_{-i}(I^*)$. If $\nu^*(S^*_{-i}(I^*)|B_{\mu^*}(J^*)) >$ 0, by Lemma 2 with $\{F_1^*\} = \{S^*_{-i}(I^*)\}$ implies $S^*_{-i}(I^*) \geq S^*_{-i}(J^*)$, contradiction: thus, $\nu^*(S^*_{-i}(I^*)|B_{\mu^*}(J^*)) =$ 0. Since $U_i^*((s_i, b), (s_{-i}, w_{-i}), s_c^*) = U_i^*((s_i, p), (s_{-i}, w_{-i}, s_c^*))$ for $s_{-i} \notin S_{-i}(I)$ and all $w_{-i} \in W_{-i}$, $s_c^* \in \{h, t\}$, it follows that $\mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(J^*))} U_i^*((s_i, b), \cdot) = \mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(J^*))} U_i^*((s_i, p), \cdot)$. Hence, not $(s_i, b) \not\succsim^{\mu^*} (s_i, p)$.

Finally, consider $J^* \in \mathscr{I}_i^*$ such that $\mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(J^*))} U_i^*((s_i, b), \cdot) > \mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(J^*))} U_i^*((s_i, p), \cdot)$. Again because $U_i^*((s_i, b), (s_{-i}, w_{-i}), s_c^*) = U_i^*((s_i, p), (s_{-i}, w_{-i}, s_c^*))$ for $s_{-i} \notin S_{-i}(I)$ and all $w_{-i} \in W_{-i}$, $s_c^* \in$

$\{h,t\}$, it must be that $\nu^*(S^*_{-i}(I^*)|B^*_{\mu^*}(J^*)) > 0$. Lemma 2 implies that $S^*_{-i}(I^*) \geq^{\mu^*} S^*_{-i}(J^*)$. If $S^*_{-i}(I^*) =^{\mu^*} S^*_{-i}(J^*)$, then $B^*_{\mu}(I^*) = B^*_{\mu^*}(J^*)$, which contradicts the fact that $\mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(I^*))} U^*_i((s_i,b),\cdot) < \mathrm{E}_{\nu^*(\cdot|B^*_{\mu^*}(I^*))} U^*_i((s_i,p),\cdot)$. Hence $S^*_{-i}(I^*) >^{\mu^*} S^*_{-i}(J^*)$. But then, since J^* was arbitrary, $(s_i,p) \not\succsim^{\mu^*} (s_i,b)$. Thus, $(s_i,p) \succ^{\mu^*} (s_i,b)$, as claimed.

The case $\mu(E|S_{-i}(E)) > p$ is analogous, so the proof is omitted.

Finally, suppose that $Q_i = (I,E,p)$ and (s_i,b) is structurally rational in the elicitation game. Suppose that there is $t_i \in S_i$ such that $t_i \succ^{\mu} s_i$. Then, by (1), $(t_i,b) \succ^{\mu^*} (s_i,b)$: contradiction. Thus, s_i is structurally rational in the original game. Furthermore, suppose that $\mu(E|S_{-i}(I)) < p$: then (2) implies that $(s_i,p) \succ^{\mu^*} (s_i,b)$, contradiction. Thus, $\mu(E|S_{-i}(I)) \geq p$. The case of (s_i,p) structurally rational is analogous, so the proof is omitted. ■

References

Frank J. Anscombe and Robert J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

R.J. Aumann and J.H. Dreze. Assessing strategic risk. *American Economic Journal: Microeconomics*, 1(1):1–16, 2009.

P. Battigalli and M. Siniscalchi. Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games. *Journal of Economic Theory*, 88(1):188–230, 1999.

P. Battigalli and M. Siniscalchi. Strong Belief and Forward Induction Reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002.

Pierpaolo Battigalli. Structural consistency and strategic independence in extensive games. *Ricerche Economiche*, 48(4):357–376, 1994.

G.M. Becker, M.H. DeGroot, and J. Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 1964.

E. Ben-Porath. Rationality, Nash equilibrium and backwards induction in perfect-information games. *The Review of Economic Studies*, pages 23–46, 1997.

Elchanan Ben-Porath and Eddie Dekel. Signaling future actions and the potential for sacrifice. *Journal of Economic Theory*, 57(1):36–51, 1992.

Mariana Blanco, Dirk Engelmann, Alexander K Koch, and Hans-Theo Normann. Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, 13(4):412–438, 2010.

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.

J. Brandts and G. Charness. The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398, 2011.

David J Cooper and John B Van Huyck. Evidence on the equivalence of the strategic and extensive form representation of games. *Journal of Economic Theory*, 110(2):290–308, 2003.

Russell Cooper, Douglas V DeJong, Robert Forsythe, and Thomas W Ross. Forward induction in the battle-of-the-sexes games. *American Economic Review*, 83(5):1303–1316, 1993.

Miguel A Costa-Gomes and Georg Weizsäcker. Stated beliefs and play in normal-form games. *The Review of Economic Studies*, 75(3):729–762, 2008.

David Dillenberger. Preferences for one-shot resolution of uncertainty and allais-type behavior. *Econometrica*, 78(6):1973–2004, 2010.

Larry G. Epstein and Stanley E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57:937–969, 1989.

Urs Fischbacher, Simon Gächter, and Simone Quercia. The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4):897–913, 2012.

Itzhak Gilboa and David Schmeidler. A derivation of expected utility maximization in the context of a game. *Games and Economic Behavior*, 44(1):172–182, 2003.

Steffen Huck and Wieland Müller. Burning money and (pseudo) first-mover advantages: an experimental study on forward induction. *Games and Economic Behavior*, 51(1):109–127, 2005.

E. Kohlberg and J.F. Mertens. On the strategic stability of equilibria. *Econometrica: Journal of the Econometric Society*, 54(5):1003–1037, 1986.

David Kreps and Garey Ramey. Consistency, structural consistency, and sequential rationality. *Econometrica: Journal of the Econometric Society*, 55:1331–1348, 1987.

David M. Kreps and Evan L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46:185–200, 1978.

D.M. Kreps and R. Wilson. Sequential equilibria. *Econometrica: Journal of the Econometric Society*, 50(4):863–894, 1982.

R. Duncan Luce and Howard Raiffa. *Games and Decisions.* Wiley, New York, 1957.

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

R.B. Myerson. Multistage games with communication. *Econometrica*, 54(2):323–358, 1986. ISSN 0012-9682.

Roger B Myerson. Refinements of the Nash equilibrium concept. *International journal of game theory*, 7(2):73–80, 1978.

Yaw Nyarko and Andrew Schotter. An experimental study of belief learning using elicited beliefs. *Econometrica*, 70(3):971–1005, 2002.

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

P.J. Reny. Backward induction, normal form perfection and explicable equilibria. *Econometrica*, 60(3):627–649, 1992. ISSN 0012-9682.

A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Hungarica*, 6(3):285–335, 1955.

Pedro Rey-Biel. Equilibrium play and best response to (stated) beliefs in normal form games. *Games and Economic Behavior*, 65(2):572–585, 2009.

A. Rubinstein. Comments on the interpretation of game theory. *Econometrica*, 59(4):909–924, 1991. ISSN 0012-9682.

Leonard J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

Andrew Schotter, Keith Weigelt, and Charles Wilson. A laboratory investigation of multiperson rationality and presentation effects. *Games and Economic behavior*, 6(3):445–468, 1994.

R. Selten. Ein oligopolexperiment mit preisvariation und investition. *Beiträge zur experimentellen Wirtschaftsforschung, ed. by H. Sauermann, JCB Mohr (Paul Siebeck), Tübingen*, pages 103–135, 1967.

R. Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, 4(1):25–55, 1975. ISSN 0020-7276.

Marciano Siniscalchi. Foundations for structural preferences. mimeo, Northwestern University, 2016.

Eric Van Damme. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *International Journal of Game Theory*, 13(1):1–13, 1984.

John B Van Huyck, Raymond C Battalio, and Richard O Beil. Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, 80(1):234–248, 1990.

Paul Weirich. Causal decision theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2016 edition, 2016.