How Many Levels Do Players Reason?
An Observational Challenge and Solution*

Adam Brandenburger† Amanda Friedenberg‡

November 16, 2014
Extremely Preliminary

Abstract

How can a researcher identify the number of levels of reasoning undertaken by players in a game?
We focus on the case where the researcher cannot observe players' beliefs. Instead, the researcher has
access to data which serve as a signal of the players' strategies. In the case of simultaneous-move games,
standard results relate levels of reasoning to rounds of elimination of dominated strategies and, thereby,
allow the researcher to partition the data to answer this question. However, in extensive-form games,
levels of reasoning cannot be directly related to elimination of conditionally dominated strategies. The
main theorem of this paper shows how to solve the researcher's inference problem in extensive-form
games.

1 Introduction

Interactive reasoning is an important aspect of how players behave. To determine whether a particular
course of action is good or bad, Ann may need to form a theory about Bob's play of the game. In forming
such a theory, she may reason that Bob is 'strategically sophisticated' — if so, she may reason that Bob
forms a belief about her own play to determine whether a particular strategy is good or bad for him. That
is, Ann may want to form a second-order theory about Bob's play of the game. Of course, Ann may then
reason that Bob uses a second-order theory to choose his strategy. In this case, Ann may want to form a
third-order theory about Bob's play of the game. And so on.

How many levels of reasoning do players undertake? This is an important question in many fields. In
economics, an answer can be used by researchers as an empirical input in addressing substantive issues.
The answer can serve to generate new (i.e., out-of-sample) predictions and, in turn, can have important

*We thank Jessy Hsieh, Terri Kneeland, Xiaomin Li, and Elliot Lipnowski for valuable comments, and Alexander Danieli for
excellent research assistance. We also thank audiences at Bristol, the 2013 Asian Meeting of the Econometric Society, National
University of Singapore, and the 2013 Workshop on Logical Structure of Correlated Information Change for important input.
Adam Brandenburger thanks the NYU Stern School of Business for financial support. Amanda Friedenberg thanks NSF
SES-1358008 for financial support and the UCL Economics Department for its unbounded hospitality.
†Stern School of Business, Polytechnic School of Engineering, Institute for the Interdisciplinary Study of Decision Making,
Center for Data Science, New York University, New York, NY 10012, email: adam.brandenburger@stern.nyu.edu, website:
adambrandenburger.com
‡W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287, email: amanda.friedenberg@asu.edu, website:
www.public.asu.edu/~afrieden
normative implications for the design of markets, institutions, and policies. Section 7 discusses why the answer is also important for work in cognitive science.

This paper focuses on the case where the players’ processes of reasoning are not observable. Instead, the researcher observes only the behavior of the players — or, perhaps, only the outcome of the game. We address the question: Can the researcher use this information to identify — or provide bounds on — levels of reasoning?

A Motivating Example

Figure 1.1 depicts the game of Battle-of-the-Sexes with an Outside Option: Ann can either choose to exercise an outside option or choose to play Battle-of-the-Sexes with Bob. The standard argument is that, if the players are ‘strategically sophisticated,’ then Ann will choose $In-U$ and Bob will choose $L$. Here is the argument. The strategy $In-D$ is dominated for Ann by the outside option. Thus, if Ann does play $In$, Bob should reason that she will play the undominated $In-U$; in this case, his best response is to play $L$. Ann should understand that Bob will reason this way and expect Bob to play $L$. With this expectation, Ann should play $In-U$.

Thus, there appears to be a clear prediction: Ann will play $In-U$ and Bob will play $L$. Yet, in the lab, $Out$ is played with significant frequency. (See, inter alia, Cooper, DeJong, Forsythe and Ross, 1993 and Brandts and Holt, 1995.) One might then draw the conclusion that there is limited reasoning: If Bob engages in one level of reasoning, he may choose either $L$ or $R$ depending on what he believes about Ann’s play. So, if Ann engages in two levels of reasoning, she might choose to play $Out$. It is only if Ann reasons three (or more) levels that she would not play $Out$.

But, in fact, the behavior need not be an artifact of limited reasoning. There is another reason why Ann might choose to play $Out$ — one based on the idea that there is a “context” to the game. In particular, suppose that it is commonly understood that “Bob is a bully” and, so, whenever a Battle of the Sexes game is played, he attempts to go for his best option and play $R$. To be specific, suppose:

Bully-1: at each information set, Ann believes that Bob plays $R$,

Bully-2: at each information set, Bob believes “Bully-1,”

Bully-3: at each information set, Ann believes “Bully-2.”
If this is the context under which the game is played, Ann may play Out, even if she reasons three levels. In fact, she may play Out, even if she engages in common reasoning about the play of the game.

Section 2 will give a formal analysis of this fact. To preview: When the game is played in this context, if Ann reasons at least one level, then she must play Out. (Ex ante, she expects that Bob will play R, and so Out is the unique best choice.) If Bob reasons at least two levels, then he will reason that Ann reasons at least one level, provided he has not observed information that contradicts this hypothesis. Thus, if Bob reasons at least two levels, he must begin the game believing Ann plays Out. Conditional upon Ann’s playing In, he is forced to reason that Ann does not reason one level. So, Bob can reason two levels and conclude (after observing In) that Ann is playing In-U, but he can also conclude that Ann is playing In-D. Thus, if Bob reasons two levels, he can play either of L or R. It follows that believing that Bob play will R is consistent with three levels of reasoning for Ann.

In light of the above, it appears premature to conclude that observing the play of Out indicates that Ann is necessarily only a level $m$-reasoner, for some $m \leq 2$. This paper focuses on the case where the players’ beliefs are unobservable by the researcher and, in a similar vein, she also cannot observe which beliefs players consider possible. Thus, we seek an answer to our question that is independent of the context of the game.

**Approach and Challenge**

In keeping with what we have just seen, we will need to describe what beliefs players do vs. do not consider possible in a particular game. The device we will use to describe these beliefs is an **epistemic type structure**, denoted by $\mathcal{T}$. An epistemic type structure will consist of a set of types for each player, where a type for a player will describe that player’s hierarchies of beliefs about the play of the game. Different type structures are associated with different events which are commonly believed. For instance, for Battle-of-the-Sexes with an Outside Option, there is a type structure where the event “Bob is a bully” is commonly believed and there are other type structures where it is not.

![Figure 1.2: Level-$m$ Reasoning](image)

Call a pair $(\Gamma, \mathcal{T})$ an **epistemic game**. For a given epistemic game $(\Gamma, \mathcal{T})$, we can define the set of strategy-type pairs which are consistent with $m$ levels of reasoning, to be denoted $R^m(\mathcal{T})$. Refer to Figure
1.2, which illustrates these sets. Specifically:

1. The set $R^0(T)$ is the set of all strategy-type pairs. This set captures **level-0 reasoning**, since there is no requirement on reasoning.

2. The set $R^1(T)$ is the set of strategy-type pairs where the players’ strategies are optimal given their belief (i.e., type). This set captures **level-1 reasoning**.

...  

$(m)$ The set $R^{m+1}(T)$ will be the set of strategy-type pairs in $R^m(T)$ where each player reasons that the other players engage in **level-$m$ reasoning**. This set captures **level-$(m+1)$ reasoning**.

...  

$(\infty)$ The set $R^\infty(T)$ will be the set of strategy-type pairs in $R^m(T)$ for all $m$. This set captures **level-$\infty$ or common reasoning**.

The sets $R^m(T)$ depend not only on the game $\Gamma$ but also on the type structure $T$. This fits with our informal analysis of Battle-of-the-Sexes with an Outside Option, where the behavior of a level-3 reasoner depended on whether or not the event “Bob is a bully” is commonly believed.

Observe that the sets $R^0(T), \ldots, R^m(T), \ldots$ are decreasing. This reflects the fact that, if players reason at least $(m+1)$ levels, then they reason at least $m$ levels. As a consequence, we will not be able to identify the minimum number of levels of reasoning by observing behavior alone. Instead, we seek to identify the maximum number of levels of reasoning consistent with observed behavior.

The goal then is to identify when a strategy is consistent with $m$ but not $(m+1)$ levels of reasoning. Toward this end, we seek to construct an ordered partition of the strategy set, namely $L = \{L^0, L^1, \ldots, L^m, \ldots, L^\infty\}$, that satisfies the following criteria: When $m$ is finite, $s \in L^m$ if it is

1. consistent with level-$m$ reasoning in some epistemic game $(\Gamma, T^*)$, but
2. inconsistent with level-$(m+1)$ reasoning in any epistemic game $(\Gamma, T)$.

When $m$ is infinite, $s \in L^m$ if it is consistent with level-$\infty$ reasoning in some epistemic game $(\Gamma, T^*)$. We refer to players who choose $s \in L^m$ as **Level-$m$ Reasoners** (or **$L^m$-Reasoners**).

For the case of a matrix, standard results give that $s \in L^m$ if and only if $s$ is $m$ but not $(m+1)$-rationalizable. One might conjecture that the same is true for the tree, where we now take “rationalizability” to mean “extensive-form rationalizability” as in Pearce (1984) and Battigalli (1997) (or, equivalently, “iterated conditional dominance,” as in Shimoji and Watson, 1998). This is not the case. Refer back to Battle-of-the-Sexes with an Outside Option (Figure 1.1). There, $Out$ is consistent with two but not three rounds of extensive-form rationalizability. But, the “Bob is a bully” analysis showed that $Out \in L^\infty$.

This paper provides a novel procedure which serves to construct the partition $L$ in a finite number of steps. A major challenge in providing such an procedure is the fact that the definition of $L^m$ makes reference to all type structures. For a given finite tree, there are infinitely many associated type structures and, therefore, searching across all type structures would appear to be an infinite task. Theorem 5.1 will provide a way to side-step this difficulty for generic games. Specifically, it does so by characterizing the

---

1 As the level-$k$ and cognitive hierarchy literatures make clear, it may be possible to identify the minimum number of levels of reasoning by making auxiliary assumptions about behavior or beliefs. See Section 7 for a discussion.
set of strategies consistent with \( m \) levels of reasoning as a property of the game alone — without reference to any type structure. Section 6 shows how to implement the procedure in a “simple” manner.

**Upper Bound on Reasoning**

This paper seeks to identify the maximum number of levels of reasoning consistent with observed behavior. The focus on this upper bound is inevitable, absent making auxiliary assumptions on either behavior or beliefs. But, there are also good reasons why this upper bound is of interest. To the extent that the researcher is interested in using the number of levels of reasoning as an empirical input, the researcher may care only that the player acts ‘as if’ she is an \( L^2 \)-Reasoner — even if she is, in fact, an \( L^1 \)-Reasoner or \( L^0 \)-Reasoner. A similar argument applies to higher levels of reasoning.

For certain datasets, it may be possible to distinguish an \( L^2 \)-Reasoner from an \( L^1 \)-Reasoner who acts as if she is an \( L^2 \)-Reasoner. With a large dataset, we should expect, on statistical grounds, to see more than just occasional in-sample play of strategies inconsistent with higher levels of reasoning. Moreover, the researcher may be able to design an experiment so that, even in small samples, we would expect to see play inconsistent with higher levels of reasoning. For instance, the design in Kneeland (2013) is based on “ring games,” where the payoff to any player depends only on his/her left-most neighbor’s choice. So, changing payoffs to a player two steps to the left of Ann should not affect her behavior if she is a \( L^2 \)-Reasoner, but should affect her behavior is she is a \( L^3 \)-Reasoner. As a consequence, varying payoffs across the experimental session should generate distinct behavior for \( L^2 \)- and \( L^3 \)-Reasoners, thereby allowing the experimenter to identify the \( L^3 \)-Reasoner.

2 **Epistemic Games and Epistemic Conditions**

We begin with description of the strategic situation, specified by an epistemic game. The next step will be to add epistemic conditions. These add endogenous restrictions on behavior and beliefs, i.e., restrictions based on strategic reasoning.

**Extensive-Form Game**

Write \( \Gamma \) for a finite two-player extensive-form game, in the sense of Kuhn (1953). The players of the game are \( a \) (Ann) and \( b \) (Bob).\(^2\) Write \( c \) for an arbitrary player in \( \{a, b\} \) and \( -c \) for the player in \( \{a, b\}\backslash\{c\} \). The tree is given by \((V, \succ)\), where \( V \) is a set of vertices (or nodes) and \( \succ \) is a successor relation on \( V \) satisfying the usual properties of a tree. The initial node is written as \( \phi \) and the set of terminal nodes is \( Z \).

Write \( H_c \) for the set of information sets at which player \( c \) moves. Assume the game is non-trivial, in the sense that each player has at least two distinct actions at some \( h \in H_c \).\(^3\) The set of information sets, viz. \( H = H_a \cup H_b \), forms a partition of \( V \backslash Z \). Player \( c \)'s extensive-form payoff function is given by \( \Pi_c : Z \to \mathbb{R} \).

Let \( S_c \) be the set of strategies for player \( c \) and let \( S = S_a \times S_b \). There is a mapping \( \zeta : S \to Z \) so that \( \zeta(s_a, s_b) \) is the terminal node reached by \((s_a, s_b)\). Say two strategies \( s_c \) and \( r_c \) are equivalent if they induce the same plan of action, i.e., \( \zeta(s_c, \cdot) = \zeta(r_c, \cdot) \). Write \([s_c] \) for the set of strategies that are equivalent to \( s_c \), and observe that, since the game is non-trivial, each \([s_c] \subsetneq S_c \).

\(^2\)The analysis extends to three or more players, up to issues of correlation.

\(^3\)This is weaker than the Kuhn (1953) assumption that each player has at least two distinct actions at each \( h \in H_c \).
Player $c$’s strategic-form payoff function is given by $\pi_c = \Pi_c \cdot \zeta$. If $s_c$ and $r_c$ are equivalent, then $\pi_c(s_c, \cdot) = \pi_c(r_c, \cdot)$. (The converse does not obtain.)

Say a strategy $s_c$ allows $h \in H$ if there exists some $s_{-c}$ so that the path from $\phi$ to $\zeta(s_c, s_{-c})$ passes through some node in $h$. Write $S_c(h)$ for the set of strategies $s_c \in S$ that allow $h$.

**Type Structure**

The premise of this paper is that players face uncertainty about how others play the game. So, Bob will have a belief about Ann’s choice from $S_a$. But, over the course of play, Bob may learn information about Ann’s choice and this information may be inconsistent with his initial hypothesis. For instance, in Battle-of-the-Sexes with an Outside Option (Figure 1.1), Bob may begin the game with a hypothesis that Ann exercises her outside option, but may come to learn that this is false. If so, he will be forced to revise his belief about Ann’s choice from $S_a$. Thus, we will need to specify Bob’s conditional beliefs about the play of the game.

Referring to $(\Omega, B(\Omega))$ as a probability space, when $\Omega$ is a compact metric space and $B(\Omega)$ is the Borel sigma-algebra on $\Omega$. Write $\mathcal{P}(\Omega)$ for the set of Borel probability measures on $\Omega$. Endow $\mathcal{P}(\Omega)$ with the topology of weak convergence so that it is again a compact metric space.

Call $(\Omega, B(\Omega), \mathcal{E})$ a **conditional probability space** if $(\Omega, B(\Omega))$ is a probability space and $\mathcal{E} \subseteq B(\Omega) \setminus \{\emptyset\}$ is finite. The collection $\mathcal{E}$ will be referred to as (a finite set of) **conditioning events**. Since $B(\Omega)$ is clear from the context, we will suppress reference to $B(\Omega)$ and simply write $(\Omega, \mathcal{E})$ for a conditional probability space.

**Definition 2.1.** Fix a conditional probability space $(\Omega, \mathcal{E})$. A **conditional probability system (CPS)** on $(\Omega, \mathcal{E})$ is some $p : B(\Omega) \times \mathcal{E} \to [0, 1]$ satisfying the following criteria:

(i) For each $E \in \mathcal{E}$, $p(\cdot|E) \in \mathcal{P}(\Omega)$ with $p(E|E) = 1$.

(ii) For each $E, F \in \mathcal{E}$ with $G \subseteq F \subseteq E$, $p(G|E) = p(G|F)p(F|E)$.

Write $\mathcal{C}(\Omega, \mathcal{E})$ for the set of CPS’s on $(\Omega, \mathcal{E})$. Note that $\mathcal{C}(\Omega, \mathcal{E}) \subseteq [\mathcal{P}(\Omega)]^{\mathcal{E}}$. Endow $[\mathcal{P}(\Omega)]^{\mathcal{E}}$ with the product topology and $\mathcal{C}(\Omega, \mathcal{E})$ with the relative topology, so that $\mathcal{C}(\Omega, \mathcal{E})$ is a compact metric space.

In our analysis, player $c$’s set of conditioning events will correspond to

$$\mathcal{E}_c = \{S_{-c}(h) : h \in H_c \cup \{\phi\}\}.$$ 

So, Ann has a conditioning event that corresponds to the beginning of the game, viz. $S_b(\phi) = S_b$, and a conditioning event that corresponds to each information set $h \in H_a$ at which she moves, viz. $S_b(h)$.

Ann will have a system of beliefs about Bob’s play, i.e., a first-order CPS on $(S_b, \mathcal{E}_b)$. If Ann also reasons about how Bob reasons about her (i.e., Ann’s) own play, then Ann has a system of beliefs about both (i) Bob’s choice from $S_b$ and (ii) Bob’s first-order CPS’s on $(S_a, \mathcal{E}_a)$. And so on. We will model these hierarchies of conditional beliefs via a type structure.

**Definition 2.2.** A **$\Gamma$-based type structure** is some $\mathcal{T} = (\Gamma; T_a, T_b; \beta_a, \beta_b)$ where

(i) $T_c$ is a compact metric type set for player $c$ and

(ii) $\beta_c : T_c \to \mathcal{C}(S_{-c} \times T_{-c}, \mathcal{E}_c \otimes T_{-c})$ is a continuous belief map for player $c$. 


As is standard, we will identify a simultaneous-move game with an extensive-form game in which all players move without information about past play. In that case, $E_a = \{S_b\}$ and $E_b = \{S_a\}$. Thus, $C(S_{-c} \times T_{-c}, E_c \otimes T_{-c}) = P(S_{-c})$, i.e., $\beta_c : T_c \to P(S_{-c} \times T_{-c})$.

Battigalli and Siniscalchi (1999) construct a canonical type structure, which induces all possible hierarchies of conditional beliefs. Their type structure $T^* = (\Gamma; T^*_a, T^*_b; \beta^*_a, \beta^*_b)$ has the property that it is type-complete (Brandenburger, 2003), i.e., for each CPS $p_c \in C(S_{-c} \times T^*_c, E_c \otimes T^*_c)$, there is a type $t_c$ with $\beta_c(t_c) = p_c$. Other type structures model the assumption that some event is commonly believed. The following example illustrates the concept relative to the “Bob is a bully” example.

**Example 2.1.** Consider the game $\Gamma$ of Battle-of-the-Sexes with an Outside Option. There exists a type structure $T = (\Gamma; T_a, T_b; \beta_a, \beta_b)$ satisfying the following properties:

- Each $\beta_a(t_a)(|S_b \times T_b)$ assigns probability one to $\{R\} \times T_b$.
- For each CPS $p_a$ with $p_a(\{R\} \times T_b|S_b \times T_b) = 1$, there is a type $t_a$ with $\beta_a(t_a) = p_a$.
- For each CPS $p_b$, there is a type $t_b$ with $\beta_b(t_b) = p_b$.

The fact that such a type structure exists follows from Battigalli and Friedenberg (2009). It captures the contextual assumption that there is common belief that Bob plays $\{R\}$. (See Appendix A in Battigalli and Friedenberg, 2009 for a formal treatment.)

**Epistemic Game**

For a given game $\Gamma$, write $T(\Gamma)$ for the set of $\Gamma$-based type structures. Since $\Gamma$ is non-trivial, there is an uncountable number of elements in $T(\Gamma)$. An (extensive-form) epistemic game is some pair $(\Gamma, T)$ with $T \in T(\Gamma)$. The epistemic game is the exogenous description of the strategic situation. An epistemic game induces a set of states, viz. $S_a \times T_a \times S_b \times T_b$.

**Epistemic Conditions**

Within a given epistemic game $(\Gamma, T)$, we can specify the conditions of rationality and reasoning about rationality. We begin with rationality.

**Definition 2.3.** Fix some $X_c \subseteq S_c$. Say $s_c \in X_c$ is a best response under $\mu \in P(S_{-c})$ given $X_c$ if

$$\sum_{s_{-c} \in S_{-c}} \left[ \pi_c(s_c, s_{-c}) - \pi_c(r_c, s_{-c}) \right] \mu(s_{-c}) \geq 0$$

for all $r_c \in X_c$.

**Definition 2.4.** Say $s_c$ is a sequential best response under $p_c \in C(S_{-c}, E_c)$ if, for each $h \in H_c$ with $s_c \in S_c(h)$, $s_c$ is a best response under $p_c(\cdot|S_{-c}(h))$ given $S_c(h)$.

Notice that each $\beta_c(t_c)$ induces a CPS in $C(S_{-c}, E_c)$ via marginalization. Write $\text{marg}_{S_{-c}} \beta_c(t_{-c}) \in C(S_{-c}, E_c)$ for this marginal CPS.

**Definition 2.5.** Fix some $(\Gamma, T)$. Say $(s_c, t_c)$ is rational if $s_c$ is a sequential best response under the marginal CPS $\text{marg}_{S_{-c}} \beta_c(t_c)$.
The next step is specifying the requirement that a player ‘reasons’ that the other player is rational. We take “reasons” to mean that a player maintains a hypothesis that the other player is rational, so long as she has not observed otherwise.

**Definition 2.6** (Battigalli and Siniscalchi, 2002). Say a CPS $p \in \mathcal{C}(\Omega, \mathcal{E})$ **strongly believes** an event $F$ if, for each conditioning event $E \in \mathcal{E}$, $E \cap F \neq \emptyset$ implies $p(F|E) = 1$.

**Definition 2.7.** A type $t_c$ **strongly believes** an event $E \subseteq S_{-c} \times T_{-c}$ if $\beta_c(t_c)$ strongly believes $E$.

Notice, in the specific case of a simultaneous-move game, “strong belief” coincides with “belief,” i.e., a type believes an event $E$ if its single probability measure assigns probability one to the event $E$.

It will be convenient to set $R^0_c(\Gamma, T) = S_c \times T_c$. Let $R^1_c(\Gamma, T)$ be the set of rational strategy-type pairs $(s_c, t_c)$ in $(\Gamma, T)$. Inductively define sets $R^m_c(\Gamma, T)$ and $R^n_c(\Gamma, T)$ so that

$$R^{m+1}_c(\Gamma, T) = R^m_c(\Gamma, T) \cap [S_c \times \{t_c : t_c \text{ strongly believes } R^m_{-c}(\Gamma, T)\}]$$

Write $R^m(\Gamma, T) = R^m_0(\Gamma, T) \times R^n_c(\Gamma, T)$. The set $R^{m+1}(\Gamma, T)$ is the set of strategy-type pairs (in $(\Gamma, T)$) at which there is rationality and $m^{th}$-order strong belief of rationality ($R_mSBR$). The set $R^\infty(\Gamma, T) = \cap_{m \in N} R^m(\Gamma, T)$ is the set of strategy-type pairs (in $(\Gamma, T)$) at which there is rationality and common strong belief of rationality. When it is clear that the associated epistemic game is $(\Gamma, T)$, we suppress reference to the epistemic game, simply writing $R^m_c$.

**Example 2.2.** Consider the game $\Gamma$ of Battle-of-the-Sexes with an Outside Option. First, consider an associated epistemic game $(\Gamma, T^*)$ where $T^* = (\Gamma; T^*_a, T^*_b, \beta^*_a, \beta^*_b)$ is type-complete. Then,

- $\text{proj}_{S_a \times S_b} R^1(\Gamma, T^*) = \{\text{Out}, \text{In-U}\} \times \{L, R\}$,
- $\text{proj}_{S_a \times S_b} R^2(\Gamma, T^*) = \{\text{Out}, \text{In-U}\} \times \{L\}$, and
- $\text{proj}_{S_a \times S_b} R^3(\Gamma, T^*) = \{\text{In-U}\} \times \{L\}$.

Next, consider an associated epistemic game $(\Gamma, T)$ where $T$ represents the assumption that “Bob is a bully” is common belief, i.e., as in Example 2.1. Then, for each $m \geq 1$, $\text{proj}_{S_a \times S_b} R^m(\Gamma, T) = \{\text{Out}\} \times \{L, R\}$. \hfill \Box

3 Identifying Level-$m$ Reasoners

The exercise we have in mind is as follows: The researcher has access to a dataset, where the dataset is indicative of players’ choices in a game $\Gamma$. Write $\mathcal{D}$ for the set of possible realizations in this dataset and write $\delta_c : S_a \times S_b \rightarrow \mathcal{D}$ for the mapping from strategies of $c$ to data realizations. For instance, under “the strategy method” in experimental economics, the dataset would consist of reported strategies and the mapping $\delta_c$ would be the projection of $S_a \times S_b$ onto $S_c$. Alternatively, the dataset might consist of information on the path of play (resp. outcome) of the game — e.g., the researcher may observe a path of bids in an auction (resp. who won the auction and at what price). In that case, $\mathcal{D}$ is the set of terminal nodes (resp. outcomes) and $\delta_c$ maps strategy profiles to their induced paths of play (resp. induced outcomes).

The researcher will use the mapping $\delta_c$ to back out the maximum level of reasoning consistent with the data, from the maximum level of reasoning consistent with a given strategy. With this in mind, say a
strategy $s_c$ is **consistent with** $m$ levels of reasoning (resp. common reasoning) for $c$ if there exists some epistemic game $(\Gamma, T^*)$ so that $s_c \in \text{proj} \bigcap_{m} R^m_m(\Gamma, T^*)$ (resp. $s_c \in \text{proj} S^*_m(\Gamma, T^*)$). Say the observed data $d$ is **consistent with** $m$ levels of reasoning (resp. common reasoning) for $c$ if there exists some $(s^*_a, s^*_b)$ with $d = s^*(s^*_a, s^*_b)$ and $s^*_c$ is consistent with $m$ levels of reasoning (resp. common reasoning) for $c$.

**Definition 3.1.** Say the observed data $d \in \mathbb{D}$ reflects the fact that $c$ is a **Level-$m$ Reasoner** if $d$ is consistent with $m$ levels of reasoning for $c$ but inconsistent with $(m + 1)$ levels of reasoning for $c$. Say the observed data $d \in \mathbb{D}$ reflects the fact that $c$ is a **Level-$\infty$ Reasoner** if, for each $m$, the data is consistent with $m$ levels of reasoning for $c$.

The data reflects the fact that $c$ is a Level-$m$ Reasoner if (i) there is some strategy $s^*_c$ consistent with both the data and $m$ levels of reasoning, and (ii) there is no strategy $s_c$ consistent with both the data and $(m + 1)$ levels of reasoning. The data reflects the fact that $c$ is a Level-$\infty$ Reasoner if there is some strategy consistent with both the data and all levels of reasoning—that is, there is no bound on the number of levels of reasoning consistent with the data. If the data is consistent with common reasoning, then certainly there is no bound on the number of levels of reasoning consistent with the data. In Section 6, we will show that the converse also holds: If there is no bound on the number of levels of reasoning consistent with the data, then the data must, in fact, be consistent with common reasoning. (See Corollary 6.1.)
In the discussion of extensive-form best response sets (EFBRS), we have a fixed point requirement: If $s_c \in S_c$ is a sequential best response under $p_c \in C(S_{-c}, E_c)$, then $s_c$ is $p_c$-justifiable if $s_c \in B[R[p_c]]$. Write $J_s[c] / X_{-c}$ for the set of strategies $s_c$ that are sequential best responses under $p_c \in C(S_{-c}, E_c)$. (See Battigalli and Friedenberg, 2012.) We begin by recalling this concept.

Given a CPS $p_c \in C(S_{-c}, E_c)$, write $B[R[p_c]]$ for the set of strategies $s_c$ that are sequential best responses under $p_c \in C(S_{-c}, E_c)$. Say $s_c$ is $p_c$-justifiable if $s_c \in B[R[p_c]]$. Write $J_s[c] / X_{-c}$ for the set of CPS's $p_c \in C(S_{-c}, E_c)$ so that (i) $s_c \in B[R[p_c]]$, and (ii) $p_c$ strongly believes $X_{-c}$.

**Definition 4.1.** Call $Q_a \times Q_b \subseteq S_a \times S_b$ an **extensive-form best response set (EFBRS)** if, for each $s_c \in Q_c$, there exists some $p_c \in J_s[c] / X_{-c}$ so that $B[R[p_c]] \subseteq Q_c$.

In Battle-of-the-Sexes with an Outside Option, there are three EFBRS's, viz. $\{\text{Out}\} \times \{L, R\}$, $\{\text{Out}\} \times \{R\}$, and $\{\text{In-U}\} \times \{L\}$.

**Proposition 4.1** (Battigalli and Friedenberg, 2012). Fix a game $\Gamma$.

(i) For each epistemic game $(\Gamma, T)$, $\text{proj}_S R^\infty(\Gamma, T)$ is an EFBRS.

(ii) Given an EFBRS $Q_a \times Q_b$, there exist an epistemic game $(\Gamma, T)$, so that $\text{proj}_S R^\infty(\Gamma, T) = Q_a \times Q_b$.

**Corollary 4.1.** For each game $\Gamma$,

$$S^\infty = \bigcup_{Q_a \times Q_b \text{ is an EFBRS}} (Q_a \times Q_b).$$

This result says that EFBRS's characterize the behavior consistent with RCSBR. The EFBRS concept has a fixed point requirement: If $Q_a \times Q_b$ is an EFBRS then, for each $a \in Q_a$, there is a CPS $p_a$ under which $s_a$ is optimal, so that $p_a$ strongly believes the 'prediction' $Q_b$. This fixed-point requirement naturally arises from characterizing RCSBR behavior: Fix some strategy-type pair $(s_a, t_a)$ at which there is RCSBR. First,
$s_a$ is optimal under $\text{marg}_{S_a} \beta_a(t_a)$. Second, since $\text{marg}_{S_a} \beta_a(t_a)$ assigns probability one to each $\text{proj}_{S_a} R^m_b$, $\text{marg}_{S_a} \beta_a(t_a)$ assigns probability one to $\text{proj}_{S_a} R^\infty_b$. Third, if $r_a$ is also optimal under $\text{marg}_{S_a} \beta_a(t_a)$, then $(r_a, t_a)$ is also rational. It then follows that $(r_a, t_a) \in R^m$ for all $m$, i.e., $r_a \in \text{proj}_{S_a} R^\infty$.

A natural starting point is to turn this fixed-point definition into an iterative definition. In this section, we present a natural starting point, based on what is known from characterizing common reasoning. We will then see that, for an arbitrary game, this first step is insufficient.

```
Fix an epistemic game $(\Gamma, \mathcal{T})$. We have the following basic properties of the sets $R^0, R^1, R^2, \ldots$:
```

- Each $S \times \text{proj}_{S} R^m$ is a product set, i.e., $\text{proj}_{S} R^m = \text{proj}_{S} R^m \times \text{proj}_{S} R^m$.
- The sequence $(\text{proj}_{S} R^0, \text{proj}_{S} R^1, \ldots)$ is decreasing, i.e., $\text{proj}_{S} R^{m+1} \subseteq \text{proj}_{S} R^m$.

With this in mind, we will abuse terminology and say that $(S^0, \ldots, S^m)$ is a decreasing sequence of strategy profiles if (i) for each $n = 0, \ldots, m$, $S^n = S^n_a \times S^n_b$, (ii) $S^0 = S_a \times S_b$, and (iii) for each $n = 0, \ldots, m-1$, $S^{n+1} \subseteq S^n$.

We take the convention that $X = X_a \times X_b = \emptyset$ implies $X_a = \emptyset$ and $X_b = \emptyset$.

**Definition 4.2.** Say $Q = Q_a \times Q_b$ satisfies the (extensive-form) best response property relative to $(S^0, \ldots, S^m)$ if $(S^0, \ldots, S^m, Q)$ is a decreasing sequence of strategy profiles satisfying the following property: for each $s_c \in Q_c$, there is a CPS $p_c \in \bigcap_{n=1}^m J[s_c] S^n - c$ so that $\mathcal{B}R[p_c] \subseteq Q_c$.

**Definition 4.3.** Let $m \geq 1$. Say $(S^0, \ldots, S^m)$ satisfies the $m$-(extensive-form) best response property (m-BRP) if

- (i) $S^1$ is non-empty and
- (ii) for each $n = 1, \ldots, m-1$, $S^{n+1}$ satisfies the best response property relative to $(S^0, \ldots, S^n)$.

Notice that, when $m \geq 2$, $(S^0, \ldots, S^m)$ satisfies the m-BRP if and only if $(S^0, \ldots, S^{m-1})$ satisfies the $(m-1)$-BRP and $S^m$ satisfies the extensive-form best response property relative to $(S^0, \ldots, S^{m-1})$.

**Proposition 4.2.** Fix an epistemic game $(\Gamma, \mathcal{T})$. The sequence $(\text{proj}_{S} R^0, \ldots, \text{proj}_{S} R^m)$ satisfies the m-BRP.

Note the following implication of Proposition 4.2. If a set $Q = Q_a \times Q_b$ is consistent with $\text{RmSBR}$, then $Q$ is consistent with the $(m+1)$-BRP, i.e., there exists a decreasing sequence of sets $(S^0, \ldots, S^m)$ so that $(S^0, \ldots, S^m, Q)$ satisfies the $(m+1)$-BRP. Thus, the $(m+1)$-BRP provides an upper bound on the behavior consistent with $\text{RmSBR}$. The argument is standard, but we include it for completeness.

**Proof of Proposition 4.2.** Fix an epistemic game $(\Gamma, \mathcal{T})$. We will show that, for each $m \geq 1$, $(\text{proj}_{S} R^0, \ldots, \text{proj}_{S} R^m)$ satisfies the m-BRP. The proof is by induction on $m$.

- $m = 1$: If $s_c \in \text{proj}_{S} R^1_c$, then there exists some $t_c \in T_c$ so that $(s_c, t_c) \in R^1_c$. Take $p_c = \text{marg}_{S_{-c}} \beta_c(t_c)$. Note that $p_c \in J[s_c]$. Moreover, if $r_c \in \mathcal{B}R[p_c]$, then $(r_c, t_c) \in R^1_c$ and so $r_c \in \text{proj}_{S} R^1_c$.

- $m \geq 2$: Assume the claim holds for $m$ and fix some $(\text{proj}_{S} R^0, \ldots, \text{proj}_{S} R^m, \text{proj}_{S} R^{m+1})$. Then, by the induction hypothesis, $(\text{proj}_{S} R^0, \ldots, \text{proj}_{S} R^m)$ satisfies the m-BRP. Thus, it suffices to show that
Let $\Gamma$ be the simultaneous-move game in Figure 4.1. Take $s_c \in \text{proj}_{S_c} R_{m+1}$ satisfies the extensive-form best response property relative to $(\text{proj}_{S_0} R^0, \ldots, \text{proj}_{S_m} R^m)$.

Fix some $s_c \in \text{proj}_{S_c} R_{m+1}$. There exists some $t_c \in T_c$ so that $(s_c, t_c) \in R^m_{c+1}$. Take $p_c = \text{marg}_{S_c} \beta_c(t_c)$. Since $(s_c, t_c) \in R^1_c$, $p_c \in J[s_c]$. Moreover, $\beta_c(t_c)$ strongly believes $R^0_{c-1}, \ldots, R^m_{c-1}$. So applying Lemma A.1, marg $s_c \beta_c(t_c)$ strongly believes $\text{proj}_{S_{c-1}} R^0_{c-1}, \ldots, \text{proj}_{S_{c-1}} R^m_{c-1}$. Finally, if $r_c \in \mathbb{B}[p_c]$, then $(r_c, t_c) \in R^m_{c+1}$ and so $r_c \in \text{proj}_{S_c} R_{m+1}$.

We next show that a set $Q = Q_a \times Q_b$ may be consistent with the $(m+1)$-BRP even though it is inconsistent with $R_{m+1}$-BR, i.e., even though there is no epistemic game $(\Gamma, T)$ with

$$(\text{proj}_{S_0} R^0(\Gamma, T), \ldots, \text{proj}_{S_m} R^m(\Gamma, T), \text{proj}_{S_{m+1}} R^{m+1}(\Gamma, T)) = (S^0, \ldots, S^m, Q).$$

This is different from the common reasoning analogue, viz. Proposition 4.1.

**Example 4.1.** Let $\Gamma$ be the simultaneous-move game in Figure 4.1. Take $S^0 = S^1 = \{U, D\} \times \{L, R\}$ and $S^2 = \{U\} \times \{L, R\}$. It is readily verified that $(S^0, S^1, S^2)$ satisfies the 2-BRP.

```
<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>1,1</td>
<td>1,0</td>
</tr>
<tr>
<td>D</td>
<td>1,1</td>
<td>0,1</td>
</tr>
</tbody>
</table>
```

Figure 4.1

We will show that there is no $(\Gamma, T)$ so that $\text{proj}_{S_0} R^2 = \{U\} \times \{L, R\}$. Suppose, contra hypothesis, there is some $(\Gamma, T)$ so that $\text{proj}_{S_0} R^2(\Gamma, T) = \{U\} \times \{L, R\}$. Then, there exists some $t_b$ so that $(R, t_b) \in R^2_b(\Gamma, T)$. Since $(R, t_b) \in R^1_b(\Gamma, T)$, $\beta_b(t_b)(\{D\} \times T_a) > 0$. Since $t_b$ believes $R^1_a$, $D \in \text{proj}_{S_a} R^1_a$. Using the fact that $\text{proj}_{S_a} R^2_a(\Gamma, T) \subseteq \text{proj}_{S_a} R^1_a(\Gamma, T)$, $\text{proj}_{S_a} R^1_a(\Gamma, T) = \{U, D\} \times \{L, R\} = S^1 = S^0$.

Now, since $D \in \text{proj}_{S_b} R^1_b(\Gamma, T)$, there exists some $t_a$ so that $(D, t_a) \in R^1_b(\Gamma, T)$. It follows that $\beta_a(t_a)(\{L\} \times T_b) = 1$. But note that $(L) \times T_b \subseteq R^1_a(\Gamma, T)$ and so $\beta_a(t_a)$ assigns probability 1 to $R^1_a(\Gamma, T)$. It follows that $(D, t_a) \in R^1_b(\Gamma, T)$, a contradiction.

Proposition 4.2 states that $R_{m+1}$BR is characterized by a subset of the sets consistent with the $(m+1)$-BRP. Example 4.1 illustrates that this subset is strict, i.e., the $R_{m+1}$BR predictions may not include all subsets consistent with the $(m + 1)$-BRP. But, the game in Example 4.1 was not generic. The next section shows that this failure of genericity was important.

## 5 Generic Games

Say $s_c$ is **justifiable** if there exists some CPS $p_c$ with $s_c \in \mathbb{B}[p_c]$. Note, if $s_c \in \mathbb{B}[p_c]$ then $[s_c] \subseteq \mathbb{B}[p_c]$. So $s_c$ is justifiable if and only if $[s_c]$ is justifiable.

**Definition 5.1.** Call a game $\Gamma$ **generic** if, for each justifiable strategy $s_c \in S_c$, there exists some CPS $p_c$ so that $[s_c] = \mathbb{B}[p_c]$.
Note, Example 4.1 is a non-generic game. The strategy $D$ is justifiable. But, for any $p_a$, either $B\mathbb{R}[p_a] = \{U\}$ or $B\mathbb{R}[p_a] = \{U, D\}$. A perfect-information game satisfying “no relevant ties” (Battigalli, 1997) satisfies the genericity requirement. (See Lemma C.1 in the Appendix.) But, the definition of genericity adopted here may fail even if a game has no ties. (See Example C.1 in the Appendix.) Moreover, the definition does not require no ties.

When a game is generic, in the sense of this paper, the predictions of $R_mSBR$ are exactly captured by the sets consistent with the $(m + 1)$-BRP:

**Theorem 5.1.** Fix a generic game $\Gamma$. The following hold for each $m$.

(i) For each epistemic game $(\Gamma, T)$, $(\text{proj}_S R^0(\Gamma, T), \ldots, \text{proj}_S R^m(\Gamma, T))$ satisfies the $m$-BRP.

(ii) If $(S^0, \ldots, S^m)$ satisfies the $m$-BRP, then there exists some epistemic game $(\Gamma, T)$ so that $(\text{proj}_S R^0(\Gamma, T), \ldots, \text{proj}_S R^m(\Gamma, T)) = (S^0, \ldots, S^m)$.

Part (i) is a special case of Proposition 4.2. Part (ii) is specific to generic games. It says that, for a generic game and an associated $m$-BRP, we can construct a type structure so that, for each $n = 0, \ldots, m - 1$, the predictions of $R_nSBR$ are exactly captured by $S^{n+1}$. As a Corollary of Theorem 5.1, we have a characterization of $S^m$ for generic games.

**Corollary 5.1.** Fix a generic game $\Gamma$. For each finite $m$,

$$S^m = \{(a, b) : (a, b) \in Q_0 \times Q_1 \text{ and } Q_0 \times Q_1 \text{ is consistent with the } m\text{-BRP}\}.$$ 

In the case of simultaneous-move games, the union over sets $Q_0 \times Q_1$ that satisfy the $m$-BRP is the set of strategies that survive $m$ rounds of elimination of strongly dominated strategies (or $m$ rounds of rationalizability). Thus, Theorem 5.1 and Corollary 5.1 collapse to standard results. The innovation here is that Corollary 5.1 applies to a general extensive-form game. In that case, $S^m$ need not be the set of strategies that survive $m$ rounds of eliminating conditionally dominated strategies (or $m$ rounds of extensive-form rationalizability). For instance, in Battle-of-the-Sexes with an Outside Option, $S^m = \{\text{Out, In-U}\} \times \{L, R\}$ for all $m \geq 1$, whereas two rounds of eliminating conditionally dominated strategies would give $\{\text{Out, In-U}\} \times \{L\}$.

We conclude this section by providing a sketch of the proof of Theorem 5.1-(ii). The aim is to highlight the role of genericity. Throughout, fix a generic game $\Gamma$ and a 2-BRP $(S^0, S^1, S^2)$. The goal is to construct a type structure $T$ so that $\text{proj}_S R^1 = S^1$ and $\text{proj}_S R^2 = S^2$.

Begin by setting the type set for player $c$ to be $T_c = S^1_c \cup S^2_c$. For each $s_c \in S^2_c \subseteq S^1_c$, there will be two associated types: a 1-type labeled $s^1_c$ and a 2-type labeled $s^2_c$. The idea will be to construct belief maps so that 1-strategy-type pairs $(s_c, s^1_c)$ are rational but do not strongly believe rationality, and 2-strategy-type pairs are rational and strongly believe rationality. In fact, we will ask for somewhat more. Refer to Figure 5.1. We will seek to construct belief maps satisfying two requirements: First, $R^1_c \setminus R^2_c$ is ‘along the diagonal’ of $S^1_c \times S^1_c$ modulo equivalence classes, i.e.,

$$R^1_c \setminus R^2_c = \{[s_c] \times [s^1_c] : s_c \in S^1_c\}.$$ 

This diagonal is illustrated as the red squares in Figure 5.1. Second, $R^2_c$ contains the diagonal of $S^2_c \times S^2_c$ and is contained in the square $S^2_c \times S^2_c$. The blue rectangles in Figure 5.1 illustrate this set.
Begin by constructing the beliefs associated with 2-types. By definition of a 2-BRP, for each \( s \in S_c^2 \), there is a CPS \( j_c(s^2) \) on \((S_{-c}, \mathcal{E}_c)\) so that \([s] \subseteq \mathbb{BR}[j_c(s^2)] \subseteq S_c^2 \) and \( j_c(s^2) \) strongly believes \( S_{-c}^1 \). (This follows from the definition of a 2-BRP.) Now choose \( \beta_c(s^2) \) so that (i) marg \( S_{-c} \beta_c(s^2) = j_c(s^2) \) and (ii) if \( s_{-c} \in S_{-c}(h) \cap S_{-c}^1 \), then \( \beta_c(s^2)((s_{-c}, S_{-c}^1(h) \times T_{-c}) = j_c(s^2)((s_{-c}, S_{-c}(h))) \). So, if \( S_{-c}(h) \cap S_{-c}^1 \neq \emptyset \), then \( \beta_c(s^2)(\cdot|S_{-c}(h) \times T_{-c}) \) is concentrated on the diagonal of \( S_{-c}^1 \times S_{-c}^1 \).

Next construct the beliefs associated with 1-types. Since the game is generic, for each \( s \in S_c^1 \), there is a CPS \( j_c(s^1) \) on \((S_{-c}, \mathcal{E}_c)\) so that \([s] = \mathbb{BR}[j_c(s^1)]\). In fact, if \( S_{-c}^1 \neq S_{-c} \), we can and do choose this CPS so that \( j_c(s^1) \) does not strongly believe \( S_{-c}^1 \). (Lemma B.3 in the Appendix shows how this is done.) We choose \( \beta_c(s^1) \) so that marg \( S_{-c} \beta_c(s^1) = j_c(s^1) \) and each \( \beta_c(s^1)(\cdot|S_{-c}(h) \times T_{-c}) \) is concentrated on 1-types (i.e., on \( S_{-c} \times S_{-c}^1 \)). Moreover, if \( S_{-c}^1 = S_{-c} \), then we also add an additional requirement—one that is, in a sense, an anti-diagonal construction. Refer to Figure 5.2. Observe that each \( \beta_c(s^2)(\cdot|S_{-c}(h) \times T_{-c}) \) is concentrated on the diagonal of \( S_{-c}^1 \times S_{-c}^1 \). Specifically, if \( s_{-c} \in [t_{-c}] \subseteq S_{-c}^1 \), then \( \beta_c(s^2)((s_{-c}, t_{-c})|S_{-c}(h) \times T_{-c}) = 0 \).

The fact that this ‘off the diagonal’ construction can be achieved follows from the fact that the game is non-trivial and the fact that \( S_{-c}^1 = S_{-c} \).

Observe that, under the construction, we have

\[
R_c^1 = \{[s] \times [s]: s \in S_c^1\} \cup \bigcup_{s^2 \in S_c^2} (\mathbb{BR}[j_c(s^2)] \times \{s^2\}).
\]

Now, by the diagonal construction, types in \( S_c^2 \) strongly believe \( R_{-c}^1 \). Moreover, types in \( S_c^1 \) do not strongly believe \( R_{-c}^1 \). If \( S_{-c}^1 \neq S_{-c} \), then this follows from the fact that each \( \beta_c(s^2)(\cdot|S_{-c}(h) \times T_{-c}) \) is concentrated on 1-types and each \( j_c(s^1) \) does strongly believe \( S_{-c}^1 \). If \( S_{-c}^1 = S_{-c} \) this follows from the anti-diagonal construction. Thus, \( R_c^1 = \bigcup_{s^2 \in S_c^2} (\mathbb{BR}[j_c(s^2)] \times \{s^2\}) \).
6 Termination of the Procedure

Theorem 5.1 provides a key step in identifying a bound on reasoning. Fix a game $\Gamma$. It gives that for each finite $m$,

$$S^m = \{(s_a, s_b) \in S_a \times S_b : (s_a, s_b) \in Q_a \times Q_b \text{ and } Q_a \times Q_b \text{ is consistent with the } m\text{-BRP}\}.$$ 

Observe that the sets $S^0, S^1, \ldots, S^m, \ldots$ are decreasing. Thus, we can think of these sets as implicitly defining an elimination procedure. Since the game is finite, there exists some finite number $\overline{M}$ so that $S^m = S^\overline{M}$ for all finite $m \geq \overline{M}$. (Observe that $\overline{M}$ may depend on the game.)

This suggests a procedure for computing $L_c$ within a finite number of steps: For each $m = 0, \ldots, \overline{M} - 1$, set $L^m_c = S^m_c \setminus S^{m+1}_c$. Set $L^\infty_c = S^\overline{M}_c$. At this point, the procedure terminates, taking $L^m_c = \emptyset$ for all finite $m \geq \overline{M} + 1$.

But, observe, to apply this procedure in practice, the researcher must be able to determine that the elimination procedure $(S^0, S^1, \ldots, S^m, \ldots)$ has stopped shrinking, i.e., to determine that $\overline{M}$ satisfies $S^m = S^\overline{M}$ for all finite $m \geq \overline{M}$. This is crucial for determining that the final step of the procedure, setting $L^\infty_c = S^\overline{M}_c$, can be implemented.

At first glance, there might appear to be straightforward routes to determine $\overline{M}$. One possibility would be to find some value $m$ so that $S^m = S^\infty$. By Proposition 4.1 the strategies in $S^m = S^\infty$ would then be consistent with common reasoning. However, it is not obvious that such a number $m$ exists: Recall, $S^\overline{M}$
is the set of strategy profiles for which there is no bound on the number of levels of reasoning consistent with that profile. That is, for any \((s_a, s_b) \in S^m\) and any \(m\), there is some epistemic game \((\Gamma, T^m)\) with \((s_a, s_b) \in \text{proj}_{S_a \times S_b} R^m_a(\Gamma, T^m)\). This set contains the union of the EFBRS's. But, in principle, there may be some profile \((s_a, s_b) \in S^m\) so that, for any epistemic game \((\Gamma, T)\), \((s_a, s_b) \not\in \text{proj}_{S_a \times S_b} \bigcap_m R^m_a(\Gamma, T)\).²

A second possibility is to draw a lesson from other elimination procedures: In the case of, say, rationalizability, iterated weak dominance, etc., the procedure stops shrinking at the first round where no strategy is eliminated for either player. (Thus, the number of steps required is bounded by the cardinality of the strategy set.) The next example illustrates that, because the current elimination procedure is derived from an epistemic analysis, it is not obvious that a similar property necessarily holds.

---

**Example 6.1.** Consider the simultaneous-move game \(\Gamma\), given by Figure 6.1. Fix some \(m \geq 1\). We will construct some \((\Gamma, T^m)\) so that

- \(\text{proj}_{S} R^{n}(\Gamma, T^m) = \{(U, D) \times \{L, R\}\} \) for each \(n \leq 2(m - 1)\) and
- \(\text{proj}_{S} R^{2m-1}(\Gamma, T^m) = \{(U, D) \times \{R\}\} \).

Since we can choose \(m\) arbitrarily, this says that there is no finite \(M\) so that, for each \(m \geq M\) and each \(m\)-BRP \((S^0, \ldots, S^M, \ldots, S^m)\), \(S^m = S^M\).

We now turn to construct \(T^m\): The type sets are \(T_a = \{t_{a1}^1, \ldots, t_{am}^m, u_a\}\) and \(T_b = \{t_{b1}^1, \ldots, t_{bm}^m, u_b\}\). The belief maps are described in Figure 6.2: For each \(n = 1, \ldots, m-1\), \(\beta_a(t_{a}^n)(L,t_{b}^{n+1}) = 1\) and \(\beta_a(u_{a})(R,u_{b}) = 1\). For each \(n = 1, \ldots, m-1\), \(\beta_b(t_{b}^n)(U,t_{a}^{n+1}) = 1\) and \(\beta_b(t_{b}^m)(D,u_{a}) = \beta_b(u_{b})(D,u_{a}) = 1\).

---

Suppose we could show that there is an \(S^m = S^\infty\). In that case, implementing the procedure would involve also computing the EFBRS's of the game. Since the definition of an EFBRS involves a fixed-point operation, this step may be computationally taxing.
The set of rational strategy-type pairs are
\[ \begin{align*}
R_b^1(\Gamma, T^m) &= (\{U\} \times \{t_b^1, \ldots, t_b^m\}) \cup \{(D, u_b)\}, \\
R_a^1(\Gamma, T^m) &= (\{L\} \times \{t_a^1, \ldots, t_a^{m-1}\}) \cup \{(R) \times \{t_a^m, u_a\}\}.
\end{align*} \]

Type \( t_b^m \) assigns probability one to \((L, t_b^m)\) and so does not believe \( R_b^1(\Gamma, T^m) \). But each of Ann’s other types, viz. \( t_a^1, \ldots, t_a^{m-1}, u_a \), believes \( R_a^1(\Gamma, T^m) \) and each of Bob’s types believes \( R_a^1 \). Thus, \( R_a^2(\Gamma, T^m) = R_a^1(\Gamma, T^m) \setminus \{(U, t_a^m)\} \) and \( R_b^2(\Gamma, T^m) = R_b^1(\Gamma, T^m) \). Now notice that type \( t_b^{m-1} \) assigns probability one to \((U, t_b^m)\) and so does not believe \( R_b^2(\Gamma, T^m) \). It follows that \( R_b^3(\Gamma, T^m) = R_b^2(\Gamma, T^m) \) and \( R_a^3(\Gamma, T^m) = R_a^2(\Gamma, T^m) \setminus \{(L, t_a^{m-1})\} \). And so on. Proceeding inductively we see that \( R_a^{2m-1}(\Gamma, T^m) = \{(U, t_a^m), (D, u_a)\} \) and \( R_b^{2m-1}(\Gamma, T^m) = \{R\} \times \{t_b^m, u_b\} \). From this, \( \text{proj}_S R^{2m-1}(\Gamma, T^m) = \{U, D\} \times \{R\} \).

\[ \text{Figure 6.3: m-BRP Elimination Procedure} \]

Fix a sequence \((S^0, S^1, S^2, \ldots)\), where each \((S^0, \ldots, S^m)\) specifies an m-BRP. We can also think of \((S^0, S^1, S^2, \ldots)\) as defining an elimination procedure. Since the strategy set is finite, this procedure must terminate. But, referring to Figure 6.3, the procedure can involve pauses before eliminations occur. Figure 3.1 explains why this can occur. We may have that \( \text{proj}_S R^2(\Gamma, T) = \text{proj}_S R^3(\Gamma, T) \) even though \( R^3(\Gamma, T) \subseteq R^2(\Gamma, T) \).

With this in mind, the mere fact that \( S^1 = S^0 \) does not mean that the \((S^0, S^1, S^2, \ldots)\) procedure has terminated. Moreover, Example 6.1 shows that there is no (even game-dependent) bound \( M \) so that the procedure necessarily terminates within \( M \) steps: In that game, for any \( M \), we can construct some m-BRP \((S^0, \ldots, S^M, \ldots, S^m)\) so that the \((S^0, S^1, S^2, \ldots)\) procedure has not terminated within \( M \) steps.

Since \( \overline{S}^m \) is the union of all sets consistent with the m-BRP, it is not obvious how to determine that the elimination procedure \((\overline{S}^0, \overline{S}^1, \ldots, \overline{S}^m, \ldots)\) has stopped shrinking. Yet, the next result shows that \( \overline{M} \) is bounded by the cardinality of the strategy set. (The proof will also point to a tighter method to determine \( \overline{M} \).) It also shows that when the procedure \((\overline{S}^0, \overline{S}^1, \ldots, \overline{S}^m, \ldots)\) stops shrinking, the resulting solution is the union of the EFBRS’s, i.e., \( S^{\infty} = \overline{S}^{\infty} \).

**Proposition 6.1.** Fix a game \( \Gamma \) and set
\[ \overline{M} = \begin{cases} 
2 \min\{|S_a|, |S_b|\} - 1 & \text{if } |S_a| \neq S_b, \\
2 \min\{|S_a|, |S_b|\} - 2 & \text{if } |S_a| = S_b.
\end{cases} \]

Then, for all \( m \geq \overline{M} \), \( S^m = \overline{S}^{\infty} \).

Before coming to the proof, we observe a consequence of Propositions 4.1 and 6.1:

**Corollary 6.1.** The data is consistent with Level-\( \infty \) Reasoning for \( c \) if and only if the data is consistent with common reasoning for \( c \).

**Proof of Proposition 6.1.** Fix some game \( \Gamma \). Let \( S = (S^0, S^1, \ldots) \) be a BRP-sequence, i.e., for each finite \( m \), \((S^0, \ldots, S^m)\) is an m-BRP. Since the game is finite, there is some \( M(S) \) so that, \( S^{M(S)} = S^{M(S)+1} \).
We can and do choose $M(\mathcal{S})$ so that

$$M(\mathcal{S}) = \begin{cases} 
2 \min(|S_a|,|S_b|) - 1 & \text{if } |S_a| \neq |S_b|, \\
2 \min(|S_a|,|S_b|) - 2 & \text{if } |S_a| = |S_b|.
\end{cases}$$

Then take $\overline{M}$ to be the maximum of all such $M(\mathcal{S})$ and observe that it, too, is less than or equal to $2 \min\{|S_a|,|S_b|\} - 1$ (resp. $2 \min\{|S_a|,|S_b|\} - 2$) if $|S_a| \neq |S_b|$ (resp. $|S_a| = |S_b|$).

It remains to show that $\overline{\mathcal{S}}^{\overline{M}} = \mathcal{S}^\infty$. Certainly $\mathcal{S}^\infty \subseteq \overline{\mathcal{S}}^{\overline{M}}$. Observe that that

$$\overline{\mathcal{S}}^{\overline{M}} \subseteq \bigcup_{\text{BRP-sequences } \mathcal{S}} \mathcal{S}^{M(\mathcal{S})}.$$  

For each BRP-sequence $\mathcal{S}$, $\mathcal{S}^{M(\mathcal{S})} = \mathcal{S}^{M(\mathcal{S})+1}$ and so $\mathcal{S}^{M(\mathcal{S})}$ is itself an EFBRS. With this $\mathcal{S}^{M(\mathcal{S})} \subseteq \mathcal{S}^\infty$, establishing that $\overline{\mathcal{S}}^{\overline{M}} \subseteq \mathcal{S}^\infty$. □

The proof of Proposition 6.1 provides a tight method to determine the lowest number $\overline{M}$ so that $\mathcal{S}^0 = \mathcal{S}^\infty$ for all $m \geq \overline{M}$:

- Set $\mathcal{S}^0$ to be the full strategy set.
- Identify all the $\mathcal{S}^1 \subseteq \mathcal{S}^0$ so that $(\mathcal{S}^0, \mathcal{S}^1)$ satisfies 1-BRP.
- For each 1-BRP $(\mathcal{S}^0, \mathcal{S}^1)$, identify all the $\mathcal{S}^2 \subseteq \mathcal{S}^1$ so that $(\mathcal{S}^0, \mathcal{S}^1, \mathcal{S}^2)$ satisfies 2-BRP.
- ... We can think of this procedure as a series of paths to identify the $m$-BRP’s. (Figure 6.3 illustrated one potential path.) Along any path there may be pauses, i.e., it may be that along some path $\mathcal{S}^2 \subsetneq \mathcal{S}^1 = \mathcal{S}^0$ while along another path $\mathcal{S}^2 = \mathcal{S}^1 \subsetneq \mathcal{S}^0$, etc. After a finite number of steps, we will have had a pause along every path. The proof of Proposition 6.1 says that the number $\overline{M}$ can be taken to be the first point at which there has been a pause along every path.

7 Discussion

a. Modeling Finite Levels of Reasoning  Type structures induce infinite hierarchies of beliefs about the play of the game — i.e., $m^{th}$-order beliefs about play, for all $m$. A literal interpretation of our formal treatment, then, is that our players reason to an infinite level. This might seem to contradict our earlier terminology, where we talked about $L^m$-Reasoners. However, the contradiction is illusory. The key observation is that if Ann is an $L^m$-Reasoner (about Bob’s rationality), then hierarchies of beliefs beyond level $m$ do not affect her behavior. Formally, consider two types $t_a$ and $u_a$ with the same $(m+1)^{th}$-order beliefs about the strategies played in the game. For any strategy $s_a$, the strategy-type pair $(s_a, t_a)$ is consistent with $RmSBR$ if and only if $(s_a, u_a)$ is consistent with $RmSBR$. The higher-order beliefs become an artifact of our formalism and do not have any behavioral significance.

To better understand this point, compare two scenarios:
Scenario 1 Ann has a belief about the strategy Bob plays. But, she does not have a belief about Bob’s belief about the strategy she plays. This might be because Ann cannot reason two levels, or because, while she can reason two levels, she thinks Bob cannot reason even one level.

Scenario 2 Ann reasons ad infinitum about how the game is played. But she does not assign probability 1 to the event that Bob is rational.

The second scenario is directly expressible in our framework. The first scenario is not. But because, in the first scenario, Ann does not have a belief about Bob’s belief, she cannot reason about whether Bob is rational or irrational. So, the two scenarios are identical in terms of their implications for Ann’s behavior.

Now consider a third scenario:

Scenario 3 Ann has a belief about the strategy Bob plays and chooses her strategy accordingly. She thinks it possible that Bob cannot reason even one level (as in Scenario 1) and also thinks it possible he reasons one level (but no higher).

This scenario is also not directly expressible in our framework. But, from the perspective of Ann’s behavior, this too can be captured in our framework. It corresponds to the case where Ann assigns a certain positive probability — viz., the probability she assigns to Bob’s reasoning one level — to Bob’s being rational.

b. Lower Bound on Reasoning The main theorem of this paper identifies an upper bound on the number of levels of reasoning consistent with the data. We do not provide a lower bound: If a strategy-type pair \((s_a, t_a) \in R^m_a\), then it is also contained in \(R^1_a\) and even \(R^0_a\). This raises a problem for the identification of a non-trivial lower bound on the number of levels of reasoning.

The researcher may be able to impose specific assumptions about behavior and beliefs which, in certain games, do identify both a non-trivial lower bound and an upper bound on the number of levels of reasoning. This is a lesson familiar from the “level-k” and “cognitive hierarchy” literatures. (See Nagel, 1995, Stahl and Wilson, 1995, Costa-Gomes, Crawford and Brosella, 2001, and Camerer, Ho and Chong, 2004, among many others.)

These literatures focus on simultaneous-move games. They begin by anchoring the behavior of level-0 players. (For instance, these players may choose at random or may choose some focal action.) First-order beliefs are then constructed under the assumption that level-0 players behave this way, and the behavior of level-1 players can thereby be deduced. Second-order beliefs are then constructed under the assumption that level-1 players behave as deduced — or, in the case of the cognitive hierarchy literature, that level-0 and level-1 players behave as assumed and deduced. Behavior of level-2 players can then be deduced. In laboratory experiments, games are typically constructed so that the behavior of level-2 players is distinct from behavior of level-1 players. And so on, inductively. Thus, in the context of specific games, assumptions about behavior and beliefs allow the researcher to derive both non-trivial lower and upper bounds on the number of levels of reasoning consistent with observed behavior.

By contrast, we do not anchor behavior of level-0 reasoners. We only require that they not choose conditionally-dominant strategies. As a consequence, the beliefs of level-1 reasoners face fewer constraints. This is natural from the perspective of identifying a level \(m\) so that players act ‘as if’ they are level-\(m\) reasoners.

\[5\] It is directly expressible in the frameworks of Kets (2010) and Heifetz and Kets (2013).
c. Connections to Cognitive Science  Reasoning about reasoning is an important area of study in cognitive science, where it goes under several names, including theory of mind, intentionality, and mentalizing (Dunbar, 2004). A common exercise in this area is to present subjects with stories (verbal or visual) involving social situations with several characters and then to test their theory-of-mind ability by asking them to answer questions about the stories (see, e.g., Kinderman, Dunbar and Bentall, 1998).

In neuroscience, fMRI investigation of subjects engaged in exercises such as the above has identified regions of the brain active in theory-of-mind processing. These regions overlap significantly with regions activated when human subjects are engaged in playing a game against another person (Gallagher and Frith, 2003). (See McCabe, Houser, Ryan, Smith and Trouard, 2001, Gallagher, Jack, Roepstorff and Frith, 2002, and Rilling, Sanfey, Aronson, Nystrom and Cohen, 2004, among others, for neuroimaging studies of game-playing subjects.) This finding offers strong evidence that players in a game do indeed engage in reasoning about reasoning, when other players are known to be intentional.

The result in our paper could be used to design experiments in which subjects play extensive games under neural monitoring. Differences in how the games are played could then be related, via our result, to different levels of reasoning, and the latter, in turn, correlated to neural activity. This way it may be possible to investigate the neural counterparts to different levels of reasoning about reasoning in social situations. Coricelli and Nagel (2009) is a step in this direction in the case of a simultaneous-move game.

d. Game Dependence  Some of the language in this paper may give the impression that we assume that a particular player engages in a certain number of levels of reasoning, independent of the game being played. We do not make this assumption. We show how to infer from behavior the number of levels of reasoning a player undertakes in a given game. Whether or not this number is the same or similar across games is an empirical question — in fact, precisely the kind of empirical question which could be addressed using our result.6

e. Weak Dominance  We mentioned that, in the case of simultaneous-move games, our partition of strategies reduces to the one induced by iterated elimination of strongly dominated strategies. This follows from our definition of rationality, which is expected payoff maximization at each (non-precluded) information set, and which therefore reduces to simple expected payoff maximization in a simultaneous-move game. We could consider a different definition of rationality that includes an admissibility requirement, that is, the avoidance of weakly dominated strategies. This definition of rationality is usually employed on the strategic form of a game tree, where a suitable epistemic treatment yields behavior with ‘good’ properties in the tree. (See Brandenburger, Friedenberg and Keisler, 2008 and Brandenburger and Friedenberg, 2010.) It would be of interest to derive the appropriate partition of strategies according to levels of reasoning in this case. But this will not be immediate, for an analogous reason to why the derivation of the partition in the current paper was not immediate. We leave this to future work.

Appendix A  Preliminaries

Marginalization

6Non-behavioral data, specifically, visual attention patterns, is suggestive that there may be some stability within a given class of games. See Polonio, Di Guida and Coricelli (2014). Alaoui and Penta (2014) bring behavioral data to bear on the question, in the context of a model that endogenizes the depth of reasoning via a cost-benefit analysis, where the benefit depends on a player’s payoff function.
Lemma A.1. Fix epistemic game $(\Gamma, T)$. If $\beta_c(t_c)$ strongly believes the event $E_{-c} \subseteq S_{-c} \times T_{-c}$, then $\text{marg}_{S_{-c}} \beta_c(t_c)$ strongly believes $\text{proj}_{S_{-c}} E_{-c}$.

Proof. Suppose $\beta_c(t_c)$ strongly believes the event $E_{-c} \subseteq S_{-c} \times T_{-c}$. Fix some $S_{-c}(h) \times T_{-c} \in E_c \otimes T_{-c}$. If $\text{proj}_{S_{-c}} E_{-c} \cap S_{-c}(h) \neq \emptyset$, then there exists $(s_{-c}, t_{-c}) \in E_{-c}$ so that $s_{-c} \in S_{-c}(h)$. It follows that $E_{-c} \cap (S_{-c}(h) \times T_{-c}) \neq \emptyset$ and so $\beta_c(E_{-c}|S_{-c}(h) \times T_{-c}) = 1$. Now note that

$$\text{marg}_{S_{-c}} \beta_c(\text{proj}_{S_{-c}} E_{-c}|S_{-c}(h) \times T_{-c}) = \beta_c(\text{proj}_{S_{-c}} E_{-c} \times T_{-c}|S_{-c}(h) \times T_{-c}) \geq \beta_c(E_{-c}|S_{-c}(h) \times T_{-c}).$$

It follows that $\text{marg}_{S_{-c}} \beta_c(\text{proj}_{S_{-c}} E_{-c}|S_{-c}(h) \times T_{-c}) = 1$, as desired. ■

Image CPS’s: Fix a CPS $p_c \in C(S_{-c}, E_c)$ and some (measurable) mapping $\tau_{-c} : S_{-c} \rightarrow S_{-c} \times T_{-c}$. Define $q_c$ as follows: For each conditioning event $S_{-c}(h) \times T_{-c} \in E_c \otimes T_{-c}$, set

$$q_c(E_{-c}|S_{-c}(h) \times T_{-c}) = p_c((\tau_{-c})^{-1}(E_{-c})|S_{-c}(h))$$

for each Borel $E_{-c} \subseteq S_{-c} \times T_{-c}$. We refer to $q_c$ at the image CPS of $p_c$ under $\tau_{-c}$. So defined, $q_c$ is indeed a CPS. See Battigalli, Brandenburger, Friedenberg and Siniscalchi (2012, Part III, Chapter 4). Moreover, if $\tau_{-c}(s_{-c}) \in \{s_{-c}\} \times T_{-c}$ for each $s_{-c}$, then the image CPS of $p_c$ under $\tau_{-c}$, viz. $q_c$, has marg $S_{-c}q_c = p_c$. As a consequence, for any given CPS $p_c \in C(S_{-c}, E_c)$, we can find some CPS $q_c \in C(S_{-c} \times T_{-c}, E_c \otimes T_{-c})$ so that $\text{marg}_{S_{-c}} q_c = p_c$.

Appendix B  Proof of Theorem 5.1

Strong Justification We will focus on $m$-BRP’s that satisfy a strong justification property. To define this concept, refer to a set $X_c \subseteq S_c$ as an effective singleton if there exists some $s_c$ so that $X_c = [s_c]$. If $X_c \subseteq S_c$ is not effectively a singleton, then we simply say it is non-singleton.

Definition B.1. Fix a game $\Gamma$ and an $m$-BRP $(S^0, \ldots, S^m)$. Say that the $m$-BRP satisfies the strong justification property if, for each player $c$, we can find a mapping $j_c : S_c \rightarrow C(S_{-c}, E_c)$ satisfying the following criteria:

(j.a) If $s_c \in S^0_c$ for $n = 2, \ldots, m$, then $s_c \in \mathbb{BR}[j_c(s_c)] \subseteq S^m_c$ and $j_c(s_c)$ strongly believes $S^0_{-c}, \ldots, S^{m-1}_{-c}$.

(j.b) If $s_c \in S^1_c \backslash S^2_c$, then $\mathbb{BR}[j_c(s_c)] = [s_c]$. Moreover, if $S^1_{-c}$ if effectively a singleton, then $j_c(s_c)$ does not strongly believe $S^1_{-c}$.

(j.c) If $s_c \in S_c \backslash S^1_c$, then $j_c(s_c) \in j_c(S^1_c)$.

Observe that, by definition of an $m$-BRP, we can always find a mapping $j_c : S_c \rightarrow C(S_{-c}, E_c)$ satisfying conditions (j.a) and (j.c). But, condition (j.b) is stronger than that required by an $m$-BRP. If we find a mapping $j_c : S_c \rightarrow C(S_{-c}, E_c)$ satisfying these requirements, we say that $j_c$ strongly justifies the $m$-BRP for player $c$ or $j_c$ an $j_b$ strongly justify the $m$-BRP.

Proposition B.1. Fix a game $\Gamma$ and an $m$-BRP $(S^0, \ldots, S^m)$ satisfying the strong justification property. Then there exists an associated epistemic game $(\Gamma, T)$ so that, for each $n = 1, \ldots, m$, proj $S^n_R = S^n$.
Type Structure  For each player $c$ and each $n = 1, \ldots, m$, set $U^n_c = S^n_c$ and write $v^n_c : S^n_c \rightarrow U^n_c$ for the identity map. The type set for player $c$ will be $T_c = \bigcup_{n=1}^m U^n_c$. We will refer to types in $U^n_c$ as the $n$-types for player $c$.

It will be convenient to specify the diagonal of $S^n_c \times U^n_c$. This will be given by

$$\text{diag}^n_c = \bigcup_{s_c \in U^n_c} ([s_c] \times v^n_c([s_c]))$$

Observe that, if $[s_c] = [r_c]$ then $v^n_c([s_c]) = v^n_c([r_c])$ and so $[s_c] \times v^n_c([r_c]) \subseteq \text{diag}^n_c$. Moreover, if $S^n_c$ is non-singleton then, for each $s_c \in S_c$, there exists a type $t_c \in U^n_c$ so that $(s_c, t_c) \notin \text{diag}^n_c$.

For each $n = 1, \ldots, m$, define a mapping $\tau^n_c : S_{-c} \rightarrow S_{-c} \times T_{-c}$ with $\tau^n_c(s_{-c}) \in \{s_{-c}\} \times T_{-c}$. For $n = 1$, if $S_{-c}$ is non-singleton, then the range of $\tau^1_{-c}$ is concentrated on $S_{-c} \times U^1_{-c}$ but off of $\text{diag}^1_{-c}$, i.e., each $\tau^1_{-c}(s_{-c}) \in (S_{-c} \times U^1_{-c}) \backslash \text{diag}^1_{-c}$. For $n = 2, \ldots, m$, for each $s_{-c} \in S^1_{-c}$, $\tau^n_c(s_{-c})$ is in the maximal diagonal consistent with $s_{-c}$. Specifically, for a given $s_{-c} \in S^1_{-c}$, let $\ell = \max\{k = 1, \ldots, n-1 : s_{-c} \in S^k_{-c}\}$ and set $\tau^n_{-c}(s_{-c}) = (s_{-c}, u^n_{-c}(s_{-c}))$. The belief map is such that, for each $v^n_c(s_c), \beta_c(v^n_c(s_c))$ be the image CPS of $j_c(s_c)$ under $\tau^n_{-c}$. Observe that, for each $s_c \in S^n_c$, $\text{proj}_{S_{-c}} \beta_c(v^n_c(s_c)) = j_c(s_c)$.

Analysis It will be convenient to define sets of $n$-strategy-type pairs of the players. In particular, for each player $c$ and each $n = 1, \ldots, m$, set

$$Q^n_c = \bigcup_{s_c \in S^n_c} (\mathbb{B}[j_c(s_c)] \times \{v^n_c(s_c)\}).$$

For $n = 1$, $Q^1_c = \text{diag}^1_c$. (This follows from Condition (j.b) of strong justification.) For $n = 2, \ldots, m$, $\text{diag}^n_c \subseteq Q^n_c$. (This follows from Condition (j.a) of strong justification.) Observe the following:

**Lemma B.1.** For each $n = 1, \ldots, m$, $\text{proj}_{S_n} Q^n_c = S^n_c$.

**Proof.** If $s_c \in S^n_c$, then $s_c \in \mathbb{B}[j_c(s_c)]$ and so $(s_c, v^n_c(s_c)) \in Q^n_c$. Fix some $(s_c, v^n_c(r_c)) \in Q^n_c$. Then, $r_c \in S^n_c$ and $s_c \in \mathbb{B}[j_c(r_c)]$. It follows that $s_c \in \mathbb{B}[j_c(r_c)] \subseteq S^n_c$, as required. ■

**Lemma B.2.** For each $n = 1, \ldots, m$, $R^n_a \times R^n_b = \bigcup_{k=0}^m (Q^n_a \times Q^n_b)$.

**Proof.** The case of $n = 1$ is immediate from the construction. Thus, we show $n = 2, \ldots, m$. The proof is by induction on $n$.

$n = 2$ : Fix some $k = 1, \ldots, m$ and some $(r_c, v^k_c(s_c)) \in Q^k_c = \mathbb{B}[j_c(s_c)] \times \{v^k_c(s_c)\}$. Since the claim holds for $n = 1$, to show the following:

(i) If $k = 1$, then $v^k_c(s_c)$ does not strongly believe $R^1_{-c}$.

(ii) If $k = 2, \ldots, m$, then $v^k_c(s_c)$ strongly believes $R^1_{-c}$.

For both cases, we make use of the following property: Having established this claim for $n = 1$, we have that $R^1_{-c} = \bigcup_{k=1}^m Q^k_{-c}$ and so, by Lemma B.1, $S^1_{-c} = \bigcup_{k=1}^m S^k_{-c} = \text{proj}_{S_{-c}} R^1_{-c}$.

First, suppose that $k = 1$ and $S^1_{-c}$ is an effective singleton. By Condition (j.b) of strong justification, $j_c(s_c)$ does not strongly believe $S^1_{-c}$, i.e., there exists some information set $h$ with $S^1_{-c} \cap S_{-c}(h) \neq \emptyset$. Then, $v^1_c(s_c)$ does not strongly believe $R^1_{-c}$, i.e., there exists some $h$ with $S^1_{-c} \cap S_{-c}(h) \neq \emptyset$. By Condition (j.a) of strong justification, $j_c(s_c)$ strongly believes $R^1_{-c}$, i.e., $v^1_c(s_c)$ strongly believes $R^1_{-c}$.

The proof is by induction on $n$. ■
\(\emptyset\) and \(j_c(s_c)(S_{-c}\setminus S^1_{-c}|S_{-c}(h)) > 0\). Since \(S^1_{-c} = \text{proj}_{S_{-c}} R^1_{-c},\) \(R^1_{-c} \cap (S_{-c}(h) \times T_{-c}) \neq \emptyset\). Moreover, \(\beta_c(v^1_c(s_c))((S_{-c}\setminus S^1_{-c}) \times T_{-c}) \subset S_{-c}(h) \times T_{-c}) > 0\) and, again using the fact that \(S^1_{-c} = \text{proj}_{S_{-c}} R^1_{-c},\)
\(((S_{-c}\setminus S^1_{-c}) \times T_{-c}) \cap R^1_{-c} = \emptyset\). Thus, \(v^1_c(s_c)\) does not strongly believe \(R^1_{-c}\).

Next, suppose that \(k = 1\) and \(S^1_{-c}\) is non-singleton. Observe that, in this case, for each \((s_{-c}, t_{-c})\) with \(\beta_c(v^1_c(s_c))((s_{-c}, t_{-c})|S_{-c} \times T_{-c}) > 0\), \(t_{-c} \in U^1_{-c}\) and \((s_{-c}, t_{-c}) \notin \text{diag}^1_{-c}\). Since \(\text{diag}^1_{-c} = Q^1_{-c}\) and \((s_{-c}, t_{-c}) \notin Q^k_{-c}\) for \(k = 2, \ldots, m\), it follows from this claim (established for \(n = 1\)) that \((s_{-c}, t_{-c}) \notin R^1_{-c}\). Thus, \(v^1_c(s_c)\) does not strongly believe \(R^1_{-c}\).

Finally, suppose that \(k = 2, \ldots, m\). Fix a conditioning event \(S_{-c}(h) \times T_{-c}\) so that \(R^1_{-c} \cap (S_{-c}(h) \times T_{-c}) \neq \emptyset\). It follows that \(S^1_{-c} \cap S_{-c}(h) \neq \emptyset\), since \(S^1_{-c} = \text{proj}_{S_{-c}} R^1_{-c}\). So, using the fact that \(j_c(s_c)\) strongly believes \(S^1_{-c}\) (i.e., \(k = 2, \ldots, m\)), it follows that \(j_c(s_c)(S^1_{-c}|S_{-c}(h)) = 1\). Now observe that, by construction,

\[
\beta_c(v^k_c(s_c))\left(\bigcup_{l=1}^{k-1} \text{diag}^l_{-c}|S_{-c}(h) \times T_{-c}\right) = j_c(s_c)(S^1_{-c}|S_{-c}(h)) = 1.
\]

Since \(R^1_{-c} = \bigcup_{l=1}^{m} Q^l_{-c}\) (Lemma B.2) and \(\bigcup_{l=1}^{k-1} \text{diag}^l_{-c} \subseteq \bigcup_{l=1}^{m} Q^l_{-c}\), it follows that \(\beta_c(v^k_c(s_c))(R^1_{-c}|S_{-c}(h) \times T_{-c}) = 1\), as desired.

\(n \geq 3\): Fix some \(n = 2, \ldots, m - 1\) and some \(k = n, \ldots, m\) and some \((r_c, v^n_c(s_c)) \in Q^n_c = \mathbb{B}(j_c(s_c)) \times \{v^n_c(s_c)\}\). Since the claim holds for \(n\) it suffices to show the following:

(i) If \(k = n\), then \(v^n_c(s_c)\) does not strongly believe \(R^n_{-c}\).

(ii) If \(k = n + 1, \ldots, m\), then \(v^n_c(s_c)\) strongly believes \(R^n_{-c}\).

First, suppose that \(k = n\). For each \((s_{-c}, t_{-c})\) with \(\beta_c(v^n_c(s_c))((s_{-c}, t_{-c})|S_{-c} \times T_{-c}) > 0\), \(t_{-c} \in U^{n-1}_{-c}\).

By the induction hypothesis, \((s_{-c}, t_{-c}) \notin R^{n-1}_{-c}\), establishing that \(v^n_c(s_c)\) does not strongly believe \(R^{n-1}_{-c}\).

Next, suppose that \(k = n + 1, \ldots, m\). Fix a conditioning event \(S_{-c}(h) \times T_{-c}\) so that \(R^n_{-c} \cap (S_{-c}(h) \times T_{-c}) \neq \emptyset\). By the induction hypothesis,

\[
\text{proj}_{S_{-c}} R^n_{-c} = \bigcup_{k=n}^{m} S^n_{-c} = S^n_{-c}
\]

and so \(S^n_{-c} \cap S_{-c}(h) \neq \emptyset\). Since \(j_c(s_c)\) strongly believes \(S^n_{-c}\), it follows that \(j_c(s_c)(S^n_{-c}|S_{-c}(h)) = 1\). Now observe that, by construction,

\[
\beta_c(v^n_c(s_c))\left(\bigcup_{l=n}^{k-1} \text{diag}^l_{-c}|S_{-c}(h) \times T_{-c}\right) = j_c(s_c)(S^n_{-c}|S_{-c}(h)) = 1.
\]

Since \(R^n_{-c} = \bigcup_{l=n}^{m} Q^l_{-c}\) (Lemma B.1) and \(\bigcup_{l=n}^{k-1} \text{diag}^l_{-c} \subseteq \bigcup_{l=n}^{m} Q^l_{-c}\), it follows that \(\beta_c(v^n_c(s_c))(R^n_{-c}|S_{-c}(h) \times T_{-c}) = 1\), as desired. \&
Proof of Proposition B.1. By Lemmata B.1-B.2,
\[
\text{proj}_{S_a} R_a^n \times \text{proj}_{S_b} R_b^n = \text{proj}_{S_a} [\bigcup_{k=n}^m Q^k_a] \times \text{proj}_{S_b} [\bigcup_{k=n}^m Q^k_b)]
\]
\[
= \bigcup_{k=n}^m S^k_a \times \bigcup_{k=n}^m S^k_b
\]
\[
= S_a^n \times S_b^n,
\]
as required. ■

**Generic Games and Strong Justification:** To prove Theorem 5.1, it suffices to show that, in a generic game, any \(m\)-BRP satisfies the strong justification property. This will follow from the following Lemma, shown in Appendix C.

**Lemma B.3.** Suppose \(\Gamma\) is generic and let \(X_{-c} \subset S_{-c}\). If \(s_c\) is justifiable, then there exists some CPS \(p_c\) so that \([s_c] = \mathbb{BR}[p_c]\) and \(p_c\) does not strongly believe \(X_{-c}\).

**Appendix C Generic Games**

**Which Games are Generic?**

We begin by showing that a perfect information game satisfying no relevant ties is generic.

**Definition C.1** (Battigalli, 1997). A game \(\Gamma\) satisfies no relevant ties if \(\zeta(s_c, s_{-c}) \neq \zeta(r_c, s_{-c})\) implies \(\pi_c(s_c, s_{-c}) \neq \pi_c(r_c, s_{-c})\).

**Lemma C.1.** A perfect-information game satisfying no relevant ties is generic.

To show this lemma, we will need an auxiliary definition and result.

**Definition C.2.** Given a conditional probability space \((\Omega, \mathcal{E})\), call a CPS \(p \in \mathcal{C}(\Omega, \mathcal{E})\) degenerate if, for each conditioning event \(E\), there exists some \(\omega \in E\) with \(p(\omega|E) = 1\).

The following Lemma follows almost immediately from Ben-Porath (1997, Lemma 1.2.1).

**Lemma C.2.** Fix a perfect-information game satisfying no relevant ties. If \(s_c\) is justifiable, then there exists some degenerate CPS \(p_c \in \mathcal{C}(S_{-c}, \mathcal{E}_c)\) so that \(s_c \in \mathbb{BR}[p_c]\).

In a perfect-information game, we can identify an information set \(h\) with the unique node it contains. In that case, we will say an information set \(h\) precedes an information set \(h'\) if \(h = \{v\}, h' = \{v'\}\), and \(v\) precedes \(v'\). We will say that \(h\) strictly precedes \(h'\) if \(h\) precedes \(h'\) and \(h \neq h'\). We will say that \(h\) weakly precedes \(h'\) if \(h = h'\).

**Proof.** Let \(s_c\) be a justifiable strategy. Then, by Lemma 1.2.1 in Ben-Porath (1997), for each \(S_{-c}(h) \in \mathcal{E}_c\) with \(s_c \in S_c(h)\), we can find some \(s^h_{-c} \in S_{-c}(h)\) so that \(\pi_c(s_c, s^h_{-c}) \geq \pi_c(r_c, s^h_{-c})\) for all \(r_c \in S_c(h)\). Use the collection \(\{s^h_{-c} : h \in H_c \cup \{\phi\}\}\) to form a CPS \(p_c\).

We will inductively define the measures \(p_c(\cdot|S_{-c}(h))\). For each \(S_{-c}(h)\) with \(s^h_{-c} \in S_{-c}(h)\), set \(p_c(s^h_{-c} | S_{-c}(h)) = 1\). Next, fix an information set \(h^* \in H_c\) where \(p_c(\cdot|S_{-c}(h))\) has been defined for each
that strictly precedes \(h^*\) but for which \(p_c(\cdot|S_{-c}(h^*))\) has not been defined. Set \(p_c(s^h|S_{-c}(h)) = 1\) for each \(S_{-c}(h)\) with \(s^h \in S_{-c}(h)\). Proceeding along these lines, we define \(p_c(\cdot|S_{-c}(h))\) for each conditioning event \(S_{-c}(h)\).

It can be verified that, so defined, \(p_c\) is a CPS. Moreover, \(s_c\) is optimal under \(p_c\): Given an information set \(h \in H_c\) with \(s_c \in S_c(h)\), there exists an information set \(h^*\) that precedes (perhaps weekly) \(h\) so that \(s_{-c} \in S_{-c}(h)\) and \(p_c(s_{-c}^h|S_{-c}(h)) = 1\). Then, the claim follows from the fact that \(S_c(h) \subseteq S_c(h^*)\) and the fact that \(\pi_c(s_c, s_{-c}^h) \geq \pi_c(r_c, s_{-c}^h)\) for all \(r_c \in S_c(h)\).

**Proof of Lemma C.2.** Fix a perfect-information game satisfying no relevant ties and some strategy \(s_c\) that is justifiable. Then there exists some degenerate CPS \(p_c \in \mathcal{C}(S_{-c}, \mathcal{E}_c)\) so that \(s_c \in \mathbb{R}[p_c]\). We will show that, if \(r_c \notin [s_c]\), then \(r_c \notin \mathbb{R}[p_c]\).

Fix some \(r_c \notin [s_c]\). Then there exists some \(h \in H_c\) with \(s_c, r_c \in S_c(h)\) and \(s_c(h) \neq r_c(h)\). Let \(s_{-c} \in S_{-c}(h)\) with \(p_c(s_{-c}|S_{-c}(h)) = 1\). Since \(s_c\) is sequentially optimal under \(p_c\), \(\pi_c(s_c, s_{-c}) \geq \pi_c(r_c, s_{-c})\). But, since \(\zeta(s_c, s_{-c}) \neq \zeta(r_c, s_{-c})\), no relevant ties implies \(\pi_c(s_c, s_{-c}) > \pi_c(r_c, s_{-c})\). Thus, \(r_c \notin \mathbb{R}[p_c]\).

The next example illustrates that a game may satisfy no relevant ties, but may be non-generic.

**Example C.1.** The game in Figure C.1 satisfies no relevant ties. Yet it is not generic: \(Out\) is optimal under a CPS \(p_a\) if and only if \(p_a(L|S_b) = p_a(R|S_b) = \frac{1}{2}\). But, for such a CPS \(p_a\), \(\mathbb{R}[p_a] = \{Out, U, M\}\), even though \(U, M\) do not induce the same plan of action as \(Out\).

![Figure C.1: No Relevant Ties](image)

**Proof of Lemma B.3**

This section is devoted to proving Lemma B.3. To do so, it will be useful to construct a CPS that is, in a sense, a convex combination of (different) CPS and a probability measure. In particular, fix some CPS \(p_c \in \mathcal{C}(S_{-c}, \mathcal{E}_c)\) and some measure \(\mu_c \in \mathcal{P}(S_{-c})\) with \(\text{Supp} \mu_c \cap \text{Supp} p_c(\cdot|S_{-c}) = \emptyset\). Given \(\varepsilon \in (0, 1)\), construct \(q_c^\varepsilon\) as follows: Set

\[
q_c^\varepsilon(s_{-c}|S_{-c}) = (1 - \varepsilon)p_c(s_{-c}|S_{-c}(h)) + \varepsilon \mu_c(s_{-c}).
\]
If either \( p_c(S_{-c}(h) | S_{-c}) > 0 \) or \( \mu_c(S_{-c}(h)) > 0 \), set \( q^\varepsilon_c(\cdot | S_{-c}(h)) \) so that

\[
q^\varepsilon_c(s_{-c} | S_{-c}(h)) = \frac{q^\varepsilon_c(s_{-c} | S_{-c})}{q^\varepsilon_c(s_{-c} | S_{-c}(h)| S_{-c})}
\]

for \( s_{-c} \in S_{-c}(h) \) and \( q^\varepsilon_c(s_{-c} | S_{-c}(h)) = 0 \) for \( s_{-c} \not\in S_{-c}(h) \). For all other conditioning events \( S_{-c}(h) \), set \( q^\varepsilon_c(\cdot | S_{-c}(h)) = p_c(\cdot | S_{-c}(h)) \). We will refer to \( q^\varepsilon_c \) a \( \varepsilon \)-convex combination of \( \mu_c \) and \( p_c \). The first Lemma shows that \( q^\varepsilon_c \) is indeed a CPS.

**Lemma C.3.** Fix some CPS \( p_c \in C(S_{-c}, \mathcal{E}_c) \), some measure \( \mu_c \in \mathcal{P}(S_{-c}) \) with \( \mu_c \cap \text{Supp} p_c(\cdot | S_{-c}) = \emptyset \). For any \( \varepsilon \in (0, 1) \), the \( \varepsilon \)-convex combination of \( \mu_c \) and \( p_c \) is a CPS.

**Proof.** Let \( q^\varepsilon_c \) be the \( \varepsilon \)-convex combination of \( \mu_c \) and \( p_c \). It is immediate that each \( q^\varepsilon_c(\cdot | S_{-c}(h)) \in \mathcal{P}(S_{-c}) \) with \( q^\varepsilon_c(S_{-c}(h) | S_{-c}(h)) = 1 \). Fix some \( E \subseteq S_{-c}(h) \subseteq S_{-c}(h') \) and write \( F = \text{Supp} p_c(\cdot | S_{-c}) \cup \text{Supp} \mu_c \).

First, suppose that \( F \cap S_{-c}(h) \neq \emptyset \). In that case, we also have that \( F \cap S_{-c}(h') \neq \emptyset \). From this

\[
q^\varepsilon_c(E | S_{-c}(h')) = \frac{q^\varepsilon_c(E \cap S_{-c})}{q^\varepsilon_c(S_{-c}(h') | S_{-c} )} = \frac{q^\varepsilon_c(E \cap S_{-c})}{q^\varepsilon_c(S_{-c}(h) | S_{-c}) q^\varepsilon_c(S_{-c}(h') | S_{-c})} = q^\varepsilon_c(E | S_{-c}(h)) q^\varepsilon_c(S_{-c}(h) | S_{-c}(h')), \]

as desired. Next suppose that \( F \cap S_{-c}(h) = \emptyset \) and \( F \cap S_{-c}(h') \neq \emptyset \). In that case \( q^\varepsilon_c(S_{-c}(h) | S_{-c}(h')) = 0 \) and \( q^\varepsilon_c(E | S_{-c}(h')) = 0 \), so that \( q^\varepsilon_c(E | S_{-c}(h')) = q^\varepsilon_c(E | S_{-c}(h)) \times q^\varepsilon_c(S_{-c}(h) | S_{-c}(h')) = 0 \). Finally, suppose that \( F \cap S_{-c}(h') = \emptyset \). In that case, \( q^\varepsilon_c(\cdot | S_{-c}(h)) = p_c(\cdot | S_{-c}(h)) \) and \( q^\varepsilon_c(\cdot | S_{-c}(h')) = p_c(\cdot | S_{-c}(h')) \), from which the desired equality follows. \( \blacksquare \)

**Proof.** Fix some justifiable strategy \( s_c \) and some \( X_{-c} \subseteq S_{-c} \). Since the game is generic, there exists some \( p_c \in C(S_{-c}, \mathcal{E}_c) \) so that \( \text{BR}[p_c] = [s_c] \). If \( p_c(X_{-c} | S_{-c}) \neq 1 \), then \( p_c \) does not strongly believe \( X_{-c} \). If \( p_c(X_{-c} | S_{-c}) = 1 \), choose some \( s^*_c \in S_{-c} \setminus X_{-c} \) so that \( p(s^*_c | S_{-c}) = 0 \) and set \( \mu(s^*_c) = 1 \). For each \( \varepsilon \in (0, 1) \), let \( q^\varepsilon_c \) be the \( \varepsilon \)-convex combination of \( \mu_c \) and \( p_c \). Certainly each \( q^\varepsilon_c \) does not strongly believe \( X_{-c} \).

We will show that, for each \( r_c \notin [s_c] \), there exists some \( \tau[r_c] \in (0, 1) \) so that \( r_c \not\in \text{BR}[q^\varepsilon_c] \) for each \( \varepsilon \in (0, \tau[r_c]) \). Then, since the game is finite, we can take \( \tau = \min \{ \tau[r_c] : r_c \notin [s_c] \} \) and obtain that, for all \( \varepsilon \in (0, \tau) \), \( \text{BR}[q^\varepsilon_c] = [s_c] \).

Fix \( r_c \notin [s_c] \). Then there exists some \( h \) so that \( s_c, r_c \in S_c(h), S_{-c}(h) \in \mathcal{E}_c \) and

\[
\sum_{s_{-c} \in S_{-c}(h)} [\pi_c(s_c, s_{-c}) - \pi_c(r_c, s_{-c})] p_c(s_{-c} | S_{-c}(h)) > 0. \tag{1}
\]

If \( q^\varepsilon_c(\cdot | S_{-c}(h)) = p_c(\cdot | S_{-c}(h)) \), certainly \( r_c \not\in \text{BR}[q^\varepsilon_c] \) for each \( \varepsilon \in (0, 1) \). If \( q^\varepsilon_c(\cdot | S_{-c}(h)) \neq p_c(\cdot | S_{-c}(h)) \), then

\[
\sum_{s_{-c} \in S_{-c}(h)} [\pi_c(s_c, s_{-c}) - \pi_c(r_c, s_{-c})] q^\varepsilon_c(s_{-c} | S_{-c}(h)) =
\frac{(1 - \varepsilon)}{q^\varepsilon_c(S_{-c}(h) | S_{-c})} \sum_{s_{-c} \in S_{-c}(h)} [\pi_c(s_c, s_{-c}) - \pi_c(r_c, s_{-c})] p_c(s_{-c} | S_{-c}(h)) + \frac{\varepsilon}{q^\varepsilon_c(S_{-c}(h) | S_{-c})} [\pi_c(s_c, s^*_c) - \pi_c(r_c, s^*_c)]. \tag{2}
\]

By Equation \( 1 \), this is strictly positive for \( \varepsilon \) sufficiently small. \( \blacksquare \)
References


27


