Abstract

Sequential rationality requires that players maximize their continuation payoff at all information sets, including those 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, sequential rationality is not a testable behavioral assumption in general. This paper addresses this fundamental methodological 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 using a version of the strategy method (Selten, 1967). Finally, unlike sequential rationality, 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.
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, tastes (i.e., utilities) and beliefs can be elicited via incentive-compatible “side bets” whose outcomes depend upon the strategies of coplayers (Luce and Raiffa, 1957, §13.6). Hence, it is possible to interpret 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. 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. The analysis of a player’s beliefs and actions following unexpected moves by her opponents is, of course, essential to solution concepts for dynamic games. Yet, if the assumed optimality criterion is sequential rationality, such beliefs and actions are neither directly observable, nor indirectly elicitable in an incentive-compatible way. Section A in the Appendix illustrates this point with an example.

---

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.
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 strategic rationality.\textsuperscript{3} On the other hand, a rationality criterion consistent with the noted findings can potentially address the methodological concerns that motivate this project. To account for the first set of experimental findings, such a criterion must leverage the extensive-form structure to refine strategic-form predictions. To account for the evidence on commitment, it must allow for strategies and, potentially, conditional beliefs, to be elicited in an incentive-compatible way.

The main contribution of this paper is precisely to identify such a criterion, structural rationality. 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.

\textsuperscript{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.
compatible way. Theorem 3 identifies a crucial consistency property of conditional beliefs that is required for structural rationality to be well-defined. A companion paper, Siniscalchi (2015), provides axiomatic foundations for structural rationality; in-progress work 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, and aligns it more closely with the evidence from experiments.

The remainder of this paper is organized as follows. Section 2 presents the setup. Section 3 formalizes structural preferences and structural rationality. Section 4 relates structural and sequential rationality, and Section 5 formalizes the elicitation result. Section 6 analyzes the consistency of conditional beliefs. Finally, Section 7 discusses the related literature and concludes. The Appendix contains all proofs as well as additional examples.

2 Setup

This paper considers extensive games with possibly 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; additional details can be found in Appendix B.1.

An extensive game form is a tuple $\Gamma = (N, H, P, (\mathcal{I}_i)_{i \in N})$, 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 $\phi$, $P$ is the player function, which associates with each history $h$ the player on the move at $h$, and each $\mathcal{I}_i$ is a partition of the histories where Player $i$ moves; the elements of $\mathcal{I}_i$ 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. A minor departure from OR is that chance moves are not part of the description of the game; Chance can be modeled as an additional player, and the probabilities
of chance moves can be incorporated into players’ beliefs.

Given an extensive game form, certain derived objects of interest, including strategies, can be defined. The set of actions available at a history \( h \in H \) is \( A(h) = \{ a \in A : (h, a) \in H \} \). A history \( h \) is terminal if \( A(h) = \emptyset \); the set of terminal histories is denoted \( Z \). The definition of an extensive game form requires that \( A(h) = A(h') \) whenever \( h, h' \in I \in \mathcal{I} \); thus, one can also denote by \( A(I) \) the set of actions available at an information set \( I \). For every player \( i \in N \), a strategy is a map \( s_i : \mathcal{I}_i \rightarrow A \) such that \( s_i(I) \in A(I) \) for all \( I \in \mathcal{I} \); 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 \( I \in \mathcal{I}_i \), \( S(I) \) 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 \( \mathcal{I}_i(s_i) \) allowed by strategy \( s_i \): that is, for every \( I \in \mathcal{I}_i \), \( I \in \mathcal{I}_i(s_i) \) if and only if \( s_i \in S_i(I) \).

To specify payoffs at terminal histories, fix a set \( X \) of material (i.e., physical or monetary) outcomes, ⁴ and assume that each player is characterized by a utility function \( u_i : X \rightarrow \mathbb{R} \). Then, if outcome \( x \) is attached to terminal history \( z \), player \( i \)'s numerical payoff at \( z \) is \( u_i(x) \) of outcome \( x \).

It is useful to allow for payoff uncertainty: material outcomes may be determined by the realization of some exogenous uncertainty, in addition to the terminal history reached. It is convenient to adopt a slightly more general formalism than the one used for standard incomplete-information games. Fix a set \( \Theta \), endowed with a sigma-algebra \( \mathcal{F} \). For each player \( i \), define the outcome function \( \xi_i : Z \times \Theta \rightarrow X \). When terminal history \( z \) is reached and the realization

---

⁴This two-step definition is consistent with the axiomatic approach in Siniscalchi (2015). The set of outcomes is the same for all players; this is only to streamline the notation. In Siniscalchi (2015), \( X \) is assumed to be a convex set (for example, a set of lotteries). This is not required here, except for Remark 1.
the exogenous uncertainty is \( \theta \in \Theta \), player \( i \)'s material outcome is \( \xi_i(z, \theta) \).\(^5\)

A **dynamic game** is then a tuple \((\Gamma, X, \Theta, \mathcal{F}, (\xi_i)_{i \in N})\), where \(\Gamma\) is an extensive game form with player set \( N \), \(\Theta\) characterizes payoff uncertainty, and \(\xi_i\) is \( i \)'s outcome function.

### 3 Conditional Beliefs and Structural Preferences

#### 3.1 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 additional exogenous uncertainty; let \( \Omega_i = S_{-i} \times \Theta \), and endow this set with the product sigma-algebra \( \Sigma_i = 2^{S_{-i}} \times \mathcal{F} \).

Player \( i \)'s beliefs at an information set \( I \) are conditional upon the information she possesses regarding the play of others. Since beliefs are defined over \( \Omega_i = S_{-i} \times \Theta \), this information must be expressed as a subset of \( \Omega_i \) as well. In particular, conditional upon reaching \( I \), Player \( i \) can rule out strategies of her coplayers that do not allow \( I \). Thus, for each \( I \in \mathcal{I}_i \), let

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

the class of **conditioning events** for player \( i \) is then

\[
\mathcal{F}_i = \{\Omega_i\} \cup \{[I] : I \in \mathcal{I}_i\}.
\]

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

Finally, for a measurable space \((Y, \mathcal{Y})\), \(pr(\mathcal{Y})\) denotes the set of finitely additive probability measures on \( \mathcal{Y} \). Conditional beliefs can now be formally defined.

---

\(^5\)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 7, this reflects the *interim* perspective in games of incomplete information.
Definition 1 (Rényi (1955); Ben-Porath (1997); Battigalli and Siniscalchi (1999, 2002)) A conditional probability system (CPS) for player \( i \) in the dynamic game \((\Gamma, X, \Theta, \mathcal{T}, (\xi_i)_{i \in N})\) is a collection \( \mu_i \equiv (\mu_i(\cdot|F))_{F \in \mathcal{F}_i} \) such that:

1. for every \( F \in \mathcal{F}_i \), \( \mu_i(\cdot|F) \in \text{pr}(\Sigma_i) \) and \( \mu_i(F|F) = 1 \);
2. for every \( E \in \Sigma_i \) and \( F, G \in \mathcal{F}_i \) such that \( E \subseteq F \subseteq G \),
   \[
   \mu_i(E|G) = \mu_i(E|F) \cdot \mu_i(F|G);
   \] (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.

The set of CPS for player \( i \) is denoted by \( \text{cpr}(\Sigma_i, \mathcal{F}_i) \). Denote by \( B_0(\Sigma_i) \) the set of \( \Sigma_i \)-measurable real functions with finite range\(^6\), and by \( B(\Sigma_i) \) its sup-norm closure. For any probability charge \( \pi \in \text{pr}(\Sigma_i) \) and function \( a \in B(\Sigma_i) \), let \( E_\pi[a] = \int_{\Omega_i} a \ d\pi \), the standard Dunford integral of \( a \) with respect to \( \pi \); when no confusion can arise, I will sometimes omit the square brackets.

### 3.2 Structural Preferences

To introduce the key notion of structural preferences, I proceed in three steps. First, I observe that a CPS \( \mu \) 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 \( \mu \) also uniquely identifies a collection of probabilities, interpreted as alternative ex-ante beliefs that generate \( \mu \) by conditioning. (The exact formulation of this statement is Theorem 3 in Section 6). 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 \((\Gamma, X, \Theta, \mathcal{T}, (\xi_i)_{i \in N})\).

\(^6\)Recall that, while \( S_{-i} \) is finite, the set \( \Theta \), and hence the state space \( \Omega_i \), need not be. Hence the need to define the set of simple functions explicitly.
A preorder over conditioning events. Fix two information sets $I, J \in \mathcal{I}_i$. If $I < J$, then it is easy to see that, by perfect recall, $S_{-i}(I) \supseteq S_{-i}(J)$. If $\mu$ 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, due to the structure of the game, $I$ must be reached. This intuition generalizes. For information sets $I, J \in \mathcal{I}_i$ 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 $\mu$ on $(\Sigma, \mathcal{F}_i)$. For all $D, E \in \mathcal{F}_i$, $D$ is at least as plausible as $E$ ($D \triangleright E$) if there are $F_1, \ldots, F_N \in \mathcal{E}$ such that $F_1 = E$, $F_N = D$, and for all $n = 1, \ldots, N - 1$, $\mu(F_{n+1}|F_n) > 0$.

A collection of alternative prior beliefs. Consider the game in Figure 1; 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 1: Alternative theories and plausibility](image-url)

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

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

(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 that contains both. However, I suggest that, indirectly, \( \mu \) 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 suggests that, conditional upon Bob not choosing \( o \), Ann deems \( t \) and \( b \) equally likely.

Even more can be said. Given Ann’s CPS \( \mu \), her conditioning events are ranked as follows in terms of plausibility (Def. 2):

\[
S_{Bob} \downarrow I, S_{Bob} \downarrow J, I \downarrow J, J \downarrow I.
\]

It is not the case that \( I \downarrow S_{Bob} \) or \( J \downarrow S_{Bob} \). 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 \( \mu \) 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 interpretation is in the spirit of structural consistency (Kreps and Wilson, 1982; Kreps and Ramey, 1987); see Section 7 for further discussion. The following definition formalizes it.

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

1. for every \( C, D \in \mathcal{F}_i \), \( p_C = p_D \) if and only if both \( C \uparrow D \) and \( D \uparrow C \);
2. for every \( C \in \mathcal{F}_i \), \( p_C(\cup\{D \in \mathcal{F}_i : C \uparrow D, D \uparrow C\}) = 1 \);
3. for every \( C \in \mathcal{F}_i \), \( p_C(C) > 0 \) and, for every \( E \in \Sigma \), \( \mu(E \cap C|C) = \frac{p_E(E \cap C)}{p_C(C)} \).

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

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

For the CPS \( \nu \), \( S_{\text{Bob}} \) 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 \( \nu \) 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 6 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 \( \Sigma_i \)-measurable functions \( f: \Omega_i \to X \). The set of all acts for player \( i \) is denoted \( \mathcal{A}_i \). Given the dynamic game \( (\Gamma, X, \Theta, (\xi_i)_{i \in N}) \), every strategy \( s_i \in S_i \), together with the outcome function \( \xi_i \), 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.

**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 \( \mu \) for \( i \) that admits a basis \( p = (p_F)_{F \in \mathcal{F}_i} \). For any pair of acts \( f, g \in \mathcal{A}_i \), \( f \) is (weakly) **structurally preferred** to \( g \) given \( u_i \) and \( p \), written \( f \succ^{u_i, \mu} g \), iff for every \( F \in \mathcal{F}_i \) such that \( E_{p_F} u_i \circ f < E_{p_F} u_i \circ g \), there is \( G \in \mathcal{F}_i \) such that \( G \succ F \) and \( E_{p_G} u_i \circ f > E_{p_G} u_i \circ g \).

Strict preference \( (\succ^{u_i, \mu}) \) is defined as usual: \( f \succ^{u_i, \mu} g \) iff \( f \succ^{u_i, \mu} g \) and not \( g \succ^{u_i, \mu} f \).

The definition of structural preferences is reminiscent of that of lexicographic preferences (Blume, Brandenburger, and Dekel, 1991a). The crucial difference is that both the probabilities 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 7.

**Examples.** Definition 4 characterizes an *ex-ante* ranking, before the player has observed any moves made by others. 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 1; here and throughout all examples in this paper, interpret the numbers attached to terminal nodes as monetary payoffs, and assume utility is linear. Table I below summarizes the expected payoffs given the two basis probabilities for the CPS $\mu$ in Eq. (4): the prior, $\mu(\cdot|S_{Bob})$, and the probability $p$ that assigns equal weight to $t, m, b$.

| $s_i$      | $f^{s_i}$ | EU for $\mu(\cdot|S_{Bob})$ | 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. 1 and the CPS $\mu$

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 structural preferences disregard choices prescribed at information sets that are ruled out by a player’s own prior moves; in particular, $RTT'$ is deemed just as good as $RTB'$. This is because

---

7The companion paper FSP shows that Savage’s usual update rule for preferences (Savage, 1954) defines dynamically consistent conditional structural preferences.

8Equivalent, assume they represent utility levels: this is immaterial to the discussion.

9The 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 $(f^{s_i}(t), f^{s_i}(m), f^{s_i}(b), f^{s_i}(o))$. 

11
RTT' and RTB' are realization-equivalent, and hence correspond to the same act, namely (6, 3, 0, 1).\footnote{Any ranking of strategies defined via a preference relation over acts would also not distinguish between realization-equivalent strategies.} I return to this point in the next Section, where sequential rationality is defined.

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

| $s_i$ | $f^{s_i}$ | EU for $\nu(\cdot|S_{Bob})$ | 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. 1 and the CPS $\nu$

With these beliefs, no strategy is strictly, or even weakly structurally preferred to all others. Consider for instance strategy RTT’. This strategy 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 $\nu$. This illustrates that structural preferences can be incomplete. This is a consequence of the fact that the CPS $\nu$ does not rank $[I]$ and $[J]$ in terms of their plausibility: one cannot say that one is “infinitely more likely” than the other. That said, strategies RTT’ and LTT’ are maximal—no other strategy is structurally preferred to either of them. By way of contrast, for instance, RBT’ is strictly worse than RTT’.

A caveat: basis probabilities and conditional probabilities. In some cases, basis probabilities coincide with conditional probabilities: this is true for Ann’s CPS $\nu$ (Eq. 5), but not for her CPS $\mu$ (Eq. 4). Could one define structural preference with reference to conditional probabilities directly, rather than via their basis? The answer is negative. Consider Ann’s strategy RBT’. Its conditional expected payoff given $\nu(\cdot|[J])$ is 2.5, which is strictly higher than that of
any other strategy, including \( RTT' \). The only conditioning event that is more plausible than \([J] = S_b\), and ex-ante all strategies yield 1. Thus, a variant of Definition 4 that employs conditional probabilities, rather than basis probabilities, would deem \( RBT' \) maximal. However, \( RBT' \) is not sequentially rational for the CPS \( \mu \).

The reason for this undesirable conclusion is that the variant using conditional probabilities directly leads one to compare the expected payoff of strategies \( RTT' \) and \( RBT' \) conditional upon an 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 probability 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 \( RTT' \), which is maximal, must make an optimal choice at \( I \).

## 4 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 \( \mu_i \) that admits a basis \( p_i = (p_i, F)_{F \in F_i} \), and a utility function \( u_i : X \to \mathbb{R} \). Also, for every \( i \in N \), let \( \nu_i \) be the plausibility relation induced by \( \mu_i \). 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

\[
U_i(s_i, s_{-i}, \theta) \equiv u_i(\xi_i(\zeta(s_i, s_{-i}), \theta)) = u_i(f_i^{s_i}(s_{-i}, \theta)),
\]

where the equality follows from the definition of the \( i \)-act \( f^{s_i} \) associated with strategy \( s_i \). De-
note 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 $s_i, t_i \in S_i$, $s_i \succ^{u_i, \mu_i} t_i$ if and only if, for every event $F \in \mathcal{F}_i$ such that $E_{p_i, F} U_i(s_i, \cdot) < E_{p_i, F} U_i(t_i, \cdot)$, there is $G \in \mathcal{F}_i$ such that $G \triangleright_i F$ and $E_{p_i, G} U_i(s_i, \cdot) > E_{p_i, G} U_i(t_i, \cdot)$.

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 $U_i, \mu_i$) 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 1, if Ann’s beliefs are given by $\mu$ then $RTT'$ and $RTB'$ are the only structurally rational strategies; if instead they are given by $\nu$, 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 $(U_i, \mu_i)$ if, for every $I \in \mathcal{I}(s_i)$ and $t_i \in S_i(I)$, $E_{p_i[I]} U_i(s_i, \cdot) \geq E_{p_i[I]} U_i(t_i, \cdot)$.

By standard arguments, in any finite game $\Gamma$, a sequentially rational strategy exists.

The main result of this section can now be stated:

---

11The extensive game itself is assumed to be finite; the sets $\Theta$ can have arbitrary cardinality.
Theorem 1  If a strategy is structurally rational for \((U_i, \mu_i)\), then it is sequentially rational for \((U_i, \mu_i)\).

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. 2. Bob’s strategies are \(S_b = \{d_1 d_2, d_1 a_2, a_1 d_2, a_1 a_2\}\); there is no additional uncertainty. Assume that Ann’s initial and conditional beliefs correspond to Bob’s backward-induction strategy \(d_1 a_2\). Then, \(D_1^*\) (the realization-equivalent strategies \(D_1 D_2\) and \(D_1 A_2\)) and \(A_1 A_2\) 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 2: Sequential preferences refine sequential rationality](image)

However, for any CPS \(\mu_a\) for Ann that admits a basis, \(D_1^*\) is strictly better than \(A_1 A_2\), so \(A_1 A_2\) is not structurally rational. To see this, let \(p_a = (p_{a,S_b}, p_{a,[I]}\) 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 \(A_1 D_2\) and/or \(A_1 A_2\). In this case, \(\mu_a([I]|S_b) > 0\), so \(S_b \triangleright_a [I]\) and \([I] \triangleright_a S_b\), so \(p_{a,S_b} = p_{a,[I]}\). Moreover, the expected payoff to \(A_1 A_2\) with respect to \(p_{a,S_b}\) is strictly less than 2, so \(D_1^*\) is strictly better than \(A_1 A_2\) ex-ante; thus, \(D_1^* \succ^u a \triangleright a A_1 A_2\).

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 $p_{a,b}$ 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 $S_b \succ_a [I]$, again $D_{1*} \succ_{u,a} A_1A_2$. □

5 Elicitation

This section investigates how players’ structural preferences, and hence their conditional beliefs, can be elicited. The elicitation game is a slight extension of 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. In addition, I fix a distinguished player $i$, and ask her to choose one of two acts $f$ or $g$, in addition to her strategy. I also 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. This construction can be readily modified so as to elicit additional preference rankings for player $i$, or for multiple players.

The elicitation result of this section depends crucially upon the assumption that, as the experimenter implements the strategies subjects have committed to, she conveys to them the same information about coplayers’ moves as they would receive by playing the game directly. The following definition formalizes this assumption.

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

- $\Gamma^* = (N, H^*, P^*, (\mathcal{G}^*_j)_{j \in N})$, where, writing $\Gamma = (N, H, P, (\mathcal{G}_j)_{j \in N})$,

- $h^* \in H^*$ 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 $h^*$.  

16
– \( P^*(h^*) = j \) if and only if \( h^* \) has length \( j - 1 \), or if it extends some \( h \in H \) with \( P(h) = j \)

– for \( j \neq i \), \( \mathcal{I}_j = \{ I_j^1 \} \cup \{ \langle s_j, I \rangle : s_j \in S_j(I), I \in \mathcal{I}_j \} \), where \( I_j^1 = \{ h^* : h^* \) 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 \( i \), \( \mathcal{I}_i = \{ I_i^1 \} \cup \{ \langle s_i, k, I \rangle : s_i \in S_i(I), I \in \mathcal{I}_i, k \in \{ f, g \} \} \), where \( I_i^1 = \{ h^* : h^* \) 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 \( \mathcal{T}^* = \mathcal{T} \times 2^{\{ o, a \}} \);

• \( \xi_j((s_1, \ldots, s_N, k, z), (\theta, r)) \) equals \( \xi(z, \theta) \) if \( j \neq i \) or \( r = o \), and \( k((s_{-i}, \theta)) \) if \( j = i \) and \( r = a \).

In words, the extensive game form \( \Gamma^* \) consists of histories of play that begin with a commitment phase: players choose strategies \( s_1, \ldots, s_N \) in the original game form \( \Gamma \), and player \( i \) also chooses an act \( k \in \{ f, g \} \). Then, histories continue with the path of play generated by the profile \( (s_1, \ldots, s_N) \). The moves in the commitment phase are represented as sequential choices; however, information sets in this phase are defined so that player \( j \) does not observe the prior actions of players \( 1, \ldots, j - 1 \), so these choices can also be thought of as being simultaneous. The player function is defined so that the player choosing at step \( j \) of the commitment phase is indeed player \( j \), and that otherwise players are assigned to histories in a manner consistent with the original game. Each player \( j \neq i \) has one information set in the commitment phase. Furthermore, 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. Similar considerations apply to player \( i \), except that she also remembers her choice of an act \( k \in \{ f, g \} \).

The description of the elicitation game is completed by extending the space \( \Theta \) of exogenous uncertainty so as to include the realization of the randomizing device, \( r \in \{ o, a \} \), and by extending the the definition of the outcome function \( \xi \) so that \( i \)'s payoffs are determined by her choice of act in case \( a \) obtains.
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 $I^j_j$; 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 2 in Appendix C.2). This formalizes the assumption that players are 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 $\sigma_j : S_j \rightarrow S^*_j$ such that $\sigma_j(s_j)$ is the unique strategy in $S^*_j$ that chooses $s_j$ at $I^1_j$. Similarly, for player $j = i$, and for every $k \in \{f, g\}$, there is a bijection $\sigma_{i,k} : S_i \rightarrow S^*_i$, such that $\sigma_{i,k}(s_i)$ is the unique strategy in $S^*_i$ that chooses $s_i$ and $k$ at $I^1_i$.

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 C.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.

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

$$
\mu^*_i\left(\{(\sigma_i(s_i))_{t \neq i}\} \times U \times \{r\}\right) = \frac{1}{2} \mu_i\left(\{s_{-i}\} \times U\right)_{\Omega_i}
$$

18
\[
\mu_i^*\left(\{[\tau_\ell](s_\ell)\}_{\ell \neq i}\} \times U \times \{r\} \bigg| ([s_i, I])\right) = \frac{1}{2}\mu_i\{\{s_{-i}\} \times U\}|I];
\]
and if \( j \neq i \),
\[
\mu_j^*\left(\{[\tau_\ell](s_\ell)\}_{\ell \neq (i, j)}\} \times \{\sigma_{i, f}(s_i), \sigma_{i, g}(s_i)\} \times U \times \{r\} \bigg| \Omega_j \times \{o, a\}\right) = \frac{1}{2}\mu_j\{\{s_{-j}\} \times U\}|\Omega_j),
\]
\[
\mu_j^*\left(\{[\tau_\ell](s_\ell)\}_{\ell \neq (i, j)}\} \times \{\sigma_{i, f}(s_i), \sigma_{i, g}(s_i)\} \times U \times \{r\} \bigg| ([s_j, I])\right) = \frac{1}{2}\mu_j\{\{s_{-j}\} \times U\}|I).\]

The equations in Definition 8 state that, conditional upon any event in \( F_j^* \), 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 \( \Gamma \). 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 \((\Gamma, X, \Theta, \mathcal{T}, (\xi_i)_{i \in N})\) and acts \( f, g \in \mathcal{A}_i \) of player \( i \in N \). For every \( j \in N \), fix a CPS \( \mu_j \in \text{cpr}(\Omega_i, F_j) \) that admits a basis. Then every \( \mu_j \) has an extension \( \mu_j^* \),\(^{12}\) which also admits a basis. For any choice of extensions \((\mu_j^*)_{j \in N}\) and utilities \((u_j)_{j \in N}\):

1. For all \( j \neq i \) and \( s_j, t_j \in S_j, \sigma_j(s_j) \succeq_{u_j, \mu_j} \sigma_j(t_j) \) if and only if \( s_j \succeq_{u_j, \mu_j} t_j \);
2. for all \( s_i, t_i \in S_i \) and \( k \in \{f, g\}, \sigma_{i, k}(s_i) \succeq_{u_i, \mu_i} \sigma_{i, k}(t_i) \) if and only if \( s_i \succeq_{u_i, \mu_i} t_i \);
3. for every \( s_i \in S_i, \sigma_{i, f}(s_i) \succeq_{u_i, \mu_i} \sigma_{i, g}(s_i) \) if and only if \( f \succeq_{u_i, \mu_i} g \).

Parts 1 and 2 of Theorem 2 state that, if every player has “the same” beliefs in the original game \( \Gamma \) and in the elicitation game \( \Gamma^* \), then every player’s preferences over strategies are effectively unchanged. This in turn suggests a reason why players might hold the same beliefs in \( \Gamma \).

\(^{12}\)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.
and $\Gamma^*$—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: if one disregards the elicitation of $i$’s ranking of $f$ vs. $g$, the game $\Gamma^*$ provides a way to elicit every player’s behavior in $\Gamma$ from the observation of her choices in the initial commitment stage of $\Gamma^*$.

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 $\Gamma$: otherwise, for some strategy $t_i$, the pair $(t_i, f)$ would be strictly preferred, and thus not chosen in $\Gamma^*$. Similarly, by part 3 of the Theorem, it cannot be the case that $g \succ_{\mu_i, u_i} f$, for otherwise $(s_i, g)$ would be strictly preferred. Of course, 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 $\Gamma$, or that $f$ and $g$ might be incomparable in $\Gamma$.

How to interpret or allow for such indecisiveness is a standard concern with incomplete preferences (see e.g. Eliaz and Ok, 2006, and references therein). Fortunately, by exploiting specific features of structural preferences, it is nevertheless possible to elicit the utility function $u_i$ and the CPS $\mu_i$ 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 (2015), 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 $\mu_i$. The following Remark provides the key 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 $\omega \notin E$. 

20
Remark 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 $I \in \mathcal{I}_i$ and event $G \in \Sigma$ with $G \subseteq [I]$, there is a unique number $\alpha \in [0, 1]$ such that

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

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

Leveraging Remark 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 Remark 1, surely $u_i(y) > 1 \geq \mu(G[I])$. Then, in the elicitation game $\Gamma^*$ 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 Remark 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 $\mu_i(G[I]).$

Indeed, it is straightforward to modify the elicitation game analyzed in this section so as to present player $i$ with a collection of outcomes $y_1, \ldots, y_M$ that determine the desired grid of utility / probability values in $[0, 1]$. The key modification is to alter the commitment phase so that player $i$ must choose (i) a strategy $s_i \in S_i$ and (ii) for each $m = 1, \ldots, M$, one of the acts $f_m = y_m[I]x_0$ or $g_m = x G x_0$. Then, adopt a randomizing device that generates a number $r \in \{0, \ldots, M\}$: if $r = 0$, the outcomes are determined by the strategy profile; if instead $r = m > 0$, player $i$’s payoff is determined by her choice of $f_m$ vs. $g_m$. The proof of Theorem 2 is still valid, modulo (cumbersome) notational changes. Of course, one can further extend the procedure so as to elicit the probabilities of multiple events, at multiple information sets. The resulting construction is in the spirit of random lottery incentive schemes (e.g. Grether and Plott, 1979), and again, the analysis of this section still applies.
6 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. 3.

Figure 3: Newcomb’s paradox?

Suppose that Ann’s CPS $\mu$ is as in Eq. (11):

$$\mu(\{o\}|S_b) = \mu(\{b\}|[I]) = \mu(\{c\}|[J]) = 1.$$  \hspace{1cm} (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 $\mu$ must consist of two measures, $p_\varphi$ and $p_{IJ}$, where $p_\varphi = \mu(\cdot|S_b)$ 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 $\mu$ in Eq. (11) does not admit a basis.

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

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 \textit{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 \( J \).\(^{13}\) Moreover, modify the game in Figure 3 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 \( \mu \) cannot be extended to a new CPS \( \mu' \) for the new game. A fortiori, the CPS \( \mu \) cannot be extended to a \textit{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.

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

The CPS \( \mu \) is \textit{consistent} if, for every \( \mu \)-sequence \( F_1, \ldots, F_L \), and all \( E \subseteq F_1 \cap F_L \),

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

An immediate consequence of this definition is that, if \( F_1, \ldots, F_L \) is a \( \mu \)-sequence, so is \( F_{\ell}, \ldots, F_m \), where \( 1 \leq \ell \leq m \leq L \). Furthermore, the preorder \( \succ \) in Definition 2 can be characterized in terms of \( \mu \)-sequences: \( F \succ G \) iff there is a \( \mu \)-sequence \( F_1, \ldots, F_L \) such that \( F_1 = G \) and \( F_L = F \).

\(^{13}\)In the language of Kreps and Wilson (1982), fix a convergent sequence of strictly positive behavioral strategy profiles \( \pi^k \). If \( m^k \) is the system of beliefs derived from \( \pi^k \), \( m^k(I) \) 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.
Consistency can be viewed as a strengthening of the chain rule of conditioning. Consider $F, G \in \mathcal{F}_i$ and $E \in \Sigma$, and assume that $E \subseteq F \subseteq G$. Then the ordered list $F, G$ is a $\mu$-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.

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 $\mu \in \text{cpr}(\Sigma_i, \mathcal{F}_i)$ be a CPS for player $i \in N$. Define $\mathcal{F}_\mu = \{\bigcup_{\ell=1}^L F_\ell : F_1, \ldots, F_L \text{ is a } \mu\text{-sequence}\}$.

The following are equivalent:

1. $\mu$ is consistent;
2. $\mu$ admits a unique basis;
3. there is a unique CPS $\nu \in \text{cpr}(\Sigma_i, \mathcal{F}_\mu)$ such that $\nu(\cdot|F) = \mu(\cdot|F)$ for all $F \in \mathcal{F}$;
4. there is a sequence $(p^n) \in \text{pr}(\Sigma_i)$ such that $p^n(F) > 0$ for all $m$ and $p^m(E \cap F)/p^k(F) \to \mu(E \cap F|F)$ for all $F \in \mathcal{F}$ and $E \in \Sigma_i$.\(^\text{14}\)

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

### 7 Related Literature and Conclusions

**Incomplete-information games** In the textbook model of games with incomplete information, there is a set $\Theta_i$ of “types” for each player $i$, and possibly a set $\Theta_0$ that describes resid-

\(^{14}\)Even though the state space may be infinite, convergence here is pointwise. See the proof for details.
ual 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 analy-
sis 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 $\mu_i$ introduced in Definition
1 are now interpreted as the ones characterizing the preferences of $i$’s type $\theta^*_i$. The results in
Sections 4 and 5 then state that, if the selected type $\theta^*_i$ 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.

**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 $\Omega_i$. Given acts
$f, g \in \mathcal{A}_i$, $f$ is lexicographically (weakly) preferred to $g$ if $E_{p_m} u_i \circ f < E_{p_m} u_i \circ g$ implies $E_{p_\ell} u_i \circ f > E_{p_\ell} u_i \circ g$ 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, but the same CPS may
be generated by multiple LPSs. For instance, the CPS $\nu$ 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 $\delta_\omega$ denotes the Dirac measure
concentrated on $\{\omega\}$. Intuitively, $\lambda^1$ assigns higher plausibility to $t$ than to $m$, whereas the
opposite is true of $\lambda^2$. However, this plausibility assessment is not derived from Ann’s CPS $\nu$.
By way of contrast, by design, structural preferences are defined solely in terms of informa-
tion that can be derived from the player’s conditional beliefs. For this reason, they are close

---

15For every $I \in \mathcal{I}_i$, let $\mu(\cdot[I]) = \mu_{m(I)}(\cdot[I])$, where $m(I)$ is the lowest index $m$ for which $p_m([I]) > 0$—assuming
one such index can be found.
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.

**Myerson (1986)** In an influential paper, 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 noted in the Introduction (see also the example in Appendix C.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. A more detailed discussion can be found in the companion paper Siniscalchi (2015).^16^

**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. The notion of a basis thus incorporates a notion of minimality, or parsimony: it identifies the minimal set of

---

^16^The same observability issue applies to Asheim and Perea (2005), who generalize Myerson's analysis.
alternative hypotheses that generate the CPS. Furthermore, as Theorem 3 shows, it requires that the CPS (and the alternative hypotheses derived from it) specify relative likelihoods in a consistent way.

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 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 resolution: see e.g. Kreps and Porteus (1978); Epstein and Zin (1989), and in particular Dillenberger (2010). In the latter paper, preferences are allowed to depend upon whether information is revealed gradually rather than in a single period, even if no action can be taken upon the arrival of partial information. This is close in spirit to the observation that subjects behave differently in the strategic form of a dynamic game, and when the game is played with commitment as in the strategy method. 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 sets have positive prior probability.)
A  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 4.

![Figure 4: Burning Money](image)

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. $B$) 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 4 the choices and beliefs that are actually observed and elicited may fail to provide evidence either for or against sequential rationality. Furthermore, it is not possible to verify whether actions and beliefs off the observed path of play satisfy properties of interest.17

These difficulties are unique to dynamic games: in a game with simultaneous moves, actions are observable ex post, and players cannot cause bets to be called off, or believe that they will be called off due to the actions of their opponents.

B Preliminaries on extensive game forms

B.1 Additional definitions

Begin with additional definitions and observations related to extensive game forms that will be needed in the proofs. Throughout this subsection, fix \( \Gamma = (N, H, P, (I_i)_{i \in N}) \).

Histories in \( H \) are ordered by a precedence 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 \( h’ = (h, a_{n+1}, \ldots, a_{n+k}) \).

---

17Suppose 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 \( l_n \) 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 \( r_n \) in the subgame on the left, in which case Not is indeed sequentially rational.

18For instance, if Ann chooses Not, one cannot experimentally verify whether, upon observing Burn, Bob would have believed that Ann would have continued with U, as in the forward-induction analysis of Ben-Porath and Dekel (1992).
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 $< \leq$ 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 $I < I'$.

With a slight abuse of notation, denote by $S(h)$ the strategy profiles that reach history $h \in H$; this is analogous to the notation $S(I)$ for information sets. 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 $I \in \mathcal{I}_i$, $S(I) = \bigcup_{h \in I} S(h)$.

By perfect recall, $S(I) = S_i(I) \times S_{-i}(I)$, and furthermore for all $I, J \in \mathcal{I}_i$,

$$
either S(I) \subset S(J) \text{ or } S(J) \subset S(I) \text{ or } S(I) \cap S(J) = \emptyset. \quad (12)$$

Perfect recall also implies that, if $s_i, t_i \in S_i(I), J \in \mathcal{I}_i$ and $J < I$, then $s_i(J) = t_i(J)$.

B.2 Bases

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

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

1. For all $E, F \in \mathcal{F}_i$, $p_E(E) > 0$ 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 $F \in \mathcal{F}_i$ and $E \in \Sigma$, if $p_E(E) > 0$, then there is $G \in \mathcal{F}_i$ such that $G \triangleright F$, $F \triangleright G$, and $p_F(E \cap G) > 0$. This follows because, by condition (2) in Def. 3,

$$p_F(E) = p_F \left( E \cap \bigcup \{ G \in \mathcal{F}_i : F \triangleright G, G \triangleright F \} \right) \leq \sum_{G \in \mathcal{F}_i : F \triangleright G, G \triangleright F} p_F(E \cap G).$$
1: by the Preliminary Claim, there is \( G \in \mathcal{F}_i \) with \( G \uparrow F, \, F \uparrow G \), and \( p_F(G \cap E) > 0 \). By condition (1), \( p_c(E \cap G) > 0 \), 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, \( E \uparrow G \); since the relation \( \uparrow \) is transitive, \( E \uparrow F \).

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

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

**Proposition 1** Fix a CPS \( \mu \in cpr(\Sigma_i, \mathcal{F}_i) \) for player \( i \in N \). The following are equivalent:

1. \( \mu \) is consistent;

2. for every \( \mu \)-sequence \( F_1, \ldots, F_K \in \mathcal{F} \), 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 \( E \subseteq F_\ell \),

\[
p(E) = \mu(E|F_\ell)p(F_\ell).
\]

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)\(\Rightarrow\)(2): assume that \( \mu \) is consistent. Let \( F_1, \ldots, F_K \in \mathcal{F}_i \) be a \( \mu \)-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 \( k \neq \ell \).]. Also, \( G_k \subseteq F_k \) for all \( k = 1, \ldots, K \).

Define a set function \( \rho : \Sigma_i \to \mathbb{R} \) as follows. First, for every \( \ell = 1, \ldots, K \) and \( E \in \Sigma_i \) such that \( E \subseteq G_\ell \), let

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

31
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 $\ell$ then $E \not\subseteq G_k$ for $k \neq \ell$, so $\rho(E)$ is uniquely defined; furthermore, $\emptyset \subseteq G_k$ for all $k$, but $\rho(\emptyset)$ is still well-defined and equal to 0.

To complete the definition of $\rho(\cdot)$, for all events $E \in \Sigma_i$ such that $E \not\subseteq G_k$ 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 $E_m \cap E_m = \emptyset$ 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 $\mu(\cdot|F_\ell)$,

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

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 $\sum_k \sum_m \rho(E_m \cap G_k) = \sum_k \rho(\left[ \bigcup_m E_m \right] \cap G_k)$, a sum of finitely many finite terms.

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

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

32
Multiply both sides by the positive quantity $\prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1} | F_{k+1})}{\mu(F_k \cap F_{k+1} | F_k)}$ 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 $m \in \{1, \ldots, K\}$,

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

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

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

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

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

For every $m \in \{1, \ldots, K\}$ and every event $E \subseteq F_m$, the probability measure $p$ satisfies

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

as asserted.

To show that $p$ is uniquely defined, let $q \in pr(\Sigma_i)$ be a measure that satisfies Eq.(15). I first claim that, for every $m = 1, \ldots, K$,

$$q(F_m) = \prod_{k=m}^{K-1} \frac{\mu(F_k \cap F_{k+1} | F_{k+1})}{\mu(F_k \cap F_{k+1} | F_k)} \cdot q(F_N) = \rho(F_m) q(F_K).$$
The claim is trivially true for \( m = K \), so consider \( m \in \{1, \ldots, K - 1 \} \) and assume that the claim holds for \( m + 1 \). By Eq. (15),
\[
\mu(F_m \cap F_{m+1} | F_{m+1}) q(F_{m+1}) = q(F_m \cap F_{m+1}) = \mu(F_m \cap F_{m+1} | F_m) q(F_m);
\]
since \( \mu(F_m \cap F_{m+1} | F_m) > 0 \) by assumption, solving for \( q(F_m) \) and invoking the inductive hypothesis yields
\[
q(F_m) = \frac{\mu(F_m \cap F_{m+1} | F_{m+1}) q(F_{m+1})}{\mu(F_m \cap F_{m+1} | F_m)} = \frac{\mu(F_m \cap F_{m+1} | F_{m+1})}{\mu(F_m \cap F_{m+1} | F_m)} \prod_{k=m+1}^{K-1} \frac{\mu(F_k \cap F_{k+1} | F_{k+1})}{\mu(F_k \cap F_{k+1} | F_k)} \cdot q(F_k).
\]
Since \( G_m \subseteq F_m \), Eq. (15) 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 \( \sum_k q(G_k) = q(\bigcup_k G_k) = q(\bigcup_k F_k) \), if in addition \( q \) satisfies \( q(\bigcup_k F_k) = 1 \), then
\[
1 = \sum_m \rho(G_m) \cdot q(F_k) = q(F_k) \rho(\bigcup_m G_m)
\]
which immediately implies that \( q(F_k) > 0 \), and indeed that
\[
q(F_k) = \frac{1}{\rho(\bigcup_m G_m)} = \frac{\rho(F_k)}{\rho(\bigcup_m G_m)} = p(F_k).
\]
so also \( p(F_k) > 0 \), as claimed. Furthermore, for \( m = 1, \ldots, K - 1 \),
\[
q(F_m) = \rho(F_m) q(F_N) = \rho(F_m) \frac{1}{\rho(\bigcup_m G_m)} = p(F_m).
\]

Furthermore, let \( k_0 \in \{1, \ldots, K - 1 \} \) be such that \( \mu(F_k \cap F_{k+1} | F_{k+1}) > 0 \) for all \( k > k_0 \), and \( \mu(F_{k_0} \cap F_{k_0+1} | F_{k_0+1}) = 0 \). By inspecting Eq. (13), it is clear that \( \rho(F_k) = 0 \) for \( k = 1, \ldots, k_0 \), and \( \rho(F_k) > 0 \) for \( k = k_0 + 1, \ldots, K \). Then, \( p(F_k) = 0 \) for \( k = 1, \ldots, k_0 \), and \( p(F_k) > 0 \) for \( k = k_0 + 1, \ldots, K \). From the above argument, it follows that the same is true for any \( q \in pr(\Sigma_i) \) that satisfies Eq. (15) and \( q(\bigcup_k F_k) = 1 \). Thus, the last claim of the Proposition follows.
Finally, if \( q \in pr(\Sigma_i) \) satisfies Eq. (15) and \( q(\cup F_k) = 1 \), 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. (15) and \( p(\cup F_k) = 1 \).

\( (2) \Rightarrow (1) \): assume that (2) holds. Consider a \( \mu \)-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. (15) 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 \( k = 2, \ldots, K \) as well, so

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

\[\blacksquare\]

**Corollary 1** If \( \mu \) is consistent, then for every \( \mu \)-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 \( \mu \)-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 \( \mu \)-sequence. Let \( p \) be unique the measure in (2) of Proposition 1. Per the last claim of the Proposition shows that necessarily \( p(F_{K+1}) > 0 \), but since
The following corollaries are used in the proof of Theorem 3.

**Corollary 2** Let $G_1, \ldots, G_N$ be a $\mu$-sequence and $p$ the measure in (2) of Proposition 1; consider $F \in \mathcal{F}_i$ such that $F \subset \bigcup_{k=1}^{K} G_k$. 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 $F \subseteq \bigcup_{m} G_m$ 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 $\mu$-sequence that satisfies the conditions of Corollary 1, so that in particular $\mu(G_k|F) > 0$, as claimed. Finally, suppose $m > k$. Since $p(G_k) > 0$, by the last claim of Proposition 1, $\mu(G_k|G_{\ell+1}) > 0$ 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 $\mu$-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 $\mu$-sequence. Let $p'$ be the measure delivered by Proposition 1 for this $\mu$-sequence. Notice that $p(F \cup \bigcup G_k) = p'(F \cup \bigcup G_k) = 1$, 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 $E \subseteq F$,

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

by 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

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

"
as claimed. ■

**Corollary 3** Let $G_1, \ldots, G_K$ and $F_1, \ldots, F_M$ be $\mu$-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

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

Therefore, $p' = q$, or $p(E') = p(\cup_m F_m)q(E')$ for every $m$ and $E' \in \Sigma_i$ 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 $m$,

$$p(E \cap \bar{F}_m) = p(\cup_i F_i)q(E \cap \bar{F}_m)$$

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 $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (4) \Rightarrow (1)$.

$(1) \Rightarrow (3)$: by Corollary 3, if $F_1, \ldots, F_L$ and $G_1, \ldots, G_M$ are $\mu$-sequences such that $\cup_i F_i = \cup_m G_m$, and $p$ and $q$ are the measures in condition 2 of Proposition 1, then $p = q$. Therefore, one can define an array $(\nu(\cdot|F))_{F \in \mathcal{F}_i}$ of probabilities on $\Sigma_i$ by letting $\nu(\cdot|\cup_i F_i)$ be the measure in condition 2 of Proposition 1 associated with the $\mu$-sequence $F_1, \ldots, F_L$. In particular, if $F \in \mathcal{F}_i$,
then \( \nu(F) = \mu(F) \). Again by Corollary 3, if \( F_1, \ldots, F_L \) and \( G_1, \ldots, G_M \) are \( \mu \)-sequences with \( \bigcup_m G_m \subset \bigcup_l F_l \), then for every measurable \( E \subseteq \bigcup_m G_m \), \( \nu(E \mid \bigcup_m G_m) = \nu(E \mid \bigcup_l F_l) \nu(\bigcup_m F_m \mid \bigcup_l G_l) \). Thus, \( \nu \) is a CPS on \((\Sigma, \mathcal{F}_\mu)\). Since the measure \( \nu(\cdot \mid \bigcup_l F_l) \) associated with each \( \mu \)-sequence \( F_1, \ldots, F_L \) is unique, \( \nu \) 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 \( F_1 \triangleright F_2 \triangleright \cdots \triangleright F_L \). By definition, for every \( \ell = L, L-1, \ldots, 2 \), there is a \( \mu \)-sequence \( F_1^\ell, \ldots, F_M^\ell \) such that \( F_1^\ell = F_\ell \) and \( F_M^\ell = F_{\ell-1} \). Since in addition \( F_L \triangleright F_1 \), there is also a \( \mu \)-sequence \( F_1^1, \ldots, F_M^1 \) with \( F_1^1 = F_1 \) and \( F_M^1 = F_L \). Since by construction \( F_M^{\ell+1} = F_{\ell+1} = F_\ell \) for \( \ell = L, L-1, \ldots, 2 \), the ordered list

\[
F_1^L, \ldots, F_M^{L(L)}, F_1^{L-1}, \ldots, F_M^{L(2)}, F_1^1, \ldots, F_M^1
\]

is a \( \mu \)-sequence. Since \( F_1^L = F_L = F_1^M \), Corollary 1 implies that the reverse-ordered list

\[
F_M^{1(L)}, \ldots, F_1^1, F_M^{2(L)}, \ldots, F_1^{L-1}, F_M^{L(L)}, \ldots, F_1^L
\]

is also a \( \mu \)-sequence. This has two implications. First, since any segment of the above \( \mu \)-sequences is also a \( \mu \)-sequence, it follows that, for every \( \ell = 1, \ldots, L \) and \( m = 1, \ldots, M(\ell) \), both \( F_1 = F_1^1 \triangleright \cdots \triangleright F_1^M \) and \( F_1^M \triangleright \cdots \triangleright F_1 = F_1 \). Hence \( F_1^M \) is an element of \( \{ F_1, \ldots, F_L \} \); in particular, \( \bigcup_{\ell=1}^L \bigcup_{m=1}^{M(\ell)} F_1^M = \bigcup_{\ell=1}^L F_1 \). Second, for all \( \ell = 1, \ldots, L \) and \( m = 1, \ldots, M(\ell)-1 \), both \( \mu(F_1^M \mid F_1^M) > 0 \) and \( \mu(F_1^M \mid F_1^M) > 0 \).

Since \( \nu \) is a CPS on \( \mathcal{F}_\mu \) which agrees with \( \mu \) on \( \mathcal{F}_\mu \), \( \nu(E \cap \bigcup_l F_l) = \mu(E \mid F_1^M) \nu(F_1^M \mid \bigcup_l F_l) \) for all measurable \( E \subseteq F_1^M \). Therefore, Proposition 1 implies that \( \nu(F_1^M \mid \bigcup_l F_l) > 0 \) for all \( m, \ell \).

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

Thus, consider an equivalence class \( \{ F_1, \ldots, F_L \} \) for \( \triangleright \), and every \( \ell = 1, \ldots, L \), \( p_F(F_l) = \nu(F_l \mid \bigcup_l F_l) \) if \( F_l \cap F_l > 0 \) and \( \mu(E \mid F_\ell) = \nu(E \mid \bigcup_l F_l) / \nu(F_\ell \mid \bigcup_l F_l) = p_{F_\ell}(E) / p_{F_\ell}(F_\ell) \); thus, Condition (3) in Def. 3 holds.
Moreover, $p_F(\cup \{ G : G \triangleright F_i, F_i \triangleright G \}) = p_F(\cup_i F_i) = \nu(\cup_i F_i | \cup_i F_i) = 1$, so Condition (2) holds as well. Finally, if $F \triangleright G$ and $G \triangleright F$, then $F, G$ belong to the same equivalence class $\{ F_1, \ldots, F_n \}$ and so by construction $p_F = \nu(\cup_i F_i) = p_G$. Thus, to show that Condition (1) holds, it remains to be shown that, if $F, G \in \mathcal{F}_i$ 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 $D \in \mathcal{F}_i$ 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 $E \in \mathcal{F}_i$ 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 $\mu$, 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 $\mu$. The probabilities $\{ p_F : F \in \mathcal{F} \}$ 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 $\{ p_F : F \in \mathcal{F} \}$ such that, for all $\ell, m$, $p_\ell \geq p_m$ 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 $F \in \mathcal{F}_i$, let $\ell(F)$ denote the index $\ell$ such that $p_\ell = p_F$. Finally, define a sequence $(p^n) \subset pr(\Sigma_i)$ by letting

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

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

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

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

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

$(4) \Rightarrow (1)$: consider a $\mu$-sequence $F_1, \ldots, F_L$ and an event $E \subseteq F_1 \cap F_L$. Let $(p^n) \subseteq p_{\Sigma_i}$ generate $\mu$ 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 $p^n(F_{\ell+1} \cap F_{\ell})/p^n(F_{\ell}) > 0$. For every such $n$ and measurable set $E \subseteq F_1 \cap F_L$,

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

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

### C Sequential rationality and elicitation

#### C.1 Theorem 1 (sequential optimality and sequential rationality)

Suppose that $s_i \in S_i$ is maximal for $(u_i, p_i)$, but not sequentially rational for $(U_i, \mu_i)$. Then there is $I \in \mathcal{S}(s_i)$ and $t_i \in S_i(I)$ such that $E_{\mu_i[I]} U(s_i, \cdot) < E_{\mu_i[I]} U(t_i, \cdot)$.

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 $J \in \mathcal{J}_i$, $r_i(J) = t_i(J)$ if $I \subseteq J$, and $r_i(J) = s_i(J)$ otherwise.
Notice that, for all \((s_{-i}, \theta) \in [I]\),

\[
U(r_i, s_{-i}, \theta) = u(\xi(r_i, s_{-i}), \theta) = u(\xi(t_i, s_{-i}), \theta) = U(t_i, s_{-i} \theta).
\]

To see this, note that, by perfect recall, since \(s_i, t_i \in S_i[I]\) it must be the case that \(s_i\) and \(t_i\) take the same actions at every \(J \in \mathcal{J}_i\) 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 \((s_{-i}, \theta) \notin [I]\),

\[
U(r_i, s_{-i}, \theta) = u(\xi(r_i, s_{-i}), \theta) = u(\xi(s_i, s_{-i}), \theta) = U(s_i, s_{-i}, \theta).
\]

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 \(J \in \mathcal{J}_i\) with \(I \leq J\). Therefore, \(r_i\) agrees with \(s_i\) at all \(J \in \mathcal{J}_i\) 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 \(E_{\mu_i([I])}U(s_i, \cdot) < E_{\mu_i([I])}U(t_i, \cdot)\) implies

\[
\int_{[I]} U(s_i, s_{-i}, \theta)dp_{i,[I]} = p_{i,[I]}([I])E_{\mu_i([I])}U(s_i, \cdot) < p_{i,[I]}([I])E_{\mu_i([I])}U(t_i, \cdot) = \int_{[I]} U(t_i, s_{-i}, \theta)dp_{i,[I]}.
\]

Therefore,

\[
E_{p_{i,[I]}}U(s_i, \cdot) = \int_{\mathcal{S}_{-i} \times \Theta} U(s_i, s_{-i}, \theta)dp_{i,[I]} = \int_{[I]} U(s_i, s_{-i}, \theta)dp_{i,[I]} + \int_{(\mathcal{S}_{-i} \times \Theta) \setminus [I]} U(s_i, s_{-i}, \theta)dp_{i,[I]} < \int_{[I]} U(t_i, s_{-i}, \theta)dp_{i,[I]} + \int_{(\mathcal{S}_{-i} \times \Theta) \setminus [I]} U(t_i, s_{-i}, \theta)dp_{i,[I]} = \int_{[I]} U(r_i, s_{-i}, \theta)dp_{i,[I]} + \int_{(\mathcal{S}_{-i} \times \Theta) \setminus [I]} U(r_i, s_{-i}, \theta)dp_{i,[I]} = E_{p_{i,[I]}}U(r_i, \cdot).
\]

Furthermore, consider \(F \in \mathcal{F}_i\). Two cases must be considered.
Case 1: \( p_{i,F}([I]) = 0 \). For such \( F \), trivially

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

and so

\[
E_{p_{i,F}} U(s_i, \cdot) = \int_{s_i \times \Theta} U(s_i, s_{-i}, \theta) dp_{i,F} = \\
= \int_{[I]} U(s_i, s_{-i}, \theta) dp_{i,F} + \int_{(S_i \times \Theta) \setminus [I]} U(s_i, s_{-i}, \theta) dp_{i,F} = \\
= \int_{[I]} U(r_i, s_{-i}, \theta) dp_{i,F} + \int_{(S_i \times \Theta) \setminus [I]} U(r_i, s_{-i}, \theta) dp_{i,F} = \\
= E_{p_{i,F}} U(r_i, \cdot).
\]

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

To conclude the argument, consider first \( F \in \mathcal{F}_i \) with \( F \succ [I] \) and \( F \neq [I] \). If \( p_{i,F}([I]) = 0 \), then per Case 1, \( E_{p_{i,F}} U(r_i, \cdot) = E_{p_{i,F}} U(s_i, \cdot) \); if instead \( p_{i,F}([I]) > 0 \), per Case 2 \([I] \succ F\), so by condition (1) in Def. 3 \( p_{i,F} = p_{i,[I]} \) and so \( E_{p_{i,F}} U(r_i, \cdot) > E_{p_{i,F}} U(s_i, \cdot) \). Thus, \( E_{p_{i,[I]}} U(r_i, \cdot) > E_{p_{i,[I]}} U(s_i, \cdot) \) and \( E_{p_{i,F}} U(r_i, \cdot) \geq E_{p_{i,F}} U(s_i, \cdot) \) for all \( F \in \mathcal{F}_i \) with \( F \succ [I] \); hence, \( s_i \not\succ_{u_i,p_i} r_i \).

On the other hand, consider \( F \in \mathcal{F}_i \) such that \( E_{p_{i,F}} U(r_i, \cdot) < E_{p_{i,F}} U(s_i, \cdot) \). Then \( F \) must be consistent with Case 2 above, so \([I] \succ F\). Since \( E_{p_{i,[I]}} U(r_i, \cdot) > E_{p_{i,[I]}} U(s_i, \cdot) \) and \( F \) was chosen arbitrarily, \( r_i \succ_{u_i,p_i} s_i \).

Thus, \( r_i \succ_{u_i,p_i} s_i \), which contradicts the assumption that \( s_i \) was maximal given \((u_i, p_i)\). □

C.2 Elicitation

C.2.1 Elicitation game: preliminaries

Begin by pointing out consequences of Def. 7 and introducing additional notation that will be useful in the analysis.
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 committed 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 2** Let \( h^* = (s, k, h) \in H^* \) be a history of length at least \( N \). Let \( j = P(h) = P^*(h^*) \), and let \( I \in \mathcal{I}_i \) be the infoset 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 collection of information sets for each player \( j \in N \) is denoted by \( \mathcal{I}_j^* \), with generic element \( I^* \).

The elicitation game has perfect recall because the original one does.

**Remark 3** The game form \( \Gamma^* \) 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 \( I^* \in \mathcal{I}_j^* \), \( h^*, \bar{h}^* \in I^* \) 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 \( I^* \in \mathcal{I}_j^* \setminus \{I_j^1\} \).

I analyze in detail the case \( j \in N \setminus \{i\} \): the case \( j = i \) is analogous. Write \( I^* = (s_j, I) \), where \( s_j \in S_j \) and \( I \in \mathcal{I}_j(s_j) \).
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 $\hat{h}^* \in (P^*)^{-1}(j)$ such that $\hat{h}^* < h^*$, ordered by the subhistory relation: that is, $h^*_0 < h^*_1 < \ldots h^*_n < h^*$, and $\hat{h}^* < h^*$ for no other $\hat{h}^* \in H^*$ with $P(\hat{h}^*) = j$. 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 $\tilde{h} \in P^{-1}(j)$ such that $\tilde{h} < h$; then, since $h < \zeta(s')$, also $\tilde{h} < \zeta(s')$, i.e., $s' \in S(\tilde{h})$. It follows that $(s', k, \tilde{h}) \in H^*$, and since $(s', k, \tilde{h}) < (s', k, h) = h^*$ by Observation 2, $\tilde{h} = h_m$ 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 $I^*_m \in \mathcal{I}_j$ 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 $I_m \in \mathcal{I}_j$. By the definition of information sets in $\Gamma^*$, $h_m \in I_m$. By Remark 2, $A(h^*_m) = s_j(I_m)$. Therefore,

$$X^*_j(h^*) = \big( (I^*_1, s_j), (I^*_1, s_j(I_1)), \ldots, (I^*_n, s_j(I_n)) \big), \quad X^*_j(h) = \big( (I_1, s_j(I_1)), \ldots, (I_n, s_j(I_n)) \big).$$

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

$$X^*_j(\hat{h}^*) = \big( (\hat{I}^*_1, s_j), (\hat{I}^*_1, s_j(\hat{I}_1)), \ldots, (\hat{I}^*_n, s_j(\hat{I}_n)) \big), \quad X^*_j(\hat{h}) = \big( (\hat{I}_1, s_j(\hat{I}_1)), \ldots, (\hat{I}_n, s_j(\hat{I}_n)) \big).$$

Since $\Gamma$ has perfect recall and $h, \hat{h} \in I$ by the definition of the infoset $\langle s_j, I \rangle$ 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 $I^*_1$, $i$ 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 infosets $I^*_m$.

44
C.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 4** Fix a profile $s^* \in S^*$ such that $s^*_i(I^*_i) = (s_i, k)$ and $s^*_j(I^*_j) = s_j$ for $j \neq i$. Then

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

Furthermore,

$$U^*_j(s^*, w, \theta) = \begin{cases} U_j(s, w) & \theta = o; \\ U_j(s, w) & \theta = a, j \neq i; \\ u_j(k(s_{i-1}, w)) & \theta = a, j = i, \end{cases}$$

**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 $(s, k)$ for short. 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 $\xi^*_j$, $\xi^*_j((s, k, \zeta(s)), w, o) = \xi_j(\zeta(s), w)$. Finally, by definition, $u_j(\xi_j(\zeta(s), w)) = U_j(s, w)$. The other cases are similar. 

The first key observation is that the sets $S^*_j$ in the elicitation game are in one-to-one correspondence with $S_j$ in the original game for players $j \neq i, 0$, and with $S_i \times \{f, g\}$ for player $i$. Specifically, Remark 2 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 $s^*_j(I^*_j) = s_j$; the reason $s^*_j$ is unique is that, at every infoset $I^* \in \mathcal{I}^* \setminus \{I^*_i\}$, a single action is available (this is true at every infoset containing histories of length $N$ or more, i.e.,
in the second stage of \( \Gamma^* \). Therefore, the map \( \sigma_j : S_j \rightarrow S_j^* \) 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 \( i \), but additional care is required because \( i \) also chooses a pair \( k \in \{ f, g \} \) at \( I_i^1 \). Thus, for every \( k \in \{ f, g \} \), define a bijection \( \sigma_{i,k} : S_i \rightarrow S_i^* \) 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 \( \sigma_{j,k}(s_{-j}) = [(\sigma_j(s_j))_{j \neq i}, \sigma_{i,k}(s_i)] \). It is also convenient to define the correspondence \( \sigma_{-j} : S_{-j} \rightarrow 2^{S_j^*} \) by letting \( \sigma_{-j}(s_{-j}) = \{ \sigma_{-j,i}(s_{-i}), \sigma_{-j,g}(s_{-j}) \} \) 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 infosets in \( \Gamma \) and their counterparts in \( \Gamma^* \).

**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 \( (\sigma_{i,k}(s_i), \sigma_{-i}(s_{-i})) \in S^*[(s, k, h)] \);

(iii) For every \( j \in N \setminus \{ i \} \) and \( s_j \in S_j \), \( J^*(\sigma_j(s_j)) = \{ I_j \} \cup \{ (s_j, I) : I \in J_j(s_j) \} \);

(iv) for every \( s_i \in S_i \) and \( k \in \{ f, g \} \), \( J^*(\sigma_{i,k}(s_i)) = \{ I_i \} \cup \{ (s_i, k, I) : I \in J_i(s_i) \} \).

**Proof:** Part (i) just restates Remark 4 in terms of the maps \( \sigma_j(\cdot) \). For part (ii), the “if” direction is immediate because \( (s, k, h) \in H^* \) immediately 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 \( J^* \). Consider \( I^* \in J^* \). If \( I^* = I_j^1 \), then clearly \( I^* \in J^*(\sigma_j(s_j)) \cap J^* \). Next, suppose \( I^* = (t_j, I) \) for some \( t_j \in S_j \) and \( I \in J_j(t_j) \). If \( t_j \neq s_j \), then \( I^* \notin J^*(\sigma_j(s_j)) \); 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 $I^* \notin \mathcal{F}^*$. Thus, consider $t_j = s_j$, so clearly $\langle s_j, I \rangle \in \mathcal{F}^*$. Fix $h^* \in I^*$; then by construction $h^* = (s', k, h) \in H^*$ for some $k \in \{f, g\}$ and $s' \in S(h)$; moreover, $h^* \in I^*$ implies $s_j' = s_j$ and $h = I$. Then, by part \( (ii) \), \((s_j, s_{-j}') \in S(h)\) implies $\langle \sigma_{i,k}(s'), \sigma_{j}(s_j), \sigma_{\ell}(s_{-j}') \rangle \in S^*((s, k, h)) = S^*(h^*) \subseteq S^*(I^*)$, and so $I^* \in \mathcal{G}^*_f(\sigma_f(s_j))$.

For part \( (iv) \), letting $\mathcal{G}^*$ denote the rhs of the equality in the claim, again $I_1^* \in \mathcal{G}^*_f(\sigma_{i,k}(s_i)) \cap \mathcal{G}^*$. Also, adapting the argument for part \( (iii) \), $\langle t_i, k', I \rangle \in \mathcal{G}^*_f(\sigma_{i,k}(s_i)) \cap \mathcal{G}^*$ if $t_i = s_i$ and $k' = k$, and $\langle t_i, k', I \rangle \notin \mathcal{G}^*_f(\sigma_{i,k}(s_i)) \cap \mathcal{G}^*$ otherwise. □

The second key observation is that players make the same inferences about coplayers' strategies as in the original game, except that players $j \neq i$ observe all prior moves perfectly in the "elicitation" part of the second stage. The following Lemma formalizes this.

**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 $I^* = \langle s_j, I \rangle \in \mathcal{G}^*$ then $S_{-j}^*(I^*) = \sigma_{-j}(S_{-j}(I))$;

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

**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 $\sigma_{-i} : S_{-i} \to S^*_{-j} \subseteq S^*_{-j}$ 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, $\sigma_{-j}(s_{-i}) \not\in \sigma_{-j}(D)$, and so $\sigma_{-j}(C) \not\subseteq \sigma_{-j}(D)$.

2: Fix $s^*_{-j} \in S_{-j}^*(I^*)$ arbitrarily; thus, there is $s^*_j \in S_j^* \subseteq S^*(I^*)$. In particular, \((s^*_j, s^*_{-j}) \in S^*(h^*)\) for some $h^* \equiv (s', k, h) \in I^*$. This implies that $s_j^* = s^*_j(I_j^*)$ for $\ell \neq i$ and $s_{-j}' = s^*_{-j}(I_{-j}^*)$. 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, $s^*_{-j} = \sigma_{-j,k}(s_{-j}') \in \sigma_{-j,k}(S_{-j}(I)) \subseteq \sigma_{-j}(S_{-j}(I))$.

47
Conversely, fix $s^*_i \in \sigma_{-i}(S_{-j}(I))$, and let $s_i = \sigma_{i}^{-1}(s^*_i) = s^*_i(I_i^1)$ for $\ell \neq j$, 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 4, $\zeta^*(s^*) = (s, k, \zeta(s))$; 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, $s^*_i \in S^*_i(I^*)$.

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 $\mathcal{F}_j^* = \Omega_j^* \cup \{I^* : I^* \in \mathcal{F}_j\}$ for every $j \in N$. For every player $j \in N$, let $\varphi_j : \mathcal{F}_j \rightarrow 2^{\Omega_j^*}$ be defined by

$$
\varphi_j(F) = \sigma_{-j}(\text{proj}_{S_j} F) \times \Theta^*.
$$

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

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

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

**Proof:** 1: Denote by $\mathcal{G}_j^*$ the set on the r.h.s. Fix $F^* \in \mathcal{F}_j^*$, so $F^* = S_{-j}(I^*) \times \Theta^*$ for some $I^* \in \mathcal{G}_j^*$. If $I^* = I_j^1$, then $F^* = \Omega_j^*$, since $\sigma_{-j}(\text{proj}_{-j}(\Omega_j)) \times \Theta^* = S_{-j} \times \Theta^* = \Omega_j^* \text{ and } \Omega_j \in \mathcal{F}_j$, $F^* \in \mathcal{G}_j^*$. If instead $F^* = [I^*]$ for some $I^* \in \mathcal{G}_j^* \setminus \{I_j^1\}$, there are two cases. If $j \neq i$, then $I^* = \langle s_j, I \rangle$, with $s_j \in S_j$ and $I \in \mathcal{G}_j$, 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 $I \in \mathcal{G}_j$; again, Lemma 3 implies that $S_{-j}(I^*) = \sigma_{-j}(S_{-j}(I))$. Hence, in either case, $F^* = S_{-j}(I^*) \times \Theta^* = \sigma_{-j}(S_{-j}(I)) \times \Theta^* = \varphi_{-j}(\langle I \rangle) \times \Theta^*$, so $F^* \in \mathcal{G}_j^*$.

Conversely, fix $F \in \mathcal{F}_j$. If $F = \Omega$, then $\varphi_{-j}(\Omega_j) = \sigma_{-j}(\text{proj}_{S_j} \Omega_j) \times \Theta^* = \sigma_{-j}(S_{-j}) \times \Theta^* = S_{-j} \times \Theta^* = \Omega_j^* \in \mathcal{F}_j^*$. If instead $F = [I] = S_{-j}(I) \times \Theta$, 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, $\varphi_{-j}(F) \in \mathcal{F}_j^*$.

2: immediate from the definition of $\varphi_{-j}(\cdot)$ and Lemma 3 part 1. ■
Now consider conditional beliefs. For every \( j \in N \), consider the Sigma-algebra \( \Sigma_j^* = 2^{\Sigma_j} \times \mathcal{T}^* \), where \( \mathcal{T}^* = \Theta \times 2^{[0,c]} \). Then a CPS for \( j \) in the elicitation game is an element \( \mu_j^* \in \text{cpr}(\Sigma_j^*, \mathcal{T}_j^*) \). Assume henceforth that \( \mu_j^* \) is the extension of a CPS \( \mu_j \in \text{cpr}(\Sigma_j, \mathcal{T}_j) \), 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 \( \mu_j^* \in \text{cpr}(\Sigma_j^*, \mathcal{T}_j^*) \) of \( \mu_j \); 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 \( U \in \mathcal{T} \). Therefore, a probability measure on \( \Omega_i^* \) 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 \( \Omega_i^* \) as follows: for every \( s_{-j} \in S_{-j} \), \( r \in \{o, a\} \), and \( U \in \Theta \),

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

and for every \( I \in \mathcal{I}_j \),

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

Equations 9 and 10 then imply that

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

for all \( s_{-j} \in S_{-j} \), \( r \in \{o, a\} \), \( U \in \Theta \), and \( I \in \mathcal{I}_j \). 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 \( \mu_j^*\left(\cdot|\Omega_j^*\right) \) and \( \mu_j^*\left(\cdot|\sigma_{-j}(S_{-j}(I)) \times \{o, a\} \times \Theta\right) \) are uniquely determined.
It is routine to verify that the set functions just defined are indeed probability measures on $\Sigma^*_j$. 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 $I \in \mathcal{F}_j$.

Finally, fix $I \in \mathcal{F}_j$, $F \in \mathcal{F}_j$, $s_{-j} \in S_{-j}$, $k \in \{f, g\}$, $r \in \{o, a\}$, 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 $J \in \mathcal{F}_j$, 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, $\mu_j^*$ is a CPS, per Definition 1.

Now consider player $i$. Since $\sigma_{-i}$ 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 $\mu_j$ admits a basis $p_j$. Finally, for every $j \in N$, let $\succ_j$ and $\succ_j^*$ denote the plausibility ordering induced by $\mu_j$ and $\mu_j^*$ respectively, as per Definition 2.

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

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

**Proof:** 1: if \( F = \{ I \} \) for some \( I \in \mathcal{F}_j \), then

\[
\mu_j^*(\varphi_j(F)|\varphi_j(G)) = \mu_j^*(\sigma_j(S_j(I)) \times \Theta \times \{ o, a \} | \varphi_j(G)) = \\
= \mu_j^*(\sigma_j(S_j(I)) \times \Theta \times \{ o \} | \varphi_j(G)) + \mu_j^*(\sigma_j(S_j(I)) \times \Theta \times \{ a \} | \varphi_j(G)) = \\
= \frac{1}{2} \mu_j(S_j(I) \times \Theta | G) + \frac{1}{2} \mu_j(S_j(I) \times \Theta | G) = \mu_j(F|G).
\]

If instead \( F = \Omega \), then \( \mu(F|G) = 1 \) and \( \mu_j^*(\varphi_j(F)|\varphi_j(G)) = \mu_j^*(\Omega^*|\varphi_j(G)) = 1 \), so the claim holds in this case, too.

2: suppose \( F \triangleright G \), so there is a sequence \( F_1, \ldots, F_N \in \mathcal{F}_j \) 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 \( \mathcal{F}_j^* \) 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, \( \varphi_j(F) \triangleright^* \varphi_j(G) \).

Conversely, suppose that \( \varphi_j(F) \triangleright^* \varphi_j(G) \), so there is \( F_1^*, \ldots, F_N^* \in \mathcal{F}_j^* \) 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 \( f \in \mathcal{F}_j \) 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, \( F \triangleright G \). \( \blacksquare \)

**Lemma 7** For every \( j \in \mathbb{N} \), the CPS \( \mu_j^* \) admits a basis \( p_j^* = (p_{j,F_*})_{F_* \in \mathcal{F}_j^*} \). In particular, for all \( F \in \mathcal{F}_j \), \( C \subseteq S_j \), \( U \in \Theta \),

\[
p_j^*|_{\varphi_j(F)}(\sigma_j(C) \times U \times \{ o \}) = p_j^*|_{\varphi_j(F)}(\sigma_j(C) \times U \times \{ a \}) = \frac{1}{2} p_{j,F}(C \times U).
\] (17)

**Proof:** Since \( \mathcal{F}_j^* = \{ \varphi_j(F) : F \in \mathcal{F}_j \} \) by Lemma 4, Eq. (17) defines a collection \( p_j^* = (p_{j,F_*})_{F_* \in \mathcal{F}_j^*} \).

I now show that \( p_j^* \) is a basis for \( \mu_j^* \).

Fix \( F, G \in \mathcal{F}_j \). By Lemma 2, \( F \triangleright G \) iff \( \varphi_j(F) \triangleright^* \varphi_j(G) \). Hence, \( p_{j,\varphi_j(F)} = p_{j,\varphi_j(G)} \) iff \( p_{j,F} = p_{j,G} \) iff \( F \triangleright G \) iff \( \varphi_j(F) \triangleright^* \varphi_j(G) \).
Similarly, for $F \in \mathcal{F}_j$,
\[
p_{j,\varphi_j^{-1}(F)}^*(\bigcup\{G^* : G^* \in \mathcal{F}_j, G^* \not\subseteq \varphi_j^{-1}(F), \varphi_j^{-1}(F) \not\subseteq G^*\}) = \\
p_{j,\varphi_j^{-1}(F)}^*\left(\bigcup\{\varphi_j^{-1}(G) : G \in \mathcal{F}_j, \varphi_j^{-1}(G) \not\subseteq \varphi_j^{-1}(F), \varphi_j^{-1}(F) \not\subseteq \varphi_j^{-1}(G)\}\right) = \\
p_{j,\varphi_j^{-1}(F)}^*\left(\bigcup\{\varphi_j^{-1}(G) : G \in \mathcal{F}_j, \varphi_j^{-1}(F) \not\subseteq \varphi_j^{-1}(G)\}\right) = \\
+ p_{j,\varphi_j^{-1}(F)}^*\left(\bigcup\{\varphi_j^{-1}(G) : G \in \mathcal{F}_j, \varphi_j^{-1}(F) \not\subseteq \varphi_j^{-1}(G)\}\right)
\]}

Finally, fix $F \in \mathcal{F}_j$; then
\[
p_{j,\varphi_j^{-1}(F)}^*\left(\varphi_j^{-1}(F)\right) = p_{j,\varphi_j^{-1}(F)}^*\left(\varphi_j^{-1}(F) \not\subseteq \Theta \times \{o, a\}\right) = \\
= \sum_{\theta \in \Theta \times \{o, a\}} p_{j,\varphi_j^{-1}(F)}(\varphi_j^{-1}(F) \not\subseteq \Theta \times \{\theta\}) = \frac{1}{2} \sum_{\theta \in \Theta \times \{o, a\}} p_{j,F}(\varphi_j^{-1}(F) \times \Theta) = p_{j,F}(F) > 0;
\]

furthermore, for $C \subseteq \varprojlim_{S_j} F$, $U \in \Theta$, and $\theta \in \{o, a\}$,
\[
\frac{p_{j,\varphi_j^{-1}(F)}^*(\varphi_j^{-1}(C) \times U \times \{\theta\})}{p_{j,\varphi_j^{-1}(F)}^*(\varphi_j^{-1}(F))} = \frac{1}{2} p_{j,F}(C \times U) = \frac{1}{2} \mu_j(C \times U | F) = \mu_j^*(\varphi_j^{-1}(C) \times U \times \{\theta\} | \varphi_j^{-1}(F)).
\]
Thus, \( p_j^* \) is a basis for \( \mu_j^* \).  

**Proof of Theorem 2:** Lemma 5 shows that, for every \( j \in N \), every \( \mu_j \in cpr(\Sigma_j, \mathcal{F}_j) \) admits an extension. Lemma 7 shows that, if \( \mu_j \in cpr(\Omega_i, \mathcal{F}_i) \) admits a basis, so does any extension \( \mu_j^* \) of \( \mu_j \). Eq. (17) in Lemma 7 and Remark 4 imply that, for all \( j \neq i \), \( s_j \in S_j \), and \( F \in \mathcal{F}_j \),

\[
E_{p_{j,e-j}(F)} p^*_j(s_j, \cdot) = E_{p_j,F} U_j(s_j, \cdot);
\]

similarly, for all \( s_i \in S_i \), \( k \in \{f, g\} \), and \( F \in \mathcal{F}_i \),

\[
E_{p_{i,e-i}(F)} p^*_i(s_i, \cdot) = \frac{1}{2} E_{p_i,F} U_i(s_i, \cdot) + \frac{1}{2} E_{p_i,F} u_i \circ k.
\]

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

**C.2.3 Proof of Remark 1**

To simplify the notation, denote \( \succeq_{u_i, p_i} \) by \( \succeq_i \).

Let \( F = [I] \). That the number \( a \), if it exists, is unique, follows from the fact that \( y F x_0 >_i x G x_0 \) and \( y F x_0 <_i x G x_0 \) cannot both hold.

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

\[
E_{p_i,F} u_i \circ y F x_0 = u_i(y) p_i,F(F) + u_i(x_0)[1 - p_i,F(F)] = u_i(y) p_i,F(F)
\]

and similarly

\[
E_{p_i,F} u_i \circ x G x_0 = u_i(x) p_i,F(G) + u_i(x_0)[1 - p_i,F(G)] = p_i,F(G) = \mu_i(G|F)p_i,F(F),
\]

where the last equality follows from the fact that \( p_i \) is a basis for \( \mu_i \). The same assumption implies that \( p_i,F(F) > 0 \). Therefore, it is immediate that

\[ u_i(y) > \mu_i(G|F) \implies E_{p_i,F} u_i \circ y F x_0 > E_{p_i,F} u_i \circ x G x_0 \quad \text{and} \quad u_i(y) < \mu_i(G|F) \implies E_{p_i,F} u_i \circ y F x_0 < E_{p_i,F} u_i \circ x G x_0. \]
Now consider $K \in \mathcal{F}_i$ such that $K \succ_i F$. 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 $E_{p_i,x} u_i \circ y F x_0 = u_i(x_0) = E_{p_i,x} u_i \circ x G x_0$, because $G \subseteq F$. Therefore,

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

Finally, suppose that $u_i(y) > \mu_i(G|F)$ and $K \in \mathcal{F}_i$ is such that $E_{p_i,x} u_i \circ y F x_0 < E_{p_i,x} u_i \circ x G x_0$. Then $p_K(F) > 0$, so by Lemma 1 part 1, $F \succ_i K$. Since $E_{p_i,x} u_i \circ y F x_0 > E_{p_i,x} u_i \circ x G x_0$ and $K$ was chosen arbitrarily, $y F x_0 \succ_i x G x_0$. Similarly, $u_i(y) < \mu_i(G|F)$ implies that $y F x_0 \prec_i x G x_0$. Thus,

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

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

### References


