

Structural Rationality in Dynamic Games

Marciano Siniscalchi

May 2, 2020

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. 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 weak sequential rationality (Reny, 1992). 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 is consistent with experimental evidence about play in the extensive and strategic form, and provides a theoretical rationale for 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 Bart Lipman and three anonymous referees for their comments and suggestions. I also thank Amanda Friedenber, as well as Pierpaolo Battigalli, Gabriel Carroll, Francesco Fabbri, Drew Fudenberg, Ben Golub, 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, 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 possibly the most 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 both on and off the predicted path of play. The motivation is both methodological and practical: the results in this paper strengthen the foundations of dynamic game theory, but also broaden the range of 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 path of play poses additional challenges. For instance, in the game of [Figure 1](#) (cf. [Ben-Porath and Dekel, 1992](#)), the profile (Out, S, S) is a subgame-perfect equilibrium: Ann chooses *Out* at

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

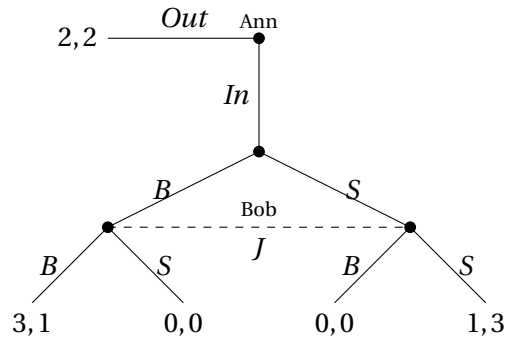


Figure 1: The Battle of the Sexes with an Outside Option

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 6.2.) 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 bet 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 payoff is as in Figure 1, independently of Bob’s betting choice.² If Bob assigns positive probability

²This is a simplified version of the elicitation mechanism in Section 6.2; it is based on De Finetti (2017), §4.

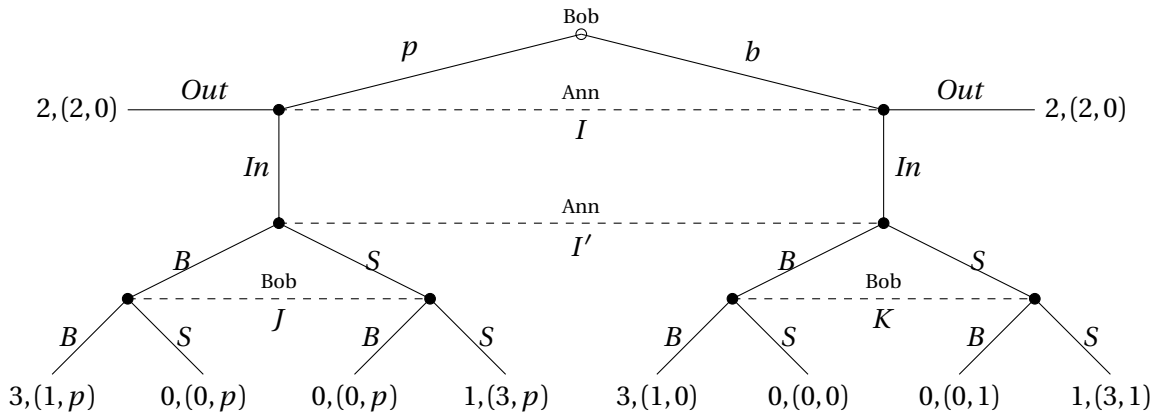


Figure 2: Eliciting Bob's conditional on *In* with ex-ante side bets.

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) maximize the player's expected payoff given the beliefs she holds at that point in the game; payoffs contingent upon zero-probability events (specifically, Bob's betting payoff in case Ann unexpectedly plays *In*) are simply not taken into account. Hence, under sequential rationality,

³This 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*.

on-path choices cannot convey any information about off-path beliefs. A stronger notion of rationality is required for the elicitation scheme in Figure 2 to succeed.

Such a notion must satisfy two requirements. First, it must reflect Bob’s *ex-ante* perspective—the one that is relevant when he chooses between b and p . Second, it must be *cautious*—it must take into account the possibility that Ann might unexpectedly play In . This paper proposes a novel rationality criterion that uses “trembles” (Selten, 1975), i.e., perturbations of the player’s beliefs, to formalize these two requirements. The proposed criterion, *structural rationality*, is defined as the *minimal* (i.e., most permissive) notion of best reply that employs trembles to formalize a player’s cautious, *ex-ante* perspective on the game: given the player’s beliefs, a strategy is ruled out as a possible best reply if and only if another strategy does strictly better against all belief perturbations.

In the subgame-perfect equilibrium ($OutS, S$) of the game in Figure 1, Ann’s strategy $OutS$ also represents Bob’s beliefs: he assigns prior probability one to Ann choosing Out , and conditional probability one to her playing S if the subgame is reached. (The definitions and results in this paper allow for, but are not restricted to equilibrium analysis.) A *perturbation* of these beliefs is a sequence of probability distributions $(p_k)_{k \geq 1}$ over Ann’s strategies that assigns (i) positive but vanishing probability to In , and (ii) probability converging to one to S , conditional on Ann having played In .⁴ A strategy s_b of Bob is a *structural best reply* to $OutS$ if there is no alternative strategy t_b with the property that, for *all* perturbations $(p_k)_{k \geq 1}$ of $OutS$, and all k large enough, t_b yields a strictly higher expected payoff under p_k than s_b .

In Figure 1, Bob’s unique structural best reply to $OutS$ is his sequential best reply, namely S . Under *any* perturbation $(p_k)_{k \geq 1}$, the probability of Ann playing In followed by S is positive and “infinitely greater” than that of Ann playing In followed by B . Thus, eventually, S yields a strictly higher *ex-ante* payoff than B given p_k . Theorem 2 shows that this holds more generally: structural rationality implies sequential rationality in arbitrary dynamic games.

⁴The probabilities p_k need not have full support, so long as they satisfy (i) and (ii).

In addition, structural rationality allows the elicitation of Bob's belief in the subgame. Assume that Bob's beliefs about Ann in the elicitation game of Figure 2 are also represented by *OutS*. Then, in that game, under any perturbation $(p_k)_{k \geq 1}$ of *OutS*, for k sufficiently large, choosing b at the initial node and S at K yields a strictly higher *ex-ante* expected payoff given p_k than any other strategy of Bob. Hence, this is the unique structural best reply to *OutS* in Figure 2. Symmetrically, if Bob assigned probability one to Ann choosing B in the subgame, structural rationality would imply that he should play p followed by B (see Section 6.2.1). Thus, Bob's initial choice conveys information about the probability he assigns to Ann playing S after unexpectedly playing *In*. Theorem 4 shows that, again, this holds for general dynamic games, and for beliefs at arbitrary information sets.

Since, by Theorem 2, structural rationality implies sequential rationality, it is also the minimal refinement of sequential rationality based on perturbations. Indeed, in games without “relevant ties” (Battigalli, 1997), sequential and structural rationality coincide (Theorem 3). Also, informally, structural rationality can be viewed as the most permissive criterion consistent with the view that players do not assign “truly zero probability” to any information set.

Minimality also implies that structural rationality depends solely on the player's beliefs at every information set, just like sequential rationality. However, this is achieved indirectly, by quantifying over all perturbations. Conceptually, a more direct characterization of structural rationality in terms of the player's system of beliefs is desirable. Theorem 1 provides such a characterization. From a practical standpoint, this characterization also simplifies the task of computing structural best replies in games. Structural rationality can also be characterized via lexicographic optimality (Blume, Brandenburger, and Dekel, 1991a,b): see Section 7.G.

The companion paper Siniscalchi (2020) provides an axiomatic behavioral analysis of structural rationality. Siniscalchi (2020) also indicates another sense in which structural preferences are minimal: they constitute the coarsest relation that still allows the behavioral identification of beliefs and utilities (see Theorem 2 in Siniscalchi, 2020).

Finally, structural preferences can rationalize the evidence on the *strategy method* (Sel-

ten, 1967). According to this broadly used experimental protocol, subjects playing a dynamic game are required to commit to extensive-form strategies, which the experimenter then implements, acting as their delegate. Sequential rationality does not distinguish between the resulting “commitment game” and the strategic form of the original dynamic game. Thus, under sequential rationality, one should expect choices made under the strategy method to resemble strategic-form, rather than extensive-form behavior. Yet, the evidence suggests that subjects make qualitatively similar on-path choices when they play a dynamic game directly and when the strategy method is employed (Brandts and Charness, 2011; Fischbacher, Gächter, and Quercia, 2012; Schotter, Weigelt, and Wilson, 1994). At the same time, there is ample evidence that subjects play differently in the strategic and extensive form (Cooper, DeJong, Forsythe, and Ross, 1993; Schotter et al., 1994; Cooper and Van Huyck, 2003; Huck and Müller, 2005). Structural rationality can account for both aspects of the evidence. Under a suitable implementation of the strategy method, subjects should indeed exhibit the same behavior as in the original game (Corollary 2 in Section 6.2). 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.⁵

The present paper focuses on rationality, and not on specific solution concepts. However, Section 7.J draws a connection with trembling-hand perfect equilibrium.

Organization. Section 2 introduces the required notation. Section 3 formalizes beliefs and sequential rationality. Section 4 defines structural rationality via trembles, and Section 5 characterizes it via conditional beliefs. Section 6 contains the main results. Section 7 discusses the related literature, as well as extensions. All proofs are in the Appendix. The Online Appendix contains additional results, examples, the proofs of Theorems 3 and 5, and a discussion of alternative, unsatisfactory definitions of preferences in dynamic games.

⁵ To the best of my knowledge, no known theory of play can account for both findings. For instance, invariance (Kohlberg and Mertens, 1986) predicts that behavior should be the same in all presentations of the game.

2 Basic Notation

This paper considers dynamic games with imperfect information. The analysis only requires that certain familiar reduced-form objects be defined. Online Appendix B 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 7 indicates how to extend the notation to allow for incomplete information.

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

- N is the set of **players**.
- S_i is the set of **strategies** of player i ; as usual, $S_{-i} = \prod_{j \neq i} S_j$ and $S = S_i \times S_{-i}$.
- \mathcal{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.
- $U_i : S_i \times S_{-i} \rightarrow \mathbb{R}$ is the reduced-form **payoff function** for player i (see Section 7); as usual, for $p \in \Delta(S_{-i})$, $U_i(s_i, p) = \sum_{s_{-i}} U_i(s_i, s_{-i}) \cdot p(\{s_{-i}\})$.
- For every $i \in N$ and $I \in \mathcal{I}_i$, $S(I)$ is the set of strategy profiles $(s_j)_{j \in N} \in \prod_j S_j$ that **reach** I . In particular, $S(\phi) = S$.

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 $I \in \mathcal{I}_i$, $S(I) = S_i(I) \times S_{-i}(I)$, where $S_i(I) = \text{proj}_{S_i} S(I)$ and $S_{-i}(I) = \text{proj}_{S_{-i}} S(I)$. If $s_{-i} \in S_{-i}(I)$, say that s_{-i} **allows** I .⁶ Sets of the form $S_{-i}(I)$, for $I \in \mathcal{I}_i$, are called **conditioning events**. The collection $S_{-i}(\mathcal{I}_i) = \{S_{-i}(I) : I \in \mathcal{I}_i\}$ plays an important role.

In games with perfect recall, for every $i \in N$ and $I \in \mathcal{I}_i$, the set $S(I)$ satisfies **strategic independence** (Mailath, Samuelson, and Swinkels, 1993, Definition 2 and Theorem 1): 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 at I and all subsequent information sets, where it coincides with t_i .

⁶That 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 $S_i(I)$.

3 Beliefs and Sequential Rationality

I represent player i 's beliefs as a collection $(\mu(\cdot|I))_{I \in \mathcal{I}_i}$ of probability distributions over coplayers' strategies, indexed by her information sets $I \in \mathcal{I}_i$ (Rényi, 1955; Myerson, 1986; Ben-Porath, 1997; Kohlberg and Reny, 1997; Battigalli and Siniscalchi, 2002). These probabilities have a dual interpretation. From an *interim* perspective, every $\mu(\cdot|I)$ can be interpreted as the beliefs that player i would hold upon reaching I . This is the interpretation that best fits the notion of sequential rationality. Alternatively, the entire probability array $(\mu(\cdot|I))_{I \in \mathcal{I}_i}$ can be viewed as a description of player i 's *prior* beliefs, according to which every information set is reached with positive, but possibly “infinitesimal” probability. In this interpretation, $\mu(\{s_{-i}\}|I)$ describes the likelihood of strategy profile s_{-i} relative to that of information set I , which may itself be infinitely unlikely a priori. This interpretation is particularly apt from the perspective of structural rationality.

Definition 1 A **consistent conditional probability system (CCPS)** for player i is an array $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i} \in \Delta(S_{-i})^{\mathcal{I}_i}$ such that

- (1) for all $I \in \mathcal{I}_i$, $\mu(S_{-i}(I)|I) = 1$;
- (2) for every $I_1, \dots, I_L \in \mathcal{I}_i$ and $E \subseteq S_{-i}(I_1) \cap S_{-i}(I_L)$,

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

Denote the set of CCPSs for player i by $\Delta(S_{-i}, \mathcal{I}_i)$.

Take $L = 2$ in property (2), and assume that $S_{-i}(I_1) \subseteq S_{-i}(I_2)$. Then Eq. (1) reduces to

$$\mu(E|I_1) \cdot \mu(S_{-i}(I_1)|I_2) = \mu(E|I_2), \quad (2)$$

which, together with property (1), characterizes “conditional probability systems” with conditioning events $S_{-i}(\mathcal{I}_i)$, as defined in Rényi (1955). Eq. (2) can be interpreted from an *interim* perspective as requiring that player i update her beliefs in the usual way whenever possible:

if $E \subseteq S_{-i}(I_1)$ and $\mu(S_{-i}(I_1)|I_2) > 0$, then $\mu(E|I_1) = \frac{\mu(E|I_2)}{\mu(S_{-i}(I_1)|I_2)}$. Eq. (1) imposes additional restrictions, which are motivated by the *ex-ante* interpretation of the probability array μ . Again, take $L = 2$, but no longer assume that $S_{-i}(I_1)$ and $S_{-i}(I_2)$ are ordered by inclusion. Suppose that $\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_1) > 0$.⁷ Then Eq. (1) can be rewritten as

$$\mu(E|I_1) \cdot \frac{\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_2)}{\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_1)} = \mu(E|I_2). \quad (3)$$

Consistently with the interpretation of $\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_2)$ and $\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_1)$ as ex-ante relative likelihoods, the fraction $\frac{\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_2)}{\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_1)}$ is an indirect measure of the ex-ante relative likelihood of $S_{-i}(I_1)$ vs. $S_{-i}(I_2)$. To aid intuition, if $\mu(\cdot|I_1)$ and $\mu(\cdot|I_2)$ are the updates of some $P \in \Delta(S_{-i})$ with $P(S_{-i}(I_1) \cap S_{-i}(I_2)) > 0$, then $\frac{\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_2)}{\mu(S_{-i}(I_1) \cap S_{-i}(I_2)|I_1)} = \frac{P(S_{-i}(I_1))}{P(S_{-i}(I_2))}$.⁸ Then, Eq. (3) imposes a “cancellation” or “product” rule: the relative likelihood of E vs. $S_{-i}(I_1)$ times that of $S_{-i}(I_1)$ vs. $S_{-i}(I_2)$ equals the relative likelihood of E vs. $S_{-i}(I_2)$. The interpretation for $L > 2$ is analogous.⁹

Eq. (1) is key in establishing a formal connection between CCPSs and a specific, familiar representation of “infinitesimal” probabilities—trembles, or perturbations.

Definition 2 Fix $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i} \in \Delta(S_{-i})^{\mathcal{I}_i}$. A **perturbation** of μ is a sequence $(p^k)_{k \geq 1} \subset \Delta(S_{-i})$ such that, for every $I \in \mathcal{I}_i$, (i) $p^k(S_{-i}(I)) > 0$ for every k , and (ii) $\lim_{k \rightarrow \infty} p^k(\cdot|S_{-i}(I)) = \mu(\cdot|I)$.

The probabilities p_k in Definition 2 need *not* have full support. In particular, in games with simultaneous moves, wherein $\mathcal{I}_i = \{\phi\}$, the constant sequence defined by $p_k = \mu(\cdot|\phi)$ for all k is a perturbation of player’s CCPS $\mu = \{\mu(\cdot|\phi)\}$.

Proposition 1 An array $\mu \in \Delta(S_{-i})^{\mathcal{I}_i}$ is a CCPS if and only if it admits a perturbation.

⁷If not, then a fortiori $\mu(E|I_1) = 0$, so Eq. (1) holds trivially (i.e. it imposes no substantive restriction).

⁸This is also the case if P takes values in a non-Archimedean ordered field that extends \mathbb{R} , as e.g. in Hammond (1999). Indeed, the definitions and results of this section and the next can be translated in terms of non-Archimedean probabilities; this is not pursued in this paper.

⁹Assuming that Eq. (1) holds for $L = 2$ does *not* imply that it holds for $L > 2$; counterexamples exist with $L = 3$.

The contribution of Proposition 1 is to prove necessity. Sufficiency is immediate: given a perturbation $(p^k)_{k \geq 1}$ of μ , it is readily verified that the array $(p^k(\cdot|S_{-i}(I)))_{I \in \mathcal{I}_i}$ satisfies Eq. (1) for every k ; hence, Eq. (1) holds for μ in the limit as $k \rightarrow \infty$. Since furthermore $\mu(S_{-i}(I)|I) = \lim_k p^k(S_{-i}(I)|S_i(I)) = 1$ for every $I \in \mathcal{I}_i$, μ is a CCPS.

Section 7.D relates CCPSs to other representations of beliefs in dynamic games.

Finally, I formalize the notion of sequential rationality used in this paper. Following Reny (1992) and Rubinstein (1991), the definition I adopt does not restrict the actions specified by a strategy s_i of player i at information sets that s_i does not allow. As these authors have argued, such restrictions are best seen as assumptions on the beliefs of i 's coplayers which reflect the logic of backward induction: they do not characterize player i 's rational decision-making. I follow Reny (1992) and call the resulting notion “weak sequential rationality,” to distinguish it from the definition in Kreps and Wilson (1982).

Definition 3 Fix a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$. A strategy $s_i \in S_i$ is **weakly sequentially rational given μ** if, for every $I \in \mathcal{I}_i$ with $s_i \in S_i(I)$, and all $t_i \in S_i(I)$, $U_i(s_i, \mu(\cdot|I)) \geq U_i(t_i, \mu(\cdot|I))$.

4 Structural Rationality via Perturbations

The following definition generalizes the description of structural rationality given in the Introduction. For conciseness, all definitions and results in this section apply to a fixed dynamic game $(N, (S_i, \mathcal{I}_i, U_i)_{i \in N}, S(\cdot))$, a player $i \in N$, and a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$ for player i .

Definition 4 For all strategies $s_i, t_i \in S_i$, t_i is **structurally strictly preferred to s_i given μ** , written $t_i \succ^\mu s_i$, if $U_i(t_i, p^k) > U_i(s_i, p^k)$ eventually¹⁰ for all perturbations $(p^k)_{k \geq 1}$ of μ . A strategy s_i is **structurally rational given μ** if there is no $t_i \in S_i$ with $t_i \succ^\mu s_i$.

¹⁰That is, for all sufficiently large k .

Definition 4 is in the spirit of Bewley (2002)’s representation of ambiguity, or Knightian uncertainty (Ellsberg, 1961). Expected payoffs are computed with respect to perturbations, rather than individual probabilities: intuitively, the structurally rational agent perceives ambiguity about “infinitesimal” deviations from her CCPS.

Remark 1 *Strategy s_i is structurally rational given μ if and only if, for every $t_i \in S_i$, there is a perturbation $(p^k)_{k \geq 1}$ of μ such that $U_i(s_i, p^k) \geq U_i(t_i, p^k)$ for all k .*

The perturbations in Remark 1 may depend upon the specific strategy t_i that is being compared to s_i : see Example 3.

Structural rationality depends upon (i) the extensive-form structure of the game, and specifically on the collection $S_{-i}(\mathcal{I}_i)$ of conditioning events; and (ii) on player i ’s entire CCPS. This is because conditioning events and the associated conditional beliefs characterize the set of perturbations. Hence, structural rationality is not invariant with respect to the strategic form.

That said, in simultaneous-move (“strategic-form”) games, one particular perturbation of μ is given by $p_k = \mu(\cdot|\phi)$ for all k . This is also the case in general dynamic games, if player i ’s prior $\mu(\cdot|\phi)$ assigns positive probability to every $I \in \mathcal{I}_i$. By Remark 1, *in these cases, a strategy is structurally rational given μ if and only if maximizes player i ’s ex-ante expected payoff.*

Example 1 The game in Figure 3 is an extension of “Matching Pennies” in which Bob has an additional choice, o , following which Ann moves again. Denote Ann’s CCPS by μ , and assume that, as in the unique subgame-perfect equilibrium of this game, Ann initially expects Bob to play h and t with probability $\frac{1}{2}$: $\mu(\{h\}|\phi) = \mu(\{t\}|\phi) = \frac{1}{2}$. For conciseness, I denote by T any one of the realization-equivalent strategies of Ann that choose T at ϕ . I adopt similar notation throughout this paper.

Any perturbation $(p^k)_{k \geq 1}$ of μ must satisfy $p^k(\{o\}) > 0$, $p^k(\{h\}) \rightarrow \frac{1}{2}$, and $p^k(\{t\}) \rightarrow \frac{1}{2}$. Since $p^k(\{o\}) > 0$ implies $U_a(HL, p^k) > U_a(HR, p^k)$, HR is not structurally rational given μ . But how about HL and T ? If $2p^k(\{o\}) + p^k(\{h\}) > -p^k(\{o\}) + p^k(\{t\})$, then $U_a(HL, p^k) > U_a(T, p^k)$;

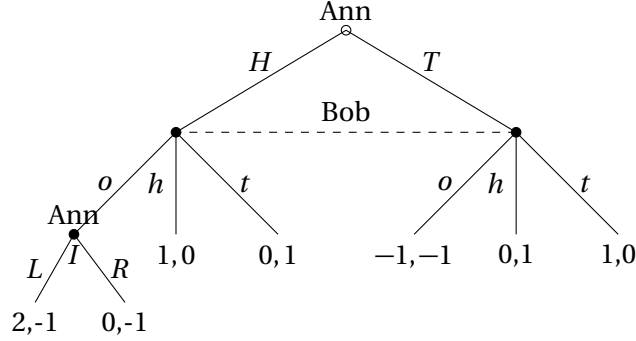


Figure 3: Modified Matching Pennies

for example, let $p^k(\{o\}) = \frac{1}{k}$ and $p^k(\{h\}) = p^k(\{t\}) = \frac{1}{2} - \frac{1}{2k}$. If however $2p^k(\{o\}) + p^k(\{h\}) < -p^k(\{o\}) + p^k(\{t\})$, then $U_a(HL, p^k) < U_a(T, p^k)$; for instance, let $p^k(\{o\}) = \frac{1}{8k}$, $p^k(\{h\}) = \frac{1}{2} - \frac{1}{2k}$, and $p^k(\{t\}) = \frac{1}{2} + \frac{3}{8k}$. Thus, neither $HL \succ^\mu T$ nor $T \succ^\mu HL$, so both HL and T are structurally rational given μ . Of course, these are also the weakly sequentially rational best replies to μ .

Both HL and T are cautious choices: T avoids any further subgames, whereas HL makes the conditionally optimal choice if I is reached. Different perturbations of Ann's beliefs μ select one or the other strategy. Minimality—the fact that Definition 4 takes into account all perturbations of Ann's CCPS—ensures that both strategies are deemed structurally rational.

Example 2 In the game of Figure 4, Ann initially expects Bob to play d , and can either end the game by playing O , or choose which “signal” about Bob's action to observe and respond to. If Bob does play d , the game ends. Otherwise, if Ann chooses L (resp. R) at the initial node, the game ends if Bob chooses c (resp. a), and otherwise Ann is informed that Bob chose a or b (resp. b or c). Ann's CCPS μ satisfies $\mu(\{d\}|\phi) = 1$ and $\mu(\{a\}|I) = \mu(\{b\}|J) = \frac{1}{2}$.

The ex-ante expected payoff from all strategies of Ann is 2. This is also the expected payoff of LA and RA' conditional upon reaching I and J . Thus, O , LA , and RA' are all weakly sequentially rational given μ . However, only O is structurally rational given μ . Any perturbation $(p^k)_{k \geq 1}$ of μ must assign positive probability to $\{a, b\}$ and $\{b, c\}$; furthermore, $p^k(\{d\}) \rightarrow 1$,

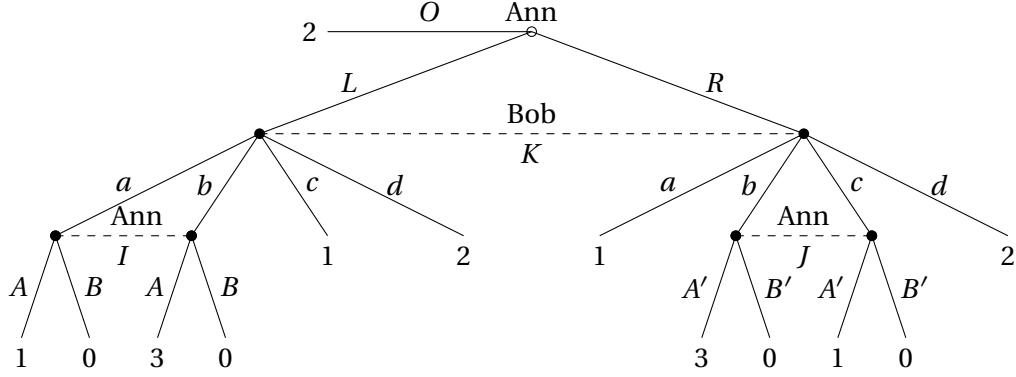


Figure 4: A signal-choice game. Only Ann's payoffs are shown.

$p^k(\{a\}|S_b(I)) = p^k(\{a\}|\{a, b\}) \rightarrow \frac{1}{2}$ and $p^k(\{b\}|S_b(J)) = p^k(\{b\}|\{b, c\}) \rightarrow \frac{1}{2}$. It follows that $p^k(\{s_b\}|\{a, b, c\}) \rightarrow \frac{1}{3}$ for $s_b \in \{a, b, c\}$.¹¹ Consequently, $U_a(LA, p^k) = 2 \cdot p^k(\{d\}) + U_a(LA, p^k(\cdot|\{a, b, c\})) \cdot p^k(\{a, b, c\}) < 2 = U_a(O, p^k)$ eventually, because $U_a(LA, p^k(\cdot|\{a, b, c\})) \rightarrow \frac{5}{3} < 2$. Similarly, $U_a(RA', p^k) < U_a(O, p^k)$ eventually. Thus, $O \succ^\mu LA$ and $O \succ^\mu RA'$. Since, as is readily verified, $LA \succ^\mu LB$ and $RA' \succ^\mu RB'$, O is indeed the unique structural best reply to μ .

This example emphasizes that structural rationality reflects an ex-ante perspective on the game. While the expected payoff of LA (resp. RA') conditional upon reaching I (J) is 2, which equals the payoff of O , this reflects Ann's knowledge that Bob did not play c (a). Ex-ante, she does not have this knowledge. And, under any perturbation of μ , there is roughly a $\frac{2}{3}$ chance of receiving a payoff of 1 by playing LA or RA' , and only a $\frac{1}{3}$ chance of receiving 3. This motivates Ann's ex-ante preference for O . All refinements of equilibrium based on trembles, such as trembling-hand perfection, also select O . This is no accident: see Section 7.J.

Example 3 In the preceding examples, every structurally rational strategy happened to be a best reply to some perturbation of the player's belief. As noted above, Definition 4 and Remark 1 do not require this. The game in Figure 5 illustrates this possibility. Ann's CCPS μ satisfies

¹¹ $p^k(\{a\}|\{a, b\}) \rightarrow \frac{1}{2}$ implies $\frac{p^k(\{a\})}{p^k(\{b\})} \rightarrow 1$, and similarly $\frac{p^k(\{b\})}{p^k(\{c\})} \rightarrow 1$. Then $\frac{p^k(\{a\})}{p^k(\{c\})} = \frac{p^k(\{a\})}{p^k(\{b\})} \cdot \frac{p^k(\{b\})}{p^k(\{c\})} \rightarrow 1$ as well, which implies the claim.

$\mu(\{o\}|\phi) = 1$, $\mu(\{r\}|I) = 1$, and $\mu(\{c\}|J) = 1$. Thus, LB , M , and RD are all weakly sequentially rational on \mathcal{A}_a . Any perturbation $(p^k)_{k \geq 1}$ of μ satisfies $p^k(\{o\}) \rightarrow 1$, $p^k(\{\ell, c\}) > 0$, $p^k(\{r\}) > 0$, and $p^k(\{c\})/p^k(\{\ell, c\}) \rightarrow 1$. Depending on the relative weight of $p^k(\{c\})$ vs. $p^k(\{r\})$, either LB or RB' is a best reply to p^k , for large enough k . Furthermore, C is *not* a best reply to *any* perturbation of μ . Yet, LB , RB' , and C are *all* structurally rational. I omit the argument for LB and RB' , and focus on C .

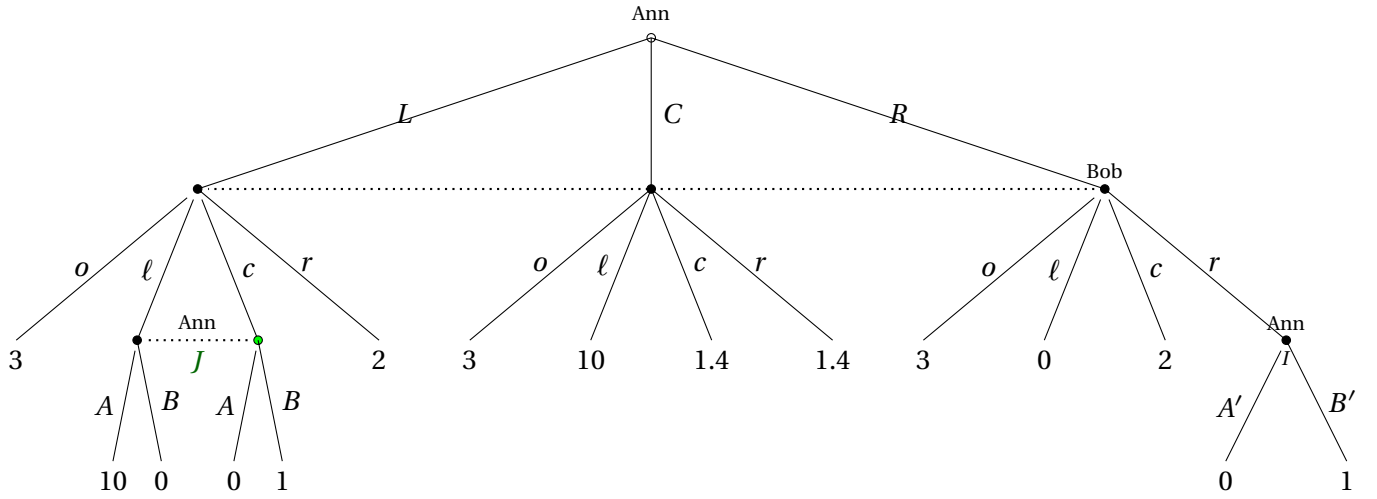


Figure 5: No single justifying perturbation (Ann's payoffs shown)

By the definition of a perturbation, $p^k(\{\ell\})/p^k(\{c\}) \rightarrow 0$. Thus, $U_a(C, p^k) \geq U_a(LB, p^k)$ requires that $p^k(\{c\})/p^k(\{r\}) \geq \frac{3}{2}$. On the other hand, $U_a(C, p^k) \geq U_a(RB', p^k)$ requires that $p^k(\{c\})/p^k(\{r\}) \leq \frac{2}{3}$. Clearly, no single perturbation can satisfy both inequalities. However, the same inequalities show how to construct perturbations under which C has a weakly higher expected payoff than LB (and LA), and *different* perturbations under which it has a higher expected payoff than RB' (and RA'). Thus, by Remark 1, C is structurally rational.

5 Structural Rationality via CCPSs

This section characterizes structural rationality in terms of the player’s belief system. Conceptually, as noted in the Introduction, it is desirable to characterize structural rationality directly in terms of the player’s CCPS, rather than indirectly via perturbations. Pragmatically, the examples of Section 4 indicate that applying Definition 4 directly requires ad-hoc, game-specific, and sometimes delicate arguments that do not lend themselves to a simple algorithmic implementation. Theorem 1 in this Section addresses both concerns.

The analysis builds upon two key observations. The first is that, for a given CCPS μ , there may be information sets I, J for which, under *every* perturbation of μ , the probability of reaching I vanishes no faster than the probability of reaching J . Intuitively, all perturbations rank the magnitudes of the “infinitesimal” probabilities of I and J in the same way. For instance, in Example 2, for any perturbation $(p^k)_{k \geq 1}$ of μ ,

$$\frac{p^k(S_b(I))}{p^k(S_b(J))} = \frac{p^k(S_b(I))}{p^k(\{b\})} \cdot \frac{p^k(\{b\})}{p^k(S_b(J))} \rightarrow \frac{\mu(\{b\}|J)}{\mu(\{b\}|I)} > 0,$$

i.e., the probability of I vanishes no faster than that of J . On the other hand, in Example 3, the limit of $\frac{p^k(S_b(I))}{p^k(S_b(J))}$ can be zero or positive depending on the specific perturbation $(p^k)_{k \geq 1}$ one considers. Definition 5 and Proposition 2 show that whether the ranking of “infinitesimal probabilities” is uniform across all perturbations is ultimately determined by the CCPS itself.

Definition 5 For $I, J \in \mathcal{I}_i$, let $I \geq_0^\mu J$ iff $\mu(S_{-i}(I)|J) > 0$; let \geq^μ be the transitive closure of \geq_0^μ .

The symmetric and asymmetric parts of \geq^μ are denoted $=^\mu$ and $>^\mu$ respectively.

Proposition 2 For $I, J \in \mathcal{I}_i$, $I \geq^\mu J$ (resp. $I >^\mu J$) if and only if $\liminf_k \frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} > 0$ (resp. $\lim_k \frac{p^k(S_{-i}(J))}{p^k(S_{-i}(I))} = 0$) for all perturbations $(p^k)_{k \geq 1}$ of μ .

The second observation is that, if $I =^\mu J$ in the notation of Definition 5, then *every* perturbation induces the same limiting probability distribution P supported on $S_{-i}(I) \cup S_{-i}(J)$.

Intuitively, if $I =^\mu J$, for every perturbation $(p^k)_{k \geq 1}$ the probabilities $\mu(\cdot|I) = \lim_k p^k(\cdot|S_b(I))$ and $\mu(\cdot|J) = \lim_k p^k(\cdot|S_b(J))$ represent “infinitesimals of the same magnitude,” which can be combined into a unique probability $P = \lim_k p^k(\cdot|S_b(I) \cup S_b(J))$.¹² Moreover, again, P is fully pinned down by the CCPS μ itself: it is the *only* probability with $P(S_b(I) \cup S_b(J)) = 1$ that yields $\mu(\cdot|I)$ and $\mu(\cdot|J)$ when conditioning on $S_{-i}(I)$ and $S_{-i}(J)$ respectively. For instance, in Example 2, for every perturbation $(p^k)_{k \geq 1}$ of Ann’s CCPS μ , $p^k(\cdot|S_b(I) \cup S_b(J))$ converges to the uniform distribution P on $S_b(I) \cup S_b(J) = \{a, b, c\}$; furthermore, the *only* probability distribution P with the property that $\mu(\cdot|I)$ and $\mu(\cdot|J)$ are the updates of P is precisely the uniform distribution on $\{a, b, c\}$. Proposition 3 shows that these properties hold more generally.

Proposition 3 *Fix $I \in \mathcal{I}_i$. There is a unique $P \in \Delta(S_{-i})$ such that*

$$P(\cup_{J:I=^\mu J} S_{-i}(J)) = 1 \quad \text{and} \quad \forall J \in \mathcal{I}_i \text{ s.t. } I =^\mu J, E \subseteq S_{-i}(J): P(E) = \mu(E|J) \cdot P(S_{-i}(J)). \quad (4)$$

Moreover, $\prod_{J:I=^\mu J} P(S_{-i}(J)) > 0$. For all perturbations $(p^k)_{k \geq 1}$ of μ , $p^k(\cdot|\cup_{J:I=^\mu J} S_{-i}(J)) \rightarrow P$.

Proposition 3 ensures that the following definition is well-posed:

Definition 6 *For every $I \in \mathcal{I}_i$, let $P_\mu(I)$ be the unique probability that satisfies Eq. (4).*

In particular, $P_\mu(\phi) = \mu(\cdot|\phi)$. Section 5.1 provides a condition under which, for every $I \in \mathcal{I}_i$, there is $J \in \mathcal{I}_i$ (which is explicitly identified, but not necessarily equal to I) such that $P_\mu(I) = \mu(\cdot|J)$. Again, Example 2 shows that this is not the case for arbitrary games and CCPS.

Kreps and Wilson (1982, p. 873) motivate their definition of consistent assessments by assuming that players entertain a collection of alternative “hypotheses” about their coplayers’ behavior, (linearly) ordered in terms of “likelihood.” Thus, abstracting from differences in formalism, Definitions 5 and 6 may also be interpreted as eliciting player i ’s (partial) ordering and alternative hypotheses from her CCPS.

¹²Proposition 5 in Appendix A implies that, more generally, this holds whenever if $I \geq^\mu J$. If $I >^\mu J$, then $P_b(S_b(I)) = 1$ and $\mu(\cdot|I) = P(\cdot|S_b(I)) = P$, but now $P(S_b(J)) = 0$. However this case is not relevant for this section.

I can now provide the promised characterization of structural preferences, which is reminiscent of lexicographic expected-payoff maximization (see Section 7.G): $t_i \succ^\mu s_i$ requires that, if $U_i(t_i, s_{-i}) < U_i(s_i, s_{-i})$ for some strategy profile $s_{-i} \in S_i$, then there is an information set I that is “at least as likely” as s_{-i} (formally: such that $I \geq^\mu J$ for some $J \in \mathcal{I}_i$ with $s_{-i} \in S_{-i}(J)$), for which $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$. In addition, to avoid degeneracies, there must be at least one I^* for which $U_i(t_i, P_\mu(I^*)) > U_i(s_i, P_\mu(I^*))$. Theorem 1 restates this criterion more concisely.

Theorem 1 *For $s_i, t_i \in S_i$, $t_i \succ^\mu s_i$ if and only if there are $I_1, \dots, I_M \in \mathcal{I}_i$ with $M \geq 1$, $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$ for $m = 1, \dots, M$, and $U_i(t_i, s_{-i}) \geq U_i(s_i, s_{-i})$ for $s_{-i} \notin \bigcup_{m=1}^M \bigcup_{J \in \mathcal{I}_i: I_m \geq^\mu J} S_{-i}(J)$.*

Given player i 's payoff function U_i , structural preferences are thus characterized by the probabilities $\{P_\mu(I) : I \in \mathcal{I}_i\}$. The fact that, by Proposition 3, for every $I \in \mathcal{I}_i$, $P_\mu(I)(S_{-i}(I)) > 0$, is an alternative formalization of caution that does not depend on trembles.

I now illustrate how Theorem 1 streamlines the analysis of the examples in Section 3.

Example 1: since $\mu(S_b(I)|\phi) = 0$, not $I \geq^\mu \phi$; and since $\mu(S_{-i}(\phi)|I) \geq \mu(S_{-i}(I)|I) = 1$, so $\phi \geq^\mu I$: hence, $\phi \succ^\mu I$. Then $P_\mu(\phi) = \mu(\cdot|\phi)$ and $P_\mu(I) = \mu(\cdot|I)$. Taking $M = 1$ and $I_1 = I$, one obtains $U_a(HL, P_\mu(I)) = 2 > 1 = U_a(HR, P_\mu(I))$, and $U_a(HL, s_b) = U_a(HR, s_b)$ for $s_b \notin S_b(I) = \{o\}$ (the only information set J with $I \geq^\mu J$ is $J = I$). Thus, by Theorem 1, $HL \succ^\mu HR$. However, HL and T are unranked: $U_a(HL, P_\mu(\phi)) = U_a(T, P_\mu(\phi))$ and, while $U_a(HL, P_\mu(I)) = 2 > -1 = U_a(T, P_\mu(I))$, one has $t \notin S_b(I)$ and $U_a(HL, P_\mu(I)) = 0 < 1 = U_a(T, P_\mu(I))$.

Example 2: as in Example 1, $\phi \succ^\mu I$ and $\phi \succ^\mu J$. However, now $\mu(S_b(I)|J) = \frac{1}{2} = \mu(S_b(J)|I)$, so $I \equiv^\mu J$. Furthermore, as argued above, $P_\mu(I) = P_\mu(J)$ must place probability $\frac{1}{3}$ on a , b , and c . Then $U_a(O, P_\mu(I)) = 2 > \frac{5}{3} = U_a(LA, P_\mu(I)) = U_a(RA', P_\mu(I)) > \frac{1}{3} = U_a(LB, P_\mu(I)) = U_a(RB', P_\mu(I))$, and $U_a(s_a, d) = 2$ for all strategies s_a of Ann. Thus, O is the unique structural best reply to μ .

Example 3: here $\phi \succ^\mu I$ and $\phi \succ^\mu J$, but I and J are not ranked by \geq^μ . Thus, $P_\mu(I) = \mu(\cdot|I)$ and $P_\mu(J) = \mu(\cdot|J)$. Taking $M = 1$ and $I_1 = I$, $U_a(LB, P_\mu(I)) = 2 > 1.4 = U_a(C, P_\mu(I))$; however, $c \notin S_b(I)$ (the only information set K with $I \geq^\mu K$ is $K = I$), and $U_a(LB, P_\mu(I)) = 1 < 1.4 =$

$U_a(C, P_\mu(I))$. Thus, not $LB \succ^\mu C$. Similarly, taking $M = 1$ and $I_1 = J$, one has $U_a(RB', P_\mu(J)) = 2 > 1.4 = U_a(C, P_\mu(J))$ but $r \notin S_b(J)$ and $U_a(RB', P_\mu(J)) = 1 < 1.4 = U_a(C, P_\mu(J))$. Thus, also not $RB' \succ^\mu M$. A fortiori, not $LA \succ^\mu C$ and not $RA' \succ^\mu C$, so C is structurally rational given μ .

5.1 Computational Considerations

The characterization of structural preferences in Theorem 1 is similar in complexity to the definition of lexicographic preferences, given \geq^μ and $P_\mu(\cdot)$. Identifying these objects is also computationally tractable. The pair $(\mathcal{I}_i, \geq^\mu)$ can be viewed as a directed graph: the set of vertices is \mathcal{I}_i , and for all $I, J \in \mathcal{I}_i$, there is an edge from J to I if and only if $I \geq^\mu J$. Equivalence classes for \geq^μ are strongly connected components of this directed graph, and can thus be computed efficiently (e.g. Tarjan, 1972). The probabilities $P_\mu(\cdot)$ can then be derived from the player's CCPS μ by solving Eq. (4), which is a system of linear equations.

A simple restriction on player i 's CCPS ensures that, for every information set $I \in \mathcal{I}_i$, $P_\mu(I) = \mu(\cdot|J)$ for some $J \in \mathcal{I}_i$ (where J is not necessarily equal to I).

Definition 7 A CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$ has **nested supports** if, for every $I, J \in \mathcal{I}_i$, $\mu(S_{-i}(I)|J) > 0$ and $\mu(S_{-i}(J)|I) > 0$ imply that either $\text{supp } \mu(\cdot|I) \subseteq \text{supp } \mu(\cdot|J)$ or $\text{supp } \mu(\cdot|J) \subseteq \text{supp } \mu(\cdot|I)$.

Proposition 4 If a CCPS μ has nested supports, then for all $I \in \mathcal{I}_i$, $P_\mu(I) = \mu(\cdot|J)$, where $J \in \mathcal{I}_i$ satisfies $J =^\mu I$ and, for all $K \in \mathcal{I}_i$ with $K =^\mu I$, $\text{supp } \mu(\cdot|J) \supseteq \text{supp } \mu(\cdot|K)$.

Online Appendix C proves this simple result, and also provides a sufficient condition that only depends upon the game form, and not the specific CCPS. It then shows that the nested-support condition holds in any signalling game, and more generally any game in which every player moves only once on each path of play (but may move on different paths), as well as in centipede games, and other games of theoretical or experimental interest, such as Battle of the Sexes with an outside option (Figure 1), the Burning Money game of Ben-Porath and Dekel (1992), and Selten's Horse (Selten, 1975).

6 Main Results

Throughout subsections 6.1 and 6.2, fix an arbitrary dynamic game $(N, (S_i, \mathcal{I}_i, U_i)_{i \in N}, S(\cdot))$.

6.1 Structural and Weak Sequential Rationality

Theorem 2 *Fix a player $i \in N$ and a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$ for i . If strategy $s_i \in S_i$ is structurally rational given μ , then it is weakly sequentially rational given μ .*

This result follows directly from Theorem 1, and the proof provides further insight into the relationship between structural and weak sequential rationality. Thus, I present it here. The basic structure of the argument is reminiscent of the familiar proof 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 $I \in \mathcal{I}_i$: if s_i is not conditionally optimal at I , and $S_{-i}(I)$ has positive prior probability, then there is a strategy t_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 . With structural preferences, this argument extends to the case in which $S_{-i}(I)$ has zero prior probability by leveraging caution—specifically, the property that $P_\mu(I)(S_{-i}(I)) > 0$ for every $I \in \mathcal{I}_i$.

Proof: Suppose that $s_i \in S_i$ is structurally, but not weakly sequentially rational on \mathcal{I}_i given μ . Then there are an information set $I \in \mathcal{I}_i$ with $s_i \in S_i(I)$ and another strategy $r_i \in S_i(I)$ such that $U_i(r_i, \mu(\cdot|I)) > U_i(s_i, \mu(\cdot|I))$. By strategic independence (cf. Sec. 2), there is $t_i \in S_i$ such that

$U_i(t_i, s_{-i}) = U_i(r_i, s_{-i})$ for $s_{-i} \in S_{-i}(I)$, and $U_i(t_i, s_{-i}) = U_i(s_i, s_{-i})$ for $s_{-i} \notin S_{-i}(I)$. Then

$$\begin{aligned}
U_i(t_i, P_\mu(I)) &= \sum_{s_{-i} \in S_{-i}(I)} U_i(t_i, s_{-i}) P_\mu(I)(\{s_{-i}\}) + \sum_{s_{-i} \notin S_{-i}(I)} U_i(t_i, s_{-i}) P_\mu(I)(\{s_{-i}\}) = \\
&= P_\mu(I)(S_{-i}(I)) \sum_{s_{-i} \in S_{-i}(I)} U_i(r_i, s_{-i}) \mu(\{s_{-i}\} | I) + \sum_{s_{-i} \notin S_{-i}(I)} U_i(s_i, s_{-i}) P_\mu(I)(\{s_{-i}\}) > \\
&> P_\mu(I)(S_{-i}(I)) \sum_{s_{-i} \in S_{-i}(I)} U_i(s_i, s_{-i}) \mu(\{s_{-i}\} | I) + \sum_{s_{-i} \notin S_{-i}(I)} U_i(s_i, s_{-i}) P_\mu(I)(\{s_{-i}\}) = \\
&= \sum_{s_{-i} \in S_{-i}(I)} U_i(s_i, s_{-i}) P_\mu(I)(\{s_{-i}\}) + \sum_{s_{-i} \notin S_{-i}(I)} U_i(s_i, s_{-i}) P_\mu(I)(\{s_{-i}\}) = U_i(s_i, P_\mu(I)).
\end{aligned}$$

The second equality follows from the definition of t_i and the fact that, by Proposition 3, $P_\mu(I)(\{s_{-i}\}) = P_\mu(I)(S_{-i}(I)) \cdot \mu(\{s_{-i}\} | I)$; the strict inequality follows from the assumption that $U_i(r_i, \mu(\cdot | I)) > U_i(s_i, \mu(\cdot | I))$ and the fact that $P_\mu(I)(S_{-i}(I)) > 0$. Now apply Theorem 1 with $M = 1$ and $I_1 = I$: $U_i(t_i, P_\mu(I)) > U_i(s_i, P_\mu(I))$, and for all $s_{-i} \notin S_{-i}(I)$ —hence, a fortiori, for all $s_{-i} \notin \bigcup_{J: I \geq^\mu J} S_{-i}(J)$ — $U_i(t_i, s_{-i}) = U_i(s_i, s_{-i})$. Thus, $t_i \succ^\mu s_i$, contradiction. ■

Example 2 shows that the converse to Theorem 2 does not hold. However, structural and weak sequential rationality are “generically” equivalent. The proof of this result requires a detailed description of extensive-form games that goes beyond the notation introduced in Section 2. To state the result, however, it is sufficient to augment the description of a dynamic game in Section 2 with a specification of **terminal histories** Z and an **outcome map** $\zeta : S \rightarrow Z$ that specifies, for every strategy profile $s \in S$, the terminal history $\zeta(s)$ that s induces. A dynamic game has a **relevant tie for player i** (Battigalli, 1997) if there is an information set $I \in \mathcal{I}_i$, strategies $s_i, t_i \in S_i(I)$, and a profile $s_{-i} \in S_{-i}(I)$, such that $\zeta(s_i, s_{-i}) \neq \zeta(t_i, s_{-i})$ and $U_i(s_i, s_{-i}) = U_i(t_i, s_{-i})$. That is: starting from I , when coplayers play according to s_{-i} , strategies s_i and t_i lead to different terminal histories, but player i receives the same payoff at those histories. “Not having relevant ties” is a particularly simple form of genericity, which in particular does not depend upon any particular CCPS under consideration.

Theorem 3 Fix a player $i \in N$, and a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$. If a strategy is weakly sequentially rational given μ , but not structurally rational for μ , then Γ admits a relevant tie for player i .

Corollary 1 If a game has no relevant ties for player i , then for any CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$, a strategy is structurally rational given μ if and only if it is weakly sequentially rational given μ .

The proof of this result is in Online Appendix B.2.

6.2 Eliciting Conditional Beliefs

6.2.1 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, $\mathcal{I}_b = \{\phi, J, K\}$ with $S_a(J) = S_a(K) = \{InB, InS\}$. Assume that Bob's CCPS μ satisfies $\mu(\{Out\}|\phi) = 1$ and $\mu(\{S\}|J) = \mu(\{S\}|K) = \pi \in [0, 1]$ (the Introduction focused on the case $\pi = 1$). Bob's expected payoffs are depicted in Table I. 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 I is thus the expectation with respect to the relevant belief on S_a as well as the lottery probabilities, indexed by information set.

s_b	<i>Out</i>	<i>InB</i>	<i>InS</i>	ϕ	J, K
pB	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot p$	$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot p$	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot (1 - \pi) + \frac{1}{2} \cdot p$
pS	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot p$	$\frac{1}{2} \cdot 3 + \frac{1}{2} \cdot p$	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 3\pi + \frac{1}{2} \cdot p$
bB	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1$	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot (1 - \pi) + \frac{1}{2} \cdot \pi$
bS	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 1$	$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 3\pi + \frac{1}{2} \cdot \pi$

Table I: Bob's payoffs and expected payoffs for the game in Figure 2.

Conditional upon Ann choosing In , 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).¹³ In particular, the best game action is S if and only if $\pi > \frac{1}{4}$ and the best betting action is b if and only if $\pi > p$. Then, caution delivers the intended result. All strategies yield $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0$ if Ann plays In , and for all perturbations $(p^k)_{k \geq 1}$ of μ , $p^k(S_a(J)) > 0$ and $p^k(\cdot | S_a(J)) \rightarrow \mu(\cdot | J)$. Hence, for all $s_b, t_b \in S_b$, $U_b(t_b, \mu(\cdot | J)) > U_b(s_b, \mu(\cdot | J))$ implies that $U_b(t_b, p^k) > U_b(s_b, p^k)$ eventually for all perturbations, so $t_b \succ^\mu s_b$. Thus, for instance, if $\pi = 1$, the unique structurally rational strategy for Bob is bS ; if $\pi = 0$, it is pB . In particular, Bob’s choice at ϕ reveals whether or not he assigns probability greater than p to S conditional upon Ann choosing In .

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

6.2.2 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 , the experimenter cannot observe Bob’s move and make Ann’s *betting* payoffs contingent upon it. Thus,

¹³ 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.

whatever Ann's *betting* choice may be, it is not in response to real incentives.¹⁴

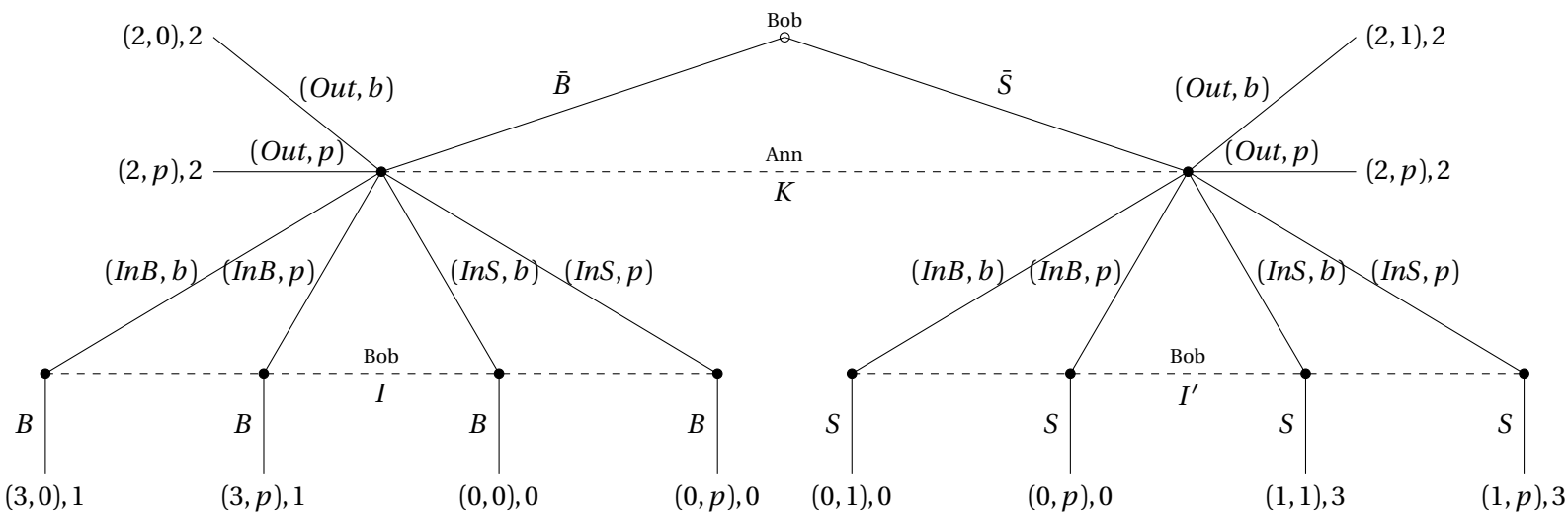


Figure 6: Eliciting Ann's initial beliefs in the game of Figure 1.

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 strategies; the experimenter then implements them. Figure 6 depicts a simplified¹⁵ 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 by the experimenter (though not by Bob) *and* has actual payoff consequences: betting incentives are real.¹⁶

At information sets *I* and *I'*, Bob learns that Ann chose *In*. This is exactly what he learns at *J* in Figure 1.¹⁷ Thus, the conditioning events for Bob in Figure 6 “correspond to” his condi-

¹⁴Modifying 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 §7.F.

¹⁵Figure 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.

¹⁶Strictly speaking, Ann bets on \bar{S} in Figure 6; however, \bar{S} commits Bob to choosing *S* at *I'*.

¹⁷The 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.

tioning events in Figure 1. Hence, any CCPS for Bob in Figure 1 can be used to define a CCPS in Figure 6 that preserves Bob's conditional beliefs about Ann's choices of *In* vs. *Out* and *B* vs. *S* (see Definitions 8 and 9). The same is true for Ann's conditioning events and beliefs. Since structural rationality is fully characterized by a player's conditioning events and CCPS, defining Ann's and Bob's beliefs in Figure 6 in this way pins down their preferences. Indeed Bob's preferences will be "the same" as in Figure 1, up to relabelling; the same is true for Ann's preferences over the "game" component of her strategies: see Theorem 4 part (i).

Now caution delivers the intended result. All perturbations of Bob's CCPS must assign positive, though vanishing probability to $S_a(I) = S_a(I')$; also, conditional on this event, in the limit they must assign probability one to Ann's ("game") action \bar{S} . But then, Bob 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 $p < 1$ ¹⁸ (see Online Appendix D). 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 (weak) sequential rationality, is crucial to this conclusion. For instance, with $p = \frac{2}{3}$ in Figure 6, there is a sequential equilibrium¹⁹ 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, (weak) sequential rationality allows Bob's behavior to differ in the two games. Consequently, while Ann's betting behavior reveals her (prior) beliefs in Figure 6, these do not correspond to her beliefs in the posited equilibrium of the game in Figure 1.

¹⁸Indeed, since \bar{S} is Bob's assumed equilibrium action in the original game, if Ann's beliefs about Bob in Figure 6 are as in the $(OutS, S)$ equilibrium in Figure 1, she *does* assign probability one to \bar{S} .

¹⁹This is a sequential equilibrium in the sense of Kreps and Wilson (1982): full sequential rationality holds. This shows that the distinction between weak and full sequential rationality is immaterial to the analysis.

6.2.3 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.²⁰

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.²¹

Definition 8²² A *questionnaire* is a collection $Q = (I_i, W_i)_{i \in N}$ such that, for every $i \in N$, $I_i \in \mathcal{I}_i$ 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 $(N \cup \{c\}, (S_i^*, \mathcal{I}_i^*, U_i^*)_{i \in N \cup \{c\}}, S^*(\cdot))$, where $S_c^* = \{h, t\}$, $\mathcal{I}_c^* = \{\phi^*\}$, $U_c^* \equiv 0$, and for all $i \in N$:

1. (Strategies) $S_i^* = S_i \times W_i$;
2. (Information) $\mathcal{I}_i^* = \{\phi^*, I_i^1\} \cup \{(s_i, w_i, I) : (s_i, w_i) \in S_i^*, I \in \mathcal{I}_i, s_i \in S_i(I)\}$;
3. (First stage) $S^*(I_i^1) = S^*$
4. (Second stage) for all $(s_i, w_i, I) \in \mathcal{I}_i^*$, $S^*((s_i, w_i, I)) = \{(s_i, w_i)\} \times S_{-i}(I) \times W_{-i} \times S_c^*$,²³
5. (Payoffs) for all $((s_i, w_i), (s_{-i}, w_{-i}), s_c^*) \in S^*$: if $s_c^* = h$ or $W_i = \{*\}$, then $U_i^*((s_i, w_i), (s_{-i}, w_{-i}), s_c^*) =$

²⁰For notational simplicity, in Definition 8 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 9).

²¹Thus, a justification for the use of the strategy method without belief elicitation follows as a corollary.

²²I use the formalism of Section 2. Online Appendix B.3 formalizes the extensive form of the elicitation game.

²³Here 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

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

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.²⁴

Next, I formalize the assumptions that players (a) hold 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 I (cf. Lemma 8).

Definition 9 Let $(N^*, (S_i^*, \mathcal{G}_i^*, U_i^*)_{i \in N^*}, S^*(\cdot))$ be the elicitation game associated with questionnaire $(I_i, W_i)_{i \in N}$. For $i \in N$ and $\mu \in \Delta(S_{-i}, \mathcal{G}_i)$, the CCPS $\mu^* \in \Delta(S_{-i}^*, \mathcal{G}_i^*)$ **agrees with** μ if

$$\text{marg}_{S_{-i} \times S_c^*} \mu^*(\cdot | \phi^*) = \frac{1}{2} \mu(\cdot | \phi) \quad \text{and} \quad \forall I^* = (s_i, w_i, I) \in \mathcal{G}_i^*, \quad \text{marg}_{S_{-i} \times S_c^*} \mu^*(\cdot | I^*) = \frac{1}{2} \mu(\cdot | I). \quad (5)$$

More than one CCPS for player i in the elicitation game may agree with her CCPS 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 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.

²⁴Part 4 of the definition also indicates that i has a single action available at (s_i, w_i, I) ; see Appendix B.3.

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

- (1) *for all $(s_i, w_i), (t_i, w_i) \in S_i^*$, $(s_i, w_i) \succ^{\mu_i^*} (t_i, w_i)$ if and only if $s_i \succ_i^\mu t_i$;*
- (2) *if $W_i = (E, p)$, then for all $s_i \in S_i$, $p > \mu_i(E|I_i)$ implies $(s_i, p) \succ^{\mu_i^*} (s_i, E)$ and $p < \mu_i(E|I_i)$ implies $(s_i, E) \succ^{\mu_i^*} (s_i, p)$.*

Hence, if $W_i = (E, p)$ and (s_i, E) (resp. (s_i, p)) is structurally rational in the elicitation game, then s_i is structurally rational in the original game, and $\mu_i(E|I_i) \geq p$ (resp. $\mu_i(E|I_i) \leq p$).²⁵

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

Corollary 2 *Under the assumptions of Theorem 4, suppose that $W_i = \{*\}$ for all $i \in N$. Then, for all $i \in N$ and all $s_i, t_i \in S_i$, $s_i \succ^{\mu_i} t_i$ if and only if $(s_i, *) \succ^{\mu_i^*} (t_i, *)$. 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 4 and Corollary 2 depend crucially on the assumption that players are structurally rational. (Weak) sequential rationality is not sufficient to deliver these results, *even if players' conditional beliefs are the same as in the original game:*

Remark 2 *Under the assumptions of Theorem 4, for every player $i \in N$, $(s_i, w_i) \in S_i^*$ is weakly sequentially rational given μ_i^* if and only if (i) $s_i \in \arg \max_{t_i \in S_i} E_{\mu_i(\cdot|S_{-i})} U_i(t_i, \cdot)$, and (ii) if $W_i = (E, p)$ and $w_i = b$ (resp. $w_i = p$), then $\mu_i(E|\phi) \geq p \cdot \mu_i(S_{-i}(I)|\phi)$ (resp. $\mu_i(E|\phi) \leq p \cdot \mu_i(S_{-i}(I)|\phi)$).*

²⁵A weak inequality is needed because, if $p = \mu_i(E|I)$, the strategies (s_i, b) and (s_i, p) may be incomparable.

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 4 and Remark 2 with the generic equivalence result described in Section 6.1, notice that elicitation games feature numerous relevant ties *by construction*. For instance, take the perspective of Bob at the initial node in Figure 2. If Ann chooses *Out* at I , then (in the formalism of Definition 8) Bob receives a payoff of 2 if Chance chooses h , and 0 otherwise, *regardless of Bob's strategy*. This is by design: the conditional bet on Ann playing S after In is “called off” if Ann plays *Out*. “Calling off” the bet is implemented by making Bob’s choices payoff-irrelevant—i.e., by creating a relevant tie. To sum up, by construction, elicitation games are non-generic and such that structural rationality is strictly stronger than weak sequential rationality.

7 Discussion

7.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 $U_i : S \rightarrow \mathbb{R}$ with (i) a set of *material consequences* X_i , (ii) a *consequence function* $C_i : S \rightarrow X_i$, and (iii) a (von Neumann-Morgenstern) *utility function* $u_i : X_i \rightarrow \mathbb{R}$: 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 2, 3, and 4 are modified in the obvious way, Propositions 1 and 3, and Theorems 1, 2 and 3 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 4 can be adapted so that both beliefs and utilities can be elicited in the game. In particular, one may fix “good” and “bad” prizes $x_g, x_b \in X_i$, and stipulate that, if i chooses $w_i = E$ and event E obtains, i receives x_g , etc.; and if i chooses $w_i = p$, then he receives x_g with probability p and x_b with probability $1 - p$.

7.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 \mathcal{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 $U_i : S \times \Theta \rightarrow \mathbb{R}$, where $\Theta = \Theta_0 \times \prod_{j \in N} \Theta_j$. The set of conditioning events for player i is $\mathcal{F}_i = \{S_{-i}(I) \times \Theta_{-i} : I \in \mathcal{I}_i\}$, 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 CCPS μ_{θ_i} on $S_{-i} \times \Theta_{-i}$, with conditioning events \mathcal{F}_i . If the sets Θ_j are finite, Definitions 2, 3, and 4 can be applied to each type $\theta_i \in \Theta_i$ separately; Theorems 1, 2, 3 and 4 then have straightforward extensions. Otherwise, it is more convenient to take the characterization in Theorem 1 as the definition of structural preferences, in which case, again, the remaining results go through unmodified.

7.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 6.2. 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 $\Theta_i = \Delta(S_{-i}, \mathcal{I}_i)$ for each player i . The incomplete-information extension of Theorem 4 ensures that they can be elicited in an incentive-compatible way. The argument extends to beliefs of higher orders.

7.D CCPSs and other representations of beliefs Recall from Section 3 that a conditional probability system in the sense of Rényi (1955) (henceforth Renyi CPS), defined on S_{-i} and with conditioning events $S_{-i}(\mathcal{I}_i)$, is an array $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i}$ that satisfies property (1) in Definition 1, and Eq. (2), which is a special case of Eq. (1). Thus, not every Renyi CPS is CCPS. To see that the inclusion is strict, consider the game in Figure 7. Note that $S_b(I) = \{a, b, c\}$ and $S_b(J) = \{b, c, d\}$. Define an array μ for Ann by $\mu(\{o\}|\phi) = 1$, $\mu(\{b\}|I) = \frac{1}{3} = 1 - \mu(\{c\}|I)$, and $\mu(\{b\}|J) = \frac{2}{3} = 1 - \mu(\{c\}|J)$. Then μ is a Renyi CPS; in particular, Eq. (2) holds trivially because $\mu(S_b(I)|\phi) =$

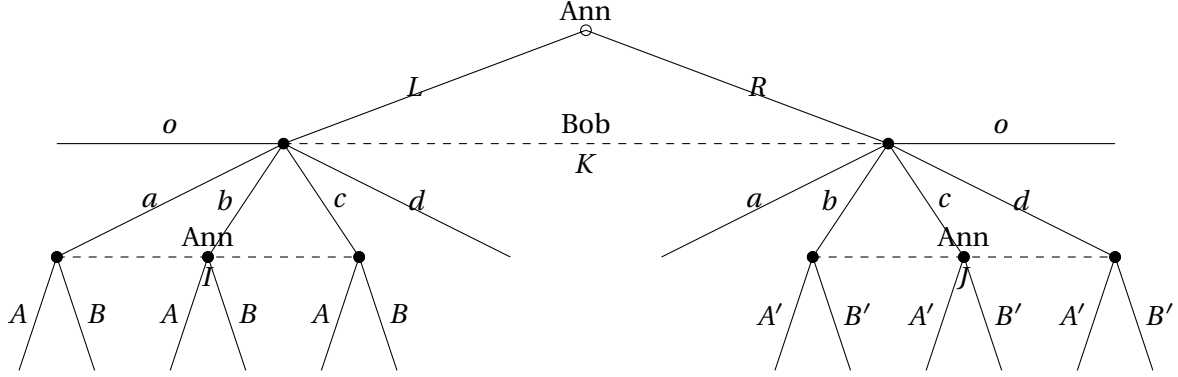


Figure 7: Renyi CPS vs. CCPS

$\mu(S_b(J)|\phi) = 0$, and $S_b(I)$ and $S_b(J)$ are not nested. However, it is not a CCPS: taking $L = 2$, $I_1 = I$, $I_2 = J$, and $E = \{b\}$ in Eq. (1), one has $\mu(\{b\}|I) \cdot \mu(\{b, c\}|J) = \frac{1}{3}$, but $\mu(\{b\}|J) \cdot \mu(\{b, c\}|I) = \frac{2}{3}$. Informally, under the Renyi CPS μ , the relative likelihood of b vs. c at I and at J is different, which is inconsistent with the assumption that Ann assigns well-defined, though vanishing ex-ante probabilities to b and c .

Myerson (1986) introduces *complete* conditional probability systems (complete CPSs): these are Renyi CPSs on the set of coplayers' strategies S_{-i} , in which every non-empty subset $F \subseteq S_{-i}$ is a conditioning event. Myerson shows that an array of probabilities on S_{-i} is a complete CPS if and only if it can be obtained from a sequence of full-support probabilities $(p_k)_{k \geq 1}$, by taking the limits of the sequences $(p_k(\cdot|F))_{k \geq 1}$ for every $F \in 2^{S_{-i}} \setminus \emptyset$. Proposition 1 thus implies that every complete CPS is a CCPS. Furthermore, every CCPS can be extended to a complete CPS: if $(p_k)_{k \geq 1}$ is a perturbation of a CCPS μ and q is any full-support probability on S_{-i} , then $(\frac{k-1}{k} p_k + \frac{1}{k} q \min_{I \in \mathcal{I}_i} p_k(S_{-i}(I)))_{k \geq 1}$ is a full-support perturbation of μ which, per Myerson's result, also generates a complete CPS. However, such an extension is not unique in general.

Kohlberg and Reny (1997) introduce *relative probabilities* on finite spaces. These allow one to specify the relative likelihood of any two events, including those having zero prior probability. Relative probabilities are in 1:1 correspondence with complete CPSs, so the above comments apply to them as well. In addition, Eq. (1) in Definition 1 is closely related to condition

(iv) in the definition of a relative probability (cf. Kohlberg and Reny, 1997, Definition 2.1).

Kreps and Wilson (1982) define a *consistent assessment* as a pair (β, π) where β is a profile of behavioral strategies, π is a “belief system,” i.e., a map from information sets to probability distributions over the corresponding nodes, and there is a sequence $(\beta_k)_{k \geq 1}$ of completely mixed behavioral strategy profiles such that $\beta_k \rightarrow \beta$ and $\pi_k \rightarrow \pi$, where π_k is the belief system derived from β_k . For every player i , a consistent assessment induces a CCPS: for every k , the behavioral profile β_k determines a unique $p_k \in \Delta(S_{-i})$; letting $\mu(\cdot|I) = \lim_k p_k(\cdot|S_{-i}(I))$ for every $I \in \mathcal{I}_i$ defines a probability array μ that admits $(p_k)_{k \geq 1}$ as perturbation, and is thus a CCPS.

Since complete CPSs, relative probabilities, and consistent assessments are (or induce) CCPSs, they, too, rule out the beliefs in the example of Figure 7. Thus, CCPSs formalize consistency conditions that are implicit in known representations of beliefs in dynamic games.

7.E 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 weak 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 .²⁶ 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.

7.F 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 rea-

²⁶That 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.

soning and behavior, and therefore on elicited beliefs. For instance, in the game of Figure 1, *forward-induction* reasoning selects the equilibrium (InB, 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 . However, suppose action Out is removed. Then the game 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 6.2 only modifies the game in ways that, as per Statement (1) of Theorem 4, are inessential for each player's structural preferences.

7.G Structural Preferences via Lexicographic Probabilities Trembles are closely related to lexicographic probability systems, or LPSs (Blume et al., 1991a,b). This makes it possible to characterize structural rationality in terms of lexicographic preferences. Fix a dynamic game, a player i , and a CCPS for i . An LPS on S_{-i} is a finite ordered list $\lambda = (p_1, \dots, p_L) \in \Delta(S_{-i})^L$, with $L \geq 1$; $t_i \in S_i$ is lexicographically strictly preferred to $s_i \in S_i$ given an LPS $\lambda = (p_1, \dots, p_L)$, written $t_i \succ^\lambda s_i$, if there is ℓ such that $U_i(t_i, p_\ell) > U_i(s_i, p_\ell)$ and $U_i(t_i, p_k) = U_i(s_i, p_k)$ for $k = 1, \dots, \ell - 1$.

Definition 10 An LPS $\lambda = (p_1, \dots, p_L)$ on S_{-i} **generates** μ if, for every $I \in \mathcal{I}_i$, there is $\ell \in \{1, \dots, L\}$ such that $p_\ell(S_{-i}(I)) > 0$, $p_\ell(\cdot | S_{-i}(I)) = \mu(\cdot | I)$, and $p_m(S_{-i}(I)) = 0$ for all $m = 1, \dots, \ell - 1$.

The following characterization result is proved in Online Appendix E.1.

Theorem 5 Fix strategies $s_i, t_i \in S_i$. Then $t_i \succ^\mu s_i$ if $t_i \succ^\lambda s_i$ for all LPSs λ that generate μ .

7.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 sensitivity to the timing of uncertainty resolution: see e.g. Kreps and Porteus (1978); Epstein

and Zin (1989), and especially 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 partial information has zero prior probability—that is, when there is *unexpected* partial information.

7.I Caution, elicitation, and triviality Recall that, in Figure 6, Bob has a single action available at I and I' ; furthermore, I and I' are “informationally equivalent”—at both I and I' , Bob learns that Ann played In at K . However, for any strategy $s_a \in S_a(I) = S_a(I')$ of Ann, Bob’s payoffs following I and I' are different, and whether I or I' is reached depends upon Bob’s own choice of \bar{B} or \bar{S} at the initial node ϕ . Therefore, it makes sense for Bob to exercise caution at ϕ and, per Definition 2, assign positive (though vanishing) probability to $S_a(I) = S_a(I')$. As noted above, caution is the driving force behind the elicitation result of Section 6.2.

That said, one can construct games in which a player’s actions can only allow or preclude an information set I , or other informationally equivalent ones, but not influence the payoffs she obtains conditional on the event $S_{-i}(I)$. The game in Figure 8a is one such example.²⁷

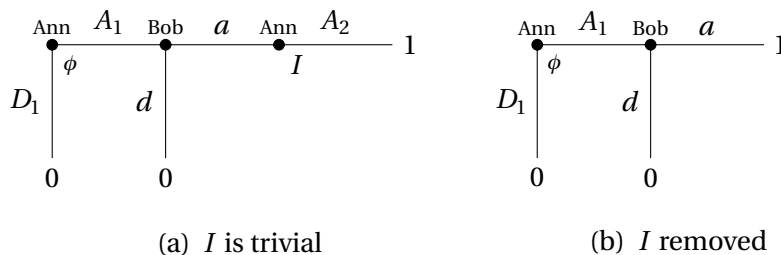


Figure 8: A trivial information set; Ann’s payoffs shown.

²⁷I thank a referee for providing this example, which motivated the analysis in this subsection.

Ann has a single action available at I , which yields a payoff of 1. Unlike in Figure 6, Ann can allow or preclude I , but not influence the payoff she receives if I is reached.

Definition 2 treats such “trivial” information sets no differently from other information sets. One implication is that in the game in Figure 8a, A_1A_2 is the only structurally rational strategy for Ann, if she assigns prior probability 1 to d . On the other hand, both D_1 and A_1A_2 are structurally rational, under the same belief, in the game in Figure 8b. Formally, this reflects the fact that Ann has different conditioning events in the two games in Figure 8. However, one may wish players to only be cautious about conditioning events whose associated payoffs they can influence. To do so, say that an information set $I \in \mathcal{I}_i$ is **trivial for i** if

$$\exists \gamma \in \mathbb{R} : \forall J \in \mathcal{I}_i, \quad S_{-i}(J) = S_{-i}(I), s \in S(J) \Rightarrow U_i(s) = \gamma. \quad (6)$$

Thus, I in Figure 8a is trivial, but $S_a(I) = S_a(I')$ in Figure 6 is not.

The only assumption in Section 2 that concerns \mathcal{I}_i is that it must contain ϕ , the initial node. If one replaces \mathcal{I}_i with the collection $\mathcal{I}_i^{\text{nt}}$ of non-trivial $I \in \mathcal{I}_i$ in the definitions of this paper, all results continue to hold with $\mathcal{I}_i^{\text{nt}}$ in lieu of \mathcal{I}_i . The resulting modified definition does treat the two games in Figure 8 identically. Beliefs at trivial information sets can no longer be elicited. On the other hand, such beliefs are immaterial for (weak) sequential rationality, and indeed for sequential equilibrium, and for refinements based on belief restrictions.

7.J Equilibrium and structurally rational strategies Incorporating structural rationality into solution concepts is left to future research. That said, *only structurally rational strategies are played in a trembling-hand perfect equilibrium* (Selten, 1975). In the notation of this paper, a (strategic-form) **(trembling-hand) perfect equilibrium** is a profile $\sigma \in \prod_{i \in I} \Delta(S_i)$ such that, for every $i \in N$, there exists a sequence $(\sigma_i^k)_{k \geq 1}$ such that $\sigma_i^k \rightarrow \sigma_i$ and every $s_i \in \text{supp } \sigma_i$ is a best reply to each product measure $p_{-i}^k \equiv \otimes_{j \neq i} \sigma_j^k$, $k \geq 1$. By Proposition 1, each sequence $(p_{-i}^k)_{k \geq 1}$ defines a CCPS $\mu_{-i} \in \Delta(S_{-i}, \mathcal{I}_i)$ (possibly considering subsequences), and by Remark 1, every $s_i \in \text{supp } \sigma_i$ is structurally rational given μ_{-i} .

A Appendix: Preliminary results on CCPs

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

For every player i and array $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i} \in \Delta(S_{-i})^{\mathcal{I}_i}$, a μ -**sequence** is an ordered list $I_1, \dots, I_M \in \mathcal{I}_i$ such that $\mu(S_{-i}(I_{m+1})|I_m) > 0$ for all $m = 1, \dots, M-1$. Thus, for all $I, J \in \mathcal{I}_i$, $I \geq^\mu J$ iff there is a μ -sequence I_1, \dots, I_M with $I_1 = J$ and $I_M = I$.

The following result states that every equivalence class of \geq^μ 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 , array $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i} \in \Delta(S_{-i})^{\mathcal{I}_i}$, and information set $I \in \mathcal{I}_i$, there is a μ -sequence $I_1, \dots, I_M \in \mathcal{I}_i$ such that $I_1 = I_M = I$ and, for all $J \in \mathcal{I}_i$, $J =^\mu I$ if and only if $J = I_m$ for some $m = 1, \dots, M$.*

Proof: let $\{I_1, \dots, I_L\}$ be an enumeration of the equivalence class of \geq^μ containing I ; assume without loss that $I_1 = I$. Since $\{I_1, \dots, I_L\}$ is an equivalence class, in particular $I_1 \geq^\mu I_2 \geq^\mu \dots \geq^\mu I_L$ and $I_L \geq^\mu I_1$. Therefore, for every $\ell = 1, \dots, L-1$, there is a μ -sequence $I_1^\ell, \dots, I_{M(\ell)}^\ell$ such that $I_1^\ell = I_{\ell+1}$ and $I_{M(\ell)}^\ell = I_\ell$; furthermore, there is a μ -sequence $I_1^L, \dots, I_{M(L)}^L$ such that $I_1^L = I_1$ and $I_{M(L)}^L = I^L$. Then the ordered list

$$I_1^L, I_2^L, \dots, I_{M(L)}^L = I_1^{L-1}, \dots, I_{M(L-1)}^{L-1} = I_1^{L-2}, \dots, I_{M(1)}^1. \quad (7)$$

is a μ -sequence, with $I_1^L = I_1 = I$ and $I_{M(1)}^1 = I_1 = I$.

By construction, $I_\ell = I_{M(\ell)}^\ell$ for every $\ell = 1, \dots, L$, so this μ -sequence contains the equivalence class $\{I_1, \dots, I_L\}$ for I . Finally, notice that, for every $\ell = 1, \dots, L$ and $m = 1, \dots, M(\ell)$, the sublist of the list in Eq. (7) beginning with I_1^L and ending with I_m^ℓ , and the sublist beginning with I_m^ℓ and ending with $I_{M(1)}^1$, are both μ -sequences, so $I_m^\ell \geq^\mu I_1^L$ and $I_{M(1)}^1 \geq^\mu I_m^\ell$. Furthermore, $I_1^L = I_{M(1)}^1 = I_1 = I$, so in fact $I_m^\ell \geq^\mu I$ and $I \geq^\mu I_m^\ell$, so $I_m^\ell =^\mu I_{\bar{\ell}}$ for some $\bar{\ell} = 1, \dots, L$. ■

The following result highlights the main property of CCPs, and emphasizes the role of the consistency condition in (2) of Definition 1.

Proposition 5 Fix an array $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i} \in \Delta(S_{-i})^{\mathcal{I}_i}$ for player $i \in N$ such that $\mu(S_{-i}(I)|I) = 1$ for all $I \in \mathcal{I}_i$. The following are equivalent:

1. μ satisfies Eq. (1) for every $I_1, \dots, I_K \in \mathcal{I}_i$ and $E \subseteq S_{-i}(I_1) \cap S_{-i}(I_K)$;
2. for every μ -sequence $I_1, \dots, I_K \in \mathcal{I}_i$, there exists $p \in \Delta(S_{-i})$ with $p(\cup_{\ell} S_{-i}(I_{\ell})) = 1$, such that, for every $\ell = 1, \dots, K$ and $E \subseteq S_{-i}(I_{\ell})$,

$$p(E) = \mu(E|I_{\ell})p(S_{-i}(I_{\ell})). \quad (8)$$

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

Proof: (1) \Rightarrow (2): assume that μ satisfies the stated condition, which is just Property (2) in Definition 1. Let $I_1, \dots, I_K \in \mathcal{I}_i$ be a μ -sequence.

Define $G_1 = S_{-i}(I_1)$ and, inductively, $G_k = S_{-i}(I_k) \setminus (S_{-i}(I_1) \cup \dots \cup S_{-i}(I_{k-1}))$ for $k = 2, \dots, K$. Note that $S_{-i}(I_1) \cup \dots \cup S_{-i}(I_k) = G_1 \cup \dots \cup G_k$ for all $k = 1, \dots, K$, [for $k = 1$ this is by definition. By induction, $G_1 \cup \dots \cup G_{k+1} = (G_1 \cup \dots \cup G_k) \cup G_{k+1} = (S_{-i}(I_1) \cup \dots \cup S_{-i}(I_k)) \cup G_{k+1} = (S_{-i}(I_1) \cup \dots \cup S_{-i}(I_k)) \cup [S_{-i}(I_{k+1}) \setminus (S_{-i}(I_1) \cup \dots \cup S_{-i}(I_k))] = S_{-i}(I_1) \cup \dots \cup S_{-i}(I_{k+1})$] and $G_k \cap G_{\ell} = \emptyset$ for all $k \neq \ell$. [Let $\ell > k$: then $G_{\ell} = S_{-i}(I_{\ell}) \setminus (S_{-i}(I_1) \cup \dots \cup S_{-i}(I_{\ell-1})) = S_{-i}(I_{\ell}) \setminus (G_1 \cup \dots \cup G_{\ell-1})$, and $k \in \{1, \dots, \ell-1\}$.] Also, $G_k \subseteq S_{-i}(I_k)$ for all $k = 1, \dots, K$.

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

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

with the usual convention that the product over an empty set of indices equals 1. By assumption, the denominators of the above fractions are all strictly positive. Also, since the sets G_1, \dots, G_k are disjoint by construction, if $\emptyset \neq E \subseteq G_{\ell}$ for some ℓ then $E \not\subseteq G_k$ for $k \neq \ell$, so $\rho(E)$ is uniquely defined; furthermore, $\emptyset \subseteq G_k$ for all k , but $\rho(\emptyset)$ is still well-defined and equal to 0.

To complete the definition of $\rho(\cdot)$, for all events $E \subseteq S_{-i}$ such that $E \not\subseteq G_k$ for $k = 1, \dots, K$

[i.e., E intersects two or more events G_k , or none], let

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

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

$$\begin{aligned} \rho\left(\bigcup_m E_m\right) &= \mu\left(\bigcup_m E_m \middle| F_k\right) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} = \left(\sum_m \mu(E_m|I_k)\right) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} = \\ &= \sum_m \left(\mu(E_m|I_\ell) \cdot \prod_{k=\ell}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)}\right) = \sum_m \rho(E_m). \end{aligned}$$

Thus, for a general ordered list $E_1, \dots, E_M \subseteq S_{-i}$ of pairwise disjoint events,²⁸

$$\begin{aligned} \rho\left(\bigcup_m E_m\right) &= \sum_k \rho\left(\left[\bigcup_m E_m\right] \cap G_k\right) = \sum_k \rho\left(\bigcup_m [E_m \cap G_k]\right) = \\ &= \sum_k \sum_m \rho(E_m \cap G_k) = \sum_m \sum_k \rho(E_m \cap G_k) = \sum_m \rho(E_m). \end{aligned}$$

Now consider $E \subseteq S_{-i}$ with $E \subseteq S_{-i}(I_m)$ and $E \subseteq G_\ell$ for some $\ell, m \in \{1, \dots, K\}$ with $\ell \neq m$. Since $S_{-i}(I_m) \subseteq S_{-i}(I_1) \cup \dots \cup S_{-i}(I_m) = G_1 \cup \dots \cup G_m$, $\ell < m$. Consider the ordered list $I_\ell, \dots, I_m \in \mathcal{C}_i$: since I_1, \dots, I_K is a μ -sequence, so is I_ℓ, \dots, I_m , so by Eq. (1), since by assumption $E \subseteq S_{-i}(I_m) \cap G_\ell \subseteq S_{-i}(I_m) \cap S_{-i}(I_\ell)$,

$$\mu(E|I_\ell) \prod_{k=\ell}^{m-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} = \mu(E|I_m).$$

Multiply both sides by the quantity $\prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)}$ to get

$$\rho(E) = \mu(E|I_\ell) \prod_{k=\ell}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} = \mu(E|I_m) \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)}.$$

²⁸For future reference, if the set S_{-i} is an arbitrary measurable space, and probabilities in i 's CPS are countably additive, this step of the proof still holds, and shows that ρ is countably additive. Specifically, the derivation holds as written for a countable collection E_1, E_2, \dots (I purposely omitted limits from the summations). In particular, interchanging the order of the summation in the second line is allowed because all summands are non-negative and the derivation shows that $\sum_k \sum_m \rho(E_m \cap G_k) = \sum_k \rho([\cup_m E_m] \cap G_k)$, a sum of finitely many finite terms.

Therefore, for all $E \subseteq S_{-i}$ with $E \subseteq S_{-i}(I_m)$ for some $m \in \{1, \dots, K\}$,

$$\begin{aligned} \mu(E|I_m) \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} &= \sum_{\ell=1}^K \mu(E \cap G_\ell|I_m) \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} = \\ &= \sum_{\ell=1}^K \rho(E \cap G_\ell) = \rho(E).^{29} \end{aligned}$$

In particular, for all $m \in \{1, \dots, K\}$, since $\mu(S_{-i}(I_m)|I_m) = 1$ by assumption,

$$\rho(S_{-i}(I_m)) = \mu(S_{-i}(I_m)|I_m) \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} = \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)}, \quad (9)$$

and therefore, for all $E \subseteq S_{-i}$ with $E \subseteq S_{-i}(I_m)$,

$$\rho(E) = \mu(E|I_m) \rho(S_{-i}(I_m)). \quad (10)$$

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

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

For every $\ell \in \{1, \dots, K\}$ and every event $E \subseteq S_{-i}(I_\ell)$, p satisfies Eq. (8), as asserted. Furthermore, $p(S_{-i}(I_K)) > 0$.

To show that p is uniquely defined, let $q \in \Delta(S_{-i})$ be a measure that satisfies Eq.(8). I first claim that, for every $m = 1, \dots, K$,

$$q(S_{-i}(I_m)) = \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} \cdot q(S_{-i}(I_K)) = \rho(S_{-i}(I_m)) q(S_{-i}(I_K)).$$

The claim is trivially true for $m = K$ because $\rho(S_{-i}(I_K)) = 1$, so consider $m \in \{1, \dots, K-1\}$ and assume that the claim holds for $m+1$. By Eq.(8),

$$\mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_{m+1}) q(S_{-i}(I_{m+1})) = q(S_{-i}(I_m) \cap S_{-i}(I_{m+1})) = \mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_m) q(S_{-i}(I_m));$$

since $\mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_m) > 0$ by assumption, solving for $q(S_{-i}(I_m))$ and invoking the inductive hypothesis yields

$$\begin{aligned} q(S_{-i}(I_m)) &= \frac{\mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_{m+1})}{\mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_m)} q(S_{-i}(I_{m+1})) = \\ &= \frac{\mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_{m+1})}{\mu(S_{-i}(I_m) \cap S_{-i}(I_{m+1})|I_m)} \cdot \prod_{k=m+1}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} \cdot q(S_{-i}(I_K)) \\ &= \prod_{k=m}^{K-1} \frac{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1})}{\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_k)} \cdot q(S_{-i}(I_K)) = \rho(S_{-i}(I_m)) q(S_{-i}(I_K)). \end{aligned}$$

Since $G_m \subseteq S_{-i}(I_m)$, Eq. (8) implies that

$$q(G_m) = \mu(G_m|I_m) q(S_{-i}(I_m)) = \mu(G_m|I_m) \cdot \rho(S_{-i}(I_m)) \cdot q(S_{-i}(I_K)) = \rho(G_m) \cdot q(S_{-i}(I_K)),$$

where the last equality follows from Eq.(10). Since $\sum_k q(G_k) = q(\cup_k G_k) = q(\cup_k S_{-i}(I_k))$, if in addition q satisfies $q(\cup_k S_{-i}(I_k)) = 1$, then

$$1 = \sum_m \rho(G_m) \cdot q(S_{-i}(I_K)) = q(S_{-i}(I_K)) \rho(\cup_m G_m)$$

which implies that $q(S_{-i}(I_K)) > 0$, and indeed that

$$q(S_{-i}(I_K)) = \frac{1}{\rho(\cup_m G_m)} = \frac{\rho(S_{-i}(I_K))}{\rho(\cup_m G_m)} = p(S_{-i}(I_K)).$$

Then, for $m = 1, \dots, K-1$,

$$q(S_{-i}(I_m)) = \rho(S_{-i}(I_m)) q(S_{-i}(I_N)) = \rho(S_{-i}(I_m)) \frac{1}{\rho(\cup_m G_m)} = p(S_{-i}(I_m)). \quad (11)$$

Furthermore, let $k_0 \in \{1, \dots, K-1\}$ be such that $\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1})|I_{k+1}) > 0$ for all $k > k_0$, and $\mu(S_{-i}(I_{k_0}) \cap S_{-i}(I_{k_0+1})|I_{k_0+1}) = 0$. By inspecting Eq. (9), it is clear that $\rho(S_{-i}(I_k)) = 0$ for $k = 1, \dots, k_0$, and $\rho(S_{-i}(I_k)) > 0$ for $k = k_0 + 1, \dots, K$. Then, $p(S_{-i}(I_k)) = 0$ for $k = 1, \dots, k_0$, and $p(S_{-i}(I_k)) > 0$ for $k = k_0 + 1, \dots, K$. From Eq. (11), it follows that the same is true for any $q \in \Delta(S_{-i})$ that satisfies Eq. (8) and $q(\cup_k S_{-i}(I_k)) = 1$. Thus, the last claim of the Proposition follows.

Finally, if $q \in \Delta(\mathcal{S}_{-i})$ satisfies Eq.(8) and $q(\cup_k \mathcal{S}_{-i}(I_k)) = 1$, for every $k = k_0 + 1, \dots, K$ and $E \subseteq \mathcal{S}_{-i}$ such that $E \subset \mathcal{S}_{-i}(I_k)$,

$$q(E) = \mu(E|I_k)q(\mathcal{S}_{-i}(I_k)) = \mu(E|I_k)p(\mathcal{S}_{-i}(I_k)) = p(E)$$

and therefore, for every $E \subseteq \mathcal{S}_{-i}$,

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

In other words, p is the unique probability measure that satisfies Eq. (8) and $p(\cup_k \mathcal{S}_{-i}(I_k)) = 1$.

(2) \Rightarrow (1): assume that (2) holds. Consider a collection $I_1, \dots, I_K \in \mathcal{I}_i$ and fix an event $E \subseteq \mathcal{S}_{-i}(I_1) \cap \mathcal{S}_{-i}(I_K)$. If neither ordered list I_1, I_2, \dots, I_K nor I_K, I_{K-1}, \dots, I_1 is a μ -sequence, then there are $k', k'' \in \{1, \dots, K-1\}$ such that $\mu(\mathcal{S}_{-i}(I_{k'}) \cap \mathcal{S}_{-i}(I_{k'+1})|I_{k'}) = 0$ and $\mu(\mathcal{S}_{-i}(I_{k''}) \cap \mathcal{S}_{-i}(I_{k''+1})|I_{k''+1}) = 0$, so $\prod_{k=1}^{K-1} \mu(\mathcal{S}_{-i}(I_k) \cap \mathcal{S}_{-i}(I_{k+1})|I_{k+1}) = \prod_{k=1}^{K-1} \mu(\mathcal{S}_{-i}(I_k) \cap \mathcal{S}_{-i}(I_{k+1})|I_{k+1}) = 0$ and Eq. (1) holds trivially. Otherwise, assume wlog that I_1, \dots, I_K is a μ -sequence (if not, then the reverse ordered list I_K, I_{K-1}, \dots, I_1 is, and the argument is symmetric). By assumption, there exists $p \in \Delta(\mathcal{S}_{-i})$ that satisfies Eq. (8) for $k = 1, \dots, K$, with $p(\cup_{k=1}^K \mathcal{S}_{-i}(I_k)) = 1$.

Since $p(\mathcal{S}_{-i}(I_K)) > 0$, $\mu(E|I_K) = \frac{p(E)}{p(\mathcal{S}_{-i}(I_K))}$. Suppose first that $p(\mathcal{S}_{-i}(I_1)) = 0$. Then a fortiori $p(E) = 0$, so $\mu(E|I_K) = 0$ and therefore

$$\mu(E|I_K) \cdot \prod_{k=1}^{K-1} \mu(\mathcal{S}_{-i}(I_k) \cap \mathcal{S}_{-i}(I_{k+1})|I_k) = 0;$$

also, by the last claim in the Proposition, which follows from (2), $p(\mathcal{S}_{-i}(I_1)) = 0$ implies that there is $k' \in \{1, \dots, K-1\}$ with $\mu(\mathcal{S}_{-i}(I_{k'}) \cap \mathcal{S}_{-i}(I_{k'+1})|I_{k'+1}) = 0$, so

$$\mu(E|I_1) \cdot \prod_{k=1}^{K-1} \mu(\mathcal{S}_{-i}(I_k) \cap \mathcal{S}_{-i}(I_{k+1})|I_{k+1}) = \mu(E|I_1) \cdot 0 = 0.$$

Thus, Eq. (1) holds. If instead $p(\mathcal{S}_{-i}(I_1)) > 0$, then $\mu(E|I_1) = \frac{p(E)}{p(\mathcal{S}_{-i}(I_1))}$; furthermore, by the last

claim in the Proposition, $p(S_{-i}(I_k)) > 0$ for all $k = 2, \dots, K - 1$ as well, so

$$\begin{aligned} \mu(E|I_1) \cdot \prod_{k=1}^{K-1} \mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1}) | I_{k+1}) &= \frac{p(E)}{p(S_{-i}(I_1))} \cdot \prod_{k=1}^{K-1} \frac{p(S_{-i}(I_k) \cap S_{-i}(I_{k+1}))}{p(S_{-i}(I_{k+1}))} = \\ &= \frac{p(E)}{p(S_{-i}(I_K))} \cdot \prod_{k=1}^{K-1} \frac{p(S_{-i}(I_k) \cap S_{-i}(I_{k+1}))}{p(S_{-i}(I_k))} = \\ &= \mu(E|I_K) \cdot \prod_{k=1}^{K-1} \mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1}) | I_k). \end{aligned}$$

Thus, again, Eq. (1) holds. ■

Corollary 3 *If μ is a CCPS, then for every μ -sequence I_1, \dots, I_K such that $\mu(S_{-i}(I_1) | I_K) > 0$, the reverse-ordered list I_K, I_{K-1}, \dots, I_1 is also a μ -sequence: that is, $\mu(S_{-i}(I_k) | I_{k+1}) > 0$ for all $k = 1, \dots, K - 1$.*

In particular, this Corollary applies if $I_1 = I_K$.

Proof: Let I_1, \dots, I_K be as in the statement, and consider the ordered list I_1, \dots, I_K, I_{K+1} with $I_{k+1} = I_1$. Then I_1, \dots, I_{K+1} is also a μ -sequence. Let p be the unique measure in (2) of Proposition 5. The last claim of that Proposition shows that necessarily $p(S_{-i}(I_{K+1})) > 0$, but since $I_{K+1} = I_1$, also $p(S_{-i}(I_1)) > 0$. Again, the last claim in Proposition 5 implies that then $p(S_{-i}(I_k)) > 0$ for all $k = 1, \dots, K$. Then, for all $k = 1, \dots, K - 1$, $\mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1}) | I_k) > 0$ implies that $p(S_{-i}(I_k) \cap S_{-i}(I_{k+1})) > 0$, and so

$$\mu(S_{-i}(I_k) | I_{k+1}) = \mu(S_{-i}(I_k) \cap S_{-i}(I_{k+1}) | I_{k+1}) = \frac{p(S_{-i}(I_k) \cap S_{-i}(I_{k+1}))}{p(S_{-i}(I_{k+1}))} > 0.$$

■

The first two claims of Proposition 3 in Section 5 follow readily from Proposition 5.

Proof of Proposition 3, first two claims: Fix a CCPS μ for player i , and $I \in \mathcal{I}_i$. By Lemma 1, there is a μ -sequence $I_1, \dots, I_K = I_1$ such that $J \stackrel{\mu}{=} I$ iff $J = I_k$ for some $k \in \{1, \dots, K\}$.

Let $P \in \Delta(\mathcal{S}_{-i})$ be the unique probability delivered by Proposition 5, so Eq. (8) holds and $P(\{\mathcal{S}_{-i}(J) : J =^\mu\}) = P(\cup_k \mathcal{S}_{-i}(I_k)) = 1$. By Corollary 3, $\mu(\mathcal{S}_{-i}(I_k)|I_{k+1}) > 0$ for all $k = 1, \dots, K-1$. Thus, by the last part of Proposition 5, $P(\mathcal{S}_{-i}(I_k)) > 0$ for all $k = 1, \dots, K$. Hence, by Eq. (8), $\mu(\cdot|J) = P(\cdot|\mathcal{S}_{-i}(J))$ for all $J \in \mathcal{I}_i$ with $J =^\mu I$. ■

The next result is useful to analyze the probabilities $P_\mu(I)$ associated to different information sets $I \in \mathcal{I}_i$ by Eq. (4) in Proposition 3, as well as to relate CCPs with perturbations.

Lemma 2 *Fix a player $i \in N$, a CCPS $\mu \in \Delta(\mathcal{S}_{-i}, \mathcal{I}_i)$, and an information set $I \in \mathcal{I}_i$. Consider a collection $J_1, \dots, J_L \in \mathcal{I}_i$ such that $J_\ell =^\mu J_m$ for all $\ell, m \in \{1, \dots, L\}$, and $P_\mu(I)(\cup_\ell \mathcal{S}_{-i}(J_\ell)) > 0$. Then there are $\hat{\ell} \in \{1, \dots, L\}$ and $\hat{I} \in \mathcal{I}_i$ such that $\hat{I} =^\mu I$ and $\mu(\mathcal{S}_{-i}(J_{\hat{\ell}})|\hat{I}) > 0$. Hence $J_\ell \geq^\mu I$ for all ℓ .*

Proof: Denote by I_1, \dots, I_M the \geq^μ -equivalence class for I . By the second claim of Proposition 3, $P_\mu(I)(\cup_m \mathcal{S}_{-i}(I_m)) = 1$, so

$$0 < P_\mu(I)(\cup_\ell \mathcal{S}_{-i}(J_\ell)) \leq \sum_{\hat{m}} P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}}) \cap [\cup_\ell \mathcal{S}_{-i}(J_\ell)]),$$

so there must be \hat{m} with $P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}}) \cap [\cup_\ell \mathcal{S}_{-i}(J_\ell)]) > 0$. Furthermore,

$$0 < P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}}) \cap [\cup_\ell \mathcal{S}_{-i}(J_\ell)]) \leq \sum_{\hat{\ell}} P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}}) \cap \mathcal{S}_{-i}(J_{\hat{\ell}})),$$

so there is $\hat{\ell}$ with $P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}}) \cap \mathcal{S}_{-i}(J_{\hat{\ell}})) > 0$. A fortiori, $P_\mu(I)(\mathcal{S}_{-i}(J_{\hat{\ell}})) > 0$ and $P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}})) > 0$. By Eq. (4), $0 < P_\mu(I)(\mathcal{S}_{-i}(J_{\hat{\ell}}) \cap \mathcal{S}_{-i}(I_{\hat{m}})) = \mu(\mathcal{S}_{-i}(J_{\hat{\ell}}) \cap \mathcal{S}_{-i}(I_{\hat{m}})|I_{\hat{m}}) \cdot P_\mu(I)(\mathcal{S}_{-i}(I_{\hat{m}}))$, so $\mu(\mathcal{S}_{-i}(J_{\hat{\ell}})|I_{\hat{m}}) = \mu(\mathcal{S}_{-i}(J_{\hat{\ell}}) \cap \mathcal{S}_{-i}(I_{\hat{m}})|I_{\hat{m}}) > 0$, so the claim holds with $\hat{I} = I_{\hat{m}}$. In turn, this implies that $J_\ell =^\mu J_{\hat{\ell}} \geq^\mu \hat{I} =^\mu I$ for all ℓ . ■

Corollary 4 *For all $I, J \in \mathcal{I}_i$, $P_\mu(I) \cup \{\mathcal{S}_{-i}(J') : J' =^\mu J\} > 0$ implies $J \geq^\mu I$. Hence $I \neq^\mu J$ implies $\text{supp } P_\mu(I) \cap \text{supp } P_\mu(J) = \emptyset$.*

Proof: Let J_1, \dots, J_M be the \geq^μ -equivalence class of J . Then by assumption $P_\mu(I)(\cup_m S_{-i}(J_m)) > 0$, and Lemma 2 implies that $J_m \geq^\mu I$; in particular, $J \geq^\mu I$. The second claim is proved by contradiction: if there is $t_{-i} \in \text{supp } P_\mu(I) \cap \text{supp } P_\mu(J)$, then $P_\mu(J)(\{S_{-i}(J') : J' =^\mu J\}) = 1$ implies that $t_{-i} \in \{S_{-i}(J') : J' =^\mu J\}$, and $t_{-i} \in \text{supp } P_\mu(I)$ implies that $P_\mu(I)(\cup_m S_{-i}(J_m)) > 0$, so $J \geq^\mu I$. By a symmetric argument, $I \geq^\mu J$, so $I =^\mu J$, contradiction. ■

The final set of results concerns the connection between CCPSs and perturbations. Given a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$, a **representative collection for μ** is a tuple $I_1, \dots, I_M \in \mathcal{I}_i$ such that (i) for all $J \in \mathcal{I}_i$ there exists $m \in \{1, \dots, M\}$ such that $J =^\mu I_m$, and (ii) for all $\ell, m \in \{1, \dots, M\}$, $I_\ell =^\mu I_m$ iff $\ell = m$. That is, I_1, \dots, I_M are arbitrarily chosen elements of each equivalence class of \geq^μ . Given a representative collection I_1, \dots, I_M for μ , a function $f : \{1, \dots, M\} \rightarrow \mathbb{R}$ **agrees with $>^\mu$** if, for all $\ell, m \in \{1, \dots, M\}$, $I_\ell >^\mu I_m$ implies $f(\ell) < f(m)$; the **canonical map for I_1, \dots, I_M** is the function $g : \{1, \dots, M\} \rightarrow \mathbb{R}$ defined by

$$\forall m \in \{1, \dots, M\}, \quad g(m) = |\ell : I_\ell >^\mu I_m| + \frac{m}{M+1}.$$

Lemma 3 *Fix a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$ and a representative collection I_1, \dots, I_M for μ . Then the canonical map for I_1, \dots, I_M is strictly positive, one-to-one and agrees with $>^\mu$.*

Proof: Strict positivity is immediate by definition. Next, fix $\ell, m \in \{1, \dots, M\}$ with $\ell \neq m$. If $|n : I_n >^\mu I_\ell| = |n : I_n >^\mu I_m|$, then $g(\ell) - g(m) = \frac{\ell}{M+1} - \frac{m}{M+1} \neq 0$. Otherwise, assume wlog that $|n : I_n >^\mu I_\ell| < |n : I_n >^\mu I_m|$; thus, $|n : I_n >^\mu I_\ell| - |n : I_n >^\mu I_m| \leq -1$. Since $\frac{\ell}{M+1} - \frac{m}{M+1} = \frac{\ell-m}{M+1} \in (\frac{1-M}{M+1}, \frac{M-1}{M+1}) \subsetneq (-1, 1)$, $g(\ell) - g(m) \leq -1 + \frac{\ell-m}{M+1} < 0$. Thus, g is one-to-one.

Furthermore, if $I_\ell >^\mu I_m$, then by transitivity, for every $n \in \{1, \dots, M\}$, $I_n >^\mu I_\ell$ implies $I_n >^\mu I_m$; and in addition, $I_\ell \not>^\mu I_\ell$ and $I_\ell >^\mu I_m$. Thus, $|n : I_n >^\mu I_\ell| < |n : I_n >^\mu I_m|$, which, as just shown, implies $g(\ell) - g(m) < 0$, or $g(\ell) < g(m)$. Thus g agrees with $>^\mu$. ■

Lemma 4 Fix a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$, a representative collection I_1, \dots, I_M for μ , and a function $f : \{1, \dots, M\} \rightarrow \mathbb{R}_+$ that agrees with $>^\mu$. Then the sequence $(p^k)_{k \geq 1}$ defined by

$$\forall k \geq 1, E \subseteq S_{-i}, \quad p^k(E) = \frac{\sum_{m=1}^M \frac{P_\mu(I_m)(E)}{k^{f(m)}}}{\sum_{m=1}^M \frac{1}{k^{f(m)}}$$

is a perturbation of μ .

Proof: For every $k \geq 1$, p^k is a weighted average of the probabilities $P_\mu(I_1), \dots, P_\mu(I_M)$, with weight $k^{-f(\ell)} / \sum_m k^{-f(m)}$ on $P_\mu(I_\ell)$. Thus, $p^k \in \Delta(S_{-i})$. Now fix $I \in \mathcal{I}_i$ and let $\ell \in \{1, \dots, M\}$ be such that $I =^\mu I_\ell$. By the second claim of Proposition 3, $P_\mu(I_\ell)(S_{-i}(I)) > 0$, so $p^k(S_{-i}(I)) > 0$ for all $k \geq 1$. Finally, consider $E \subseteq S_{-i}(I)$.

If $\mu(E|I) = 0$, then since $\mu(\cdot|I) = P_\mu(I_\ell)(\cdot|S_{-i}(I))$, by Eq. (4), $P_\mu(I_\ell)(E) = 0$.

Suppose that $P_\mu(I_m)(E) > 0$ for some $m \in \{1, \dots, M\} \setminus \{\ell\}$. Then $P_\mu(I_m)(S_{-i}(I)) > 0$, so $P_\mu(I_m)(\cup\{S_{-i}(J) : J =^\mu I_\ell\}) > 0$, and by Corollary 4, $I_\ell \geq^\mu I_m$. By the definition of a representative collection, $\ell \neq m$ implies that $I_\ell \neq^\mu I_m$, so $I_\ell >^\mu I_m$. Since f agrees with $>^\mu$, $f(\ell) < f(m)$.

It follows that

$$\begin{aligned} p^k(E) &= P_\mu(I_\ell)(E) \cdot \frac{k^{-f(\ell)}}{\sum_{n=1}^M k^{-f(n)}} + \sum_{m \neq \ell} P_\mu(I_m)(E) \cdot \frac{k^{-f(m)}}{\sum_{n=1}^M k^{-f(n)}} = \\ &= P_\mu(I_\ell)(E) \cdot \frac{k^{-f(\ell)}}{\sum_{n=1}^M k^{-f(n)}} + \sum_{m \neq \ell: P_\mu(I_m)(E) > 0} P_\mu(I_m)(E) \cdot \frac{k^{-f(m)}}{\sum_{n=1}^M k^{-f(n)}} = \\ &= P_\mu(I_\ell)(E) \cdot \frac{k^{-f(\ell)}}{\sum_{n=1}^M k^{-f(n)}} + \sum_{m \neq \ell: P_\mu(I_m)(E) > 0, f(\ell) < f(m)} P_\mu(I_m)(E) \cdot \frac{k^{-f(m)}}{\sum_{n=1}^M k^{-f(n)}}. \end{aligned}$$

Hence, since the above conclusion holds in particular for $E = S_{-i}(I)$,

$$\begin{aligned}
p^k(E|S_{-i}(I)) &= \frac{P_\mu(I_\ell)(E) \cdot \frac{k^{-f(\ell)}}{\sum_{n=1}^M k^{-f(n)}} + \sum_{m \neq \ell: P_\mu(I_m)(E) > 0, f(\ell) < f(m)} P_\mu(I_m)(E) \cdot \frac{k^{-f(m)}}{\sum_{n=1}^M k^{-f(n)}}}{P_\mu(I_\ell)(S_{-i}(I)) \cdot \frac{k^{-f(\ell)}}{\sum_{n=1}^M k^{-f(n)}} + \sum_{m \neq \ell: P_\mu(I_m)(S_{-i}(I)) > 0, f(\ell) < f(m)} P_\mu(I_m)(S_{-i}(I)) \cdot \frac{k^{-f(m)}}{\sum_{n=1}^M k^{-f(n)}}} = \\
&= \frac{P_\mu(I_\ell)(E) \cdot k^{-f(\ell)} + \sum_{m \neq \ell: P_\mu(I_m)(E) > 0, f(\ell) < f(m)} P_\mu(I_m)(E) \cdot k^{-f(m)}}{P_\mu(I_\ell)(S_{-i}(I)) \cdot k^{-f(\ell)} + \sum_{m \neq \ell: P_\mu(I_m)(S_{-i}(I)) > 0, f(\ell) < f(m)} P_\mu(I_m)(S_{-i}(I)) \cdot k^{-f(m)}} = \\
&= \frac{P_\mu(I_\ell)(E) + \sum_{m \neq \ell: P_\mu(I_m)(E) > 0, f(\ell) < f(m)} P_\mu(I_m)(E) \cdot k^{f(\ell)-f(m)}}{P_\mu(I_\ell)(S_{-i}(I)) + \sum_{m \neq \ell: P_\mu(I_m)(S_{-i}(I)) > 0, f(\ell) < f(m)} P_\mu(I_m)(S_{-i}(I)) \cdot k^{f(\ell)-f(m)}} \rightarrow \\
&\rightarrow P_\mu(I_\ell)(E|S_{-i}(I)) = \mu(E|I);
\end{aligned}$$

the second equality follows by multiplying numerator and denominator by $\sum_{n=1}^M k^{-f(n)}$, the third by multiplying numerator and denominator by $k^{f(\ell)}$, the limit by noting that, in the sums in both numerator and denominator, $f(\ell) - f(m) < 0$ for each summand, and so $k^{f(\ell)-f(m)} \rightarrow 0$, and the last equality from Eq. (4). Thus, (p^k) is a perturbation of μ . ■

It is now possible to prove Proposition 1 in Section 3.

Proof of Proposition 1: the argument showing that, if an array $\mu = (\mu(\cdot|I))_{I \in \mathcal{I}_i} \in \Delta(S_{-i})^{\mathcal{I}_i}$ admits a perturbation, then it is a CCPS was given in Section 3. For the converse, suppose μ is a CCPS and let I_1, \dots, I_M be a representative collection for μ . By Lemma 3, the canonical map g for I_1, \dots, I_M agrees with $>^\mu$. Hence, taking $f = g$ in Lemma 4 yields a perturbation of μ . ■

Lemma 5 Fix a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$. For $I, J \in \mathcal{I}_i$:

1. $I \geq^\mu J$ if and only if that there is $c \in (0, \infty)$ such that $\frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} \geq c$ eventually for all perturbations $(p^k)_{k \geq 1}$ of μ .
2. $I >^\mu J$ if and only if $\lim_k \frac{p^k(S_{-i}(J))}{p^k(S_{-i}(I))} = 0$ for all perturbations $(p^k)_{k \geq 1}$ of μ .
3. $P_\mu(I) = \lim_{k \rightarrow \infty} p^k(\cdot | \cup \{S_{-i}(I') : I \geq^\mu I'\}) = \lim_{k \rightarrow \infty} p^k(\cdot | \cup \{S_{-i}(I') : I =^\mu I'\})$ for all perturbations $(p^k)_{k \geq 1}$ of μ .

Proof: (1): suppose that $I \geq^\mu J$. Then there is a μ -sequence $I_1, \dots, I_L \in \mathcal{I}_i$ with $I_1 = J$ and $I_L = I$. Therefore, for any perturbation $(p^k)_{k \geq 1}$ of μ ,

$$0 < d \equiv \prod_{\ell=1}^{L-1} \mu(S_{-i}(I_{\ell+1})|I_\ell) = \prod_{\ell=1}^{L-1} \lim_{k \rightarrow \infty} \frac{p^k(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1}))}{p^k(S_{-i}(I_\ell))} = \lim_{k \rightarrow \infty} \prod_{\ell=1}^{L-1} \frac{p^k(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1}))}{p^k(S_{-i}(I_\ell))}; \quad (12)$$

if $0 < c < d$ then, for any perturbation $(p^k)_{k \geq 1}$ of μ , there is $K \geq 1$ such that $k \geq K$ implies

$$c \leq \prod_{\ell=1}^{L-1} \frac{p^k(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1}))}{p^k(S_{-i}(I_\ell))} \leq \prod_{\ell=1}^{L-1} \frac{p^k(S_{-i}(I_{\ell+1}))}{p^k(S_{-i}(I_\ell))} = \frac{p^k(S_{-i}(I_L))}{p^k(S_{-i}(I_1))} = \frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))}. \quad (13)$$

Conversely, suppose that there is $c > 0$ such that $p^k(S_{-i}(I))/p^k(S_{-i}(J)) \geq c$ eventually for all perturbations $(p^k)_{k \geq 1}$ of μ . By contradiction, assume that not $I \geq^\mu J$. Let I_1, \dots, I_M be a representative collection for μ and g the canonical map for I_1, \dots, I_M . Without loss of generality, let $I_1 = I$ and $I_M = J$ (this is possible because not $I \geq^\mu J$ implies $I_1 \neq^\mu J$). Finally, let $f : \{1, \dots, M\} \rightarrow \mathbb{R}_+$ by letting $f(n) = g(n) + g(M) + 1$ if $I = I_1 \geq^\mu I_n$, and $f(n) = g(n)$ otherwise. Consider $n, n' \in \{1, \dots, M\}$ with $I_n >^\mu I_{n'}$. If $I \geq^\mu I_n$, then also $I \geq^\mu I_{n'}$ and so $f(n) = g(n) + g(M) + 1 < g(n') + g(M) + 1 = f(n')$. Otherwise, $f(n) = g(n) < g(n')$ and either $f(n') = g(n')$, or $f(n') = g(n') + g(M) + 1 > g(n')$. In either case, $f(n) < f(n')$. Thus, f agrees with $>^\mu$. By Lemma 4, letting $p^k = (\sum_m k^{-f(m)})^{-1} \sum_m k^{-f(m)} P_\mu(I_m)$ defines a perturbation $(p^k)_{k \geq 1}$ of μ . Moreover

$$\frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} = \frac{p^k(S_{-i}(I_1))}{p^k(S_{-i}(I_M))} = \frac{\sum_{m: I_1 \geq^\mu I_m} k^{-f(m)} P_\mu(I_m)(S_{-i}(I_1))}{\sum_{m: I_M \geq^\mu I_m} k^{-f(m)} P_\mu(I_m)(S_{-i}(I_1))} \leq \frac{|m : I_1 \geq^\mu I_m| \cdot k^{-(g(1)+g(M)+1)}}{k^{-g(M)} P_\mu(I_M)(S_{-i}(I_M))} \rightarrow 0.$$

The second equality follows by cancelling the denominators in the definition of p^k and recalling that, by Corollary 4, $P_\mu(I_m)(S_{-i}(I_1)) > 0$ implies $I_1 \geq I_m$ and, respectively, $P_\mu(I_m)(S_{-i}(I_M)) > 0$ implies $I_M \geq I_m$. The inequality follows by noting that, in the numerator, $I_1 >^\mu I_m$ implies $f(1) < f(m)$, therefore $k^{-f(1)} \geq k^{-f(m)}$, and in addition $f(1) = g(1) + g(M) + 1$; and, in the denominator, the terms corresponding to m such that $I_M >^\mu I_m$ are all non-negative, so eliminating them weakly increases the value of the fraction, and in addition not $I_1 \geq^\mu I_M$ implies that $f(M) = g(M)$. Finally, the limit follows because $g(1) + 1 \geq 1$. This contradicts the assumption that $p^k(S_{-i}(I))/p^k(S_{-i}(J)) \geq c > 0$ eventually for all perturbations $(p^k)_{k \geq 1}$ of μ . Thus, $I \geq^\mu J$.

(2): since $I >^\mu J$ implies $I \geq^\mu J$, there is a μ -sequence $I_1, \dots, I_L \in \mathcal{I}_i$ with $I_1 = J$ and $I_L = I$. By contradiction, suppose that $p^{k(m)}(S_{-i}(J))/p^{k(m)}(S_{-i}(I)) \geq c > 0$ for some subsequence $(p^{k(m)})_{m \geq 1}$. By Eq. (12), $I \geq^\mu J$, $I_1 = J$ and $I_L = I$ imply that

$$\begin{aligned} 0 < \lim_{m \rightarrow \infty} \prod_{\ell=1}^{L-1} \frac{p^{k(m)}(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1}))}{p^{k(m)}(S_{-i}(I_\ell))} \cdot c &\leq \lim_{m \rightarrow \infty} \prod_{\ell=1}^{L-1} \frac{p^{k(m)}(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1}))}{p^{k(m)}(S_{-i}(I_\ell))} \cdot \frac{p^{k(m)}(S_{-i}(I_1))}{p^{k(m)}(S_{-i}(I_L))} = \\ &= \lim_{m \rightarrow \infty} \prod_{\ell=1}^{L-1} \frac{p^{k(m)}(S_{-i}(I_\ell) \cap S_{-i}(I_{\ell+1}))}{p^{k(m)}(S_{-i}(I_{\ell+1}))} = \prod_{\ell=1}^{L-1} \mu(S_{-i}(I_{\ell+1})|I_{\ell+1}), \end{aligned}$$

which implies that I_L, I_{L-1}, \dots, I_1 is a μ -sequence, and hence $J \geq^\mu I$, contradiction. Thus, for all $c > 0$, eventually $p^k(S_{-i}(J))/p^k(S_{-i}(I)) \leq c$, so $\lim_k p^k(S_{-i}(J))/p^k(S_{-i}(I)) = 0$.

Conversely, suppose that $\lim_k p^k(S_{-i}(J))/p^k(S_{-i}(I)) = 0$ for all perturbations $(p^k)_{k \geq 1}$ of μ . Then, for any such perturbation, eventually $p^k(S_{-i}(J))/p^k(S_{-i}(I)) \leq 1$, and so $p^k(S_{-i}(I))/p^k(S_{-i}(J)) \geq 1$. Thus, by part 1, $I \geq^\mu J$. Furthermore, for any perturbation, and any $c > 0$, $\lim_k p^k(S_{-i}(J))/p^k(S_{-i}(I)) = 0$ implies that eventually $p^k(S_{-i}(J))/p^k(S_{-i}(I)) < c$: thus, again by part 1, not $J \geq^\mu I$.

(3): let $F = \cup\{S_{-i}(I') : I \geq^\mu I'\}$. Since $S_{-i}(I) \subseteq F$ and $p^k(S_{-i}(I)) > 0$ for all k , $p^k(F) > 0$ for all k as well. Consider a subsequence $(p^{k(m)})_{m \geq 1}$ such that $\lim_m p^{k(m)}(\cdot|F)$ exists, and denote this limit by $P \in \Delta(S_{-i}(I))$. For every $I' \in \mathcal{I}_i$ with $I' =^\mu I$, part (1) implies that there are real numbers $c, d > 0$ such that $c \leq \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(S_{-i}(I))} \leq d$ eventually. Then

$$P(S_{-i}(I')) = \lim_m p^{k(m)}(S_{-i}(I')|F) = \lim_m \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(F)} = \lim_m \frac{p^{k(m)}(S_{-i}(I))}{p^{k(m)}(F)} \cdot \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(S_{-i}(I))} \in [P(S_{-i}(I)) \cdot c, P(S_{-i}(I)) \cdot d].$$

If instead $I >^\mu I'$, then by part (2), $\lim_m \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(S_{-i}(I))} = 0$, so

$$P(S_{-i}(I')) = \lim_m p^{k(m)}(S_{-i}(I')|F) = \lim_m \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(F)} = \lim_m \frac{p^{k(m)}(S_{-i}(I))}{p^{k(m)}(F)} \lim_m \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(S_{-i}(I))} = 0.$$

Then

$$1 = P(F) \leq \sum_{I': I =^\mu I'} P(S_{-i}(I')) + \sum_{I': I >^\mu I'} P(S_{-i}(I')) = \sum_{I': I =^\mu I} P(S_{-i}(I')) \leq P(S_{-i}(I)) \cdot d \cdot |\{I' : I' =^\mu I\}|,$$

so $P(S_{-i}(I)) > 0$ and therefore $P(S_{-i}(I')) \geq P(S_{-i}(I)) \cdot c > 0$ for all I' with $I' \stackrel{\mu}{=} I$. Finally, each such I' and $E \subseteq S_{-i}(I')$

$$P(E|S_{-i}(I)) = \frac{P(E)}{P(S_{-i}(I'))} = \frac{\lim_m \frac{p^{k(m)}(E)}{p^{k(m)}(F)}}{\lim_m \frac{p^{k(m)}(S_{-i}(I'))}{p^{k(m)}(F)}} = \lim_m \frac{p^{k(m)}(E)}{p^{k(m)}(S_{-i}(I'))} = \mu(E|I').$$

By the first claim of Proposition 3 and Definition 6, $P_\mu(I)$ is the only probability that satisfies these properties for all $I' \stackrel{\mu}{=} I$, so $P = P_\mu(I)$. Since this is true for all subsequences $(p^{k(m)})_{m \geq 1}$ for which $p^{k(m)}(\cdot|F)$ converges, the first equality follows in the statement follows.

For the second inequality, let $G = \cup\{S_{-i}(I') : I \stackrel{\mu}{=} I'\}$. For every k and $E \subseteq F$,

$$p^k(E|F) = p^k(E \cap G|F) + p^k(E \cap [F \setminus G]|F) = p^k(E|G) \cdot p^k(G|F) + p^k(E \cap [F \setminus G]|F).$$

But, as was shown above, $0 = P(S_{-i}(I')) = \lim_k p^k(S_{-i}(I')|F)$ for I' such that $I \stackrel{\mu}{>} I'$. Hence, $\lim_k p^k(F \setminus G|F) \leq \sum_{I': I \stackrel{\mu}{>} I'} \lim_k p^k(S_{-i}(I')|F) = 0$. Thus, a fortiori $\lim_k p^k(E \cap [F \setminus G]|F) = 0$, and furthermore $\lim_k p^k(G|F) = \lim_k [1 - p^k(F \setminus G|F)] = 1$. Thus, $\lim_k p^k(E|F) = \lim_k p^k(E|G)$, as asserted. ■

Proof of Proposition 2 Follows from Lemma 5 parts 1 and 2. In particular, for part 1, if $\frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} \geq c > 0$ eventually, then if some subsequence $(p^{k(m)})_{m \geq 1}$ satisfies $\lim \frac{p^{k(m)}(S_{-i}(I))}{p^{k(m)}(S_{-i}(J))} = d \in (0, \infty)$, then $d \geq c > 0$, so $\liminf_k \frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} \geq c > 0$. Conversely, if $\liminf_k \frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} = d > 0$, then eventually $\frac{p^k(S_{-i}(I))}{p^k(S_{-i}(J))} \geq c = \frac{1}{2}d$, or there would be a subsequence $(p^{k(m)})_{m \geq 1}$ such that $\lim \frac{p^{k(m)}(S_{-i}(I))}{p^{k(m)}(S_{-i}(J))} \leq c < d$.

Proof of Proposition 3, third claim: this is part 3 of Lemma 5.

B Appendix: Proofs of the main results

B.1 Proof of Theorem 1

(If:) Fix a perturbation $(p^k)_{k \geq 1}$ of μ . Let I_1, \dots, I_M be as in the statement. For every $m = 1, \dots, M$, let $C_m = \cup\{S_{-i}(J) : I_m \stackrel{\mu}{\geq} J\}$. It is wlog to assume that there are no $m, \ell \in \{1, \dots, M\}$ such that

$C_\ell \supseteq C_m$: if there are such ℓ, m , then $\cup_{n=1}^M \cup_{J: I_n \geq^\mu J} S_{-i}(J) = \cup_{n \in \{1, \dots, M\} \setminus \{m\}} \cup_{J: I_n \geq^\mu J} S_{-i}(J)$, so C_m can be disregarded.

Let $D_1 = C_1$ and, inductively, for $m = 2, \dots, M$, let $D_m = C_m \setminus \cup_{\ell=1}^{m-1} D_\ell$. Then D_1, \dots, D_M are pairwise disjoint, and for every $m = 1, \dots, M$, $\cup_{\ell=1}^m D_\ell = \cup_{\ell=1}^m C_\ell$. Fix a perturbation $(p^k)_{k \geq 1}$ of μ .

By Lemma 5 part 3, $p^k(\cdot|C_m) \rightarrow P_\mu(I_m)$ for every m . I claim that $p^k(D_m) > 0$ for k large and $p^k(\cdot|D_m) \rightarrow P_\mu(I_m)$ as well. The claim is trivially true for $m = 1$. Suppose it is true for some $m \geq 1$. *Subclaim:* $P_\mu(I_{m+1})(\cup_{\ell=1}^m D_m) = 0$. By contradiction, suppose $P_\mu(I_{m+1})(\cup_{\ell=1}^m D_m) = P_\mu(I_{m+1})(\cup_{\ell=1}^m C_m) > 0$. Then $P_\mu(I_{m+1})(S_{-i}(J)) > 0$ for some $J \in \mathcal{J}_i$ with $I_\ell \geq^\mu J$ for some $\ell \in \{1, \dots, m\}$. By Corollary 4, this implies that $J \geq^\mu I_{m+1}$, so $I_\ell \geq^\mu I_{m+1}$, and therefore, for all $J' \in \mathcal{J}_i$, $I_{m+1} \geq^\mu J'$ implies $I_\ell \geq^\mu J'$. But then $C_{m+1} \subseteq C_\ell$, which contradicts the choice of C_1, \dots, C_M . Therefore, $P_\mu(I_{m+1})(\cup_{\ell=1}^m D_m) = 0$. This proves the subclaim.

Then, $P_\mu(I_{m+1})(D_{m+1}) = P_\mu(I_{m+1})(C_{m+1}) \geq P_\mu(I_{m+1})(\cup\{S_{-i}(J) : I_{m+1} =^\mu J\}) = 1$. Thus, since $p^k(\cdot|C_{m+1}) \rightarrow P_\mu(I_{m+1})$, $p^k(D_{m+1}|C_{m+1}) \rightarrow 1$ as well, and in particular $p^k(D_{m+1}) > 0$ for all large k . Moreover, for all $E \subseteq D_{m+1}$,

$$\lim_k p^k(E|D_{m+1}) = \lim_k \frac{p^k(E)}{p^k(D_{m+1})} = \lim_k \frac{p^k(E)}{p^k(C_{m+1})} \cdot \frac{p^k(C_{m+1})}{p^k(D_{m+1})} = \lim_k \frac{p^k(E)}{p^k(C_{m+1})} \cdot \lim_k \frac{p^k(C_{m+1})}{p^k(D_{m+1})} = P_\mu(I_{m+1})(E).$$

This completes the proof of the inductive step.

Finally, since $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$ for all m , and $U_i(t_i, s_{-i}) \geq U_i(s_i, s_{-i})$ for all $s_{-i} \notin \cup_m \cup_{J: I_m \geq^\mu J} S_{-i}(I_m) = \cup_m C_m = \cup_m D_m$, it follows that, for k large,

$$\begin{aligned} U_i(t_i, p^k) &= \sum_{m=1}^M p^k(D_m) U_i(t_i, p^k(\cdot|D_m)) + \sum_{s_{-i} \notin \cup_m D_m} p^k(\{s_{-i}\}) U_i(t_i, s_{-i}) > \\ &> \sum_{m=1}^M p^k(D_m) U_i(s_i, p^k(\cdot|D_m)) + \sum_{s_{-i} \notin \cup_m D_m} p^k(\{s_{-i}\}) U_i(s_i, s_{-i}) = U_i(s_i, p^k). \end{aligned}$$

Since $(p^k)_{k \geq 1}$ was an arbitrary perturbation of μ , $t_i \succ^\mu s_i$.

(Only if:) suppose that $t_i \succ^\mu s_i$. Let I_1, \dots, I_M be a representative collection for μ . I claim that $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$ for at least one $m \in \{1, \dots, M\}$. By contradiction, suppose

$U_i(t_i, P_\mu(I_m)) \leq U_i(s_i, P_\mu(I_m))$ for all m . By Lemma 4, if g is the canonical map for I_1, \dots, I_M , then $(p^k)_{k \geq 1}$ defined by $p^k = (\sum_m k^{-g(m)})^{-1} \sum_m k^{-g(m)} P_\mu(I_m)$ is a perturbation of μ . Then

$$U_i(t_i, p^k) = \frac{\sum_m k^{-g(m)} U_i(t_i, P_\mu(I_m))}{\sum_m k^{-g(m)}} \leq \frac{\sum_m k^{-g(m)} U_i(s_i, P_\mu(I_m))}{\sum_m k^{-g(m)}} = U_i(s_i, p^k),$$

for all $k \geq 1$, which contradicts the assumption that $t_i \succ^\mu s_i$.

Thus, wlog assume that $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$ for $m = 1, \dots, M' \leq M$, with $M' \geq 1$, and $U_i(t_i, P_\mu(I_m)) \leq U_i(s_i, P_\mu(I_m))$ for $m = M' + 1, \dots, M$. By contradiction, suppose that $U_i(t_i, s_{-i}) < U_i(s_i, s_{-i})$ for some $s_{-i} \notin \cup_{m=1}^{M'} \cup_{J: I_m \geq^\mu J} S_{-i}(J)$.

Intuition: I construct a perturbation of μ such, loosely speaking, that the measures $P_\mu(I_m)$ for which t_i does strictly better than s_i (so $m \in \{1, \dots, M'\}$) have infinitely less weight than s_{-i} , which in turn has infinitely less weight than the measures $P_\mu(I_m)$ for which s_i does at least as well as t_i ($m \in \{M' + 1, \dots, M\}$). This is done in two steps. First, we use Lemma 4 to construct an intermediate perturbation in which the measures corresponding to $m \in \{1, \dots, M'\}$ have infinitely less weight than those for $m \in \{M' + 1, \dots, M\}$.

As above, let g be the canonical map for I_1, \dots, I_M and define $L = 2 + \max_{n \in \{M'+1, \dots, M\}} g(n)$. Define $f : \{1, \dots, M\} \rightarrow \mathbb{R}_+$ by letting $f(n) = g(n) + L$ if $I_m \geq^\mu I_n$ for some $m \in \{1, \dots, M'\}$, and $f(n) = g(n)$ otherwise. Consider $n, n' \in \{1, \dots, M\}$ with $I_n \succ^\mu I_{n'}$. If $I_m \geq^\mu I_n$ for some $m \in \{1, \dots, M'\}$, then also $I_m \geq^\mu I_{n'}$ and so $f(n) = g(n) + L < g(n') + L = f(n')$. Otherwise, $f(n) = g(n) < g(n')$, whereas $f(n') = g(n') + L$ if $I_m \geq^\mu I_{n'}$ for some $m \in \{1, \dots, M'\}$, and $f(n') = g(n')$ otherwise. In either case, $f(n) < g(n') \leq f(n')$. Thus, f agrees with \succ^μ . By Lemma 4, letting $p^k = (\sum_m k^{-f(m)})^{-1} \sum_m k^{-f(m)} P_\mu(I_m)$ for every $k \geq 1$ yields a perturbation of μ .

Intuition: now I “insert” s_{-i} between the measures $P_\mu(I_m)$ for $m \in \{M' + 1, \dots, M\}$ and the ones for $m \in \{1, \dots, M'\}$. The key step is to verify that adding weight to s_{-i} does not influence the limits of the resulting measure conditional on each $S_{-i}(I)$.

For every $k \geq 1$, let $q^k = \frac{p^k + k^{-(L-1)} \delta_{s_{-i}}}{1 + k^{-(L-1)}}$, where $\delta_{s_{-i}}$ is the Dirac measure concentrated on s_{-i} . I claim that $(q^k)_{k \geq 1}$ is also a perturbation of μ . First, for every $I \in \mathcal{I}_i$, $q^k(S_{-i}(I)) \geq \frac{p^k(S_{-i}(I))}{1 + k^{-(L-1)}} > 0$.

Second, fix $I \in \mathcal{I}_i$ and $E \subseteq S_{-i}(I)$. Then

$$\frac{q^k(E)}{q^k(S_{-i}(I))} = \frac{\frac{p^k(E) + k^{-(L-1)}\delta_{s_{-i}}(E)}{1 + k^{-(L-1)}}}{\frac{p^k(S_{-i}(I)) + k^{-(L-1)}\delta_{s_{-i}}(S_{-i}(I))}{1 + k^{-(L-1)}}} = \frac{p^k(E) + k^{-(L-1)}\delta_{s_{-i}}(E)}{p^k(S_{-i}(I)) + k^{-(L-1)}\delta_{s_{-i}}(S_{-i}(I))}. \quad (14)$$

If $s_{-i} \notin S_{-i}(I)$, so a fortiori $s_{-i} \notin E$, then $\delta_{s_{-i}}(E) = \delta_{s_{-i}}(S_{-i}(I)) = 0$ and so $\frac{q^k(E)}{q^k(S_{-i}(I))} = \frac{p^k(E)}{p^k(S_{-i}(I))} \rightarrow \mu(E|I)$. If instead $s_{-i} \in S_{-i}(I)$, then dividing numerator and denominator of the right-hand side of Eq. (14) by $p^k(S_{-i}(I)) > 0$ yields

$$\frac{q^k(E)}{q^k(S_{-i}(I))} = \frac{\frac{p^k(E)}{p^k(S_{-i}(I))} + \frac{k^{-(L-1)}\delta_{s_{-i}}(E)}{p^k(S_{-i}(I))}}{1 + \frac{k^{-(L-1)}\delta_{s_{-i}}(S_{-i}(I))}{p^k(S_{-i}(I))}}.$$

Since $p^k(E)/p^k(S_{-i}(I)) \rightarrow \mu(E|I)$ and $\delta_{s_{-i}}(\cdot) \in [0, 1]$, it is enough to show that $\frac{k^{-(L-1)}}{p^k(S_{-i}(I))} \rightarrow 0$.

Since I_1, \dots, I_M is a representative collection, there is ℓ such that $I \stackrel{\mu}{=} I_\ell$; but by the choice of s_{-i} , not $I_m \stackrel{\mu}{\geq} I$ for all $m = 1, \dots, M'$. Hence $f(\ell) = g(\ell)$, and furthermore $\ell \in \{M' + 1, \dots, M\}$. Since $L = 2 + \max_{n \in \{M'+1, \dots, M\}} g(n)$, $f(\ell) = g(\ell) < L - 1$. Moreover, if $P_\mu(I_n)(\{s_{-i}\}) > 0$ for some $n \in \{1, \dots, M\}$, so that $P_\mu(I_n)(S_{-i}(I)) > 0$, then $I \stackrel{\mu}{\geq} I_n$ by Corollary 4. Thus, $I_\ell \stackrel{\mu}{\geq} I_n$, so

$$p^k(S_{-i}(I)) = \frac{\sum_{n=1}^M k^{-f(n)} P_\mu(I_n)(S_{-i}(I))}{\sum_n k^{-f(n)}} = \frac{\sum_{n: I_\ell \stackrel{\mu}{\geq} I_n} k^{-f(n)} P_\mu(I_n)(S_{-i}(I))}{\sum_n k^{-f(n)}} \geq \frac{k^{-f(\ell)} P_\mu(I_\ell)(S_{-i}(I))}{\sum_n k^{-f(n)}}.$$

Furthermore, $\sum_n k^{-f(n)} \rightarrow 0$ because $f(n) \geq g(n) > 0$ for all n , and since it was just shown that $f(\ell) < L - 1$, one finally obtains

$$\frac{k^{-(L-1)}}{p^k(S_{-i}(I))} \leq \frac{\sum_n k^{-f(n)}}{P_\mu(I_\ell)(S_{-i}(I))} \cdot \frac{k^{-(L-1)}}{k^{-f(\ell)}} \rightarrow 0.$$

Thus $(q^k)_{k \geq 1}$ is a perturbation of μ .

Intuition: the last step is to show that, for this perturbation t_i does strictly worse than s_i eventually. The key step is to argue that, since s_{-i} has infinitely more weight than all measures $P_\mu(I_m)$ for which $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$, eventually the fact that $U_i(t_i, s_{-i}) < U_i(s_i, s_{-i})$ will determine the sign of $U_i(t_i, q^k) - U_i(s_i, q^k)$.

Finally,

$$\begin{aligned}
U_i(t_i, q^k) - U_i(s_i, q^k) &= \frac{1}{1+k^{-(L-1)}} [U_i(t_i, p^k) - U_i(s_i, p^k)] + \frac{k^{-(L-1)}}{1+k^{-(L-1)}} [U_i(t_i, s_{-i}) - U_i(s_i, s_{-i})] = \\
&= \frac{1}{1+k^{-(L-1)}} \sum_{n=1}^{M'} \frac{k^{-f(n)}}{\sum_{n'} k^{-f(n')}} [U_i(t_i, P_\mu(I_n)) - U_i(s_i, P_\mu(I_n))] + \\
&\quad + \frac{1}{1+k^{-(L-1)}} \sum_{n=M'+1}^M \frac{k^{-f(n)}}{\sum_{n'} k^{-f(n')}} [U_i(t_i, P_\mu(I_n)) - U_i(s_i, P_\mu(I_n))] + \\
&\quad + \frac{k^{-(L-1)}}{1+k^{-(L-1)}} [U_i(t_i, s_{-i}) - U_i(s_i, s_{-i})] \leq \\
&\leq \frac{1}{1+k^{-(L-1)}} \sum_{n=1}^{M'} \frac{k^{-f(n)}}{\sum_{n'} k^{-f(n')}} [U_i(t_i, P_\mu(I_n)) - U_i(s_i, P_\mu(I_n))] + \\
&\quad + \frac{k^{-(L-1)}}{1+k^{-(L-1)}} [U_i(t_i, s_{-i}) - U_i(s_i, s_{-i})] = \\
&= \frac{k^{-(L-1)}}{1+k^{-(L-1)}} \left(\sum_{n=1}^{M'} \frac{k^{-[f(n)-(L-1)]}}{\sum_{n'} k^{-f(n')}} [U_i(t_i, P_\mu(I_n)) - U_i(s_i, P_\mu(I_n))] + [U_i(t_i, s_{-i}) - U_i(s_i, s_{-i})] \right).
\end{aligned}$$

I claim that the term in parentheses is negative for large k . By assumption, $U_i(t_i, s_{-i}) < U_i(s_i, s_{-i})$. On the other hand, for each $n = 1, \dots, M'$, $U_i(t_i, P_\mu(I_n)) - U_i(s_i, P_\mu(I_n)) > 0$, but this difference is multiplied by the weight $\frac{k^{-[f(n)-(L-1)]}}{\sum_{n'} k^{-f(n')}}$. Recall that, for $n = 1, \dots, M'$, $f(n) = g(n) + L$, and by construction $g(n) > 0$, so $f(n) - (L-1) > 1$. Furthermore, since $s_{-i} \notin \cup_{m=1}^{M'} \cup_{J: I_m \geq^\mu J} S_{-i}(J)$, $\phi \neq^\mu I_m$ for $m = 1, \dots, M'$. Hence $\phi =^\mu I_{n_\phi}$ for some $n_\phi \in \{M'+1, \dots, M\}$, so $f(n_\phi) = g(n_\phi) = \frac{n_\phi}{M+1} < 1$. Therefore,

$$\frac{k^{-[f(n)-(L-1)]}}{\sum_{n'} k^{-f(n')}} \leq \frac{k^{-[f(n)-(L-1)]}}{k^{-f(n_\phi)}} = k^{-[f(n)-(L-1)-f(n_\phi)]} \rightarrow 0$$

because the term in square brackets in the exponent is positive. This proves the claim.

Then, for k large, $U_i(t_i, q^k) < U_i(s_i, q^k)$, which contradicts $t_i \succ^\mu s_i$. Therefore, for all $s_{-i} \notin \cup_{m=1}^{M'} \cup_{J: I_m \geq^\mu J} S_{-i}(J)$, $U_i(t_i, s_{-i}) \geq U_i(s_i, s_{-i})$.

B.2 Elicitation

Throughout this section, fix a dynamic game $(N, (S_i, \mathcal{I}_i, U_i)_{i \in N}, S(\cdot))$, a questionnaire $Q = (Q_i)_{i \in N}$, and the corresponding elicitation game $(N \cup \{c\}, (S_i^*, \mathcal{I}_i^*, U_i^*)_{i \in N \cup \{c\}}, S^*(\cdot))$, according to Definition 8. It is convenient to let $N^* = N \cup \{c\}$. Also, as in part 1 of Definition 8, for every $i \in N$, let $W_i = \{\emptyset\}$ if $Q_i = \emptyset$ and $W_i = \{b, p\}$ if $Q_i = (I, E, p)$.

B.2.1 Preliminaries

I first verify that the elicitation game satisfies two properties in Section 2. This is necessary to ensure that the characterization of structural rationality in Section 5 applies.

It is immediate by inspecting Definition 8 that, for every $i \in N^*$ and $I^* \in \mathcal{I}_i^*$, $S^*(I) = S_i^*(I^*) \times S_{-i}^*(I^*)$. Second, fix $i \in N$ (so $i \neq c$) and $I^*, J^* \in \mathcal{I}_i^*$: 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 8, $S^*(I^*) = \{(s_i, w_i)\} \times S_{-i}(I) \times W_{-i} \times S_c^*$ 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 8 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 $I^* = (s_i, w_i, I) \in \mathcal{I}_i^*$, 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 8 implies that $S_i^*(I^*) = \{(s_i, w_i)\}$, a singleton set, so strategic independence holds trivially.

B.2.2 Proof of Theorem 4

Throughout this subsection, fix a player $i \in N$ and a CCPS $\mu \in \Delta(S_{-i}, \mathcal{I}_i)$.

Lemma 6 Let $\mu^* \in \Delta(S_{-i}^*, \mathcal{I}_i^*)$ agree with μ (Definition 9). Then:

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

Furthermore, for every $I^*, J^* \in \mathcal{I}_i^*$ such that $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$ and $S_{-i}^*(J^*) = S_{-i}(J) \times W_{-i} \times S_c^*$, with $I, J \in \mathcal{I}_i$,

(1) $\mu^*(S_{-i}^*(I^*)|J^*) = \mu(S_{-i}(I)|J)$,

(2) $I^* \geq^{\mu^*} J^*$ if and only if $I \geq^\mu J$, and

(3) $\cup\{S_{-i}^*(\hat{I}^*) : \hat{I}^* =^{\mu^*} I^*\} = \cup\{S_{-i}(\hat{I}) : \hat{I} =^\mu I\} \times S_{-i} \times S_c^*$.

Proof: (0): Fix $I^* \in \mathcal{I}_i^*$. 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 $I \in \mathcal{I}_i$, then $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$. Conversely, for every $I \in \mathcal{I}_i$, $s_i \in S_i(I)$ and $w_i \in W_i$, $I^* = (s_i, w_i, I) \in \mathcal{I}_i^*$ satisfies $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$.

(1): if $I^* = \phi^*$ or $I^* = I_i^1$, then $I = \phi$, so both conditional probabilities equal 1. Otherwise,

$$\begin{aligned} \mu^*(S_{-i}^*(I^*)|J^*) &= \mu^*(S_{-i}(I) \times W_{-i} \times S_c^*|J^*) = [\text{marg}_{S_{-i} \times S_c^*} \mu^*(\cdot|J^*)] (S_{-i}(I) \times S_c^*) = \\ &= \mu(S_{-i}(I)|J^*) = \mu(\text{proj}_{S_{-i}} S_{-i}^*(I^*)|J^*), \end{aligned}$$

where the second equality follows from marginalization and the third from Definition 9.

(2): suppose that $I^* \geq^{\mu^*} J^*$. Then there are $I_1^*, \dots, I_L^* \in \mathcal{I}_i^*$ such that $I_1^* = J^*$, $I_L^* = I^*$, and $\mu^*(S_{-i}^*(I_{\ell+1}^*)|I_\ell^*)$ for $\ell = 1, \dots, L-1$. By (0), $S_{-i}^*(I_\ell^*) = S_{-i}(I_\ell) \times W_{-i} \times S_c^*$ for suitable $I_1, \dots, I_L \in \mathcal{I}_i$; in particular, one can choose $I_1 = J$ and $I_L = I$. Hence (1) implies that $\mu(S_{-i}(I_{\ell+1})|I_\ell) > 0$ for $\ell = 1, \dots, L-1$, which implies that $I^* = I_L^* \geq^\mu I_1^* = J^*$.

Conversely, suppose that $I \geq^\mu J$, so there are $I_1, \dots, I_L \in \mathcal{I}_i$ such that $I_1 = J$, $I_L = I$, and $\mu(S_{-i}(I_{\ell+1})|I_\ell) > 0$ for all $\ell = 1, \dots, L-1$. 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 $\ell = 2, \dots, L-1$. Then, for all $\ell = 1, \dots, L$, $S_{-i}^*(I_\ell^*) = S_{-i}(I_\ell) \times W_{-i} \times S_c^*$, so part (1) implies that $\mu^*(S_{-i}^*(I_{\ell+1}^*)|I_\ell^*) > 0$ for all $\ell = 1, L-1$. Therefore $I^* \geq^{\mu^*} J^*$.

(3) By part (2), if $\hat{I} \in \mathcal{I}_i$ satisfies $S_{-i}^*(\hat{I}^*) = S_{-i}(\hat{I}) \times W_{-i} \times S_c^*$, then $\hat{I}^* =^{\mu^*} I^*$ iff $\hat{I} =^\mu I$. Therefore, if $I_1^*, \dots, I_L^* \in \mathcal{I}_i^*$ is an enumeration of $\{\hat{I}^* : \hat{I}^* =^{\mu^*} I^*\}$, and $I_1, \dots, I_L \in \mathcal{I}_i$ satisfy $S_{-i}^*(I_\ell^*) =$

$S_{-i}(I_\ell) \times W_{-i} \times S_c^*$ for every ℓ , then I_1, \dots, I_L is an enumeration of $\{\hat{I} : \hat{I} =^\mu I\}$.³⁰ It follows that $\cup_\ell S_{-i}^*(I_\ell^*) = [\cup_\ell S_{-i}(I_\ell)] \times W_{-i} \times S_c^*$. ■

Lemma 7 *There is a CCPS $\mu^* \in \Delta(S_{-i}^*, \mathcal{I}_i^*)$ that agrees with μ .*

Proof: Since μ is a CCPS, by Proposition 1 there is a perturbation (p^n) of μ . Fix an arbitrary element $w_{-i} \in W_{-i}$ (cf. part 1 of Definition 8) and define a sequence $(q^n) \subset \Delta(S_{-i}^*)$ by letting

$$q^n(\{(s_{-i}, w_{-i}, s_c^*)\}) = \frac{1}{2} p^n(\{s_{-i}\}) \quad \forall n \geq 1, s_{-i} \in S_{-i}, s_c^* \in S_c^*.$$

Fix $I^* \in \mathcal{I}_i^*$. By Lemma 6 part (0), there is $I \in \mathcal{I}_i$ such that $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$. 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 $s_c^* \in S_c^*$,

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

If instead $w'_{-i} \in W_{-i} \setminus \{w_{-i}\}$, then $\frac{q^n(\{(s_{-i}, w'_{-i}, s_c^*)\})}{q^n(S_{-i}^*(I^*))} = 0$. Define an array $\mu^* = (\mu^* | I^*)_{I^* \in \mathcal{I}_i^*}$ by fixing, for every $I^* \in \mathcal{I}_i^*$, an element $I \in \mathcal{I}_i$ with $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$ per Lemma 6 part (0), and letting $\mu^*(\{(s_{-i}, w'_{-i}, s_c^*)\} | I^*) = 1_{w'_{-i}=w_{-i}} \frac{1}{2} \mu(\{s_{-i}\} | I)$ for all $s_{-i} \in S_{-i}$, $w'_{-i} \in W_{-i}$ and $s_c^* \in S_c^*$. By Proposition 1, μ^* is a CCPS. ■

Lemma 8 *Let $\mu^* \in \Delta(S_{-i}^*, \mathcal{I}_i^*)$ a CCPS that agrees with μ . Fix $I^* \in \mathcal{I}_i^*$ and let $I \in \mathcal{I}_i$ be such that $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$ (cf. Lemma 6 part 0). Then, for all $s_{-i} \in S_{-i}$ and $s_c^* \in S_c^*$,*

$$P_{\mu^*}(I^*)(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}) = \frac{1}{2} P_\mu(I)(\{s_{-i}\}). \quad (15)$$

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

$$U_i^*((s_i, E), P_{\mu^*}(I^*)) = \frac{1}{2} U_i(s_i, P_\mu(I)) + \frac{1}{2} P_\mu(I)(S_{-i}(\hat{I})) \cdot \mu(E | S_{-i}(\hat{I})) \quad (16)$$

$$U_i^*((s_i, p), P_{\mu^*}(I^*)) = \frac{1}{2} U_i(s_i, P_\mu(I)) + \frac{1}{2} P_\mu(I)(S_{-i}(\hat{I})) \cdot p, \quad (17)$$

³⁰This may include some duplicates: e.g., ϕ^* and I_i^1 both map to ϕ .

whereas, if $Q_i = \emptyset$, then

$$U_i^*((s_i, *), P_{\mu^*}(I^*)) = U_i(s_i, P_{\mu}(I)). \quad (18)$$

Proof: Fix $s_{-i} \in S_{-i}$ and $s_c^* \in S_c^*$. If there is no $J \in \mathcal{J}_i$ with $J =^\mu I$ and $s_{-i} \in S_{-i}(J)$, then by Lemma 6 part (3) $[\{s_{-i}\} \times W_{-i} \times \{s_c^*\}] \cap \bigcup \{S_{-i}(J^*) : J^* =^{\mu^*} I^*\} = \emptyset$, so $P_{\mu^*}(I^*)(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}) = 0 = P_{\mu}(I)(\{s_{-i}\})$. Thus, assume that there is $J \in \mathcal{J}_i$ such that $s_{-i} \in S_{-i}(J)$ and $J =^\mu I$. By Proposition 3, $P_{\mu}(I)(S_{-i}(J)) > 0$. Finally, by Lemma 6 parts (0) and (2) there is $J^* \in \mathcal{J}_i^*$ such that $S_{-i}^*(J^*) = S_{-i}(J) \times W_{-i} \times S_c^*$ and $J^* =^{\mu^*} I^*$. Then

$$\begin{aligned} P_{\mu^*}(I^*)(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}) &= \mu^*(\{s_{-i}\} \times W_{-i} \times \{s_c^*\} | J^*) \cdot P_{\mu^*}(I^*)(S_{-i}^*(J^*)) = \\ &= \frac{1}{2} \mu(\{s_{-i}\} | J) \cdot P_{\mu^*}(I^*)(S_{-i}^*(J^*)) = \\ &= \frac{1}{2} P_{\mu}(I)(\{s_{-i}\}) \cdot \frac{P_{\mu^*}(I^*)(S_{-i}^*(J^*))}{P_{\mu}(I)(S_{-i}(J))} \equiv \frac{1}{2} P_{\mu}(I)(\{s_{-i}\}) \cdot \kappa_J; \end{aligned} \quad (19)$$

the first equality follows from the properties of $P_{\mu^*}(\cdot)$, the second from Definition 9, and the third from the properties of $P_{\mu}(\cdot)$. It must thus be shown that $\kappa_J = 1$.

By Lemma 1, there is a μ -sequence J_1, \dots, J_M such that $\{J_1, \dots, J_M\}$ is the \geq -equivalence class containing I . For every $m = 1, \dots, M-1$, $\mu(S_{-i}(J_{m+1}) | J_m) > 0$, so there is $s_{-i}^m \in S_{-i}(J_m) \cap S_{-i}(J_{m+1})$. Invoking Eq. (19) for $J = J_m$ and $J = J_{m+1}$, and adding over all $s_c^* \in \{h, t\}$,

$$P_{\mu}(I)(\{s_{-i}^m\}) \cdot \kappa_{J_m} = P_{\mu^*}(I^*)(\{s_{-i}\} \times W_{-i} \times S_c^*) = P_{\mu}(I)(\{s_{-i}^m\}) \cdot \kappa_{J_{m+1}},$$

which implies that $\kappa_m = \kappa_{m+1}$. Therefore, there is $\kappa \in \mathbb{R}$ such that $\kappa_J = \kappa$ for all $J \in \mathcal{J}_i$ with $I =^\mu J$. Then, by the properties of $P_{\mu}(\cdot)$ and $P_{\mu^*}(\cdot)$, Lemma 6 part (3), and Eq. (19), this implies

$$1 = \sum_{s_{-i} \in \bigcup \{S_{-i}(J) : I =^\mu J\}, s_c^* \in \{h, t\}} P_{\mu^*}(I^*)(\{s_{-i}\} \times W_{-i} \times \{s_c^*\}) = \sum_{s_{-i} \in \bigcup \{S_{-i}(J) : I =^\mu J\}, s_c^* \in \{h, t\}} \frac{1}{2} P_{\mu}(I)(\{s_{-i}\}) \cdot \kappa = \kappa,$$

which completes the proof of Eq. (15).

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

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

i.e., Eq. (16) holds. The other equations are proved similarly. ■

The proof of Theorem 4 can now be completed. For part (1), suppose that $t_i \succ^{\mu} s_i$. By Theorem 1, there are $I_1, \dots, I_M \in \mathcal{I}_i$ such that $U_i(t_i, P_{\mu}(I_m)) > U_i(s_i, P_{\mu}(I_m))$ for each m , and $U_i(t_i, s_{-i}) \geq U_i(s_i, s_{-i})$ for all $s_{-i} \notin \cup_m \cup_{J: I_m \geq^{\mu} J} S_{-i}(J)$. By Lemma 6 part (0), there are $I_1^*, \dots, I_M^* \in \mathcal{I}_i^*$ with $S_{-i}^*(I_m^*) = S_{-i}(I_m) \times W_{-i} \times S_c^*$ for each m , and by Eqs. (16)–17 in Lemma 8, since μ^* agrees with μ and (s_i, w_i) and (t_i, w_i) choose the same element of W_i , $U_i^*((t_i, w_i), P_{\mu^*}(I_m^*)) > U_i^*((s_i, w_i), P_{\mu^*}(I_m^*))$ for each m . Furthermore, consider $s_{-i}^* = (s_{-i}, w_{-i}, s_c^*) \notin \cup_m \cup_{J^*: I_m^* \geq^{\mu^*} J^*} S_{-i}^*(J^*)$. If $s_c^* = t$, then by part 5 of Definition 8 $U_i^*((t_i, w_i), s_{-i}^*) = U_i^*((s_i, w_i), s_{-i}^*)$. Otherwise, suppose by contradiction that $s_{-i} \in S_{-i}(J)$ for some $J \in \mathcal{I}_i$ such that $I_m \geq^{\mu} J$ for some m : by Lemma 6 part (0), there is $J^* \in \mathcal{I}_i^*$ with $S_{-i}^*(J^*) = S_{-i}(J) \times W_{-i} \times S_c^*$, so $s_{-i}^* \in S_{-i}^*(J^*)$, and by part (2) of the same Lemma $I_m^* \geq^{\mu^*} J^*$, contradiction: thus, $s_{-i} \notin \cup_m \cup_{J: I_m \geq^{\mu} J} S_{-i}(J)$, and so, by part 5 of Definition 8, $U_i^*((t_i, w_i), s_{-i}^*) = U_i(t_i, s_{-i}) \geq U_i(s_i, s_{-i}) = U_i^*((s_i, w_i), s_{-i}^*)$. Hence, by Theorem 1, $(t_i, w_i) \succ^{\mu^*} (s_i, w_i)$.

Conversely, suppose $(t_i, w_i) \succ^{\mu^*} (s_i, w_i)$ so, by Theorem 1, there are $I_1^*, \dots, I_M^* \in \mathcal{I}_i^*$ such that $U_i^*((t_i, w_i), P_{\mu^*}(I_m^*)) > U_i^*((s_i, w_i), P_{\mu^*}(I_m^*))$ for each m and $U_i^*((t_i, w_i), s_{-i}^*) \geq U_i^*((s_i, w_i), s_{-i}^*)$

for each $s_{-i}^* \notin \cup_m \cup_{J^*: I_m^* \geq \mu^* J^*} S_{-i}^*(J^*)$. By Lemma 6 part (0), there are $I_1, \dots, I_M \in \mathcal{I}_i$ such that $S_{-i}^*(I_m^*) = S_{-i}(I_m) \times W_{-i} \times S_c^*$ for each m , and by Eqs. (16)–17 in Lemma 8, since μ^* agrees with μ and (s_i, w_i) and (t_i, w_i) choose the same element of W_i , $U_i(t_i, P_\mu(I_m)) > U_i(s_i, P_\mu(I_m))$ for each m . Furthermore, consider $s_{-i} \notin \cup_m \cup_{J: I_m \geq \mu J} S_{-i}(J)$. Fix $(w_{-i}, s_c^*) \in W_{-i} \times S_c^*$ arbitrarily and let $s_{-i}^* = (s_{-i}, w_{-i}, s_c^*)$. By contradiction, if there is $J^* \in \mathcal{I}_i^*$ such that $s_{-i}^* \in S_{-i}^*(J^*)$ and $I_m^* \geq \mu^* J^*$ for some m , then by Lemma 6 parts (0) and (2) there is $J \in \mathcal{I}_i$ such that $S_{-i}^*(J^*) = S_{-i}(J) \times W_{-i} \times S_c^*$, so $s_{-i} \in S_{-i}(J)$, and $I_m \geq \mu J$, contradiction. Thus, $s_{-i}^* \notin \cup_m \cup_{J^*: I_m^* \geq \mu^* J^*} S_{-i}^*(J^*)$, which together with part 5 of Definition 8 implies that $U_i(t_i, s_{-i}) = U_i^*((t_i, w_i), s_{-i}^*) \geq U_i^*((s_i, w_i), s_{-i}^*) = U_i(s_i, s_{-i})$. By Theorem 1, $t_i \succ^\mu s_i$, and the proof of part (1) is complete.

For part (2), let $Q_i = (I, E, p)$, and fix $s_i \in S_i$. Suppose that $p > \mu(E|I)$. Let $I^* \in \mathcal{I}_i^*$ be such that $S_{-i}^*(I^*) = S_{-i}(I) \times W_{-i} \times S_c^*$, which exists by part (0) of Lemma 6. By Eqs. (16) and (17), and the fact that $P_\mu(I)(S_{-i}(I)) > 0$ by the properties of $P_\mu(\cdot)$, $U_i^*((s_i, b), P_{\mu^*}(I^*)) < U_i^*((s_i, p), P_{\mu^*}(I^*))$.

Now fix $s_{-i}^* = (s_{-i}, w_{-i}, s_c^*) \in S_{-i}^*$ and suppose that $U_i^*((s_i, p), s_{-i}^*) < U_i^*((s_i, b), s_{-i}^*)$. By part 5 of Definition 8, it must be that $s_c^* = t$ and $s_{-i} \in E$. (If $p = 1$ or $E = \emptyset$, there can be no such s_{-i}^* , and the proof is complete.) But then $s_{-i}^* \in S_{-i}(I^*)$. A fortiori, for all $s_{-i}^* \notin \cup_{J^*: I_m^* \geq \mu^* J^*} S_{-i}^*(J^*)$, $U_i^*((s_i, p), s_{-i}^*) \geq U_i^*((s_i, b), s_{-i}^*)$, and Theorem 1, with $M = 1$ and $I_1^* = I^*$, implies $(s_i, p) \succ^{\mu^*} (s_i, b)$.

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, E) 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), $(t_i, E) \succ^{\mu^*} (s_i, E)$: contradiction. Thus, s_i is structurally rational in the original game. Furthermore, suppose that $\mu(E|I) < p$: then (2) implies that $(s_i, p) \succ^{\mu^*} (s_i, E)$, contradiction. Thus, $\mu(E|I) \geq p$. The case of (s_i, p) structurally rational is analogous, so the proof is omitted. ■

References

- R.J. Aumann and J.H. Dreze. Assessing strategic risk. *American Economic Journal: Microeconomics*, 1(1):1–16, 2009.
- P. Battigalli. On rationalizability in extensive games. *Journal of Economic Theory*, 74(1):40–61, 1997.
- P. Battigalli and M. Siniscalchi. Strong Belief and Forward Induction Reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002.
- 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.
- Truman Bewley. Knightian decision theory: Part I. *Decisions in Economics and Finance*, 25(2): 79–110, November 2002. (first version 1986).
- 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.
- Bruno De Finetti. *Theory of probability: A critical introductory treatment*, volume 6. John Wiley & Sons, 2017.
- David Dillenberger. Preferences for one-shot resolution of uncertainty and allais-type behavior. *Econometrica*, 78(6):1973–2004, 2010.
- Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75: 643–669, 1961.
- 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.
- Peter J Hammond. Non-archimedean subjective probabilities in decision theory and games. *Mathematical Social Sciences*, 38(2):139–156, 1999.

- 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.
- Elon Kohlberg and Philip J Reny. Independence on relative probability spaces and consistent assessments in game trees. *Journal of Economic Theory*, 75(2):280–313, 1997.
- 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.
- 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. ISSN 0236-5294.
- 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 sequential preferences. mimeo, Northwestern University, 2020.
- Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2):146–160, 1972.
- 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.