Structural Rationality in Dynamic Games

Marciano Siniscalchi*

May 3, 2016

Abstract

The analysis of dynamic games hinges on assumptions about players' actions and beliefs at information sets that are not actually reached during game play, and that players themselves do not expect to reach. However, it is not obvious how to elicit intended actions and conditional beliefs at such information sets. Hence, key concepts such as sequential rationality, backward induction, and forward induction do not readily translate to testable behavioral assumptions. This paper addresses this concern by introducing a novel optimality criterion, *structural rationality*. In any dynamic game, structural rationality implies sequential rationality. In addition, if players are structurally rational, their intended actions and conditional beliefs can be elicited via the *strategy method* (Selten, 1967). Finally, structural rationality is consistent with experimental evidence indicating that subjects behave differently in the strategic and extensive form, but take the extensive form into account even if they are asked to commit to strategies ahead of time.

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

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

1 Introduction

The analysis of simultaneous-move games is grounded in single-person choice theory. Players are assumed to maximize their expected utility (EU)—a decision rule characterized by well-known, testable properties of observed choices (Savage, 1954; Anscombe and Aumann, 1963). Furthermore, beliefs can be elicited via incentive-compatible "side bets" whose outcomes depend upon the strategies of coplayers (Luce and Raiffa, 1957, §13.6).[1] Hence, it is possible to intepret assumptions about players' rationality and beliefs in simultaneous-move games as testable restrictions on behavior.

On the other hand, the textbook treatment of dynamic games involves assumptions that are intrinsically difficult, if not impossible, to translate into testable behavioral restrictions. The prevalent notion of best response for dynamic games is sequential rationality (Kreps and Wilson, 1982). Each player is assumed to hold well-defined conditional beliefs at every one of her information sets—including those she does not expect to reach. A strategy is sequentially rational if it maximizes the player's conditional expected payoff at every information set.[2] The key difficulty is how to elicit a player's conditional beliefs, and the action she would take, at information sets that she does not expect to reach, and that indeed are not reached in the observed play of the game. If the assumed optimality criterion is sequential rationality, such beliefs and actions are neither directly observable, nor indirectly elicitable in an incentive-compatible way from observed choices. Hence, one cannot verify whether a player is indeed sequentially rational. A fortiori, key assumptions such as backward or forward induction cannot be tested, because by definition they impose restrictions on beliefs and actions at un-

[1] The experimental literature illustrates how to implement side-bets in practice: see e.g. Van Huyck, Battalio, and Beil (1990); Nyarko and Schotter (2002); Costa-Gomes and Weizsäcker (2008); Rey-Biel (2009). See also Aumann and Dreze, 2009 and Gilboa and Schmeidler, 2003.

[2] I abstract from differences in the representation of conditional beliefs, and/or in the optimality requirement, that are inessential to the present argument.

reached information sets. Section 2 illustrates these points with an example.

This paper proposes to address this fundamental methodological concern by taking a cue from two experimental findings that appear to be contradictory from the perspective of standard game-theoretic analysis. On one hand, subjects appear to behave differently in a dynamic game and in the associated strategic (i.e., matrix) form (Cooper, DeJong, Forsythe, and Ross, 1993; Schotter, Weigelt, and Wilson, 1994; Cooper and Van Huyck, 2003; Huck and Müller, 2005). On the other hand, in a broad meta-analysis of dynamic-game experiments, Brandts and Charness (2011) report that qualitatively similar findings are obtained when subjects play a dynamic game directly, and when they are required to simultaneously commit to an extensive-form strategy—a protocol known as the *strategy method* (Selten, 1967); see also Fischbacher, Gächter, and Quercia (2012). These findings cannot be reconciled with standard notions of rationality.[3] Instead, they suggest that subjects may follow a different rationality criterion—one that can potentially address the methodological concerns that motivate this project. After all, subjects in the noted experiments take the extensive form into account, and yet their strategies (and, potentially, their beliefs) can be elicited.

The main contribution of this paper is to identify a criterion, *structural rationality*, that exhibits these features. This notion evaluates strategies from the ex-ante perspective, but takes a player's conditional beliefs into account. Theorem 1 shows that, if a strategy is structurally rational given a player's conditional beliefs, then it is also sequentially rational for the same beliefs. Theorem 2 then shows that, if players are structurally rational, a version of the strategy method can be used to elicit their conditional beliefs and planned strategies in an incentive-compatible way. Theorem 3 identifies a crucial consistency property of conditional beliefs

[3]When players commit to extensive-form strategies ex-ante, sequential rationality yields the same predictions as ex-ante payoff maximization in the strategic form, and hence weaker predictions than in the original dynamic game. Alternatively, the invariance hypothesis (Kohlberg and Mertens, 1986) predicts that behavior should be the same in all three presentations of the game. Thus, neither textbook analysis based on sequential rationality, nor theories based on invariance, can accommodate the noted evidence.

that is required for structural rationality to be well-defined. A companion paper, Siniscalchi (2016a), provides axiomatic foundations for structural rationality; in-progress work (Siniscalchi, 2016b) demonstrates how structural rationality can be incorporated in solution concepts. Taken together, these results offer an approach that builds upon the received theory of dynamic games, but places it upon solid choice-theoretic foundations.

While the motivation for this paper is mainly methodological, structural rationality is more closely aligned with the evidence from experiments than the received theory. In particular, it provides a theoretical rationale for the strategy method

The remainder of this paper is organized as follows. Section 3 presents the setup. Section 4 formalizes structural preferences and structural rationality. Section 5 relates structural and sequential rationality, and Section 6 formalizes the elicitation result. Section 7 analyzes the consistency of conditional beliefs. Section 8 comments on the results and discusses the related literature. All proofs are in the Appendix.

2 Heuristics: sequential rationality is not testable

To illustrate the difficulties inherent in eliciting beliefs and verifying sequential rationality, consider the "burning money" game of Ben-Porath and Dekel (1992), depicted in Figure 1.

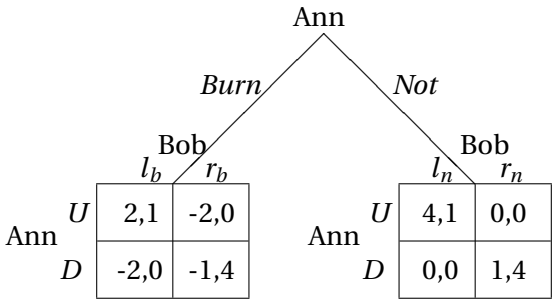

Figure 1: Burning Money

In any given play of the game, only one of the two simultaneous-moves subgames will be reached. Following Ann's initial choice, an experimenter can offer side bets on the actions in the subgame that is actually reached, and thus elicit players' conditional beliefs in that subgame. But how about their conditional beliefs in the other subgame? The experimenter might offer "conditional" side bets at the beginning of the game. For instance, before Ann makes her initial choice, the experimenter might offer Ann (resp. Bob) a bet on l_b vs. r_b (resp. U vs. D) following *Burn*, with the understanding that the bet will be called off if Ann chooses *Not*. However, two difficulties arise. First, Ann's own choice of *Burn* vs. *Not* determines which of the two subgames will be reached. If, for instance, she decides to choose *Not*, then she effectively causes the conditional side bet on Bob's actions in the subgame on the left to be called off—so whether she accepts or rejects such a bet conveys no information about her beliefs. Second, suppose that Bob initially assigns zero probability to Ann's choice of *Burn*. Then, at the beginning of the game, Bob expects the conditional side bet on U vs. D following *Burn* to be called off; therefore, again, whether he accepts or rejects such a bet conveys no information about the conditional beliefs he would hold, were Ann to unexpectedly choose *Burn*.

Thus, neither Ann's nor Bob's beliefs can be fully elicited via side bets. In addition, their strategies are not fully observable, as only one of the two proper subgames will be reached in a given play. As a consequence, in the game of Figure 1 the choices and beliefs that are *actually* observed and elicited may fail to provide evidence either for or against sequential rationality.[4] Furthermore, it is not possible to verify whether actions and beliefs off the observed path of play satisfy properties of interest—for instance, whether they are consistent with backward- or forward-induction reasoning.

[4]Suppose that Ann chooses *Not* followed by D. Suppose further that Ann believes that Bob will choose r_n following *Not*, and that this is elicited via suitable side bets. It may be that Ann additionally believes that Bob would have chosen l_b following *Burn*, in which case Ann's choice of *Not* would not be sequentially rational. Alternatively, it may be that Ann believes that Bob would have chosen r_b in the subgame on the left, in which case *Not* is indeed sequentially rational.

To summarize, by definition, only choices or side bets made by players simultaneously at the beginning of the game are guaranteed to be observable, regardless of how play continues. However, sequential rationality only requires that these choices maximize ex-ante expected payoffs. Hence, under sequential rationality, such initial choices can convey no information about a player's intended actions or conditional beliefs at information sets that the player does not expect to reach, or causes not to reach.[5]

3 Setup

This paper considers dynamic games with imperfect information, defined essentially as in Osborne and Rubinstein (1994, Def. 200.1, pp. 200-201; OR henceforth). This section only introduces notation and definitions that are essential to state the main results of this paper.

An **extensive game form** is a tuple $\Gamma = (N, H, P, (\mathscr{I}_i)_{i \in N})$, where N is the set of players, H is a finite collection of histories, i.e., finite sequences $(a_1, \ldots, a_n)$ of actions drawn from some set A and containing the empty sequence ϕ, P is the player function, which associates with each history h the player on the move at h, and each $\mathscr{I}_i$ is a partition of the histories where Player i moves; the elements of $\mathscr{I}_i$ are player i's information sets. Information sets are ordered by a precedence relation, denoted "$<$." The game form is assumed to have **perfect recall**, as per Def. 203.3 in OR. For simplicity, chance moves are omitted.

Given an extensive game form, certain derived objects of interest, including strategies, can be defined. A history is **terminal** if it is not the initial segment of any other history; the set of terminal histories is denoted Z. The set of **actions available at an information set** $I \in \mathscr{I}_i$ is denoted $A(I)$. For every player $i \in N$, a **strategy** is a map $s_i : \mathscr{I}_i \to A$ such that $s_i(I) \in A(I)$ for

[5] One might consider adding "trembles" to the game, so that, for instance, when Ann chooses *Burn*, there is a small probability that *Not* will be played instead (and Bob cannot tell whether the move was intentional or not). This does eliminate zero-probability information sets, but also defeats the purpose of the elicitation—testing assumptions about what players would do (and believe) following *unexpected* moves of their opponents.

all $I \in \mathscr{I}$; the set of strategies for i is denoted S_i, and the usual conventions for product sets are employed. The **terminal history induced by strategy profile** $s \in S$ is denoted $\zeta(s)$.

For every information set $I \in \mathscr{I}_i$, $S(I)$ is the set of **strategy profiles that reach** I; its projections on S_i and S_{-i} respectively are denoted $S_i(I)$ and $S_{-i}(I)$; perfect recall implies that $S(I) = S_i(I) \times S_{-i}(I)$. For histories $h \in H$, the notation $S(h)$ has a similar interpretation. It is also useful to define **player i's information sets $\mathscr{I}_i(s_i)$ allowed by strategy** s_i: that is, for every $I \in \mathscr{I}_i$, $I \in \mathscr{I}_i(s_i)$ if and only if $s_i \in S_i(I)$.

A dynamic game adds to the extensive game form a specification of the material (i.e., physical or monetary) consequences for each player at every terminal history; it is also useful to allow for exogenous uncertainty. Formally, fix a set X of **outcomes**, and a (finite or infinite) set Θ, endowed with a sigma-algebra $\mathscr{T}$. For each player i, define the **outcome function** $\xi_i : Z \times \Theta \to X$. When terminal history z is reached and the realization the exogenous uncertainty is $\theta \in \Theta$, player i's material outcome is $\xi_i(z, \theta)$. A **dyamic game** is then a tuple $(\Gamma, X, \Theta, \mathscr{T}, (\xi_i)_{i \in N})$, where Γ is an extensive game form with player set N, Θ characterizes payoff uncertainty, and ξ_i is i's outcome function.

Each player is characterized by a **utility function** $u_i : X \to \mathbb{R}$, and conditional beliefs, defined below. The payoff of player i at terminal history z, when the exogenous uncertainty is θ, is then $u_i(\xi_i(z, \theta))$.[6]

4 Conditional Beliefs and Structural Preferences

4.1 Conditional Probability Systems

At any point in the game, the domain of player i's uncertainty comprises the strategies of her coplayers, as well as the exogenous uncertainty; let $\Omega_i = S_{-i} \times \Theta$, and endow this set with the

[6]The utility function u_i does *not* depend upon $\theta \in \Theta$; as shall be seen momentarily, neither do i's conditional beliefs. As discussed in Section 8, this reflects the *interim* perspective in games of incomplete information.

product sigma-algebra $\Sigma_i = 2^{S_{-i}} \times \mathscr{T}$.

Player i's beliefs at an information set I are conditional upon the information she receives at I regarding the play of others. Since beliefs are defined over $\Omega_i = S_{-i} \times \Theta$, this information is formalized as a subset of Ω_i as well. Upon reaching I, Player i can rule out strategies of her coplayers that do not allow I. Thus, for each $I \in \mathscr{I}_i$, the **conditioning event** corresponding to information set I is

$$[I] = S_{-i}(I) \times \Theta; \tag{1}$$

the collection of all conditioning events for player i is then

$$\mathscr{F}_i = \{\Omega_i\} \cup \{[I] : I \in \mathscr{I}_i\}. \tag{2}$$

Observe that Ω_i is always a conditioning event, even if there is no information set $I \in \mathscr{I}_i$ such that $S_{-i}(I) \times \Theta = \Omega_i$. This is convenient (though not essential) to relate structural preferences to ex-ante expected-utility maximization.

Finally, for a measurable space $(Y, \mathscr{Y})$, $pr(\mathscr{Y})$ denotes the set of probability measures on $(Y, \mathscr{Y})$.[7] Conditional beliefs can now be formally defined.

Definition 1 (**Rényi (1955); Ben-Porath (1997); Battigalli and Siniscalchi (1999, 2002)**) *A **conditional probability system** (CPS) for player i in the dynamic game $(\Gamma, X, \Theta, \mathscr{T}, (\xi_i)_{i \in N})$ is a collection $\mu_i \equiv \left(\mu_i(\cdot|F)\right)_{F \in \mathscr{F}_i}$ such that:*

(1) for every $F \in \mathscr{F}_i$, $\mu_i(\cdot|F) \in pr(\Sigma_i)$ and $\mu_i(F|F) = 1$;

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

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

The characterizing feature of a CPS is the assumption that the chain rule of conditioning, Equation 3, holds even conditional upon events that have zero ex-ante probability.

[7] Recall that, while S_{-i} is finite, the set Θ, and hence the state space Ω_i, need not be.

The set of CPS for player i is denoted by $cpr(\Sigma_i, \mathscr{F}_i)$. For any probability measure $\pi \in pr(\Sigma_i)$ and measurable function $a : \Omega_i \to \mathbb{R}$, let $\mathrm{E}_\pi[a] = \int_{\Omega_i} a \, d\pi$; when no confusion can arise, I will sometimes omit the square brackets.

4.2 Structural Preferences

The key notion of structural preferences can now be formalized. I proceed in three steps. First, I observe that a CPS μ of player i induces an ordering over information sets (more precisely, the corresponding conditioning events) that refines the precedence ordering given by the extensive form of the game. Second, I note that any "consistent" CPS μ also uniquely identifies a collection of probabilities, interpreted as alternative ex-ante beliefs that generate μ by conditioning. (The exact formulation of this statement is Theorem 3 in Section 7). Third, and finally, I define structural preferences in terms of these ex-ante beliefs. I then illustrate the definition by means of examples, and conclude with heuristics that motivate the proposed definition. Throughout this subsection, fix a dynamic game $(\Gamma, X, \Theta, \mathscr{T}, (\xi_i)_{i \in N})$.

A preorder over conditioning events. Fix two information sets $I, J \in \mathscr{I}_i$. If $I < J$, then it is easy to see that, by perfect recall, $S_{-i}(I) \supseteq S_{-i}(J)$. If μ is a CPS for i, it must be the case that

$$\mu([I]|[J]) \geq \mu([J]|[J]) = 1 > 0.$$

One might say that, if J is reached, then it is *at least as plausible* that I is also reached—indeed, in this case, I *must* be reached if J is. This intuition generalizes. For information sets $I, J \in \mathscr{F}_i$ such that $\mu([I]|[J]) > 0$, one may say that reaching I is at least as plausible as reaching J: at J, player i believes that at least some of the strategies that her coplayers are following allow I as well. This intuition generalizes further, by appealing to transitivity.

Definition 2 *Fix a CPS μ on $(\Sigma, \mathscr{F}_i)$. For all $D, E \in \mathscr{F}_i$, D* ***is at least as plausible as*** *E ($D \triangleright E$) if there are $F_1, \ldots, F_N \in \mathscr{E}$ such that $F_1 = E$, $F_N = D$, and for all $n = 1, \ldots, N-1$, $\mu(F_{n+1}|F_n) > 0$.*

By construction, the plausibility relation $\triangleright$ is a preorder (i.e., it is reflexive and transitive). However, it is not complete: an example is given below (see the discussion following Eq. 5).

A collection of alternative prior beliefs. Consider the game in Figure 2; Bob's payoffs are omitted as they are not relevant to the discussion. Ann and Bob choose an action simultaneously. If Bob chooses o, the game ends. Otherwise, Ann's action determines what she learns about Bob's choice.

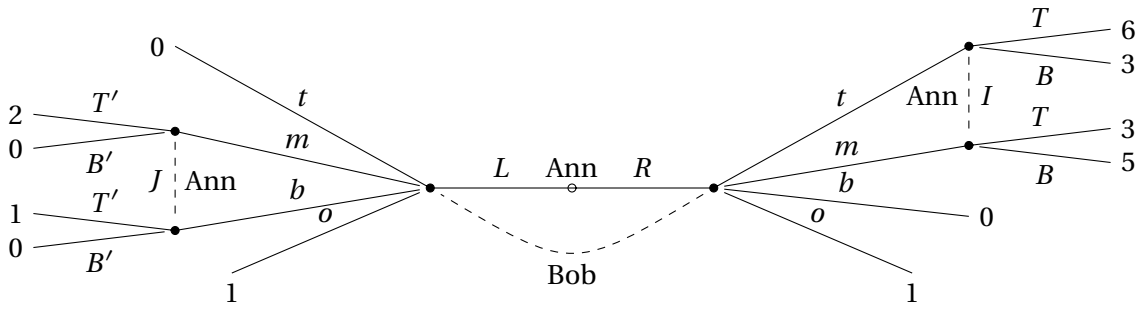

Figure 2: Alternative theories and plausibility

Assume that there is no payoff uncertainty, so $\Omega_{\text{Ann}} = S_{\text{Bob}}$. Ann's CPS μ is given by

$$\mu(\{o\}|[\phi]) = 1, \quad \mu(\{t\}|[I]) = \mu(\{m\}|[I]) = \mu(\{m\}|[J]) = \mu(\{b\}|[J]) = \frac{1}{2}. \tag{4}$$

Ann's CPS does not directly convey any information about the relative likelihood of t and b: ex ante, both actions have probability zero, and there is no further conditioning event thta contains both. However, I suggest that, indirectly, μ does pin down their relative likelihood: since $\mu(\{t\}|[I]) = \mu(\{t\}|\{t,m\}) = \frac{1}{2}$, Ann deems t and m equally likely, conditional on Bob not choosing o; similarly, $\mu(\{m\}|[J]) = \mu(\{m\}|\{m,b\}) = \frac{1}{2}$, Ann deems m and b equally likely, again conditional on Bob not choosing o. This suggest that, conditional upon Bob not choosing o, Ann deems t and b equally likely.

Even more can be said. Given Ann's CPS μ, her conditioning events are ranked as follows in terms of plausibility (Def. 2): $S_{\text{Bob}} \triangleright [I]$, $S_{\text{Bob}} \triangleright [J]$, $[I] \triangleright [J]$ and $[J] \triangleright [I]$. It is *not* the case that $[I] \triangleright S_{\text{Bob}}$

or $[I] \triangleright S_{\text{Bob}}$. Thus, S_{Bob} is strictly more plausible than $[I]$ and $[J]$, which are equally plausible. The distribution p on S_{Bob} with $p(\{t\}) = p(\{m\}) = p(\{b\} = \frac{1}{3}$ is the unique probability that (i) generates Ann's beliefs given $[I]$ and $[J]$ by conditioning, and (ii) assigns probability one to $[I] \cup [J]$, a union of equally plausible events.

This suggests that Ann's CPS μ conveys the following information. Ann entertains two *alternative prior hypotheses* about Bob's play. One is that Bob will choose o for sure; the other is that Bob is equally likely to choose t, m, b, but does not choose o. Furthermore, the first hypothesis is the more plausible one. At any information set K, Ann's beliefs are obtained by updating the most plausible belief that assigns positive probability to K. This interpetation is in the spirit of *structural consistency* (Kreps and Wilson, 1982; Kreps and Ramey, 1987); see Section 8 for further discussion. The following definition formalizes it.

Definition 3 *Fix a CPS μ on $(\Sigma_i, \mathscr{F}_i)$. A **basis** for μ is a collection $(p_C)_{C \in \mathscr{F}_i} \subset pr(\Sigma_i)$ such that*

(1) for every $C, D \in \mathscr{F}_i$, $p_C = p_D$ if and only if both $C \triangleright D$ and $D \triangleright C$;

(2) for every $C \in \mathscr{F}_i$, $p_C(\cup\{D \in \mathscr{F}_i : C \triangleright D, D \triangleright C\}) = 1$;

(3) for every $C \in \mathscr{F}_i$, $p_C(C) > 0$ and, for every $E \in \Sigma$, $\mu(E \cap C|C) = \frac{p_C(E \cap C)}{p_C(C)}$.

As was just argued, the basis of the CPS μ in Eq. (4) comprises of the prior belief $\mu(\cdot|S_{\text{Bob}})$ and another probability that is not also an element of μ. For other CPSs, all basis elements are also elements of the CPS: for instance, consider the CPS ν for Ann defined by

$$\nu(\{o\}|S_{\text{Bob}}) = \nu(\{t\}|[I]) = \nu(\{b\}|J]) = 1. \tag{5}$$

For the CPS ν, S_{Bob} is strictly more plausible than $[I]$ and $[J]$, and the latter two events are not comparable for the plausibility relation. Definition 3 implies that the basis for ν consists of the measures $\nu(\cdot|S_{\text{Bob}})$, $\nu(\cdot|[I])$, and $\nu(\cdot|[J])$.

Not all CPSs admit a basis. However, I show in Section 7 that those that do not have the "pathological" feature that a player can, essentially, choose her own future beliefs. Formally, I identify a property of CPSs, *consistency*, that ensures the existence and uniqueness of a basis.

Structural preferences over acts. It is finally possible to formalize the notion of *structural preferences*. For the purposes of establishing the connection with sequential rationality, it would be enough to define a preference ranking over strategies. However, the elicitation of beliefs requires comparisons of bets, as well as conditional bets, over arbitrary events. For this reason, I define preferences over the collection of **acts** à la Savage (1954), i.e., simple Σ_i-measurable functions $f : \Omega_i \to X$. The set of all acts for player i is denoted $\mathscr{A}_i$. Given the dynamic game $(\Gamma, X, \Theta, (\xi_i)_{i\in N})$, every strategy $s_i \in S_i$, together with the outcome function ξ_i, determines an act f^{s_i}, defined by $f^{s_i}(s_{-i}, \theta) = \xi_i(\zeta(s_i, s_{-i}), \theta)$ for all $(s_{-i}, \theta) \in \Omega_i$. Thus, a preference over acts induces a preference over strategies; however, there are acts that do not correspond to strategies.

Definition 4 *Fix a dynamic game $(\Gamma, X, \Theta, (\xi_i)_{i\in N})$, a player $i \in N$, a utility function $u_i : X \to \mathbb{R}$, and a CPS μ for i that admits a basis $\mathbf{p} = (p_F)_{F\in\mathscr{F}_i}$. For any pair of acts $f, g \in \mathscr{A}_i$, f is (weakly)* ***structurally preferred*** *to g given u_i and $\mathbf{p}$, written $f \succcurlyeq^{u_i,\mu} g$, iff for every $F \in \mathscr{F}_i$ such that $\mathrm{E}_{p_F} u_i \circ f < \mathrm{E}_{p_F} u_i \circ g$, there is $G \in \mathscr{F}_i$ such that $G \triangleright F$ and $\mathrm{E}_{p_G} u_i \circ f > \mathrm{E}_{p_G} u_i \circ g$.*

Strict preference ($\succ^{u_i,\mu}$) is defined as usual: $f \succ^{u_i,\mu} g$ iff $f \succcurlyeq^{u_i,\mu} g$ and not $g \succcurlyeq^{u_i,\mu} f$. Structural preferences are reflexive and transitive: see Siniscalchi (2016a), Appendix B.

Structural preferences reduce to EU in simultaneous-move games: in this case, $\mathscr{F}_i = \{\Omega\}$, so a CPS and, therefore, its basis, is a single probability measure. Hence, as stated in the Introduction, in general structural preferences predict different behavior in a dynamic game and in its associated matrix form. Furthermore, for any player i, structural preferences also reduce to EU if every conditioning event has positive prior probability (that is, $\mu(E|\Omega) > 0$ for every $E \in \mathscr{F}_i$). This is because, in this case, $E \triangleright \Omega$ and $\Omega \triangleright E$ both hold, so if $(p_E)_{E\in\mathscr{F}_i}$ is a basis for player i's CPS μ, then $p_E = p_\Omega$ for all $E \in \mathscr{F}_i$. Thus, structural preferences only differ from EU if one or more information sets is not expected to be reached.

The definition of structural preferences is reminiscent of that of lexicographic preferences (Blume, Brandenburger, and Dekel, 1991a). The crucial difference is that both the probabili-

ties involved in the definition (the basis), and their ordering (the plausibility ranking) are not exogenously given, but rather derived from the player's CPS, as per Definitions 2 and 3. I elaborate on this point in Section 8.

Examples. Definition 4 characterizes an *ex-ante* ranking, before the player has observed any moves made by others.[8] However, as noted in the Introduction, its definition utilizes the entire CPS of the player, via its basis. To see how the definition is applied in a non-trivial example, consider the game of Figure 2; here and throughout all examples in this paper, interpret the numbers attached to terminal nodes as monetary payoffs, and assume utilty is linear.[9] Table I below summarizes the expected payoffs given the two basis probabilities for the CPS μ in Eq. (4): the prior, $\mu(\cdot|S_{\text{Bob}})$, and the probability p that assigns equal weight to t, m, b.[10]

s_i	f^{s_i}	EU for $\mu(\cdot\|S_{\text{Bob}})$	EU for p
RTT', RTB'	$(6,3,0,1)$	1	3
RBT', RBB'	$(3,5,0,1)$	1	$\frac{8}{3}$
LTT', LBT'	$(0,2,1,1)$	1	$\frac{1}{3}$
LTB', LBB'	$(0,0,0,1)$	1	0

Table I: Expected payoffs for the strategies in Fig. 2 and the CPS μ

Applying Definition 4, the strategies RTT', RTB' are structurally strictly preferred to any other strategy: while all strategies yield the same ex-ante expected payoff of 1, choosing R followed by T secures the highest expected payoff for the basis probability p. Observe that any two *realization-equivalent* strategies induce the same act, and so a player with structural

[8] The companion paper Siniscalchi (2016a) shows that Savage's usual update rule for preferences (Savage, 1954) defines dynamically consistent conditional structural preferences.

[9] Equivalent, assume they represent utility levels: this is immaterial to the discussion.

[10] The second column indicates the act f^{s_i} associated with strategy s_i, listing the states in the order t, m, b, o: that is, it displays the vector $\left(f^{s_i}(t), f^{s_i}(m), f^{s_i}(b), f^{s_i}(o)\right)$.

preferences will be indifferent between them.[11] I return to this point in the next Section, where sequential rationality is defined.

Now consider the CPS ν in Equation (5). Table II repeats the calculations for the corresponding basis, which—as noted above–consists of the elements of ν.

s_i	f^{s_i}	EU for $\nu(\cdot\|S_{\text{Bob}})$	EU for $\nu(\cdot\|[I])$	EU for $\nu(\cdot\|[J])$
RTT', RTB'	$(6,3,0,1)$	1	6	0
RBT', RBB'	$(3,5,0,1)$	1	3	0
LTT', LBT'	$(0,2,1,1)$	1	0	1
LTB', LBB'	$(0,0,0,1)$	1	0	0

Table II: Expected payoffs for the strategies in Fig. 2 and the CPS ν

With these beliefs, strategy RTT' delivers the highest expected payoff given $\nu(\cdot|[I])$, but LTT' yields a strictly higher expected payoff given $\nu(\cdot|[J])$. These strategies are thus *not* ranked by the structural preferences induced by ν. Thus, structural preferences can be incomplete. This is a consequence of the fact that the CPS ν does not rank $[I]$ and $[J]$ in terms of their plausibility: one cannot say that one is "infinitely more likely" than the other. Yet, strategies RTT' and LTT' are maximal—no other strategy is structurally preferred to either of them. They are also sequentially rational. By way of contrast, for instance, RBT' is strictly worse than RTT'.

A caveat: basis probabilities and conditional probabilities. Structural preferences are defined using the basis of a CPS, rather than the CPS itself. This is essential to ensure that structural rationality implies sequential rationality. Suppose one instead defines a binary relation $\succcurlyeq^{\dagger}$ as follows: given acts $f, g \in \mathscr{A}_i$ and a CPS $\bar{\mu}$,

> $f \succcurlyeq^{\dagger} g$ if, for every $F \in \mathscr{F}_i$ such that $\mathrm{E}_{\bar{\mu}(\cdot|F)} u_i \circ f < \mathrm{E}_{\bar{\mu}(\cdot|F)} u_i \circ g$, there is $G \in \mathscr{F}_i$ such that $G \triangleright F$ and $\mathrm{E}_{\bar{\mu}(\cdot|F)} u_i \circ f > \mathrm{E}_{\bar{\mu}(\cdot|F)} u_i \circ g$.

[11] The same is true for any other ranking of preferences in terms of induced acts—including expected-payoff maximization and lexicographic preferences.

Now consider Ann's strategy RBT' and the CPS μ in Eq. (4). The conditional expected payoff of RBT' given $\mu(\cdot|[J])$ is 2.5, which is strictly higher than that of any other strategy, including RTT'. The only conditioning event that is more plausible than $[J]$ is S_b, and ex-ante all strategies yield 1. Hence, RBT' is maximal for the relation $\succcurlyeq^\dagger$. However, it is not sequentially rational for μ.

The reason for this undesirable conclusion is that the definition of $\succcurlyeq^\dagger$ leads one to compare the expected payoff of strategies RTT' and RBT' conditional upon the event, $[J]=\{m,b\}$, even though the information set J is *not* allowed by either strategy. By using basis probabilities, Definition 4 avoids this. The fact that RTT' and RBT' allow I but not J imply that these strategies yield the same payoff on $\{b\}$ (cf. Mailath, Samuelson, and Swinkels, 1990). The basis probabiility associated with both I and J has support $[I]\cup[J]=\{t,m,b\}$; since both strategies yield 0 in state b, relative to this probability RTT' and RBT' are effectively ranked as a function of their expected payoff given $[I]=\{t,m\}$. Since all strategies yield 1 given Ann's prior belief, this ensures that a maximal strategy must make an optimal choice at I; this rules out RBT'.

5 Sequential Rationality

This section relates structural and sequential rationality. Throughout, fix a dynamic game $(\Gamma, X, \Theta, (\xi_i)_{i\in N}$. For each player i, fix a CPS μ_i that admits a basis $\mathbf{p}_i=(p_{i,F})_{F\in\mathcal{F}_i}$, and a utility function $u_i : X \to \mathbb{R}$. Also, for every $i\in N$, let $\rhd_i$ be the plausibility relation induced by μ_i. In this section and the following, to streamline notation, and if no confusion can arise, I will denote the act f^{s_i} induced by strategy s_i simply by "s_i".

It is also convenient to derive a **strategic-form payoff function** for every player in the dynamic game under consideration, as follows: for every $s_i\in S_i$, $s_{-i}\in S_{-i}$, and $\theta\in\Theta$, let

$$U_i(s_i,s_{-i},\theta)\equiv u_i(\xi_i(\zeta(s_i,s_{-i}),\theta))=u_i(f_i^s(s_{-i},\theta)), \tag{6}$$

where the equality follows from the definition of the i-act f^{s_i} associated with strategy s_i. Denote by $U_i(s_i,\cdot)$ the map $(s_{-i},\theta)\mapsto U_i(s_i,s_{-i},\theta)$. With these definitions, structural preferences

over strategies can be represented in terms of strategic-form payoff functions, as follows.

Observation 1 For every player $i \in N$ and strategies $s_i, t_i \in S_i$, $s_i \succcurlyeq^{u_i,\mu_i} t_i$ if and only if, for every event $F \in \mathscr{F}_i$ such that $\mathrm{E}_{p_{i,F}} U_i(s_i,\cdot) < \mathrm{E}_{p_{i,F}} U_i(t_i,\cdot)$, there is $G \in \mathscr{F}_i$ such that $G \triangleright_i F$ and $\mathrm{E}_{p_{i,G}} U_i(s_i,\cdot) > \mathrm{E}_{p_{i,G}} U_i(t_i,\cdot)$.

Since structural preferences are not complete in general, an optimal strategy may fail to exist. However, since they are transitive, maximal strategies always exist:

Definition 5 *A strategy $s_i \in S_i$ is **structurally rational (for** U_i, μ_i**)** if there is no $t_i \in S_i$ such that $t_i \succ^{u_i,\mu_i} s_i$.*

The notion of maximality in Definition 5 is "ex-ante," like the structural preference defined in Definition 4. The analysis in the preceding section implies that, in the game of Figure 2, if Ann's beliefs are given by μ then RTT' and RTB' are the only structurally rational strategies; if instead they are given by ν, then RTT', RTB', LTT' and LBT' are structurally rational.

I now formally state the definition of sequential rationality. Following e.g. Rubinstein (1991); Reny (1992); Battigalli (1996); Battigalli and Siniscalchi (1999), I only require optimality of a strategy at information sets that it allows. Consequently, sequential rationality, thus defined, does not distinguish between realization-equivalent strategies. As noted in the preceding section, neither does structural rationality.

Definition 6 *A strategy s_i is **sequentially rational (for** (U_i, μ_i) if, for every $I \in \mathscr{I}(s_i)$ and $t_i \in S_i(I)$, $\mathrm{E}_{\mu(\cdot|[I])} U_i(s_i,\cdot) \geq \mathrm{E}_{\mu(\cdot|[I])} U_i(t_i,\cdot)$.*

By standard arguments, in any finite game Γ,[12] a sequentially rational strategy exists.

The main result of this section can now be stated:

Theorem 1 *If a strategy is structurally rational for (U_i, μ_i), then it is sequentially rational for (U_i, μ_i).*

[12] The extensive game itself is assumed to be finite; the sets Θ can have arbitrary cardinality.

A converse to this result does *not* hold: even in perfect-information games, structural preferences can refine sequential rationality.

Example 1 Consider the game in Fig. 3. Bob's strategies are $S_b = \{d_1d_2, d_1a_2, a_1d_2, a_1a_2\}$; there is no additional uncertainty. Assume that Ann's initial and conditional beliefs correspond to Bob's backward-induction strategy d_1a_2. Then, D_1* (the realization-equivalent strategies D_1D_2 and D_1A_2) and A_1A_2 are sequentially rational for Ann. Indeed, at the third node, A_2 is conditionally strictly dominant (it yields 0 or -1, while D_2 yields -2). At the first node, given that Bob is expected to choose D_1 at node 2, both A_1 and D_1 are rational for Ann.

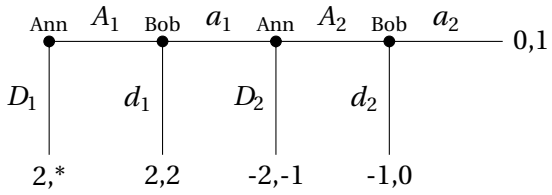

Figure 3: Sequential preferences refine sequential rationality

However, for *any* CPS μ_a for Ann that admits a basis, D_1* is strictly structurally preferred to A_1A_2, so A_1A_2 is *not* structurally rational. To see this, let $\mathbf{p}_a = (p_{a,S_b}, p_{a,[I]})$ be Ann's basis, where I denotes the third node, and it may be the case that $p_{a,S_b} = p_{a,[I]}$.

First, suppose that p_{a,S_b} assigns positive probability to a_1d_2 and/or a_1a_2. In this case, $\mu_a([I]|S_b) > 0$, so $S_b \triangleright_a [I]$ and $[I] \triangleright_a S_b$, and therefore $p_{a,S_b} = p_{a,[I]}$. Moreover, the expected payoff to A_1A_2 with respect to p_{a,S_b} is strictly less than 2, so D_1^* is strictly better than A_1A_2 ex-ante; thus, $D_1* \succ^{u_a,\mu_a} A_1A_2$.

Next, suppose that $p_{a,S_b}(d_1*) = 1$. This implies that $\mu_a([I]|S_b) = 0$, and since $\mu_a(S_b|[I]) = 1 > 0$, $S_b \triangleright_a [I]$ but the converse does not hold. Then $p_{a,S_b} \neq p_{a,[I]}$, and the expectation of D_1* and A_1A_2 with respect to p_{a,S_b} is the same, i.e., 2. Furthermore, the expectation of D_1* with respect to $p_{a,[I]}$ is 2, and that of A_1A_2 is at most 0. Since $S_b \triangleright_a [I]$, again $D_1* \succ^{u_a,\mu} A_1A_2$. □

6 Elicitation

This section investigates how players' structural preferences, and hence their conditional beliefs, can be elicited. Given a dynamic game of interest, I describe an associated "elicitation game." The main result of this section then shows that, if players' beliefs in the elicitation game are consistent with those they hold in the original game, then their *initial*—hence, observable—choices in the former reveal her strategies and preferences in the former.

The elicitation game builds upon the strategy-method procedure (Selten, 1967). As described in the Introduction, the strategy method requires players to simultaneously commit to extensive-form strategies; the experimenter then implements the resulting strategy profile. The elicitation game is defined so that, during the implementation phase, the players receive the same information they would receive if they were playing the original game. For instance, if the original game is an ascending-clock auction, players choose proxy bids, and are then required to observe the auction play out: as the price increases, they see which bidders drop out, until the winner is determined—but they cannot change their bid.[13] The key insight is that, if players make the same observations in the original game and in the elicitation game, but cannot change their actions once they have committed, then the set of strategies, conditioning events, and conditional beliefs in the two games are the same up to relabeling. Hence, every structural preference in the original game induces a unique structural preference in the elicitation game that ranks strategies the same way, again up to relabeling. In particular, any strategy a player commits to in the elicitation game is also a strategy she would follow in the original game, and conversely. Thus, structural rationality implies that the strategy method successfully elicits players' intended choices.

To elicit players' conditional beliefs as well, the elicitation game also requires players to rank pairs of acts. To keep the notation simple, I fix one distinguished player, denoted i, and ask her to rank two acts, f and g; the costruction (and Theorem 2) can be easily adapted to

[13] I thank Larry Ausubel for suggesting this example.

handle the elicitation of multiple comparisons from multiple players. In the commitment phase of the elicitation game, player i must commit to a strategy, but also choose one of the acts f or g; the other players only commit to a strategy. I then introduce a randomizing device, with outcomes denoted "o" and "a," whose realization is not observed by the players. If the outcome is o, upon reaching a terminal history the players receive the same payoffs as in the original game. If it is a, then players $j \neq i$ again receive the same payoffs as in the original game, but player i's payoff is determined by the act she has chosen. The resulting construction is in the spirit of random lottery incentive schemes (e.g. Grether and Plott, 1979).

The elicitation game is formally defined as follows.

Definition 7 *The **elicitation game** associated with* $(\Gamma, X, \Theta, \mathscr{T}, (\xi_i)_{i\in N})$ *and acts* $f, g \in \mathscr{A}_i$ *of player* $i \in N$ *is the dynamic game* $(\Gamma^*, X, \Theta^*, \mathscr{T}^*, (\xi^*_j)_{j\in N})$ *such that*

- $\Gamma^* = (N, H^*, P^*, (\mathscr{I}^*_j)_{j\in N})$*, where, writing* $\Gamma = (N, H, P, (\mathscr{I}_j)_{j\in N})$*,*
  - $h^* \in H^*$ *if and only if* $h^* = (s_1, \ldots, s_{i-1}, (s_i, k), s_{i+1}, \ldots, s_N, h)$ *for some* $k \in \{f, g\}$ *and* $h \in H$ *with* $(s_j)_{j\in N} \in S(h)$*; in this case, say that* h^* ***extends*** h*, and that* j ***plays*** s_j*, and* i ***plays*** (s_i, k) *in* h^**.*
  - $P^*(h^*) = j$ *if and only if* h^* *has length* $j-1$*, or if it extends some* $h \in H$ *with* $P(h) = j$
  - *for* $j \neq i$*,* $\mathscr{I}_j = \{I^1_j\} \cup \{\langle s_j, I\rangle : s_j \in S_j(I), I \in \mathscr{I}_j\}$*, where* $I^1_j = \{h^* : h^*$ *has length* $j-1\}$ *and* $\langle s_j, I\rangle = \{h^* : h^*$ *extends some* $h \in I$ *and* j *plays* s_j *in* $h^*\}$*;*
  - *for player* i*,* $\mathscr{I}_i = \{I^1_i\} \cup \big\{\langle s_i, k, I\rangle : s_i \in S_i(I), I \in \mathscr{I}_i, k \in \{f, g\}, \big\}$*, where* $I^1_i = \{h^* :$ h^* *has length* $i-1\}$ *and* $\langle s_i, k, I\rangle = \{h^* : h^*$ *extends some* $h \in I$ *and* i *plays* (s_i, k) *in* $h^*\}$*;*
- $\Theta^* = \Theta \times \{o, a\}$ *and* $\mathscr{T}^* = \mathscr{T} \times 2^{\{o,a\}}$*;*
- $\xi_j\big((s_1, \ldots, s_N, k, z), (\theta, r)\big)$ *equals* $\xi(z, \theta)$ *if* $j \neq i$ *or* $r = o$*, and* $k\big((s_{-i}, \theta)\big)$ *if* $j = i$ *and* $r = a$*.*

For every history $h^* = (s_1, \ldots, s_{i-1}, (s_i, k), s_{i+1}, \ldots, s_N, h)$, the initial moves $(s_1, \ldots, s_{i-1}, (s_i, k), s_{i+1}, \ldots, s_N, h)$ represent players' choices during the commitment phase of the elicitation game. Each player

has a single information set in this pahse, so these choices may as well be thought of as being simultaneous. Then, histories continue with a path of play h generated by the profile $(s_1, \ldots, s_N)$ during the implementaiton phase of the elicitation game. Each information set I of j in the original game maps to one or more information sets $\langle s_j, I \rangle$ in the elicitation game, one for each of j's strategies that allow I. This ensures that the elicitation game has perfect recall: players remember all their past choices, including the choice of a commitment strategy. Player i also remembers her choice of an act $k \in \{f, g\}$.

Definition 7 implies that there is a tight connection between strategies in the original game and in the elicitation game. For every player j, the only information set where a non-trivial choice is available is I_j^1; at that information set, the action set is S_j if $j \neq i$, ad $S_i \times \{f, g\}$ for player i. At all other information sets, players have a single available action (cf. Remark 1 in Appendix B.2). Thus, players are indeed committed to the choice of strategy they make in the first phase of the elicitation game. In addition, this property makes it possible to define, for players $j \neq i$, a bijection $\sigma_j : S_j \to S_j^*$ such that $\sigma_j(s_j)$ is the unique strategy in S_j^* that chooses s_j at I_j^1. Similarly, for player $j = i$, and for every $k \in \{f, g\}$, there is a bijection $\sigma_{i,k} : S_i \to S_i^*$, such that $\sigma_{i,k}(s_i)$ is the unique strategy in S_i^* that chooses s_i and k at I_i^1.

There is an equally tight connection between the conditioning events in the original game and in the elicitation game. Consider a player $j \neq i$ and an information set of the form $\langle s_j, I \rangle$. By Definition 7, this comprises histories h^* that extend some history $h \in I$. By the same Definition, in the commitment phase of each such history h^*, the choices $s_1, \ldots, s_N$ must be such that s reaches h, and hence I, in the original game. Hence, at $\langle s_j, I \rangle$, player j learns that, in the commitment phase of the game, her coplayers must have chosen a profile s_{-j} that allows I in the original game. This is, of course, precisely what she would learn in the original game upon reaching I. Thus, the conditioning event $[\langle s_j, I \rangle]$ in the elicitation game provides exactly the same information about coplayers as $[I]$ in the original game. (For a precise formal statement, which makes use of the bijections $\sigma_j(\cdot)$ and $\sigma_{i,k}(\cdot)$, see Lemma 4 in Appendix B.2).

Since the conditioning information in the original and elicitation games is, in a suitable

sense, "the same," if beliefs in the two games are also "the same," structural preferences should intuitively yield the same behavior. The following definition characterizes what it means for a CPS in the elicitation game to correspond to a CPS in the original game; for each player j, Σ_j^* and $\mathscr{F}_j^*$ refer to the sigma-algebra on the state space $\Omega_j^* = \Omega_j \times \{o,a\}$, and, respectively, to the set of conditioning events, in the elicitation game.

Definition 8 *A CPS $\mu_j^* \in cpr(\Sigma_j^*, \mathscr{F}_j^*)$ is an **extension** of a CPS $\mu_j \in cpr(\Sigma_j, \mathscr{F}_j)$ if, for every $s_{-j} \in S_{-j}$, $U \in \mathscr{T}$ and $r \in \{o,a\}$, the following conditions hold: if $j = i$,*

$$\mu_i^*\Big(\{(\sigma_\ell(s_\ell))_{\ell \neq i}\} \times U \times \{r\} \,\Big|\, \Omega_i \times \{o,a\}\Big) = \frac{1}{2}\mu_i\big(\{s_{-i}\} \times U \big| \Omega_i\big) \tag{7}$$

$$\mu_i^*\Big(\{(\sigma_\ell(s_\ell))_{\ell \neq i}\} \times U \times \{r\} \,\Big|\, [\langle s_i, I\rangle]\Big) = \frac{1}{2}\mu_i\big(\{s_{-i}\} \times U \big| [I]\big); \tag{8}$$

and if $j \neq i$,

$$\mu_j^*\Big(\{(\sigma_\ell(s_\ell))_{\ell \notin \{i,j\}}\} \times \{\sigma_{i,f}(s_i)), \sigma_{i,g}(s_i)\}\} \times U \times \{r\} \,\Big|\, \Omega_j \times \{o,a\}\Big) = \frac{1}{2}\mu_j\big(\{s_{-j}\} \times U \big| \Omega_j\big), \tag{9}$$

$$\mu_j^*\Big(\{(\sigma_\ell(s_\ell))_{\ell \notin \{i,j\}}\} \times \{\sigma_{i,f}(s_i)), \sigma_{i,g}(s_i)\}\} \times U \times \{r\} \,\Big|\, [\langle s_j, I\rangle]\Big) = \frac{1}{2}\mu_j\big(\{s_{-j}\} \times U \big| [I]\big). \tag{10}$$

The equations in Definition 8 state that, conditional upon any event in $\mathscr{F}_j^*$, each player j believes that the realizations of the randomizing device are independent of coplayers' strategies, and equally likely. Furthermore, j's beliefs about coplayers' strategies and exogenous uncertainty are the same as in the original game Γ. If $j \neq i$, the definition does not restrict the relative likelihood that j assigns to i choosing f or g, provided the probability she assigns to i choosing s_i is the same as in the original game.

I can finally state the main result of this section.

Theorem 2 *Fix a dynamic game $(\Gamma, X, \Theta, \mathscr{T}, (\xi_i)_{i \in N})$ and acts $f, g \in \mathscr{A}_i$ of player $i \in N$. For every $j \in N$, fix a CPS $\mu_j \in cpr(\Omega_i, \mathscr{F}_i)$ that admits a basis. Then every μ_j has an extension μ_j^*,[14] which also admits a basis. For any choice of extensions $(\mu_j^*)_{j \in N}$ and utilities $(u_j)_{j \in N}$:*

[14] For player i, this extension is unique. For players $j \neq i$, there may be different extensions, which differ in the probabilities assigned to i's choice of f vs. g. However, these differences are inconsequential to the analysis.

1. *For all* $j \neq i$ *and* $s_j, t_j \in S_j$, $\sigma_j(s_j) \succcurlyeq^{u_j,\mu_j^*} \sigma_j(t_j)$ *if and only if* $s_j \succcurlyeq^{u_j,\mu_j} t_j$;

2. *for all* $s_i, t_i \in S_i$ *and* $k \in \{f, g\}$, $\sigma_{i,k}(s_i) \succcurlyeq^{u_i,\mu_i^*} \sigma_{i,k}(t_i)$ *if and only if* $s_i \succcurlyeq^{u_i,\mu_i} t_i$;

3. *for every* $s_i \in S_i$, $\sigma_{i,f}(s_i) \succcurlyeq^{u_i,\mu_i^*} \sigma_{i,g}(s_i)$ *if and only if* $f \succcurlyeq^{u_i,\mu_i} g$.

Parts 1 and 2 of Theorem 2 state that, if every player has "the same" beliefs in the original game Γ and in the elicitation game Γ^*, then every player's preferences over strategies are effectively unchanged. This in turn suggests a reason why players might hold the same beliefs in Γ and Γ^*—if player j expects every coplayer $\ell \neq j$ not to change his beliefs, then j can conclude from Theorem 2 that they will behave similarly in the two games.

Furthermore, parts 1 and 2 provide a justification for the strategy method: setting aside i's choice of f vs. g, the game Γ^* provides a way to elicit every player's behavior in Γ from the observation of her choices in the initial commitment stage of Γ^*.

Part 3 is the elicitation result. By observing player i's preferences in the elicitation game, one can infer her preferences over arbitrary acts in the original game.

Since preferences in the original and elicitation games may be incomplete, and there may be ties, one has to be careful to interpret a *single* observed choice in the commitment phase of the elicitation game. Suppose for instance that player i chooses (s_i, f). Then, by part 2 of Theorem 2, s_i is maximal in Γ: otherwise, for some strategy t_i, the pair (t_i, f) would be strictly preferred, and thus not chosen in Γ^*. Similarly, by part 3 of the Theorem, it cannot be the case that $g \succ^{u_i,\mu_i} f$, for otherwise (s_i, g) would be strictly preferred. However, on the basis of the single observation (s_i, f), one cannot rule out the possibility that there may be multiple maximal strategies for i in Γ, or that f and g might be incomparable in Γ.

Fortunately, by exploiting specific features of structural preferences, it is nevertheless possible to elicit the utility function u_i and the CPS μ_i of any designated player i by restricting attention to specific collections of acts that player i surely *can* rank. Therefore, player i's preferences can be fully elicited.

The details are as follows. Assume, as in Anscombe and Aumann (1963) and in the companion paper Siniscalchi (2016a), that X is the set of simple lotteries over a given collection of prizes. Then, structural preferences over constant acts (i.e., effectively, over X) are consistent with von Neumann-Morgenstern expected utility under risk, and hence can be elicited by standard arguments. This pins down the utility function u_i. Thus, the key issue is how to elicit the CPS μ_i. The following result provides the main step. For outcomes $x, y \in X$ and events $E \in \Sigma$, let $x\,E\,y$ denote the act that yields x at states $\omega \in E$, and y at states $\omega \notin E$.

Proposition 1 *Fix prizes $\overline{x}, x, x_0, \underline{x} \in X$ such that $u_i(\overline{x}) > u_i(x) > u_i(x_0) > u_i(\underline{x})$ and assume without loss of generality that $u_i(x) = 1$ and $u_i(x_0) = 0$. For any information set $I \in \mathscr{I}_i$ and event $G \in \Sigma$ with $G \subseteq [I]$, there is a unique number $\alpha \in [0,1]$ such that*

$$\forall\, y \in X, \qquad u_i(y) > \alpha \quad \Rightarrow \quad y[I]x_0 \succ^{u_i,\mu_i} x\,G\,x_0, \qquad u_i(y) < \alpha \quad \Rightarrow \quad y[I]x_0 \prec^{u_i,\mu_i} x\,G\,x_0.$$

Furthermore, $\alpha = \mu_i(G|[I])$.

Leveraging Proposition 1, player i's CPS can be elicited (up to some predetermined precision) as follows. Start with a prize y that is strictly better than x, so that, under the utility normalization in Proposition 1, surely $u_i(y) > 1 \geq \mu(G|[I])$. Then, in the elicitation game Γ^* where the acts of interest are $f = y[I]x_0$ and $g = x\,G\,x_0$, the individual will choose $y[I]x_0$. Now repeat the elicitation procedure, considering successively worse prizes y. By Proposition 1 and Theorem 2, player i will consistently choose $y[I]x_0$ up to some "switching" prize y^*, and then consistently choose $x\,G\,x_0$ thereafter. The (normalized) utility of the switching prize y^* equals or approximates $\mu(G|[I])$. Finally, as noted above, it is straightforward to modify the procedure analyzed in this section so as to elicit more than one preference ranking; in particular, one can ask player i to rank M pairs of acts $f_m = y_m[I]x_0$ vs. $g_m = x\,G\,x_0$, where $y_1, \ldots, y_M$ is a grid of prizes chosen so as to identify $\mu(G|[I])$ up to a predetermined precision.

7 Consistency and bases

This section provides several characterizations of CPSs that admit a basis. The main idea can be gleaned from the game in Fig. 4.

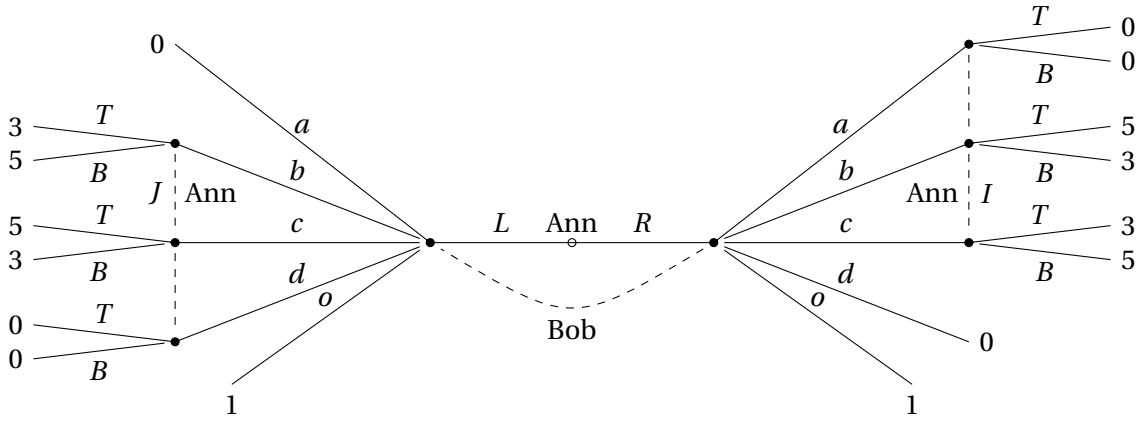

Figure 4: Newcomb's paradox?

Suppose that Ann's CPS μ is as in Eq. (11):

$$\mu(\{o\}|S_b) = \mu(\{b\}|[I]) = \mu(\{c\}|[J]) = 1. \tag{11}$$

Observe that, for this CPS, $[I] \triangleright [J]$ and $[J] \triangleright [I]$. Since both $[I]$ and $[J]$ are strictly less plausible than S_b, a basis for μ must consist of two measures, p_ϕ and p_{IJ}, where $p_\phi = \mu(\cdot|S_b)$ and p_{IJ} generates both $\mu(\cdot|[I])$ and $\mu(\cdot|[J])$ by conditioning. However, there is no such probability p_{IJ}. To see this, suppose that a suitable p_{IJ} could be found. Then in particular it must satisfy $p_{IJ}([I]) > 0$ and $p_{IJ}([J])$, or $\mu(\cdot|[I])$ and $\mu(\cdot|[J])$ could not be updates of p_{IJ}. Furthermore, since $\mu(\{a,c\}|[I]) = 0$, one must have $p_{IJ}(\{a,c\}) = 0$; and since $\mu(\{b,d\}|[J]) = 0$, one must have $p_{IJ}(\{b,d\}) = 0$. But then $p_{IJ}([I] \cup [J]) = p_{IJ}(\{a,b,c,d\}) = 0$, contradiction. Therefore, the CPS μ in Eq. (11) does not admit a basis.

A peculiar feature of this CPS μ is that Ann's own initial choice of R vs. L determines her conditional beliefs on the relative likelihood of b and c, despite the fact that Bob does not ob-

serve Ann's initial choice. (In fact, Ann's first action and Bob's move may well be simultaneous.) This phenomenon is reminiscent of Newcomb's paradox (Weirich, 2016).

Observe that, if $[I]$ and $[J]$ both had positive prior probability, the definition of conditional probability would imply that the relative likelihood of b and c must be the same at both information sets. A closely related argument implies that, in a *consistent assessment* in the sense of Kreps and Wilson (1982), Ann cannot believe that Bob played b at I, and that he played c at J.[15] Moreover, modify the game in Figure 4 so that, if Bob does not choose o, a new information K of Ann is reached; at K, Ann has a single available action, which leads to I if Bob played a, b or c, and to J if he played b, c or d. In such a game, $[K] = \{a,b,c,d\} = [I] \cup [J]$. Then, the argument given above implies that the CPS μ cannot be extended to a new CPS μ' for the new game. A fortiori, the CPS μ cannot be extended to a *complete* CPS in the sense of Myerson (1986)—a CPS in which every non-empty subset is a conditioning event.

To sum up, CPSs that do not admit a basis fail consistency requirements that are both intuitive and have well-understood counterparts in the received literature. The following definition identifies an intrinsic property of CPSs that captures this consistency requirement.

Definition 9 *Fix an extensive game form $\Gamma = (N, H, P, (\mathscr{I}_i)_{i\in N})$ and a CPS $\mu \in cpr(\Sigma_i, \mathscr{F}_i)$ for player $i \in N$. An ordered list $F_1, \ldots, F_L \in \mathscr{F}_i$ is a μ-**sequence** if $\mu(F_{\ell+1}|F_\ell) > 0$ for all $\ell = 1, \ldots, L-1$.*

*The CPS μ is **consistent** if, for every μ-sequence $F_1, \ldots, F_L$, and all $E \subseteq F_1 \cap F_L$,*

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

The preorder $\triangleright$ in Definition 2 can be characterized in terms of μ-sequences: $F \triangleright G$ iff there is a μ-sequence $F_1, \ldots, F_L$ such that $F_1 = G$ and $F_L = F$.

Consistency can be viewed as a strengthening of the chain rule of conditioning. Consider

[15] In the language of Kreps and Wilson (1982), fix a convergent sequence of strictly positive behavioral strategy profiles π^k. If m^k is the system of beliefs derived from π^k, $m^k(I)$ and $m^k(J)$ assign the same relative likelihood to the nodes corresponding to Bob's choice of b vs. c. Hence, the same holds for the limit system of beliefs.

$F, G \in \mathscr{F}_i$ and $E \in \Sigma$, and assume that $E \subseteq F \subseteq G$. Then the ordered list F, G is a μ-sequence, because $\mu(G|F) \geq \mu(F,F) = 1$, and $E \subseteq F \cap G = F$. Therefore, consistency implies that

$$\mu(E|F) \cdot \frac{\mu(F \cap G|G)}{\mu(F \cap G|F)} = \mu(E|G);$$

but since $\mu(F \cap G|G) = \mu(F|G)$ and $\mu(F \cap G|F) = \mu(F|F) = 1$, this reduces to $\mu(E|F)\mu(F|G) = \mu(E|G)$, which is precisely what the chain rule requires.

To see why consistency rules out pathological beliefs such as those in Eq. (11), consider first a μ-sequence F_1, F_2 of length 2, and assume that $\mu(F_1 \cap F_2|F_2) > 0$. Rearrange the equation in Definition 9 as follows:

$$\frac{\mu(E|F_1)}{\mu(F_1 \cap F_2|F_1)} = \frac{\mu(E|F_2)}{\mu(F_1 \cap F_2|F_2)}.$$

Suppose that $F_1 = [I_1]$ and $F_2 = [I_2]$ for two information sets I_1, I_2 of i. Since $E \subseteq F_1 \cap F_2$, this condition requires that the probability assigned to E conditional on $F_1 \cap F_2$ must be the same at I_1 and at I_2. In particular, if I_1 and I_2 are reached via different actions of i, this means that the conditional probability of E given $F_1 \cap F_2$ is independent of i's own choices—that is, no Newcomb-like paradox can arise. Definition 9 generalizes this intuition by allowing for μ-sequences of length greater than 2.

The following theorem shows that a CPS is consistent if and only if it admits a (unique) basis. Furthermore, it provides two additional equivalent characterizations of consistency that reflect the preceding discussion.

Theorem 3 *Let $\mu \in cpr(\Sigma_i, \mathscr{F}_i)$ be a CPS for player $i \in N$. Define $\mathscr{F}_\mu = \{\cup_{\ell=1}^{L} F_\ell : F_1, \ldots, F_L$ is a μ-sequence$\}$. The following are equivalent:*

1. *μ is consistent;*

2. *μ admits a unique basis;*

3. *there is a uniqe CPS $\nu \in cpr(\Sigma_i, \mathscr{F}_\mu)$ such that $\nu(\cdot|F) = \mu(\cdot|F)$ for all $F \in \mathscr{F}$;*

4. *there is a sequence* $(p^n) \in pr(\Sigma_i)$ *such that* $p^m(F) > 0$ *for all* m *and* $F \in \mathscr{F}_i$*, and* $p^m(E \cap F)/p^k(F) \to \mu(E \cap F|F)$ *for all* $F \in \mathscr{F}$ *and* $E \in \Sigma_i$*.*[16]

If $\mathbf{p} = (p_F)_{F \in \mathscr{F}_i}$ *is a basis for* μ*, and* $\nu \in cpr(\Sigma_i, \mathscr{F}_\mu)$ *satisfies* $\nu(\cdot|F) = \mu(\cdot|F)$ *for all* $F \in \mathscr{F}$*, then, for every* $F \in \mathscr{F}$*,* $p_F = \nu(\cdot|\cup\{G : F \triangleright G, G \triangleright F\})$*.*

8 Discussion, Related Literature, and Conclusions

Incomplete-information games In the textbook model of games with incomplete information, there is a set Θ_i of "types" for each player i, and possibly a set Θ_0 that describes residual uncertainty that is not captured by the realization of each player's type. Player types may affect both utilities and conditional beliefs—that is, types determine preferences. The analysis of this paper may be seen as providing a foundation for the preferences of a given player type; in other words, it concerns the *interim* stage of an incomplete-information game. To see this, fix a player i, and a type $\theta_i^* \in \Theta_i$. The exogenous uncertainty faced by this player is $\Theta = \Theta_0 \times \Theta_{-i}$. The utility function u_i and the conditional beliefs μ_i introduced in Definition 1 are now interpreted as the ones characterizing the preferences of i's type θ_i^*. The results in Sections 5 and 6 then state that, if the selected type θ_i^* of player i is structurally rational, then she is sequentially rational, and her preferences can be elicited (at the *interim* stage). One can in principle apply the analysis to each possible tuple of types $(\theta_i)_{i \in N} \in \prod_{i \in N} \Theta_i$. Thus, the results in this paper provide behavioral foundations for sequential rationality in the overall incomplete-information game.

Alternative representations of beliefs In this paper, the primary representation of conditional beliefs are consistent CPSs. To define structural preferences, I "extract" from a CPS a plausibility ordering $\triangleright$ over conditioning events $\mathscr{F}$, and a basis $(p_F)_{F \in \mathscr{F}}$. The advantage is that

[16] Even though the state space may be infinite, convergence here is pointwise. See the proof for details.

this emphasizes the primacy of conditional beliefs. However, one can alternatively represent beliefs in a format that can be used to define structural preferences directly. One way to do so is to modify the definition of a basis: see Appendix A.1 for a sketch of the approach. Theorem 3 provides another alternative: one can assume that player's beliefs are extended CPSs, defined for a larger collection of conditioning events as indicated in Section 7.

Lexicographic expected utility. A *lexicographic probability system* (LPS) on the state space $\Omega_i = S_{-i} \times \Theta$ is a linearly ordered collection of probabilities $(p_0, \ldots, p_{n-1})$ on Ω_i. Given acts $f, g \in \mathscr{A}_i$, f is lexicographically (weakly) preferred to g if $\mathrm{E}_{p_m} u_i \circ f < \mathrm{E}_{p_m} u_i \circ g$ implies $\mathrm{E}_{p_\ell} u_i \circ f > \mathrm{E}_{p_\ell} u_i \circ g$ for some $\ell < m$. This is clearly reminiscent of Definition 4. However, the LPS $(p_0, \ldots, p_{n-1})$ is defined without any reference to the underlying dynamic game. This has an important consequence: an LPS can generate a CPS by conditioning,[17] but the same CPS may be generated by *multiple* LPSs. For instance, the CPS ν in Eq. (5) can be generated by the LPS $\lambda^1 = (\delta_o, \delta_t, \delta_m)$, but also by the LPS $\lambda^2 = (\delta_o, \delta_m, \delta_t)$, where δ_ω denootes the Dirac measure concentrated on $\{\omega\}$. Intuitively, λ^1 assigns higher plausibility to t than to m, whereas the opposite is true of λ^2. However, *this plausibility assessment is not derived from Ann's CPS* ν. By way of contrast, by design, structural preferences are defined solely in terms of information that can be derived from the player's conditional beliefs. For this reason, they are close in spirit to sequential rationality, and explicitly motivated by extensive-form analysis. Lexicographic expected-utility maximization is instead a strategic-form concept; it was introduced into game theory to analyze refinements for games with simultaneous moves (Blume, Brandenburger, and Dekel, 1991b), and moreover, when coupled with a full-support assumption, it incorporates an *invariance* requirement; see Brandenburger (2007), §12. Finally, as noted in Section 4.2, structural preferences reduce to EU in simultaneous-move games; on the other hand, lexicographic preferences may of course differ from EU in such games.

[17] For every $I \in \mathscr{I}_i$, let $\mu(\cdot|[I]) = p_{m(I)}(\cdot|[I])$, where $m(I)$ is the *lowest* index m for which $p_m([I]) > 0$—assuming one such index can be found.

Conditional expected-utility maximization Myerson (1986) axiomatizes conditional expected utility maximization with respect to a CPS. The analysis assumes that a family of conditional preferences is taken as given. Preferences conditional upon nested events are related by *subjective substitution*, which is shown to characterize the chain rule of conditioning for CPSs. Just like prior beliefs do not fully determine the player's CPS due to the presence of ex-ante zero-probability events, prior preferences do not fully determine the entire system of conditional preferences. Thus, in Myerson's analysis, it is necessary to assume that all conditional preferences are observable. As shown in Section 2, this may be problematic in many dynamic games. By way of contrast, the present paper defines an *ex-ante* preference relation; Theorem 2 shows that it is elicitable by observing initial choices in suitably-designed experiments.[18]

Structural and lexicographic consistency As noted in the Introduction, bases incorporate a version of *structural consistency* (Kreps and Wilson, 1982; Kreps and Ramey, 1987): conditional beliefs should be derived from a collection of alternative prior probabilistic hypotheses about the play of opponents. CPSs also reflect structural consistency, though in a somewhat trivial sense: every conditional belief in a CPS can be interpreted as an alternative "prior" hypothesis that is adopted once an unexpected information set is reached. By way of contrast, a basis incorporates a notion of parsimony: it identifies the minimal set of alternative hypotheses that generate the CPS. Furthermore, as Theorem 3 shows, an arbitrary CPS may assign relative likelihoods in inconsistent ways, and thus represent alternative hypotheses that are not just trivial, but contradictory. The existence of a basis ensures consistency.

Kreps and Wilson (1982) also consider a notion of *lexicographic* consistency. Their definition is stated in the setting of equilibrium, rather than individual maximization; furthermore, conditional beliefs are represented by consistent assessments. Translated to the present setting and notation, lexicographic consistency requires that the player's CPS can be generated by an LPS, as described above. Hence, the above comparison with LPSs applies: in the present

[18]The same observability issue applies to Asheim and Perea (2005), who generalize Myerson's analysis.

analysis, the basis and its ordering is entirely derived from the CPS. Thus, CPSs are the starting point of the analysis. Lexicographic consistency, on the other hand, gives priority to an LPS, which adds information not present in the player's CPS.

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

A Bases

Throughout, fix an extensive game form $\Gamma = (N, H, P, (\mathcal{I}_i)_{i\in N})$.

Lemma 1 *Let μ be a CPS for player $i \in N$ that admits a basis $\mathbf{p} = (p_F)_{F\in\mathcal{F}_i}$. Denote by $\triangleright$ the plausibility relation induced by μ.*

1. *For all $E, F \in \mathcal{F}_i$, $p_F(E) > 0$ implies $E \triangleright F$.*
2. *For all $E, F \in \mathcal{F}_i$, if $E \triangleright F$ and $p_F \neq p_E$, then $p_E(F) = 0$.*

Proof: Begin with a preliminary *Claim*: for all $F \in \mathscr{F}_i$ and $E \in \Sigma$, if $p_F(E) > 0$, then there is $G \in \mathscr{F}$ such that $G \triangleright F$, $F \triangleright G$, and $p_F(E \cap G) > 0$. This follows because, by part (2) in Def. 3,

$$p_F(E) = p_F\left(E \cap \bigcup \{G \in \mathscr{F}_i : F \triangleright G, G \triangleright F\}\right) \leq \sum_{G \in \mathscr{F}_i : F \triangleright G, G \triangleright F} p_F(E \cap G).$$

1: by the Preliminary Claim, there is $G \in \mathscr{F}_i$ with $G \triangleright F$, $F \triangleright G$, and $p_F(G \cap E) > 0$. By condition (1), $p_G(E \cap G) > 0$, and by condition (3), $p_G(G) > 0$ and $\mu_i(E \cap G|G) p_{i,G}(G) = p_G(E \cap G) > 0$. Thus, by Def. 2, $E \triangleright G$; since the relation $\triangleright$ is transitive, $E \triangleright F$.

2: by contradiction, if $p_E(F) > 0$ then part 1 implies that $F \triangleright E$. Since also $E \triangleright F$, condition (1) in Def. 3 implies $p_F = p_E$, contradiction. Thus, $p_E(F) = 0$. ■

The following is the central result in the analysis of consistency and bases.

Proposition 2 *Fix a CPS $\mu \in cpr(\Sigma_i, \mathscr{F}_i)$ for player $i \in N$. The following are equivalent:*

1. *μ is consistent;*
2. *for every μ-sequence $F_1, \ldots, F_K \in \mathscr{F}$, there exists $p \in pr(\Sigma_i)$ with $p(\cup_k F_k) = 1$, such that, for every $\ell = 1, \ldots, K$ and $E \in \Sigma$ such that $E \subseteq F_\ell$,*

$$p(E) = \mu(E|F_\ell) p(F_\ell). \tag{12}$$

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

Proof: (1)⇒(2): assume that μ is consistent. Let $F_1, \ldots, F_K \in \mathscr{F}_i$ be a μ-sequence.

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

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

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

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

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

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

The function $\rho(\cdot)$ thus defined takes non-negative values. I claim that $\rho(\cdot)$ is countably additive. Consider an ordered list $E_1, E_2, \ldots \in \Sigma$ such that $E_m \cap E_{\bar{m}} = \emptyset$ for $m \neq \bar{m}$. If there is $\ell \in \{1, \ldots, K\}$ such that $E_m \subseteq G_\ell$ for all m, then by countable additivity of $\mu(\cdot|F_\ell)$,

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

Thus, for a general ordered list $E_1, E_2, \ldots \in \Sigma$ of pairwise disjoint events,

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

interchanging the order of the summation in the second line is allowed because all summands are non-negative and the derivation shows that $\sum_k \sum_m \rho(E_m \cap G_k) = \sum_k \rho([\cup_m E_m] \cap G_k)$, a sum of finitely many finite terms.

Now consider $E \in \Sigma$ with $E \subseteq F_m$ and $E \subseteq G_\ell$ for some $\ell, m \in \{1,\dots,K\}$ with $\ell \neq m$. Since $F_m \subseteq F_1 \cup \dots \cup F_m = G_1 \cup \dots \cup G_m$, it must be the case that $\ell < m$. Consider the ordered list $F_\ell, \dots, F_m \in \mathscr{F}_i$: since $F_1, \dots, F_K$ is a μ-sequence, so is $F_\ell, \dots, F_m$, so by Consistency, since by assumption $E \subseteq F_m \cap G_\ell \subseteq F_m \cap F_\ell$,

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

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

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

Therefore, for all $E \in \Sigma$ with $E \subseteq F_m$ for some $m \in \{1,\dots,K\}$,

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

It follows that, for all $m \in \{1,\dots,K\}$ and $E \in \Sigma$ with $E \subseteq F_m$,

$$\rho(F_m) = \mu(F_m|F_m)\prod_{k=m}^{K-1}\frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)} = \prod_{k=m}^{K-1}\frac{\mu(F_k \cap F_{k+1}|F_{k+1})}{\mu(F_k \cap F_{k+1}|F_k)}, \tag{13}$$

and therefore

$$\rho(E) = \mu(E|F_m)\rho(F_m). \tag{14}$$

Finally, notice that $\rho(\cup_k G_k) = \rho(\cup_k F_k) \geq \rho(F_K) = 1$; thus, one can define a probability measure $p \in pr(\Sigma_i)$ by letting

$$\forall E \in \Sigma, \qquad p(E) = \frac{\rho(E)}{\rho(\cup_k G_k)} = \frac{\rho(E)}{\rho(\cup_k F_k)}.$$

For every $\ell \in \{1,\dots,K\}$ and every event $E \subseteq F_\ell$, p satisfies Eq. (12), as asserted.

To show that p is uniquely defined, let $q \in pr(\Sigma_i)$ be a measure that satisfies Eq.(12). I first claim that, for every $m = 1,\dots,K$,

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

The claim is trivially true for $m = K$, so consider $m \in \{1,\dots,K-1\}$ and assume that the claim holds for $m+1$. By Eq.(12),

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

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

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

Since $G_m \subseteq F_m$, Eq. (12) implies that

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

where the last equality follows from Eq.(14). Since $\sum_k q(G_k) = q(\cup_k G_k) = q(\cup_k F_k)$, if in addition q satisfies $q(\cup_k F_k) = 1$, then

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

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

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

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

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

Furthermore, let $k_0 \in \{1,\dots,K-1\}$ be such that $\mu(F_k \cap F_{k+1}|F_{k+1}) > 0$ for all $k > k_0$, and $\mu(F_{k_0} \cap F_{k_0+1}|F_{k_0+1}) = 0$. By inspecting Eq. (13), it is clear that $\rho(F_k) = 0$ for $k = 1,\dots,k_0$, and $\rho(F_k) > 0$ for $k = k_0+1,\dots,K$. Then, $p(F_k) = 0$ for $k = 1,\dots,k_0$, and $p(F_k) > 0$ for $k = k_0+1,\dots,K$. From the above argument, it follows that the same is true for any $q \in pr(\Sigma_i)$ that satisfies Eq. (12) and $q(\cup_k F_k) = 1$. Thus, the last claim of the Proposition follows.

Finally, if $q \in pr(\Sigma_i)$ satisfies Eq.(12) and $q(\cup_k F_k) = 1$, for every $k = k_0 + 1, \ldots, K$ and $E \in \Sigma_i$ such that $E \subset F_k$,

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

and therefore, for every $E \in \Sigma_i$,

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

In other words, p is the unique probability measure that satisfies Eq. (12) and $p(\cup_k F_k) = 1$.

(2)⇒ (1): assume that (2) holds. Consider a μ-sequence $F_1, \ldots, F_K$. Fix an event $E \subseteq F_1 \cap F_K$. By assumption, there exists $p \in pr(\Sigma_i)$ that satisfies Eq. (12) for $k = 1, \ldots, K$, with $p(\cup_{k=1}^{K} F_k) = 1$.

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

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

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

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

■

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

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

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

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

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

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

■

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

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

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

I claim that, for any such k, $\mu(G_k|F) > 0$. Since $F \subseteq \cup_m G_m$ and $\mu(F|F) = 1$, $\mu(G_m|F) > 0$ for at least one $m \in \{1, \ldots, K\}$. If $m = k$, the claim is true. If $m < k$, then the ordered list $F, G_m, G_{m+1}, \ldots, G_k, F$ is a μ-sequence that satisfies the conditions of Corollary 1, so that in particular $\mu(G_k|F) > 0$, as claimed. Finally, suppose $m > k$. Since $p(G_k) > 0$, by the last claim of Proposition 2, $\mu(G_\ell|G_{\ell+1}) > 0$ for $\ell = k, \ldots, K-1$. Hence, since $\mu(G_m|F) > 0$, the ordered list $F, G_m, G_{m-1}, \ldots, G_{k+1}, G_k, F$ is a μ-sequence that satisfies the conditions in Corollary 1, so in particular $\mu(G_k|F) > 0$, as claimed.

This implies that the ordered list $G_1, \ldots, G_k, F, G_k, \ldots, G_K$ is a μ-sequence. Let p' be the measure delivered by Proposition 2 for this μ-sequence. Notice that $p(F \cup \bigcup G_k) = p'(F \cup \bigcup_k G_k) = 1$, and for all $\ell \in \{1, \ldots, K\}$ and $E \in \Sigma_i$ with $E \subset G_\ell$, $p'(E) = p'(G_\ell)\mu(E|G_\ell)$. Since p is the unique probability with these properties, $p = p'$. But then, for $E \in \Sigma_i$ with $E \subseteq F$,

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

as claimed. ■

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

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

Since, for every m, $F_m \subseteq \cup_k G_k$, Corollary 2 implies that, for every $E' \in \Sigma_i$ with $E' \subseteq F_m$,

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

Hence, the measure $p' \in pr(\Sigma_i)$ defined by $p'(E) = p(E \cap \cup_m F_m)/p(\cup_m F_m)$ satisfies

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

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

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

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

Finally, I prove Theorem 3.

Proof: I show (1) ⇒ (3) ⇒ (2) ⇒ (4) ⇒ (1).

(1) ⇒ (3): by Corollary 3, if $F_1,\ldots,F_L$ and $G_1,\ldots,G_M$ are μ-sequences such that $\cup_\ell F_\ell = \cup_m G_m$, and p and q are the measures in condition 2 of Propositon 2, then $p = q$. Therefore, one can define an array $(\nu(\cdot|F))_{F \in \mathscr{F}_\mu}$ of probabilities on Σ_i by letting $\nu(\cdot|\cup_\ell F_l)$ be the measure in condition 2 of Proposition 2 associated with the μ-sequence $F_1,\ldots,F_L$. In particular, if $F \in \mathscr{F}_i$,

then $\nu(\cdot|F) = \mu(\cdot|F)$. Again by Corollary 3, if $F_1,\ldots,F_L$ and $G_1,\ldots,G_M$ are μ-sequences with $\cup_m G_m \subset \cup_\ell F_\ell$, then for every measurable $E \subseteq \cup_m G_m$, $\nu(E|\cup_m G_m) = \nu(E|\cup_\ell F_\ell)\nu(\cup_m F_m|\cup_\ell G_\ell)$. Thus, ν is a CPS on $(\Sigma_i, \mathscr{F}_\mu)$. Since the measure $\nu(\cdot|\cup_\ell F_\ell)$ associated with each μ-sequence $F_1,\ldots,F_L$ is unique, ν is unique.

(3) $\Rightarrow$ (2): let $\{F_1,\ldots,F_L\}$ be an enumeration of an equivalence class of (the symmetric part of) $\triangleright$. Then in particular $F_1 \triangleright F_2 \triangleright \ldots \triangleright F_L$. By definition, for every $\ell = L, L-1,\ldots,2$, there is a μ-sequence $F_1^\ell,\ldots,F_{M(\ell)}^\ell$ such that $F_1^\ell = F_\ell$ and $F_{M(\ell)}^\ell = F_{\ell-1}$. Since in addition $F_L \triangleright F_1$, there is also a μ-sequence $F_1^1,\ldots,F_{M(1)}^1$ with $F_1^1 = F_1$ and $F_{M(1)}^1 = F_L$. Since by construction $F_{M(\ell)}^\ell) = F_{\ell-1} = F_1^{\ell-1}$ for $\ell = L, L-1,\ldots,2$, the ordered list

$$F_1^L,\ldots,F_{M(L)}^L, F_1^{L-1},\ldots,F_{M(2)}^2, F_1^1,\ldots,F_{M(1)}^1$$

is a μ-sequence. Since $F_1^L = F_L = F_{M(1)}^1$, Corollary 1 implies that the reverse-ordered list

$$F_{M(1)}^1,\ldots F_1^1, F_{M(2)}^2,\ldots,F_1^{L-1}, F_{M(L)}^L,\ldots,F_1^L$$

is also a μ-sequence. This has two implications. First, since any segment of the above μ-sequences is also a μ-sequence, it follows that, for every $\ell = 1,\ldots,L$ and $m = 1,\ldots,M(L)$, both $F_1 = F_1^1 \triangleright F_m^\ell$ and $F_m^\ell \triangleright F_1^1 = F_1$. Hence F_ℓ^m is an element of $\{F_1,\ldots,F_L\}$; in particular, $\cup_{\ell=1}^L \cup_{m=1}^{M(\ell)} F_m^\ell = \cup_{\ell=1}^L F_\ell$. Second, for all $\ell = 1,\ldots,L$ and $m = 1,\ldots,M(\ell)-1$, both $\mu(F_{m+1}^\ell|F_m^\ell) > 0$ and $\mu(F_m|F_{m+1}^\ell) > 0$. Since ν is a CPS on $\mathscr{F}_\mu$ which agrees with μ on $\mathscr{F}$, $\nu(E|\cap_\ell F_\ell) = \mu(E|F_{\bar{m}}^{\bar{\ell}})\nu(F_{\bar{m}}^{\bar{\ell}}|\cup_\ell F_\ell)$ for all measurable $E \subseteq F_{\bar{m}}^{\bar{\ell}}$. Therefore, Proposition 2 implies that $\nu(F_{\bar{m}}^{\bar{\ell}}|\cup_\ell F_\ell) > 0$ for all $\bar{m},\bar{\ell}$.

Now construct an array $\mathbf{p} = (p_F)_{F\in\mathscr{F}}$ of probabilities on Σ_i by letting $p_{F_{\bar{\ell}}} = \nu(\cdot|\cup_\ell F_\ell)$ for every equivalence class $\{F_1,\ldots,F_L\}$ of $\triangleright$ and every $\bar{\ell} = 1,\ldots,L$. Notice that this is the only candidate basis for μ; this is because, if $\mathbf{q} = (q_F)_{F\in\mathscr{F}_i}$ is a basis, then for every equivalence class $\{F_1,\ldots,F_L\}$ for $\triangleright$, every ℓ, and every $E \subseteq F_\ell$, it must satisfy $q_{F_\ell}(\cup_m F_m) = 1$ and $\mu(E|F_\ell) = q_{F_\ell}(E)/q_{F_\ell}(F_\ell)$; since, as shown above, the events $F_1,\ldots,F_L$ can be arranged into a μ-sequence, by the last claim in Proposition 2, there is at most one measure that satisfies these properties.

Thus, consider an equivalence class $\{F_1,\ldots,F_L\}$ for $\triangleright$, and every $\bar{\ell} = 1,\ldots,L$, $p_F(F_{\bar{\ell}}) = \nu(F_{\bar{\ell}}|\cup_\ell F_\ell) > 0$ and $\mu(E|F_{\bar{\ell}}) = \nu(E|\cup_\ell F_\ell)/\nu(F_{\bar{\ell}}|\cup_\ell F_\ell) = p_{F_{\bar{\ell}}}(E)/p_{F_{\bar{\ell}}}(F_{\bar{\ell}})$: thus, Condition (3) in Def. 3 holds.

Moreover, $p_{F_\ell})(\cup\{G : G \triangleright F_\ell, F_\ell \triangleright G\}) = p_{F_\ell}(\cup_\ell F_\ell) = \nu(\cup_\ell F_\ell | \cup_\ell F_\ell) = 1$, so Condition (2) holds as well. Finally, if $F \triangleright G$ and $G \triangleright F$, then F, G belong to the same equivalence class $\{F_1, \ldots, F_L\}$ and so by construction $p_F = \nu(\cdot | \cup_\ell F_\ell) = p_G$. Thus, to show that Condition (1) holds, it remains to be shown that, if $F, G \in \mathscr{F}_i$ are not in the same equivalence class, then $p_F \neq p_G$.

Suppose by contradiction that either $F \not\triangleright G$ or $G \not\triangleright F$ (or both), but $p_F = p_G$. Since $p_F(G) = p_G(G) > 0$, by the preliminary Claim in the proof of Lemma 1 (which only relies on Condition (2) of Def. 3, which was just shown to hold) there must be $D \in \mathscr{F}_i$ such that $D \triangleright F$, $F \triangleright D$ and $p_D(G \cap D) > 0$. Then, since it was just shown that Condition (3) holds, $\mu(G|D) = \mu(G \cap D|D) = p_D(G \cap D)/p_D(D) > 0$, so $G \triangleright D$. By the same argument, since $p_G(F) = p_F(F) > 0$, there is $E \in \mathscr{F}_i$ such that $E \triangleright G$, $G \triangleright E$, and $\mu(F \cap E|E > 0)$, so $F \triangleright E$. But then, since $\triangleright$ is transitive, $F \triangleright E \triangleright G$ and $G \triangleright D \triangleright F$. Thus, F and G are in the same equivalence class: contradiction. Therefore, $\mathbf{p}$ is a basis for μ, and as argued above, it is the only one.

The argument just given also establishes the last claim of Theorem 3.

(2) $\Rightarrow$ (4): let $\mathbf{p}$ be the (unique) basis of μ. The probabilities $\{p_F : F \in \mathscr{F}\}$ can be partially ordered as follows: $p_F \geq p_G$ iff $F \triangleright G$. [The ordering is clearly reflexive and transitive because so is $\triangleright$. To see that it is antisymmetric, if $p_F \geq p_G$ and $p_G \geq p_F$, then $F \triangleright G$ and $G \triangleright F$; by condition (1), this implies $p_F = p_G$.]

Let $p_1, \ldots, p_L$ be an enumeration of $\{p_F : F \in \mathscr{F}\}$ such that, for all ℓ, m, $p_\ell \geq p_m$ implies $\ell \leq m$. [This can be obtained by considering any completion of the partial order $\geq$, and assigning indices consistently with this completion, with $\ell = 1$ being the greatest element.] For every $F \in \mathscr{F}_i$, let $\ell(F)$ denote the index ℓ such that $p_\ell = p_F$. Finally, define a sequence $(p^n) \subset pr(\Sigma_i)$ by letting

$$p^n = \sum_{\ell=1}^{L} \frac{\frac{1}{n^{\ell-1}}}{\sum_{m=1}^{L} \frac{1}{n^{m-1}}} p_\ell.$$

For every $n \geq 1$ and $F \in \mathscr{F}_i$, $p_{\ell(F)}(F) = p_F(F) > 0$, and so $p^n(F) > 0$. Furthermore, consider $F \in \mathscr{F}_i$ and a measurable $E \subseteq F$. Suppose there is $G \in \mathscr{F}_i$ such that $p_G(E) > 0$; then $p_G(F) > 0$,

so by Lemma 1 part 1, $F \triangleright G$. Hence, $p_F \geq p_G$, so either $p_G = p_F$, or $\ell(F) < \ell(G)$. Thus,

$$p^n(E) = \sum_{\ell=\ell(F)}^{L} \frac{\frac{1}{n^{\ell-1}}}{\sum_{m=1}^{L} \frac{1}{n^{m-1}}} p_\ell(E).$$

This holds in particular for $E = F$. Hence,

$$\frac{p^n(E)}{p^n(F)} = \frac{\sum_{\ell=\ell(F)}^{L} \frac{\frac{1}{n^{\ell-1}}}{\sum_{m=1}^{L} \frac{1}{n^{m-1}}} p_\ell(E)}{\sum_{\ell=\ell(F)}^{L} \frac{\frac{1}{n^{\ell-1}}}{\sum_{m=1}^{L} \frac{1}{n^{m-1}}} p_\ell(F)} = \frac{\sum_{\ell=\ell(F)}^{L} \frac{1}{n^{\ell-1}} p_\ell(E)}{\sum_{\ell=\ell(F)}^{L} \frac{1}{n^{\ell-1}} p_\ell(F)} = \frac{n^{\ell(F)-1}}{n^{\ell(F)-1}} \frac{\sum_{\ell=\ell(F)}^{L} \frac{1}{n^{\ell-\ell(F)}} p_\ell(E)}{\sum_{\ell=\ell(F)}^{L} \frac{1}{n^{\ell-\ell(F)}} p_\ell(F)} =$$

$$= \frac{p_{\ell(F)}(E) + \sum_{\ell=\ell(F)+1}^{L} \frac{1}{n^{\ell-\ell(F)}} p_\ell(E)}{p_{\ell(F)}(F) + \sum_{\ell=\ell(F)+1}^{L} \frac{1}{n^{\ell-\ell(F)}} p_\ell(F)} \to \frac{p_{\ell(F)}(E)}{p_{\ell(F)}(F)} = \frac{p_F(E)}{p_F(F)} = \mu(E|F).$$

(4) $\Rightarrow$ (1): consider a μ-sequence $F_1, \ldots, F_L$ and an event $E \subseteq F_1 \cap F_L$. Let $(p^n) \subseteq pr(\Sigma_i)$ generate μ in the sense of condition (4). Since $\mu(F_{\ell+1}|F_\ell) > 0$ for all $\ell = 1, \ldots, L-1$, there is $\bar{n}$ such that $n \geq \bar{n}$ implies $p^n(F_{\ell+1} \cap F_\ell)/p(F_\ell) > 0$. For every such n and measurable set $E \subseteq F_1 \cap F_L$,

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

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

A.1 An alternative definition of beliefs

A *partially ordered probability system* (POPS) for player i is a collection $(q_F)_{F \in \mathcal{F}}$ of probabilities on Ω_i that satisfies

1. $q_F(F) > 0$ for every $F \in \mathcal{F}_i$;
2. $q_F(\cup\{G : p_G = p_F\}) = 1$;
3. for every $F, G \in \mathcal{F}_i$ with $q_F = q_G$, there exist $F_1, \ldots, F_L \in \mathcal{F}_i$ with $F_1 = F$, $F_L = G$, and, for every $\ell = 1, \ldots, L-1$, $q_{F_\ell} = q_F$ and $q_F(F_\ell \cap F_{\ell+1}) > 0$;

4. for every collection $F_1, \ldots, F_L \in \mathscr{F}_i$ such that $q_{F_\ell}(F_{\ell+1}) > 0$ for all $\ell = 1, \ldots, L-1$ and $q_{F_1} \neq q_{F_L}$, $q_{F_L}(F_1) = 0$.

It can be shown that, if μ is a CPS with basis $\mathbf{p} = (p_F)_{F \in \mathscr{F}_i}$, then $\mathbf{p}$ is a POPS; since a CPS admits at most one basis, every consistent CPS induces a unique POPS. Conversely, if $\mathbf{q} = (q_F)_{F \in \mathscr{F}_i}$ is a POPS, then it induces a unique consistent CPS μ by letting $\mu(E \cap F | F) = p_F(E \cap F)/p_F(F)$ for every $F \in \mathscr{F}_i$ and $E \in \Sigma_i$; in addition $\mathbf{q}$ is the unique basis for μ. Thus, there is a one-to-one correspondence between POPS and consistent CPSs. Furthermore, given a POPS $\mathbf{q} = (q_F)_{F \in \mathscr{F}_i}$, one can define a preorder $\triangleright$ on $\mathscr{F}_i$ by letting $F \triangleright G$ iff there is a list $F_1, \ldots, F_L \in \mathscr{F}_i$ with $F_1 = G$, $F_L = F$ and $p_{F_\ell}(F_{\ell+1}) > 0$ for all $\ell = 1, \ldots, L-1$. This coincides with the plausibility ordering derived from the CPS μ associated with $\mathbf{q}$. Thus, one can define structural preferences starting from a POPS and the corresponding plausibility ordering. The resulting preference will be identical to the one obtained by applying Def. 4. (Details available upon request.)

B Sequential rationality and elicitation

B.1 Theorem 1 (structural and sequential rationality)

Suppose that $s_i \in S_i$ is maximal for $\succcurlyeq^{u_i, \mu_i}$, but not sequentially rational for (U_i, μ_i). Then there is $I \in \mathscr{I}(s_i)$ and $t_i \in S_i(I)$ such that $\mathrm{E}_{\mu_i(\cdot|[I])} U_i(s_i, \cdot) < \mathrm{E}_{\mu_i(\cdot|[I])} U_i(t_i, \cdot)$.

Let $r_i \in S_i$ be a strategy that agrees with s_i everywhere except at information sets that weakly follow I: that is, for every $J \in \mathscr{I}_i$, $r_i(J) = t_i(J)$ if $I \leq J$, and $r_i(J) = s_i(J)$ otherwise. I claim that, for all $(s_{-i}, \theta) \in [I]$,

$$U_i(r_i, s_{-i}, \theta) = u_i(\zeta(r_i, s_{-i}), \theta) = u_i(\zeta(t_i, s_{-i}), \theta) = U_i(t_i, s_{-i}\theta).$$

To see this, note that, by perfect recall, since $s_i, t_i \in S_i(I)$, s_i and t_i take the same actions at every $J \in \mathscr{I}_i$ with $J < I$, and hence (t_i, s_{-i}) reaches the same history $h \in I$ as (s_i, s_{-i}). Hence, so does (r_i, s_{-i}). At I and all subsequent information sets, r_i takes the same actions as t_i, so

(r_i, s_{-i}) reaches the same terminal history as (t_i, s_{-i}). On the other hand, for $(s_{-i}, \theta) \notin [I]$,

$$U_i(r_i, s_{-i}, \theta) = u_i(\zeta(r_i, s_{-i}), \theta) = u_i(\zeta(s_i, s_{-i}), \theta) = U_i(s_i, s_{-i}, \theta).$$

To see this, note that, if $s_{-i} \notin S_{-i}(I)$, by perfect recall $(s_i, s_{-i}) \notin S(I)$, and hence also $(s_i, s_{-i}) \notin S(J)$ for any $J \in \mathscr{I}_i$ with $I \leq J$. Therefore, r_i agrees with s_i at all $J \in \mathscr{I}_i$ such that $(s_i, s_{-i}) \in S(J)$, and hence (r_i, s_{-i}) reaches the same terminal history as (s_i, s_{-i}).

By Definition 3, $p_{i,[I]}([I]) > 0$ and $p_{i,[I]}(E) = p_{i,[I]}([I])\mu_i(E|[I])$ for all measurable $E \subseteq [I]$, so $\mathrm{E}_{\mu_i(\cdot|[I])} U_i(s_i, \cdot) < \mathrm{E}_{\mu_i(\cdot|[I])} U_i(t_i, \cdot)$ implies

$$\int_{[I]} U_i(s_i, s_{-i}, \theta) d\, p_{i,[I]} = p_{i,[I]}([I]) \cdot \mathrm{E}_{\mu_i(\cdot|[I])} U_i(s_i, \cdot) < p_{i,[I]}([I]) \cdot \mathrm{E}_{\mu_i(\cdot|[I])} U_i(t_i, \cdot) = \int_{[I]} U_i(t_i, s_{-i}, \theta) d\, p_{i,[I]}.$$

Therefore,

$$\begin{aligned}
\mathrm{E}_{p_{i,[I]}} U_i(s_i, \cdot) &= \int_{S_{-i} \times \Theta} U_i(s_i, s_{-i}, \theta) d\, p_{i,[I]} = \\
&= \int_{[I]} U_i(s_i, s_{-i}, \theta) d\, p_{i,[I]} + \int_{(S_{-i} \times \Theta) \setminus [I]} U_i(s_i, s_{-i}, \theta) d\, p_{i,[I]} < \\
&< \int_{[I]} U_i(t_i, s_{-i}, \theta) d\, p_{i,[I]} + \int_{(S_{-i} \times \Theta) \setminus [I]} U_i(s_i, s_{-i}, \theta) d\, p_{i,[I]} = \\
&= \int_{[I]} U_i(r_i, s_{-i}, \theta) d\, p_{i,[I]} + \int_{(S_{-i} \times \Theta) \setminus [I]} U_i(r_i, s_{-i}, \theta) d\, p_{i,[I]} = \mathrm{E}_{p_{i,[I]}} U_i(r_i, \cdot).
\end{aligned}$$

Furthermore, consider $F \in \mathscr{F}_i$. Two cases must be considered.

Case 1: $p_{i,F}([I]) = 0$. For such F, trivially

$$\int_{[I]} U_i(s_i, s_{-i}, \theta) p_{i,F} = 0 = \int_{[I]} U_i(r_i, s_{-i}, \theta) p_{i,F}$$

and so

$$\begin{aligned}
\mathrm{E}_{p_{i,F}} U_i(s_i, \cdot) &= \int_{S_{-i} \times \Theta} U_i(s_i, s_{-i}, \theta) d\, p_{i,F} = \\
&= \int_{[I]} U_i(s_i, s_{-i}, \theta) d\, p_{i,F} + \int_{(S_{-i} \times \Theta) \setminus [I]} U_i(s_i, s_{-i}, \theta) d\, p_{i,F} = \\
&= \int_{[I]} U_i(r_i, s_{-i}, \theta) d\, p_{i,F} + \int_{(S_{-i} \times \Theta) \setminus [I]} U_i(r_i, s_{-i}, \theta) d\, p_{i,F} = \mathrm{E}_{p_{i,F}} U_i(r_i, \cdot).
\end{aligned}$$

Case 2: $p_{i,F}([I])>0$. In this case, Lemma 1 part 1 implies that $[I]\triangleright F$.

To conclude the argument, consider first $F\in\mathscr{F}_i$ with $F\triangleright[I]$ and $F\neq[I]$. If $p_{i,F}([I])=0$, then per Case 1, $\mathrm{E}_{p_{i,F}}U_i(r_i,\cdot)=\mathrm{E}_{p_{i,F}}U_i(s_i,\cdot)$; if instead $p_{i,F}([I])>0$, per Case 2 $[I]\triangleright F$, so by condition (1) in Def. 3 $p_{i,F}=p_{i,[I]}$ and so $\mathrm{E}_{p_{i,F}}U_i(r_i,\cdot)>\mathrm{E}_{p_{i,F}}U_i(s_i,\cdot)$. Thus, $\mathrm{E}_{p_{i,[I]}}U_i(r_i,\cdot)>\mathrm{E}_{p_{i,[I]}}U_i(s_i,\cdot)$ and $\mathrm{E}_{p_{i,F}}U_i(r_i,\cdot)\geq\mathrm{E}_{p_{i,F}}U_i(s_i,\cdot)$ for all $F\in\mathscr{F}_i$ with $F\triangleright[I]$: hence, $s_i\not\succcurlyeq^{u_i,\mu_i}r_i$.

On the other hand, consider $F\in\mathscr{F}_i$ such that $\mathrm{E}_{p_{i,F}}U_i(r_i,\cdot)<\mathrm{E}_{p_{i,F}}U_i(s_i,\cdot)$. Then Case 2 applies to F, so $[I]\triangleright F$. Since $\mathrm{E}_{p_{i,[I]}}U_i(r_i,\cdot)>\mathrm{E}_{p_{i,[I]}}U_i(s_i,\cdot)$ and F was chosen arbitrarily, $r_i\succcurlyeq^{u_i,\mu_i}s_i$.

Thus, $r_i\succ^{u_i,\mu_i}s_i$, which contradicts the assumption that s_i was maximal for $\succcurlyeq^{u_i,\mu_i}$. ■

B.2 Elicitation

B.2.1 Additional details on extensive game forms

Begin with additional definitions and observations related to extensive game forms. Fix $\Gamma=(N,H,P,(\mathscr{I}_i)_{i\in N})$.

Actions available at history $h\in H$ are denoted $A(h)$. Histories in H are ordered by the *initial-segment* relation: for $h,h'\in H$, $h<h'$ means that $h=(a_1,\ldots,a_n)$ and $h'=(a_1,\ldots,a_n,a_{n+1},\ldots,a_{n+k})$ for $a_1,\ldots,a_{n+k}\in A$, $n\geq 0$ (the case $n=0$ corresponds to $h=\phi$), and $k\geq 1$; in this case, I will also write $h'=(h,a_{n+1},\ldots,a_{n+k})$. The notation $h\leq h'$ means that either $h=h'$ (i.e. h and h' are the same) or $h<h'$; note that $\phi\leq h$ for all $h\in H$. The precedence relation $<$ extends to information sets as follows: $I<I'$ iff for every $h'\in I'$ there is $h\in I$ with $h<h'$. The notation $I\leq I'$ means that either $I=I'$ or $I<I'$. Notice that $s\in S(h)$ if there exists $z\in Z$ such that $h\leq z$ and $s\in S(z)$; furthermore, for every player $i\in N$ and $I\in\mathscr{I}_i$, $S(I)=\bigcup_{h\in I}S(h)$. Finally, perfect recall implies that, if $s_i,t_i\in S_i(I)$, $J\in\mathscr{I}_i$ and $J<I$, then $s_i(J)=t_i(J)$.

Next, I point out consequences of Def. 7 and introduce additional notation.

The set of actions is $A^*\equiv A\cup\bigcup_{j\neq i}S_j\cup(S_i\times\{f,g\})$. Strategies in the original game are actions

in the elicitation game, except that, for player i, an action specifies both a strategy $s_i \in S_i$ and a pair in $\{f, g\}$.

Whenever it is convenient to do so, I use the more compact notation (s, k, h) to denote the (partial or terminal) history of length at least N, in which i chooses $k \in \{f, g\}$, the strategies commited to are given by the profile s, and the (possibly partial) history of play h results.

Observation 2 For every $(s, k, h), (s', k', h') \in H^*$: $(s, k, h) < (s', k', h')$ iff $s = s'$, $k = k'$ and $h < h'$. Hence (s, k, h) is terminal iff h is terminal.

The set of actions available at a history h^*, denoted $A^*(h^*)$, is defined as usual from the set of histories H^*. It turns out that it is a singleton in the second stage of the game:

Remark 1 *Let $h^* = (s, k, h) \in H^*$ be a history of length at least N. Let $j = P(h) = P^*(h^*)$, and let $I \in \mathscr{I}_i$ be the information set such that $h \in I$. Then $A^*(h^*) = \{s_j(I)\}$.*

Proof: By definition, $a \in A^*(h^*)$ iff $(h^*, a) \in H^*$. Since $h^* = (s, k, h)$, $(h^*, a) \in H^*$ iff $s_j(I) = a$. Therefore, $A^*(h^*) = \{s_j(I)\}$. ■

The family of information sets for player $j \in N$ is denoted by $\mathscr{I}_j^*$, with generic element I^*.

Remark 2 *The game form Γ^* has perfect recall.*

Proof: Denote the experience function for player $j \in N$ in the elicitation game by $X_j^*(\cdot)$. It must be shown that, for all $j \in N$ and $I^* \in \mathscr{I}_j^*$, $h^*, \bar{h}^* \in I^*$ implies $X_j^*(h^*) = X_j^*(\bar{h}^*)$.

For $I^* = I_j^1$, this is immediate, as $X_j^*(h^*) = \emptyset$ for all $h^* \in I_j^1$. Thus, consider $I^* \in \mathscr{I}_j^* \setminus \{I_j^1\}$. I analyze in detail the case $j \in N \setminus \{i\}$: the case $j = i$ is analogous. Write $I^* = \langle s_j, I \rangle$, where $s_j \in S_j$ and $I \in \mathscr{I}_j(s_j)$.

Fix $h^* \in I^*$, so by definition $h^* = (s', k, h)$ for some $s' \in S$ with $s'_j = s_j$, $s' \in S(h)$ and $h \in I$. Let $h_0^*, \ldots, h_n^* \in H^*$ be the collection of all $\bar{h}^* \in (P^*)^{-1}(j)$ such that $\bar{h}^* < h^*$, ordered by the subhistory relation: that is, $h_0^* < h_1^* < \ldots h_n^* < h^*$, and $\bar{h}^* < h^*$ for no other $\bar{h}^* \in H^*$ with $P(\bar{h}^*) = j$. Then

$h_0^* = (s_1, \ldots, s_{j-1})$; furthermore, by Observation 2, for every $m = 1, \ldots, n$, $h_m^* = (s', k, h_m)$ for some $h_m \in P^{-1}(j)$, and $h_1 < h_2 < \ldots < h_n < h$. Moreover, consider an arbitrary $\bar{h} \in P^{-1}(j)$ such that $\bar{h} < h$; then, since $h < \zeta(s')$, also $\bar{h} < \zeta(s')$, i.e., $s' \in S(\bar{h})$. It follows that $(s', k, \bar{h}) \in H^*$, and since $(s', k, \bar{h}) < (s', k, h) = h^*$ by Observation 2, $\bar{h} = h_m$ for some $m = 1, \ldots, n$. Therefore, $\{h_1, \ldots, h_n\}$ is the set of all subhistories of h where j moves.

For every $m = 1, \ldots, n$, let $I_m^* \in \mathscr{I}_j^*$ be such that $h_m^* \in I_m^*$. Since j chooses s_j and Chance chooses p in each history h_m^*, it must be the case that $I_m^* = \langle s_j, I_m \rangle$ for some $I_m \in \mathscr{I}_j$. By the definition of information sets in Γ^*, $h_m \in I_m$. By Remark 1, $A(h_m^*) = s_j(I_m)$. Therefore,

$$X_j^*(h^*) = \Big((I_j^1, s_j), (I_1^*, s_j(I_1)), \ldots, (I_n^*, s_j(I_n))\Big), \quad X_j(h) = \big((I_1, s_j(I_1)), \ldots, (I_n, s_j(I_n))\big).$$

Now repeat the argument for another history $\hat{h}^* = (\hat{s}', \hat{k}, \hat{h}) \in \langle s_j, I \rangle$: then, there must be $\hat{n}$, $\hat{h}_1^*, \ldots, \hat{h}_{\hat{n}}^* \in (P^*)^{-1}(j)$ with $\hat{h}_m^* = (\hat{s}', \hat{k}, \hat{h}_m)$ for each m, and $\hat{I}_1^*, \ldots, \hat{I}_{\hat{n}}^* \in \mathscr{I}_j^*$ with $\hat{h}_m^* \in \hat{I}_m^* = \langle s_j, \hat{I}_m \rangle$ and $h_m \in I_m$ for each m, such that

$$X_j^*(\hat{h}^*) = \Big((I_j^1, s_j), (\hat{I}_1^*, s_j(\hat{I}_1)), \ldots, (\hat{I}_n^*, s_j(\hat{I}_n))\Big), \quad X_j(\hat{h}) = \big((\hat{I}_1, s_j(\hat{I}_1)), \ldots, (\hat{I}_n, s_j(\hat{I}_n))\big).$$

Since Γ has perfect recall and $h, \hat{h} \in I$ by the definition of the information set $\langle s_j, I \rangle$ and the histories $h^*, \hat{h}^*$, it must be the case that $X_j(h) = X_j(\hat{h})$. Thus, $n = \hat{n}$, and for every $m = 1, \ldots, n$, $I_m = \hat{I}_m$, so that also $s_j(I_m) = s_j(\hat{I}_m)$. But then, for every $m = 1, \ldots, n$, $I_m^* = \langle s_j, I_m \rangle = \langle s_j, \hat{I}_m \rangle = \hat{I}_m^*$. Therefore, $X_j^*(h^*) = X_j^*(\hat{h}^*)$, as required.

The argument for player $j = i$ is essentially identical, except that (i) at I_i^1, i also chooses a fixed $k \in \{f, g\}$; and (ii) I^* is of the form $\langle s_i, k, I \rangle$, and so are all information sets I_m^*. ■

B.2.2 Analysis of the elicitation game and proof of Theorem 2.

The terminal history function $\zeta^* : S^* \to Z^*$ and, for given Bernoulli utilities $u_j : X \to \mathbb{R}$, $j \in N$, the reduced-form payoff functions $U_j^* : S^* \times \Theta^* \to \mathbb{R}$, are defined as usual.

Remark 3 *Fix a profile $s^* \in S^*$ such that $s_i^*(I_i^1) = (s_i, k)$ and $s_j^*(I_j^1) = s_j$ for $j \neq i$. Then*

$$\zeta^*(s^*) = (s, k, \zeta(s)).$$

Furthermore,

$$U_j^*(s^*, w, \theta) = \begin{cases} U_j(s, w) & \theta = o; \\ U_j(s, w) & \theta = a, j \neq i; \\ u_j(k(s_{-i}, w)) & \theta = a, j = i, \end{cases}$$

Proof: Recall that $\zeta^*(s^*)$ is uniquely defined by induction on the histories $h^* \in H^*$ generated by s^*. In particular, the length-N history generated by s^* is $(s_1, \ldots, s_{i-1}, (s_i, k), s_{i+1}, \ldots, s_N)$ by assumption, denoted (s, k). There is a unique terminal history $z^* \in Z^*$ whose length-N initial segment is (s, k), namely $z^* = (s, k, \zeta(s))$. Therefore, $\zeta^*(s^*) = z^*$, and the first claim follows.

Now recall that $U_j^*(s^*, w, \theta) = u_j(\xi_j^*(\zeta^*(s^*), w, \theta))$. As was just shown, $\zeta^*(s^*) = (s, k, \zeta(s))$. From the definition of ξ_j^*, $\xi_j^*((s, k, \zeta(s)), w, o) = \xi_j(\zeta(s), w)$. Finally, by definition, $u_j(\xi_j(\zeta(s), w)) = U_j(s, w)$. The other cases are similar. ■

Remark 1 implies that, for every $j \neq i$ and $s_j \in S_j$, there is a unique strategy $s_j^* \in S_j^*$ such that $s_j^*(I_j^1) = s_j$; this is because, at every information set $I^* \in \mathscr{I}_j^* \setminus \{I_j^1\}$, a single action is available. Therefore, the map $\sigma_j : S_j \to S_j^*$ defined by letting $\sigma_j(s_j) = s_j^*$, where $s_j^*(I_j^1) = s_j$, is a bijection. A similar construction applies to player i, except that i also chooses a pair $k \in \{f, g\}$ at I_i^1. Thus, for every $k \in \{f, g\}$, define a bijection $\sigma_{i,k} : S_i \to S_i^*$ by letting $\sigma_{i,k}(s_i) = s_i^*$, where $s_i^*(I_i^1) = (s_i, k)$.

As usual, $\sigma_{-i}(s_{-i}) = (\sigma_j(s_j))_{j \neq i}$ for all $s_{-i} \in S_{-i}$. Similarly, for $k \in \{f, g\}$, $\sigma_k(s^*) = (\sigma_{-i}(s_{-i}), \sigma_{i,k}(s_i))$ and $\sigma_{-j,k}(s_{-j}) = \big((\sigma_\ell(s_\ell))_{\ell \neq i,j}, \sigma_{i,k}(s_i)\big)$. It is also convenient to define the correspondence $\sigma_{-j} : S_{-j} \to 2^{S_{-j}^*}$ by letting $\sigma_{-j}(s_{-j}) = \{\sigma_{-j,f}(s_{-j}), \sigma_{-j,g}(s_{-j})\}$ for all $s_{-j} \in S_{-j}$. For any set $T \subseteq S_{-j}$, $\sigma_{-j,k}(T) = \bigcup_{s_{-j} \in T} \{\sigma_{-j,k}(s_{-j})\}$ and $\sigma_{-j}(T) = \bigcup_{s_{-j} \in T} \sigma_{-j}(s_{-j})$.

The following result shows that the maps $\sigma_j(\cdot)$ provide a convenient link between histories or information sets in Γ and their counterparts in Γ^*.

Lemma 2

(i) For every $s \in S$ *and* $k \in \{f, g\}$, $\zeta^*(\sigma_{i,k}(s_i), \sigma_{-i}(s_{-i})) = (s, k, \zeta(s))$;

(ii) for every $h \in H$, $s \in S$, *and* $k \in \{f, g\}$: $s \in S(h)$ *iff* $(s, k, h) \in H^*$ *and* $(\sigma_{i,k}(s_i), \sigma_{-i}(s_{-i})) \in S^*\big((s, k, h)\big)$;

(iii) For every $j \in N \setminus \{i\}$ *and* $s_j \in S_j$, $\mathscr{I}_j^*(\sigma_j(s_j)) = \{I_j^1\} \cup \{\langle s_j, I\rangle : I \in \mathscr{I}_j(s_j)\}$;

(iv) for every $s_i \in S_i$ *and* $k \in \{f, g\}$, $\mathscr{I}_i^*(\sigma_{i,k}(s_i)) = \{I_i^1\} \cup \{\langle s_i, k, I\rangle : I \in \mathscr{I}_i(s_i)\}$.

Proof: Part (i) just restates Remark 3 in terms of the maps $\sigma_j(\cdot)$. For part (ii), the "if" direction is immediate because $(s, k, h) \in H^*$ implies that $h < \zeta(s)$, i.e., $s \in S(h)$; for the "only if" direction, $s \in S(h)$ implies that $h < \zeta(s)$, so $(s, k, h) \in H^*$ and $(s, k, h) < (s, k, \zeta(s)) = \zeta^*(\sigma_{i,k}(s_i), \sigma_{-i}(s_{-i}))$, where the equality follows from part (i): that is, $(\sigma_{i,k}(s_i), \sigma_{-i}(s_{-i})) \in S^*((s, k, h))$, as claimed.

For part (iii), fix $j \in N \setminus \{i\}$ and $s_j \in S_j$. Denote the rhs of the equality by $\mathscr{J}^*$. Consider $I^* \in \mathscr{I}_j^*$. If $I^* = I_j^1$, then $I^* \in \mathscr{I}_j^*(\sigma_j(s_j)) \cap \mathscr{J}^*$. Next, suppose $I^* = \langle t_j, I\rangle$ for some $t_j \in S_j$ and $I \in \mathscr{I}_j(t_j)$. If $t_j \neq s_j$, then $I^* \notin \mathscr{I}_j^*(\sigma_j(s_j))$: on one hand, since $[\sigma_j(s_j)](I_j^1) = s_j$, the j-th element of any history $h^* < \zeta^*(\sigma_j(s_j), s_{-j}^*)$ is s_j by part (i); on the other, the j-th element of every history $h^* \in I^*$ is by definition $t_j \neq s_j$. Furthermore, in this case also $I^* \notin \mathscr{J}^*$. Thus, consider $t_j = s_j$, so $\langle s_j, I\rangle \in \mathscr{J}^*$. Fix $h^* \in I^*$; then by construction $h^* = (s', k, h) \in H^*$ for some $k \in \{f, g\}$ $h \in H$, and $s' \in S(h)$; moreover, $h^* \in I^*$ implies $s'_j = s_j$ and $h \in I$. Then, by part (ii), $(s_j, s'_{-j}) \in S(h)$ implies $(\sigma_{i,k}(s'_i), \sigma_j(s_j), (\sigma_\ell(s'_\ell))_{\ell \in N \setminus \{i,j\}}) \in S^*((s, k, h)) = S^*(h^*) \subseteq S^*(I^*)$, and so $I^* \in \mathscr{I}_j^*(\sigma_j(s_j))$.

For part (iv), letting $\mathscr{J}^*$ denote the rhs of the equality in the claim, again $I_i^1 \in \mathscr{I}_i^*(\sigma_{i,k}(s_i)) \cap \mathscr{J}^*$. Also, adapting the argument for part (iii), $\langle t_i, k', I\rangle \in \mathscr{I}_i^*(\sigma_{i,k}(s_i)) \cap \mathscr{J}^*$ if $t_i = s_i$ and $k' = k$, and $\langle t_i, k', I\rangle \notin \mathscr{I}_i^*(\sigma_{i,k}(s_i)) \cap \mathscr{J}^*$ otherwise. ■

The following Lemma formalizes the intuition that information obtained during the implementation phase of the elicitation game is "the same" as in the original game.

Lemma 3

1. For all $j \in N$ *and subsets* $C, D \subseteq S_{-j}$, $C \subseteq D$ *iff* $\sigma_{-j}(C) \subseteq \sigma_{-j}(D)$;

2. *for all $j \neq i$, if $I^* = \langle s_j, I\rangle \in \mathscr{I}_j^*$ then $S_{-j}^*(I^*) = \sigma_{-j}(S_{-j}(I))$;*

3. *if $I^* = \langle s_i, k, I\rangle \in \mathscr{I}_i^*$, then $S_{-i}^*(I^*) = \sigma_{-i}(S_{-i}(I))$.*

Proof: 1: "$\Rightarrow$" is obvious; for "$\Leftarrow$," suppose that $C \not\subseteq D$, so there exists $s_{-j} \in C \setminus D$. If $j \neq i$, then $\sigma_{-j}(s_{-j}) = \{\sigma_{-j,f}(s_{-j}), \sigma_{-j,g}(s_{-j})\}$, and $\sigma_{-j,k}(\cdot) : S_{-j} \to S_{-j}^*$ is a bijection for every $k \in \{f,g\}$; if $j = i$, then $\sigma_{-i} : S_{-i} \to S_{-i}^*$ is itself a bijection. In either case, there cannot be any $t_{-j} \in D \subseteq S_{-j} \setminus \{s_{-j}\}$ such that $\sigma_{-j}(t_{-j}) = \sigma_{-j}(s_{-j})$; therefore, $\sigma_{-j}(s_{-i}) \notin \sigma_{-j}(D)$, and so $\sigma_{-j}(C) \not\subseteq \sigma_{-j}(D)$.

2: Fix $s_{-j}^* \in S_{-j}^*(I^*)$ arbitrarily; thus, there is $s_j^* \in S_j^*$ with $(s_j^*, s_{-j}^*) \in S^*(I^*)$. In particular, $(s_j^*, s_{-j}^*) \in S^*(h^*)$ for some $h^* \equiv (s', k, h) \in I^*$. This implies that $s'_\ell = s_\ell^*(I_\ell^1)$ for $\ell \neq i$ and $(s'_i, k) = s_i^*(I_i^1)$. Then, by the definition of $\sigma_{-j,k}(\cdot)$, $s_{-j}^* = \sigma_{-j,k}(s'_{-j})$. But by the definition of I^*, $h \in I$ and $s' \in S(h) \subseteq S(I)$. Therefore, $s_{-j}^* = \sigma_{-j,k}(s'_{-j}) \in \sigma_{-j,k}(S_{-j}(I)) \subset \sigma_{-j}(S_{-j}(I))$.

Conversely, fix $s_{-j}^* \in \sigma_{-j}(S_{-j}(I))$, and let $s_\ell = \sigma_\ell^{-1}(s_\ell^*) = s_\ell^*(I_\ell^1)$ for $\ell \neq j, i$, and $(s_i, k) = s_i^*(I_i^1)$. Then $s_{-j} \in S_{-j}(I)$. Since the original game has perfect recall and by assumption $s_j \in S_j(I)$, $s \equiv (s_j, s_{-j}) \in S(h)$ for some $h \in I$. Furthermore, $h^* \equiv (s, k, h) \in H^*$, so indeed $h^* \in I^*$. By Remark 3, $\zeta^*(s^*) = (s, k, \zeta(s))$; since $s \in S(h)$, $h < \zeta(s)$, so $h^* = (s, k, h) < \zeta^*(s^*)$, i.e., $s^* \in S^*(h^*) \subseteq S^*(I^*)$. Therefore, $s_{-j}^* \in S_{-j}^*(I^*)$.

3: the proof requires only minor modifications to the argument for 2, so it is omitted. ■

Now turn to the set of conditioning events, defined as usual as $\mathscr{F}_j^* = \{\Omega_j^*\} \cup \{[I^*] : I^* \in \mathscr{I}_j^*\}$ for every $j \in N$. For every player $j \in N$, let $\varphi_j : \mathscr{F}_j \to 2^{\Omega_j^*}$ be defined by

$$\varphi_j(F) = \sigma_{-j}(\mathrm{proj}_{S_{-j}} F) \times \Theta^*. \tag{15}$$

Lemma 4 *For every player $j \in N$:*

1. *$\mathscr{F}_j^* = \{\varphi_{-j}(F) : F \in \mathscr{F}_j\}$;*

2. *for all $F, G \in \mathscr{F}_j$, $F \subseteq G$ iff $\varphi_j(F) \subseteq \varphi_j(G)$.*

Proof: 1: Denote by $\mathscr{G}_j^*$ the set on the r.h.s. Fix $F^* \in \mathscr{F}_j^*$, so $F^* = S_{-j}^*(I^*) \times \Theta^*$ for some $I^* \in \mathscr{I}_j^*$. If $I^* = I_j^1$, then $F^* = \Omega_j^*$; since $\sigma_{-j}(\mathrm{proj}_{-j}(\Omega_j)) \times \Theta^* = S_{-j}^* \times \Theta^* = \Omega_j^*$ and $\Omega_j \in \mathscr{F}_j$, $F^* \in \mathscr{G}_j^*$. If

instead $F^* = [I^*]$ for some $I^* \in \mathscr{I}_j^* \setminus \{I_j^1\}$, there are two cases. If $j \neq i$, then $I^* = \langle s_j, I\rangle$, with $s_j \in S_j$ and $I \in \mathscr{I}_j$, and by Lemma 3, $S_{-j}^*(I^*) = \sigma_{-j}(S_{-j}(I))$. If instead $j = i$, then $i^* = \langle s_i, k, I\rangle$ for some $s_i \in S_i$, $k \in \{f, g\}$, and $I \in \mathscr{I}_i$; again, Lemma 3 implies that $S_{-i}^*(I^*) = \sigma_{-i}(S_{-i}(I))$. Hence, in either case, $F^* = S_{-j}^*(I^*) \times \Theta^* = \sigma_{-j}(S_{-j}(I)) \times \Theta^* = \varphi_{-j}([I]) \times \Theta^*$, so $F^* \in \mathscr{G}_j^*$.

Conversely, fix $F \in \mathscr{F}_j$. If $F = \Omega$, then $\varphi_{-j}(\Omega_j) = \sigma_{-j}(\text{proj}_{S_{-j}}\Omega_j) \times \Theta^* = \sigma_{-j}(S_{-j}) \times \Theta^* = S_{-j}^* \times \Theta^* = \Omega_j^* \in \mathscr{F}_j^*$. If instead $F = [I] = S_{-i}(I) \times \Theta$, then by Lemma 3 $\varphi_{-j}(F) = \sigma_{-j}(S_{-j}(I)) \times \Theta^* = S_{-j}^*(I^*) \times \Theta^*$, where $I^* = \langle s_j, I\rangle$ for some $s_j \in S_j(I)$ if $j \neq i$, and $I^* = \langle s_i, k, I\rangle$ for some $s_i \in S_i(I)$ and $k \in \{f, g\}$ if $j = i$. In either case, $\varphi_{-j}(F) \in \mathscr{F}_j^*$.

2: immediate from the definition of $\varphi_{-i}(\cdot)$ and Lemma 3 part 1. ■

Now consider conditional beliefs. For every $j \in N$, consider the Sigma-algebra $\Sigma_j^* = 2^{S_{-j}^*} \times \mathscr{T}^*$, where $\mathscr{T}^* = \Theta \times 2^{\{o,c\}}$. Then a CPS for j in the elicitation game is an element $\mu_j^* \in cpr(\Sigma_j^*, \mathscr{F}_j^*)$. Assume henceforth that μ_j^* is the extension of a CPS $\mu_j \in cpr(\Sigma_j, \mathscr{F}_j)$, in the sense of Definition 8. First, I verify that such an extension always exists, and is unique for player i.

Lemma 5 *For every $j \in N$, there exists an extension $\mu_j^* \in cpr(\Sigma_j^*, \mathscr{F}_j^*)$ of μ_j; furthermore, if $j = i$, then such an extension is unique.*

Proof: Observe first that, for $j \neq i$, since S_{-j} and hence S_{-j}^* is finite, every event $E \subseteq \Omega_j^*$ is a union of disjoint sets of the form $\{\sigma_{-j,k}(s_{-j}) \times \{r\} \times U$, for $s_{-j} \in S_{-j}$, $k \in \{f, g\}$, $r \in \{o, a\}$, and $U \in \Theta$ [in particular, for given s_{-j}, k, and r, $U = \{w : (\sigma_{-j,k}(s_{-j}), r, w) \in E\}$]. Similarly, for $j = i$, every $E \subseteq \Omega_i^*$ is a union of disjoint sets of the form $\{\sigma_{-i,k}(s_{-i}) \times \{r\} \times U$, for $s_{-i} \in S_{-i}$, $r \in \{o, a\}$, and $U \in \mathscr{T}$. Therefore, a probability measure on Ω_j^* is fully determined by the probabilities it assigns to sets of the form just described.

For $j \neq i$, and for all $s_{-j} \in S_{-j}$, by definition $\sigma_{-j}(s_{-j}) = \{\sigma_{-j,f}(s_{-j}), \sigma_{-j,g}(s_{-j})\}$. Thus, Equations 9 and 10 do not uniquely determine how the (marginal) probability of $\sigma_{-j}(s_{-j})$ is split between $\sigma_{-j,f}(s_{-j})$ and $\sigma_{-j,g}(s_{-j})$. To remedy this, define probability measures on Ω_i^* as fol-

lows: for every $s_{-j} \in S_{-j}$, $r \in \{o,a\}$, and $U \in \Theta$,

$$\mu_j^*\left(\{\sigma_{-j,f}(s_{-j})\} \times \{r\} \times U \middle| \Omega_j^*\right) = \frac{1}{2}\mu_j\left(\{s_{-j}\} \times U \middle| \Omega_j\right)$$

and for every $I \in \mathscr{I}_j$,

$$\mu_j^*\left(\{\sigma_{-j,f}(s_{-j})\} \times \{r\} \times U \middle| \sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta\right) = \frac{1}{2}\mu_j\left(\{s_{-j}\} \times U \middle| S_{-j}(I) \times \Theta\right).$$

Equations 9 and 10 then imply that

$$\mu_j^*\left(\{\sigma_{-j,g}(s_{-j})\} \times \{r\} \times U \middle| \Omega_j^*\right) = \mu_j^*\left(\{\sigma_{-j,g}(s_{-j})\} \times \{r\} \times U \middle| \sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta\right) = 0$$

for all $s_{-j} \in S_{-j}$, $r \in \{o,a\}$, $U \in \Theta$, and $I \in \mathscr{I}_j$. This completes the assignment of probabilities to sets of the form $\{\sigma_{-j,k}(s_{-j})\} \times \{r\} \times U$; as noted above, this implies that $\mu_j^*(\cdot|\Omega_j^*)$ and $\mu_j^*(\cdot|\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta)$ are uniquely determined.

It is routine to verify that the set functions just defined are indeed probability measures on Σ_j^*. Furthermore,

$$\mu_j^*(\Omega_j^*|\Omega_j^*) \geq \mu_j^*(\sigma_{-j,f}(S_{-j}) \times \{o\} \times \Theta|\Omega_j) + \mu_j^*(\sigma_{-j,f}(S_{-j}) \times \{a\} \times \Theta|\Omega_j) = \frac{1}{2}\mu_j(S_{-j} \times \Theta|\Omega_j) + \frac{1}{2}\mu_j(S_{-j} \times \Theta|\Omega_j) = 1$$

and similarly $\mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta|\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta) = 1$ for any $I \in \mathscr{I}_j$.

Finally, fix $I \in \mathscr{I}_j$, $F \in \mathscr{F}_j$, $s_{-j} \in S_{-j}$, $k \in \{f,g\}$, $r \in \{o,a\}$, and $U \in \Theta$. By Lemma 3, $s_{-j} \in S_{-j}(I)$ iff $\sigma_{-j,k}(s_{-j}) \in \sigma_{-j}(S_{-j}(I))$; and by Lemma 4, $[I] \subseteq F$ iff $\varphi_{-j}([I]) \subseteq \varphi_{-j}(F)$. Finally, if $F = \Omega_j$ then $\varphi_{-j}(F) = \Omega_j^*$; and if $F = [J]$ for some $J \in \mathscr{I}_j$, then $\varphi_{-j}(F) = \sigma_{-j}(S_{-j}(J)) \times \{o,a\} \times \Theta$. Now suppose that $[I] \subseteq F$ and $s_{-j} \in S_{-j}(I)$; then

$$\mu_j^*(\{\sigma_{-j,f}(s_{-j}\} \times \{r\} \times U|\varphi_{-j}(F)) = \frac{1}{2}\mu_j(\{s_{-j}\} \times U|F) = \frac{1}{2}\mu_j(\{s_{-j}\} \times U|S_{-j}(I) \times \Theta)\mu_j(S_{-j}(I) \times \Theta|F) =$$
$$= \mu_j^*(\{\sigma_{-j,f}(s_{-j})\} \times \{r\} \times U|\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta)\mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta|\varphi_{-j}(F));$$

furthermore,

$$0 = \mu_j^*(\{\sigma_{-j,g}\} \times \{r\} \times U|\varphi_{-j}(F)) = 0 \cdot \mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta|\varphi_{-j}(F)) =$$
$$= \mu_j^*(\{\sigma_{-j,g}(s_{-j})\} \times \{r\} \times U|\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta)\mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \{o,a\} \times \Theta|\varphi_{-j}(F)).$$

Thus, μ_j^* is a CPS, per Definition 1.

Now consider player i. Since σ_{-i} maps each $s_{-i} \in S_{-i}$ to a profile $s_{-i}^* \in S_{-i}^*$, Equations 7 and 8 uniquely define probability measures on $S_{-i}^* \times \{o,a\} \times \Theta = \Omega_i^*$. To verify that the conditions in Def. 1 are satisfied, one can proceed as in the case $j \neq i$. ■

Furthermore, assume that each μ_j admits a basis $\mathbf{p}_j$. Finally, for every $j \in N$, let $\triangleright_j$ and $\triangleright_j^*$ denote the plausibility ordering induced by μ_j and μ_j^* respectively, as per Definition 2.

Lemma 6 *For every $j \in N$, and every $F, G \in \mathscr{F}_j$:*

1. $\mu_j^*(\varphi_{-j}(F)|\varphi_{-j}(G)) = \mu_j(F|G)$;
2. $\varphi_{-j}(F) \triangleright_j^* \varphi_{-j}(G)$ *iff* $F \triangleright_j G$.

Proof: 1: if $F = [I]$ for some $I \in \mathscr{I}_j$, then

$$\begin{aligned}
&\mu_j^*(\varphi_{-j}(F)|\varphi_{-j}(G)) = \mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \Theta \times \{o,a\}|\varphi_{-j}(G)) = \\
=&\mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \Theta \times \{o\}|\varphi_{-j}(G)) + \mu_j^*(\sigma_{-j}(S_{-j}(I)) \times \Theta \times \{a\}|\varphi_{-j}(G)) = \\
=&\frac{1}{2}\mu_j(S_{-j}(I) \times \Theta|G) + \frac{1}{2}\mu_j(S_{-j}(I) \times \Theta|G) = \mu_j(F|G).
\end{aligned}$$

If instead $F = \Omega$, then $\mu(F|G) = 1$ and $\mu_j^*(\varphi_{-j}(F)|\varphi_{-j}(G)) = \mu_j^*(\Omega_j^*|\varphi_{-j}(G)) = 1$, so the claim holds in this case, too.

2: suppose $F \triangleright_j G$, so there is a sequence $F_1, \ldots, F_N \in \mathscr{F}_j$ such that $F_1 = G$, $F_N = F$, and $\mu_j(F_{n+1}|F_n) > 0$ for $n = 1, \ldots, N-1$. Then the sequence $\varphi_{-j}(F_1), \ldots, \varphi_{-j}(F_N)$ lies in $\mathscr{F}_j^*$ by Lemma 1, satisfies $\varphi_{-j}(F_1) = \varphi_{-j}(G)$ and $\varphi_{-j}(F_N) = \varphi_{-j}(F)$, and is such that $\mu_j^*(\varphi_{-j}(F_{n+1})|\varphi_{-j}(F_n)) = \mu_j(F_{n+1}|F_n) > 0$ by part 1 of this Lemma. Thus, $\varphi_{-j}(F) \triangleright_j^* \varphi_{-j}(G)$.

Conversely, suppose that $\varphi_{-j}(F) \triangleright_j^* \varphi_{-j}(G)$, so there is $F_1^*, \ldots, F_N^* \in \mathscr{F}_j^*$ such that $F_1^* = \varphi_{-j}(G)$, $F_N^* = \varphi_{-j}(F)$, and $\mu_j^*(F_{n+1}^*|F_n^*) > 0$ for all $n = 1, \ldots, N-1$. By Lemma 4, for every n there is $f \in \mathscr{F}_j$ such that $F_n^* = \sigma_{-j}(F_n)$; furthermore, by part 2 of the same Lemma, $F_1 = G$ and $F_N = F$. Finally, by part 1 of this Lemma, $\mu_j(F_{n+1}|F_n) = \mu_j^*(F_{n+1}^*|F_n^*) > 0$ for all $n = 1, \ldots, N-1$; thus, $F \triangleright_j G$. ■

Lemma 7 *For every $j \in N$, the CPS μ_j^* admits a basis $\mathbf{p}_j^* = (p_{j,F^*}^*)_{F^* \in \mathscr{F}_j^*}$. In particular, for all $F \in \mathscr{F}_j, C \subseteq S_{-j}, U \in \Theta$,*

$$p_{j,\varphi_{-j}(F)}^*(\sigma_{-j}(C) \times U \times \{o\}) = p_{j,\varphi_{-j}(F)}^*(\sigma_{-j}(C) \times U \times \{a\}) = \frac{1}{2} p_{j,F}(C \times U). \tag{16}$$

Proof: Since $\mathscr{F}_j^* = \{\varphi_{-j}(F) : F \in \mathscr{F}_j\}$ by Lemma 4, Eq. (16) defines a collection $\mathbf{p}_j^* = (p_{j,F^*}^*)_{F^* \in \mathscr{F}_j^*}$. I now show that $\mathbf{p}_j^*$ is a basis for μ_j^*.

Fix $F, G \in \mathscr{F}_j$. By Lemma 2, $F \triangleright_j G$ iff $\varphi_{-j}(F) \triangleright_j^* \varphi_{-j}(G)$. Hence, $p_{j,\varphi_{-j}(F)}^* = p_{j,\varphi_{-j}(G)}^*$ iff $p_{j,F} = p_{j,G}$ iff $F \triangleright_j G$ iff $\varphi_{-j}(F) \triangleright_j^* \varphi_{-j}^*(G)$.

Similarly, for $F \in \mathscr{F}_j$,

$$\begin{aligned}
&p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{G^* : G^* \in \mathscr{F}_j^*, G^* \triangleright_j^* \varphi_{-j}(F), \varphi_{-j}(F) \triangleright_j^* G^*\}\Big) = \\
=&p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{\varphi_{-j}(G) : G \in \mathscr{F}_j, \varphi_{-j}(G) \triangleright_j^* \varphi_{-j}(F), \varphi_{-j}(F) \triangleright_j^* \varphi_{-j}(G)\}\Big) = \\
=&p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{\varphi_{-j}(G) : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) = \\
=&p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{\sigma_{-j}(\mathrm{proj}_{S_{-j}} G) \times \Theta \times \{o\} : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) + \\
&+ p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{\sigma_{-j}(\mathrm{proj}_{S_{-j}} G) \times \Theta \times \{a\} : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) = \\
=&p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{\sigma_{-j}(\mathrm{proj}_{S_{-j}} G) : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\} \times \Theta \times \{o\}\Big) + \\
&+ p_{j,\varphi_{-j}(F)}^*\Big(\bigcup\{\sigma_{-j}(\mathrm{proj}_{S_{-j}} G) : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\} \times \Theta \times \{a\}\Big) = \\
=&p_{j,\varphi_{-j}(F)}^*\Big(\sigma_{-j}\Big(\bigcup\{\mathrm{proj}_{S_{-j}} G : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) \times \Theta \times \{o\}\Big) + \\
&+ p_{j,\varphi_{-j}(F)}^*\Big(\sigma_{-j}\Big(\bigcup\{\mathrm{proj}_{S_{-j}} G : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) \times \Theta \times \{a\}\Big) = \\
=&\frac{1}{2} p_{j,F}\Big(\bigcup\{\mathrm{proj}_{S_{-j}} G\} : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\} \times \Theta\Big) + \frac{1}{2} p_{j,F}\Big(\bigcup\{\mathrm{proj}_{S_{-j}} G\} : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\} \times \Theta\Big) = \\
=&\frac{1}{2} p_{j,F}\Big(\bigcup\{\mathrm{proj}_{S_{-j}} G\} \times \Theta : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) + \frac{1}{2} p_{j,F}\Big(\bigcup\{\mathrm{proj}_{S_{-j}} G\} \times \Theta : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) = \\
=&\frac{1}{2} p_{j,F}\Big(\bigcup\{G : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\} \times \Theta\Big) + \frac{1}{2} p_{j,F}\Big(\bigcup\{G : G \in \mathscr{F}_j, G \triangleright_j F, F \triangleright_j G\}\Big) = 1,
\end{aligned}$$

where the penultimate equality follows because $G = \mathrm{proj}_{S_{-j}} G \times \Theta$ for all $G \in \mathscr{F}_j$, and the last one from the fact that $\mathbf{p}_j$ is a basis for μ_j.

Finally, fix $F \in \mathscr{F}_j$; then

$$p^*_{j,\varphi_{-j}(F)}(\varphi_{-j}(F)) = p^*_{j,\varphi_{-j}(F)}(\sigma_{-j}(\mathrm{proj}_{S_{-j}} F) \times \Theta \times \{o,a\}) =$$
$$= \sum_{\theta \in \{o,a\}} p_{j,\varphi_{-j}(F)}(\mathrm{proj}_{S_{-j}}(F) \times \Theta \times \{\theta\}) = \frac{1}{2} \sum_{\theta \in \{o,a\}} p_{j,F}(\mathrm{proj}_{S_{-j}}(F) \times \Theta) = p_{j,F}(F) > 0;$$

furthermore, for $C \subseteq \mathrm{proj}_{S_{-j}}(F)$, $U \in \Theta$, and $\theta \in \{o,a\}$,

$$\frac{p^*_{j,\varphi_{-j}(F)}(\sigma_{-j}(C) \times U \times \{\theta\})}{p^*_{j,\varphi_{-j}(F)}(\varphi_{-j}(F))} = \frac{\frac{1}{2} p_{j,F}(C \times U)}{p_{j,F}(F)} = \frac{1}{2} \mu_j(C \times U | F) = \mu^*_j(\sigma_{-j}(C) \times U \times \{\theta\} | \varphi_{-j}(F)).$$

Thus, $\mathbf{p}^*_j$ is a basis for μ^*_j. ■

Proof of Theorem 2: Lemma 5 shows that, for every $j \in N$, every $\mu_j \in cpr(\Sigma_j, \mathscr{F}_j)$ admits an extension. Lemma 7 shows that, if $\mu_j \in cpr(\Omega_i, \mathscr{F}_i)$ admits a basis, so does any extension μ^*_j of μ_j. Eq. (16) in Lemma 7 and Remark 3 imply that, for all $j \neq i$, $s_j \in S_j$, and $F \in \mathscr{F}_j$,

$$\mathrm{E}_{p^*_{j,\varphi_{-j}(F)}} U^*_j(\sigma_j(s_j), \cdot) = \mathrm{E}_{p_{j,F}} U_j(s_j, \cdot);$$

similarly, for all $s_i \in S_i$, $k \in \{f, g\}$, and $F \in \mathscr{F}_i$,

$$\mathrm{E}_{p^*_{i,\varphi_{-i}(F)}} U^*_i(\sigma_{i,k}(s_i), \cdot) = \frac{1}{2} \mathrm{E}_{p_{i,F}} U_i(s_i, \cdot) + \frac{1}{2} \mathrm{E}_{p_{i,F}} u_i \circ k.$$

Finally, Lemma 6 states that, for any $j \in N$ and conditioning events $F, G \in \mathscr{F}_j$, $F \triangleright_j G$ iff $\varphi_{-j}(F) \triangleright_j \varphi_{-j}(G)$. Statements 1–3 in the Theorem now follow immediately. ■

B.2.3 Proof of Proposition 1

To simplify the notation, denote $\succcurlyeq^{u_i,\mu_i}$ by $\succcurlyeq_i$. Let $F = [I]$. That the number α, if it exists, is unique, follows from the fact that $y\,F\,x_0 \succ_i x\,G\,x_0$ and $y\,F\,x_0 \prec_i x\,G\,x_0$ cannot both hold.

Thus, consider y such that $u_i(y) > \mu_i(G|F)$. Note that

$$\mathrm{E}_{p_{i,F}} u_i \circ y\,F\,x_0 = u_i(y) p_{i,F}(F) + u_i(x_0)[1 - p_{i,F}(F)] = u_i(y) p_{i,F}(F) \quad \text{and}$$
$$\mathrm{E}_{p_{i,F}} u_i \circ x\,G\,x_0 = u_i(x) p_{i,F}(G) + u_i(x_0)[1 - p_{i,F}(G)] = p_{i,F}(G) = \mu_i(G|F) p_{i,F}(F),$$

where the last equality follows from the fact that $\mathbf{p}_i$ is a basis for μ_i. The same assumption implies that $p_{i,F}(F) > 0$. Therefore, it is immediate that

$$u_i(y) > \mu_i(G|F) \Longrightarrow \mathrm{E}_{p_{i,F}}\, u_i \circ y\, F\, x_0 > \mathrm{E}_{p_{i,F}}\, u_i \circ x\, G\, x_0 \quad \text{and} \quad u_i(y) < \mu_i(G|F) \Longrightarrow \mathrm{E}_{p_{i,F}}\, u_i \circ y\, F\, x_0 < \mathrm{E}_{p_{i,F}}\, u_i \circ x\, G\, x_0.$$

Now consider $K \in \mathscr{F}_i$ such that $K \triangleright_i F$. If $p_F = p_K$, the above implications hold for K as well. Otherwise, by Lemma 1 part 2, $p_K(F) = 0$, and so $\mathrm{E}_{p_{i,K}}\, u_i \circ y\, F\, x_0 = u_i(x_0) = \mathrm{E}_{p_{i,K}}\, u_i \circ x\, G\, x_0$, because $G \subseteq F$. Therefore,

$$u_i(y) > \mu_i(G|F) \Longrightarrow y\, F\, x_0 \not\prec_i x\, G\, x_0 \quad \text{and} \quad u_i(y) < \mu_i(G|F) \Longrightarrow y\, F\, x_0 \not\succ_i x\, G\, x_0.$$

Finally, suppose that $u_i(y) > \mu_i(G|F)$ and $K \in \mathscr{F}_i$ is such that $\mathrm{E}_{p_{i,K}}\, u_i \circ y\, F\, x_0 < \mathrm{E}_{p_{i,K}}\, u_i \circ x\, G\, x_0$. Then $p_K(F) > 0$, so by Lemma 1 part 1, $F \triangleright_i K$. Since $\mathrm{E}_{p_{i,F}}\, u_i \circ y\, F\, x_0 > \mathrm{E}_{p_{i,F}}\, u_i \circ x\, G\, x_0$ and K was arbitrary, $y\, F\, x_0 \succcurlyeq_i x\, G\, x_0$. Similarly, $u_i(y) < \mu_i(G|F)$ implies that $y\, F\, x_0 \preccurlyeq_i x\, G\, x_0$. Thus,

$$u_i(y) > \mu_i(G|F) \Longrightarrow y\, F\, x_0 \succ_i x\, F\, x_0 \quad \text{and} \quad u_i(y) < \mu_i(G|F) \Longrightarrow y\, F\, x_0 \prec_i x\, G\, x_0,$$

so one can take $\alpha = \mu_i(G|F)$. ■

References

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

Geir B Asheim and Andrés Perea. Sequential and quasi-perfect rationalizability in extensive games. *Games and Economic Behavior*, 53(1):15–42, 2005.

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

P. Battigalli. Strategic independence and perfect Bayesian equilibria. *Journal of Economic Theory*, 70(1):201–234, 1996. ISSN 0022-0531.

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

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

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.

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.

Adam Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35(4):465–492, 2007.

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

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

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

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

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

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

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

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

David M Grether and Charles R Plott. Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, 69(4):623–638, 1979.

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

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

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

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

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

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

George J. Mailath, Larry Samuelson, and Jeroen Swinkels. Extensive form reasoning in normal form games. CARESS Working Paper 90–01, University of Pennsylvania, January 1990.

Roger B. Myerson. Axiomatic foundations of bayesian decision theory. Discussion Paper 671, The Center for Mathematical Studies in Economics and Management Science, Northwestern University, January 1986.

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

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

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

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

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

Ariel Rubinstein. Comments on the interpretation of game theory. *Econometrica*, 59:909–924, 1991.

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.

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

Marciano Siniscalchi. Structural rationality: applications to epistemic game theory. mimeo, Northwestern University, 2016b.

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

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