Adam Brandenburger
H. Jerome Keisler

An Impossibility Theorem on Beliefs in Games

Abstract. A paradox of self-reference in beliefs in games is identified, which yields a game-theoretic impossibility theorem akin to Russell's Paradox. An informal version of the paradox is that the following configuration of beliefs is impossible:

Ann believes that Bob assumes that
Ann believes that Bob's assumption is wrong

This is formalized to show that any belief model of a certain kind must have a 'hole.' An interpretation of the result is that if the analyst's tools are available to the players in a game, then there are statements that the players can think about but cannot assume. Connections are made to some questions in the foundations of game theory.

Keywords: belief model, complete belief model, game, first order logic, modal logic, paradox

1. Introduction [*]

In game theory, the notion of a player's beliefs about the game—even a player's beliefs about other players' beliefs, and so on—arises naturally. Take the basic game-theoretic question: Are Ann and Bob rational, does each believe the other to be rational, and so on? To address this, we need to write down what Ann believes about Bob's choice of strategy—to decide whether she chooses her strategy optimally given her beliefs (i.e., whether she is rational). We also have to write down what Ann believes Bob believes about her strategy choice—to decide whether Ann believes Bob chooses optimally given his beliefs (i.e., whether Ann believes Bob is rational). Etc. Beliefs about beliefs about ... in games are basic.

[*] We are indebted to Amanda Friedenberg and Gus Stuart for many valuable discussions bearing on this paper. Samson Abramsky, Ken Arrow, Susan Athey, Bob Aumann, Joe Halpern, Christopher Harris, Aviad Heifetz, Jon Levin, Martin Meier, Mike Moldoveanu, Eric Pacuit, Rohit Parikh, Martin Rechenauer, Hannu Salonen, Dov Samet, Johan van Benthem, Daniel Yamins, and Noson Yanofsky provided important input. Our thanks, too, to participants in the XIII Convegno di Teoria dei Giochi ed Applicazioni (University of Bologna, June 1999), the Tenth International Conference on Game Theory (State University of New York at Stony Brook, July 1999), the 2004 Association for Symbolic Logic Annual Meeting (Carnegie Mellon University, May 2004), and seminars at Harvard University, New York University, and Northwestern University, and to referees. Financial support from Harvard Business School, the Stern School of Business, the National Science Foundation, and the Vilas Trust Fund is gratefully acknowledged.

In this paper we ask: Doesn't such talk of what Ann believes Bob believes about her, and so on, suggest that some kind of self-reference arises in games, similar to the well-known examples of self-reference in mathematical logic. If so, then is there some kind of impossibility result on beliefs in games, that exploits this self-reference?

There is such a result, a game-theoretic version of Russell's Paradox.[†] We give it first in words. By an **assumption** (or strongest belief) we mean a belief that implies all other beliefs. Consider the following configuration:

Ann believes that Bob assumes that
Ann believes that Bob's assumption is wrong.

To get the impossibility, ask: Does Ann believe that Bob's assumption is wrong? If so, then in Ann's view, Bob's assumption, namely "Ann believes that Bob's assumption is wrong", is right. But then Ann does not believe that Bob's assumption is wrong, which contradicts our starting supposition. This leaves the other possibility, that Ann does not believe that Bob's assumption is wrong. If this is so, then in Ann's view, Bob's assumption, namely "Ann believes that Bob's assumption is wrong", is wrong. But then Ann does believe that Bob's assumption is wrong, so we again get a contradiction.

The conclusion is that the configuration of beliefs in bold is impossible. But, presumably, a model of Ann's and Bob's beliefs that contained all beliefs would contain this configuration of beliefs (among many other configurations). Apparently, such a model—which we will call a complete belief model—does not exist. Alternatively put, every model of Ann's and Bob's beliefs will have a 'hole' in it; not all possible beliefs can be present.[‡]

Formal notions of belief are of central importance in both game theory and modal logic. We think that our impossibility result is best understood if it is formulated in both settings in parallel. Since our results originated from a problem in game theory, we will first formulate our results in a setting which is consistent with the literature in game theory (beginning in Section 3), and then reformulate our results in a modal logic setting (in Section 7).

As we will see in Section 9, the notion of assumption, or strongest belief,

[†] The informal statement here in the Introduction is a multi-player analog of the Liar Paradox. The formal statement we give later (Section 6) is a multi-player analog of Russell's Paradox.

[‡] For an earlier—and weaker—impossibility result, see Brandenburger [8, 2003]. We recap this result later. Other well-known paradoxes of belief include G. E. Moore's paradox ("It is raining but I don't believe it"), and the Believer's Paradox (Thomason [32, 1980]). Huynh and Szentes [21, 1999] give a one-player impossibility result.

is essential for the impossibility result. In the verbal argument, as well as in the formalization to be given later, the statement in bold must be about one particular belief for Bob but all beliefs for Ann. This interpretation seems natural when one speaks of Ann's beliefs and Bob's assumption. In the modal logic setting, belief and assumption are different modal operators. The assumption operator was introduced and analyzed (and called the "all and only" operator) in Humberstone [20, 1987].[§] The assumption operator is also closely related to the modal operator I, which means "has been informed that", in the paper of Bonanno [6, 2005].

We now add a little more precision to the verbal impossibility argument (in the game-theoretic setting). In general, a **belief model** has a set of states for each player, and a relation for each player that specifies when a state of one player considers a state of the other player to be possible. The concepts of assumption and belief have natural definitions in such a model. Given a belief model, we next consider a **language** used by the players to formulate their beliefs about the other players. We then say that a given belief model is **complete** for a language if every statement in a player's language which is possible (i.e., true for some states) can be assumed by the player. [¶] Thus completeness is relative to a language. Completeness for a given language is determined by the beliefs for each player about the other player (but not by the beliefs for the players about themselves). Depending on our choice of language, we may get a non-existence or an existence result for complete models.

The main impossibility theorem will show: *No belief model can be complete for a language that contains first order logic.* That is, every belief model has holes expressible in first order logic, where a hole is a statement about one player that is possible but is never assumed by the other player. In fact, we will show that in any model of Ann's and Bob's beliefs, a hole must occur at one of the following rather simple statements:

0. The tautologically true statement

1. Bob rules out nothing (i.e., considers everything to be possible)

2. Ann believes that Bob rules out nothing

[§] We thank Eric Pacuit for calling this to our attention.

[¶] The word "complete" has been used in this way in papers in game theory. A more descriptive term would be "assumption-complete," but we'll keep to the shorter version. Completeness in the present sense is not related to the notions of a complete formal system or a complete theory in logic.

3. Bob believes that Ann believes that Bob rules out nothing
4. Ann believes that Bob's assumption is wrong
5. Bob assumes that Ann believes that Bob's assumption is wrong

The informal argument given in this introduction shows that Ann cannot assume, or even believe, statement 5.

Section 2 gives some more motivation for the paper, by explaining connections between the completeness concept and some questions that have arisen in game theory. We return to these connections at the end of the paper, in Section 11.

The formal argument in the game-theoretic setting is developed in Sections 3 through 5. Section 3 gives the definition of a belief model, and the mathematical notions of a set being assumed and being believed. Section 4 contains the notion of a belief model being complete for a language. Section 5 contains our main impossibility theorem, Theorem 5.4. The proof of Lemma 5.6 in that section matches the informal argument above.

In Section 6 we prepare the way for a modal logic version of our impossibility results by presenting a basic "single player" modal logic with an assumption operator. This logic is a simple special case of the logic of Bonanno [6, 2005] for belief revision. Bonanno gave an axiom set and completeness theorem for his logic. To shed some light on the behavior of the assumption operator, we restate his theorem in our simpler setting. An earlier and different axiom set and completeness theorem for modal logic with an assumption operator was given by Humberstone [20, 1987].

In Section 7 we present a modal logic with assumption operators for two players, and reformulate our impossibility result in that logic. For a more detailed modal logic analysis of our impossibility result (which refers to an earlier version of this paper) see Pacuit [28, 2006].

In Section 8 we introduce belief models with additional structure, strategies, which arise naturally in game theory. These **strategic models** are not needed in our main impossibility theorem, but are of interest in the application of our results to game theory. The next two sections contain cases where complete belief models do exist. In these positive results we get strategic models.

In Section 9 we obtain two existence theorems which suggest that the notion of a statement being assumed is an essential ingredient in our main impossibility result, and that the problem is intrinsically multi-player, i.e., game-theoretic, in nature. Section 10 provides some positive results on completeness relative to restricted (but still interesting) languages. We show there are strategic models which are complete for the fragment of first order logic that is closed under finite conjunction and disjunction, the universal and existential quantifiers, and the belief and assumption operators, but not under negation. To get these models, we construct strategic models that are topologically complete, i.e., that have a topological structure and in which every nonempty compact set of states can be assumed.

2. The Existence Problem for Complete Belief Models

Belief models and languages are artifacts created by the analyst to describe a strategic situation. A long-standing question in game theory is whether these artifacts can or should be thought of as, in some sense, available to the players themselves. (For a discussion and references to other discussions of the issue, see, e.g., Brandenburger and Dekel ([9, 1993, Section 3].)

Arguably, since we, the analysts, can build belief models and use a language, such as first order logic, these same tools should indeed be available to the players. Unless we want to accord the analyst a 'privileged' position that is somehow denied to the players, it is only natural to ask what happens if a player can think about the game the same way. But then our impossibility theorem says: *If the analyst's tools are available to the players, there are statements that the players can think about but cannot assume. The model must be incomplete.* This appears to be a kind of basic limitation in the analysis of games.[‖]

This limitation notwithstanding, the question of the existence of a complete belief model turns out to be very relevant to what is known as the "epistemic program" in game theory. One aim of this program is to find conditions on the players—specifically, on their rationality, belief in one another's rationality, etc.—that lead to various well-known solution concepts (iterated dominance, Nash equilibrium, backward induction, and others). Completeness of the belief model has been found to be needed in at least two of these analyses. Battigalli and Siniscalchi [4, 2002] use completeness in their epistemic conditions for extensive-form rationalizability (a solution concept due to Pearce [29, 1984]). Brandenburger, Friedenberg, and Keisler [10, 2006] use completeness in their epistemic conditions for iterated admissibility (iterated weak dominance). These solution concepts are of independent interest, but they both also give the backward-induction outcome in

[‖] See Yanofsky [33, 2003] for a formal presentation of the idea that mathematical paradoxes indicate limitations of various systems.

perfect-information games.[**] So, we see that completeness is also very relevant (at least under these two analyses) to giving a firm foundation for the fundamental concept of backward induction in games.

Of course, given our impossibility result, both Battigalli-Siniscalchi and Brandenburger-Friedenberg-Keisler must restrict the language that the players can use to formulate their beliefs. They (effectively) do this by making various topological assumptions on the belief models they use. We'll show in Section 10 that topological assumptions can yield complete belief models for "positive languages".

Separate from the topological approach, though, this paper does pose the following open question in game theory. Can we find a logic $\mathcal{L}$ such that: (i) complete belief models for $\mathcal{L}$ exist for each game; (ii) notions such as rationality, belief in rationality, etc. are expressible in $\mathcal{L}$; (iii) the ingredients in (i) and (ii) can be combined to yield various well-known game-theoretic solution concepts, as above?

Section 11 comes back to some related papers in game theory, after our formal treatment.

3. Belief Models

In this section we will introduce belief models for two players, which are designed to allow us to state our impossibility results in the simplest form. In a belief model, each player has a set of states, and each state for one player has beliefs about the states of the other player. (Extending the definition to finitely many players would be straightforward.)

Definition 3.1. A **belief model** is a two-sorted structure

$$\mathcal{M} = (U^a, U^b, P^a, P^b, \ldots)$$

where U^a and U^b are the nonempty universe sets (for the two sorts), P^a is a proper subset of $U^a \times U^b$, P^b is a proper subset of $U^b \times U^a$, and P^a, P^b are serial, that is, the sentences

$$\forall x \exists y\, P^a(x, y), \qquad \forall y \exists x\, P^b(y, x)$$

hold. A belief model may also contain zero or more additional relations represented by the three dots. The set of relations $\{P^a, P^b, \ldots\}$ is called the **vocabulary of $\mathcal{M}$**. We place no restriction on the size of the vocabulary.

[**] Under assumptions ruling out certain ties among payoffs. See the cited papers for details.

To simplify notation, we will always use the convention that x is a variable of sort U^a and y is a variable of sort U^b. We say x **believes** a set $Y \subseteq U^b$ if $\{y : P^a(x, y)\} \subseteq Y$, and x **assumes** Y if $\{y : P^a(x, y)\} = Y$. We also use the analogous terms with a, b and x, y reversed.

Thus assumes implies believes. The members of U^a and U^b are called **states for** Ann and Bob respectively, and the members of $U^a \times U^b$ are called **states**. P^a and P^b are called the **possibility relations**. Intuitively, $P^a(x, y)$ means that state x for Ann considers state y for Bob to be possible. So x assumes the set of all states that x considers possible, and x believes the sets which contain all the states that x considers possible.

Note that every state for Ann assumes a unique subset of U^b, and every state for Bob assumes a unique subset of U^a. By the definition of a belief model, every state for Ann assumes a nonempty subset of U^b, and some state for Ann assumes a proper subset of U^b. Likewise for Bob and U^a.

Remark 3.2. This shows that the notions of belief and assumption do not collapse to the same notion under further conditions. There must be a state for Ann that assumes a proper subset of U^b, and this state believes U^b but does not assume U^b.

There is no equality relation between elements of different sorts, so we can always take the universe sets U^a and U^b of a belief model to be disjoint. That is, every belief model is isomorphic as a two-sorted structure to a belief model with U^a, U^b disjoint.

The belief models which arise naturally in game theory have additional structure, including strategies. As we explained in the Introduction, strategies will not be needed in our main impossibility theorems, so for clarity we will postpone their treatment until Section 8.

A belief model as defined here does not specify beliefs for Ann about Ann or beliefs for Bob about Bob. That is, it does not include a relation saying when a state for Ann considers another state for Ann to be possible. However, since additional relations are allowed in the vocabulary of a belief model, one can form belief models with additional possibility relations on $U^a \times U^a$ and $U^b \times U^b$ which specify beliefs for Ann about Ann and for Bob about Bob. One can also add relations on $U^a \times U^a \times U^b$ and $U^b \times U^a \times U^b$ to specify beliefs about states. Our framework allows these extra relations, but they do not play a role in the impossibility result.

4. Complete Belief Models

Given a belief model, the next step is to specify a language used by the players to think about beliefs. We'll then be able to talk about the completeness of a model, which is relative to the language. That is, a model will be complete if every statement in a player's language which is possible (i.e., true for some states) can be assumed by the player. (Otherwise, the model is incomplete.)

Conceptually, the language for a player should be a set of statements that the player can think about. We will be concerned with the family of subsets of U^b that Ann can think about, and the family of subsets of U^a that Bob can think about. The exact definition of a language will not matter much, but it will be convenient to take the statements to be first order formulas. This will give us a lot of flexibility because we are allowing a belief structure to have extra predicates in its vocabulary.

Let us first consider an arbitrary structure $\mathcal{N} = (U^a, U^b, \ldots)$, which may or may not be a belief structure. By the **first order language for** $\mathcal{N}$ we mean the two-sorted first order logic with sorts for U^a and U^b and symbols for the relations in the vocabulary of $\mathcal{N}$. Given a first order formula $\varphi(u)$ whose only free variable is u, the set **defined by** φ in $\mathcal{N}$ is the set $\{u : \varphi(u) \text{ is true in } \mathcal{N}\}$.

In general, by a **language for** $\mathcal{N}$ we will mean a subset of the set of all formulas of the first order language for $\mathcal{N}$. Given a language $\mathcal{L}$ for $\mathcal{N}$, we let $\mathcal{L}^a, \mathcal{L}^b$ be the families of all subsets of U^a, U^b respectively which are defined by formulas in $\mathcal{L}$.

Remark 4.1. (i) If the vocabulary of $\mathcal{N}$ is finite or countable, then any language $\mathcal{L}$ for $\mathcal{N}$ has countably many formulas, so the sets $\mathcal{L}^a$ and $\mathcal{L}^b$ are at most countable.

(ii) If $\mathcal{M}$ is obtained from $\mathcal{N}$ by adding additional relations to the vocabulary, then any language for $\mathcal{N}$ is also a language for $\mathcal{M}$.

We now define the notion of a belief model which is complete for a language.

Definition 4.2. Let $\mathcal{M}$ be a belief model, and let $\mathcal{L}$ be a language for $\mathcal{M}$. $\mathcal{M}$ is **complete for** $\mathcal{L}$ if each nonempty set $Y \in \mathcal{L}^b$ is assumed by some $x \in U^a$, and each nonempty $X \in \mathcal{L}^a$ is assumed by some $y \in U^b$.

In words, a belief model is complete for a language if every nonempty set of Bob's states which is definable in the language is assumed by one of Ann's states, and vice versa.

Proposition 4.3. *Suppose two belief models $\mathcal{M}$ and $\mathcal{K}$ are elementarily equivalent, that is, they satisfy the same first order sentences. Then any language $\mathcal{L}$ for $\mathcal{M}$ is also a language for $\mathcal{K}$, and $\mathcal{M}$ is complete for $\mathcal{L}$ if and only if $\mathcal{K}$ is complete for $\mathcal{L}$.*

Proof. Being complete for $\mathcal{L}$ is expressed by the set of first order sentences

$$\exists y \varphi(y) \to \exists x \forall y [P^a(x,y) \leftrightarrow \varphi(y)]$$

for each formula $\varphi(y) \in \mathcal{L}$, and similarly with a, b and x, y reversed. ■

In general, a language for a belief model $\mathcal{M}$ will contain formulas involving the possibility relations P^a, P^b as well as the symbols of the reduced structure $\mathcal{N} = (U^a, U^b, \ldots)$. In the very special case that the relations P^a, P^b do not occur in the formulas of $\mathcal{L}$, and in addition that $\mathcal{L}$ is not too large, we get an easy example of a complete belief model.

Example 4.4. Let $\mathcal{N} = (U^a, U^b, \ldots)$ be a structure where the universe sets U^a and U^b are infinite and the vocabulary (indicated by $\ldots$) is at most countable. Let $\mathcal{L}$ be the first order language of $\mathcal{N}$. Then there are relations P^a, P^b such that $\mathcal{M} = (U^a, U^b, P^a, P^b \ldots)$ is a complete belief model for the language $\mathcal{L}$.

To see this, we note that since $\mathcal{L}^a$ and $\mathcal{L}^b$ are at most countable, one can choose surjective mappings $f : U^a \to \mathcal{L}^b \setminus \{\emptyset\}$ and $g : U^b \to \mathcal{L}^a \setminus \{\emptyset\}$, and let $P^a(x,y)$ iff $y \in f(x)$ and $P^b(y,x)$ iff $x \in g(y)$.

More generally, the above example works if the universe sets U^a, U^b are infinite and of cardinality at least the number of symbols in the vocabulary of $\mathcal{N}$.

Our main result (Theorem 5.4) will show that no belief model $\mathcal{M}$ can be complete for its own first order language, regardless of the size of the vocabulary. For this reason, one is led to consider belief models which are complete for various subsets of the first order language, as in Section 10.

5. Impossibility Results

As a warm-up, we review an earlier impossibility result from Brandenburger [8, 2003], which shows that a belief model cannot be complete for a language when the family of definable sets is too large. Given a set X, the power set of X is denoted by $\mathcal{P}(X)$, and the cardinality of X is denoted by $|X|$.

Proposition 5.1. *No belief model $\mathcal{M}$ is complete for a language $\mathcal{L}$ such that $\mathcal{L}^a = \mathcal{P}(U^a)$ and $\mathcal{L}^b = \mathcal{P}(U^b)$.*

Proof. Given a belief model $\mathcal{M}$, we have $|U^a| \le |U^b|$ or $|U^b| \le |U^a|$, say the former. Since $U^a \times U^b$ has nonempty proper subsets, we must have $|U^b| > 1$. Then by Cantor's theorem, $|U^b| < |C|$ where C is the set of all nonempty subsets of U^b. It follows that $|U^a| < |C|$. Let $f : U^a \to C$ be the function where $f(x)$ is the set that x assumes. There must be a set $Y \in C \setminus range(f)$. Then $\emptyset \neq Y \in \mathcal{L}^b$ but no x assumes Y, so $\mathcal{M}$ is not complete for $\mathcal{L}$. ■

More generally, the above argument shows that if $|U^a| < |\mathcal{L}^b| - 1$ (that is, $\mathcal{L}^b$ is too large in cardinality), then $\mathcal{M}$ cannot be complete for $\mathcal{L}$.

In this paper, only the two variables x and y will be needed in formulas. We now introduce some notation that will make many formulas easier to read.

Definition 5.2. If $\varphi(y)$ is a statement about y, we will use the formal abbreviations

$$x \text{ believes } \varphi(y) \qquad \text{for} \qquad \forall y[P^a(x,y) \to \varphi(y)],$$

$$x \text{ assumes } \varphi(y) \qquad \text{for} \qquad \forall y[P^a(x,y) \leftrightarrow \varphi(y)].$$

Similarly with a, b and x, y reversed.

Note that "x believes $\varphi(y)$" and "x assumes $\varphi(y)$" are statements about x only.

Definition 5.3. The **diagonal formula** $D(x)$ is the first order formula

$$\forall y[P^a(x,y) \to \neg P^b(y,x)].$$

This is our formal counterpart to the intuitive statement "Ann believes Bob's assumption is wrong." Note that the intuitive statement contains the word "believes", but the diagonal formula is not of the form x *believes* $\varphi(y)$ in the notation of Definition 5.2.

Here is our main impossibility result, which works for countable as well as large languages.

Theorem 5.4. *Let $\mathcal{M}$ be a belief model and let $\mathcal{L}$ be the first order language for $\mathcal{M}$. Then $\mathcal{M}$ cannot be complete for $\mathcal{L}$.*

The theorem is an easy consequence of the next two lemmas.

Lemma 5.5. *In a belief model $\mathcal{M}$, suppose $\forall y P^a(x_1, y)$ and*

$$x_2 \textit{ believes } [y \textit{ believes } [x \textit{ believes} \forall x P^b(y,x)]].$$

Then $D(x_2)$.

Proof. We must show that

$$\forall y[P^a(x_2, y) \to \neg P^b(y, x_2)].$$

Suppose not. Then there is an element y_2 such that

$$P^a(x_2, y_2) \wedge P^b(y_2, x_2).$$

It then follows in turn that

$$y_2 \textit{ believes } [x \textit{ believes} \forall x\, P^b(y,x)],$$

$$x_2 \textit{ believes} \forall x\, P^b(y,x),$$

$$\forall x P^b(y_2, x),$$

$$\forall x[x \textit{ believes } \forall x\, P^b(y,x)],$$

$$x_1 \textit{ believes } \forall x\, P^b(y,x)],$$

$$\forall y \forall x\, P^b(y,x).$$

This contradicts the hypothesis that P^b is a proper subset of $U^b \times U^a$. ■

Lemma 5.6. *Suppose $\mathcal{M}$ is a belief model. Then no element x_0 of U^a satisfies the formula*

$$x \textit{ believes } (y \textit{ assumes } D(x)) \qquad (*)$$

in $\mathcal{M}$.

The proof of this lemma will closely match the argument given in the Introduction; the statement in bold type there is the informal version of the formula (*). The informal version in the Introduction is a two-player analog of the Liar Paradox. It is a semantic statement, since it involves the notion of an assumption being "wrong." Lemma 5.6, on the other hand, is a formal result, which is a two-player analog of Russell's Paradox. In full unabbreviated form, formula (*) is

$$\forall y[P^a(x,y) \to \forall x(P^b(y,x) \leftrightarrow \forall y[P^a(x,y) \to \neg P^b(y,x)\,]\,)\,].$$

Proof. We suppose that an element x_0 satisfies formula (*) in $\mathcal{M}$ and arrive at a contradiction. We ask whether $D(x_0)$.

Case 1: $D(x_0)$. Since $\forall x \exists y\, P^a(x,y)$, we may choose y_0 such that $P^a(x_0,y_0)$. Since $D(x_0)$, $\neg P^b(y_0,x_0)$. But since x_0 satisfies (*), y_0 assumes $D(x)$. Then $P^b(y_0,x_0) \leftrightarrow D(x_0)$, so $P^b(y_0,x_0)$ and we have a contradiction.

Case 2: $\neg D(x_0)$. Choose y_0 such that $P^a(x_0,y_0) \wedge P^b(y_0,x_0)$. Since $P^a(x_0,y_0)$ and x_0 satisfies (*), y_0 assumes $D(x)$. Since $P^b(y_0,x_0)$ and y_0 assumes $D(x)$, we have $D(x_0)$, contradiction. ■

Here is an English translation of the above proof, replacing x by "Ann", y by "Bob" , the relations P^a and P^b by "sees", and $D(x)$ by "Ann believes that Bob cannot see Ann". This proof, unlike the rough argument in the Introduction, involves only Ann's beliefs about Bob and Bob's beliefs about Ann.

Suppose Ann believes that Bob assumes that Ann believes that Bob cannot see Ann.

Case 1. Ann believes Bob cannot see Ann.

Ann sees (a state for) Bob who cannot see Ann. Bob sees Ann iff Ann believes Bob cannot see Ann. Since Bob cannot see Ann, Ann does not believe Bob cannot see Ann. Contradiction.

Case 2. Ann does not believe Bob cannot see Ann.

Ann sees (a state for) Bob who sees Ann. Bob sees Ann iff Ann believes Bob cannot see Ann. Since Bob sees Ann, Ann believes Bob cannot see Ann. Contradiction.

We are now ready to prove Theorem 5.4. The proof will actually give a sharper result which pinpoints the location of the holes in a belief model. We say that belief model $\mathcal{M}$ has a **hole** at a set Y if Y is nonempty but is not assumed by any element. Thus a belief model is complete for a language $\mathcal{L}$ if and only if it has no holes in $\mathcal{L}^a$ and no holes in $\mathcal{L}^b$.

Let us also say that $\mathcal{M}$ has a **big hole** at Y if Y is nonempty but is not believed by any element. Thus $\mathcal{M}$ has a big hole at Y if and only if it has a hole at every nonempty subset of Y.

Theorem 5.7. *Every belief model $\mathcal{M}$ has either a hole at U^a, a hole at U^b, a big hole at one of the formulas*

$$\forall x\, P^b(y,x), \tag{i}$$

$$x \text{ believes } \forall x\, P^b(y,x), \tag{ii}$$

$$y \text{ believes } [x \text{ believes } \forall x\, P^b(y,x)\,], \tag{iii}$$

a hole at the formula

$$D(x), \tag{iv}$$

or a big hole at the formula

$$y \textit{ assumes } D(x). \tag{v}$$

Thus there is no belief model which is complete for a language $\mathcal{L}$ which contains the tautologically true formula and formulas (i)—(v).

This immediately implies Theorem 5.4, since each of the formulas (i)—(v) is first order. Looking back at the list of statements in the Introduction, formulas (i) through (v) are the formal counterparts of the statements 1 through 5 respectively, and the sets U^a and U^b are counterparts of the tautologically true statement 0.

Proof. Suppose the theorem does not hold for a belief model $\mathcal{M}$. Since $\mathcal{M}$ does not have holes at U^a and U^b, there is an element y_1 which satisfies formula (i) and an element x_1 such that $\forall y\, P^a(x_1, y)$. Since $\mathcal{M}$ does not have big holes at formulas (i)—(iii), there is an element x_2 that believes formula (i) and thus satisfies (ii), an element y_3 that believes (ii) and thus satisfies (iii), and an element x_4 that believes formula (iii). Then by Lemma 5.5, x_4 satisfies formula (iv). Since there is no hole at (iv), there is an element y_4 which assumes the formula (iv) and thus satisfies (v). But then by Lemma 5.6, $\mathcal{M}$ must have a big hole at (v), and we have a contradiction. ■

In Section 9 we will give an example showing that the list of formulas (i)—(v) in the above theorem cannot be shortened to (i)—(iv).

6. Assumption in Modal Logic

In this section we prepare for a modal formulation of our results by presenting a basic modal logic with an assumption operator in a single-player setting, which we'll call **assumption logic**. This logic is related to the modal logic of Humberstone [20, 1987] and is also a simpler special case of Bonanno's modal logic for belief revision in [6, 2005]. Bonanno's logic has modal operators B_0 for initial belief, B_1 for final belief, I for being informed that, and A for the universal operator, which allows him to get an axiom set and completeness theorem. In our simpler case, the initial belief operator B_0 is the same as the universal operator A, and the operator I will be interpreted as assumption.

We will use the standard symbol $\Box$ for the belief operator, and the symbol $\heartsuit$ for the assumption operator.

We refer to Boolos [7, 1993] for an elementary introduction to modal logic. The models for assumption logic are (Kripke) **frames** $\mathcal{W} = (W, P)$ where P is a binary relation on W. The elements of W are called worlds, and P is called the **accessibility relation**. At a world w, $\Box\varphi$ is interpreted as "w believes φ", $\heartsuit\varphi$ as "w assumes φ", and $A\varphi$ as $\forall z\, \varphi$.

The **formulas** of assumption logic are built from a set L of proposition symbols and the false formula $\perp$ using propositional connectives and the three modal operators, $\Box$, $\heartsuit$, and A. Note that $\neg\perp$ is the true formula.

In a frame $\mathcal{W}$, a **valuation** is a function V which associates a subset of $V(\mathbf{D}) \subseteq W$ with each proposition symbol $\mathbf{D} \in L$. For a given valuation V, the notion of a formula φ being **true** at a world w, in symbols $w \models \varphi$, is defined by induction on the complexity of φ. For a proposition symbol $\mathbf{D}$, $w \models \mathbf{D}$ if $w \in V(\mathbf{D})$. The rules for connectives are as usual, and the rules for the modal operators are as follows:

$w \models \Box\varphi$ if for all $z \in W$, $P(w, z)$ implies $z \models \varphi$.

$w \models \heartsuit\varphi$ if for all $z \in W$, $P(w, z)$ if and only if $z \models \varphi$.

$w \models A\varphi$ if for all $z \in W$, $z \models \varphi$.

A formula is **valid for V in $\mathcal{W}$** if it is true at all $w \in W$, and **satisfiable for V in $\mathcal{W}$** if it is true at some $w \in W$.

Note that if the valuation assigns a first order definable set to each proposition symbol, then for each modal formula φ, $w \models \varphi$ is expressible by a formula with one free variable in the first order language of $\mathcal{W}$.

To shed more light on the behavior of the assumption operator in the modal logic setting, we restate the axioms and completeness theorem of Bonanno [6, 2005] with the simplification that comes from eliminating the initial belief operator B_0.

RULES OF INFERENCE FOR ASSUMPTION LOGIC

Modus Ponens: From $\varphi, \varphi \to \psi$ infer ψ.

Necessitation: From φ infer $A\varphi$.

AXIOMS FOR ASSUMPTION LOGIC

All propositional tautologies.

Distribution Axioms for $\Box$ and A:

$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi), \quad A(\varphi \to \psi) \to (A\varphi \to A\psi).$$

S_5 Axioms for A:

$$A\varphi \to \varphi, \quad \neg A\varphi \to A\neg A\varphi$$

Inclusion Axiom for $\Box$:

$$A\varphi \to \Box\varphi$$

Axioms for $\heartsuit$:

$$\heartsuit\varphi \wedge \heartsuit\psi \to A(\varphi \leftrightarrow \psi), \quad A(\varphi \leftrightarrow \psi) \to (\heartsuit\varphi \leftrightarrow \heartsuit\psi),$$

$$\heartsuit\varphi \wedge \Box\psi \to A(\varphi \to \psi), \quad \heartsuit\varphi \to \Box\varphi$$

Proposition 6.1. *(Bonanno [6, 2005] (Soundness and Completeness)*

(i) Assumption logic is sound, that is, every provable formula is valid in all frames.

(ii) Assumption logic is complete, that is, every formula which is valid in all frames is provable.

7. Impossibility Results in Modal Form

In this section we reformulate our two-player impossibility result in a modal logic setting. For each pair of players cd among Ann and Bob, there will be an operator $\Box^{cd}$ of beliefs for c about d, and an operator $\heartsuit^{cd}$ of assumptions for c about d. We first define the models of our modal logic for two players.

Definition 7.1. An **interactive frame** is a structure $\mathcal{W} = (W, P, U^a, U^b)$ with a binary relation $P \subseteq W \times W$ and disjoint sets U^a, U^b, such that $\mathcal{M} = (U^a, U^b, P^a, P^b)$ is a belief model, where $U^a \cup U^b = W$, $P^a = P \cap U^a \times U^b$, and $P^b = P \cap U^b \times U^a$.

In an interactive frame, the states for both a and b become members of the set W of worlds. P is the **accessibility relation**.

This definition makes no restrictions on the part of P in $U^a \times U^a$ and $U^b \times U^b$. Thus beliefs for players about themselves are allowed in the interactive frame, but the corresponding belief model does not depend on them. With this setup, it will be apparent that our impossibility phenomenon is not affected by the players' beliefs about themselves.

The requirement that $\mathcal{M}$ is a belief model means that the sets U^a, U^b are nonempty, and the relations $P \cap U^a \times U^b$ and $P \cap U^b \times U^a$ are serial proper subsets.

We are using frames with a set of states for each player but only a single accessibility relation. This is convenient for a study of the beliefs for one player about another. Another approach would be to use frames with an accessibility relation for each player, as in Lomuscio [23, 1999]. This approach gives a modal logic with a belief operator for each player about the state of the world.

We now introduce the formulas and semantical interpretation for the modal logic of interactive frames.

Interactive modal logic will have two distinguished proposition symbols $\mathbf{U}^a, \mathbf{U}^b$ and a set L of additional proposition symbols. By a **modal formula** we mean an expression which is built from proposition symbols and the false formula $\bot$ using propositional connectives, the universal modal operator A, and the modal operators $\Box^{cd}, \heartsuit^{cd}$ where c and d are taken from $\{a, b\}$.

As before, a **valuation** V associates a subset of $V(\mathbf{D}) \subseteq W$ with each proposition symbol $\mathbf{D}$ in the set L.

Given a valuation V on $\mathcal{W}$, the notion of a world w being true at a formula φ, in symbols $w \models \varphi$, is defined by induction on the complexity of φ as follows: $w \models \mathbf{U}^a$ if $w \in U^a$, and similarly for b. That is, $\mathbf{U}^a$ is true at each state for Ann, and $\mathbf{U}^b$ is true at each state for Bob. The rules for connectives are as usual, and the rules for the modal operators for each pair of players $c, d \in \{a, b\}$ are:

$w \models \Box^{cd}\varphi$ if $(w \models \mathbf{U}^c \wedge \forall z[(P(w, z) \wedge z \models \mathbf{U}^d) \rightarrow z \models \varphi])$.

$w \models \heartsuit^{cd}\varphi$ if $(w \models \mathbf{U}^c \wedge \forall z[(P(w, z) \wedge z \models \mathbf{U}^d) \leftrightarrow z \models \varphi])$.

$w \models A\varphi$ if $\forall z\, z \models \varphi$.

Validity and satisfiability are defined as in the preceding section.

We again note that if the valuation assigns a first order definable set to each proposition symbol, then for each modal formula φ, $w \models \varphi$ is expressible by a formula of the first order language of $\mathcal{W}$.

In the notation of Section 5 where x has sort U^a and y has sort U^b,

$$x \models \Box^{ab}\varphi \text{ says "}x \text{ believes } \varphi(y)\text{"},$$

$$x \models \heartsuit^{ab}\varphi \text{ says "}x \text{ assumes } \varphi(y)\text{"},$$

and similarly with a, b and x, y reversed.

In classical modal logic, one often adds axioms such as $\Box\varphi \rightarrow \varphi$ or $\Box\varphi \rightarrow \Box\Box\varphi$, which are reasonable hypotheses for beliefs about one's own beliefs. In the two-player setting, the analogue of $\Box\varphi \rightarrow \varphi$ would be

$$\Box^{aa}\varphi \rightarrow \varphi, \quad \Box^{bb}\varphi \rightarrow \varphi.$$

The analogue of $\Box\varphi \rightarrow \Box\Box\varphi$ would be

$$\Box^{aa}\varphi \rightarrow \Box^{aa}\Box^{aa}\varphi, \quad \Box^{ab}\varphi \rightarrow \Box^{aa}\Box^{ab}\varphi,$$

and similarly with a and b reversed.

However, similar properties which involve only a player's beliefs about the other player cannot be valid in an interactive frame. It is easy to see that each of the formulas

$$\Box^{ab}\mathbf{U}^b \leftrightarrow \mathbf{U}^a, \quad \Box^{ba}\mathbf{U}^a \leftrightarrow \mathbf{U}^b, \quad \Box^{ab}\mathbf{U}^a \leftrightarrow \bot, \quad \Box^{ba}\mathbf{U}^b \leftrightarrow \bot$$

is valid in all interactive frames. Since the sets U^a and U^b are nonempty and disjoint, it follows that the formulas

$$\Box^{ab}\mathbf{U}^b \to \mathbf{U}^b, \quad \Box^{ab}\mathbf{U}^b \to \Box^{ba}\Box^{ab}\mathbf{U}^b, \quad \Box^{ab}\mathbf{U}^b \to \Box^{ab}\Box^{ab}\mathbf{U}^b$$

are *never* valid in an interactive frame.

We now restate the results of Section 5 in the modal setting. The universal operator A will not be needed in these results, but is included in the language because, as in the preceding section, it makes it possible to state the properties of interactive frames as axioms in the modal language. The modal operators such as $\Box^{aa}$ of beliefs and assumptions for players about themselves will also not be needed, and are mentioned only to clarify the overall picture. What we do need are the operators $\Box^{ab}$, $\heartsuit^{ab}$ and their counterparts with ba of beliefs and assumptions for one player about the other.

Definition 7.2. For the remainder of this section we will always suppose that $\mathcal{W}$ is an interactive frame, $\mathbf{D}$ is a proposition symbol (for diagonal), and V is a valuation in $\mathcal{W}$ such that $V(\mathbf{D})$ is the set

$$D = \{w \in W : (\forall z \in W)[P(w,z) \to \neg P(z,w)]\}.$$

"Satisfiable" means satisfiable in $\mathcal{W}$ at V, and similarly for "valid".

Lemma 7.3. *(= Lemma 5.5) If* $\heartsuit^{ab}\mathbf{U}^b$ *is satisfiable then*

$$[\Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a] \to \mathbf{D}$$

is valid.

Lemma 7.4. *(= Lemma 5.6)* $\neg\Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D})$ *is valid.*

We say that $\mathcal{W}$ with V has a **hole** at a formula φ if either $\mathbf{U}^b \wedge \varphi$ is satisfiable but $\heartsuit^{ab}\varphi$ is not, or $\mathbf{U}^a \wedge \varphi$ is satisfiable but $\heartsuit^{ba}\varphi$ is not. A **big hole** is defined similarly but with $\Box$ instead of $\heartsuit$. An interactive frame $\mathcal{W}$ with valuation V is **complete** for a set $\mathcal{L}$ of modal formulas if it does not have a hole in $\mathcal{L}$.

Theorem 7.5. *(= Theorem 5.7) There is either a hole at $\mathbf{U}^a$, a hole at $\mathbf{U}^b$, a big hole at one of the formulas*

$$\heartsuit^{ba}\mathbf{U}^a, \qquad \Box^{ab}\heartsuit^{ba}\mathbf{U}^a, \qquad \Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a,$$

a hole at the formula $\mathbf{U}^a \wedge \mathbf{D}$, or a big hole at the formula $\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D})$. Thus there is no complete interactive frame for the set of all modal formulas built from $\mathbf{U}^a, \mathbf{U}^b, \mathbf{D}$.

As a corollary, we see that the impossibility result also holds for the set of modal formulas which have only the assumption operators $\heartsuit^{ab}, \heartsuit^{ba}$, without the belief operators.

Corollary 7.6. *There is a hole at one of the formulas*

$$\mathbf{U}^a, \quad \mathbf{U}^b, \quad \heartsuit^{ba}\mathbf{U}^a, \quad \heartsuit^{ab}\heartsuit^{ba}\mathbf{U}^a, \quad \heartsuit^{ba}\heartsuit^{ab}\heartsuit^{ba}\mathbf{U}^a, \quad \mathbf{U}^a\wedge\mathbf{D}, \quad \heartsuit^{ba}(\mathbf{U}^a\wedge\mathbf{D}).$$

Thus there is no complete interactive frame for the set of modal formulas built from $\mathbf{U}^a, \mathbf{U}^b, \mathbf{D}$ in which the belief operators do not occur.

Problem 7.7. Is there an interactive frame $\mathcal{W}$ which is complete for the set of modal formulas φ built from $\mathbf{U}^a, \mathbf{U}^b, \mathbf{D}$ such that the assumption operators $\heartsuit^{cd}$ do not occur in φ (that is, for each such formula φ, $\mathcal{W}$ with V does not have a hole at φ)?

Following the pattern in Section 6, one can build a set of axioms for interactive modal logic, and then prove a completeness theorem using the methods of [6, 2005]. There would be axiom schemes analogous to those of the preceding section for the two-player belief and assumption operators, plus a finite set of axioms saying that U^a, U^b are nonempty and partition W, that P^a and P^b are proper, and that $\forall x \exists y\, P^a(x, y)$ and similarly for b.

8. Strategic Belief Models

Strategic models are a particular class of belief models, used in applications to games. In a strategic model, each player has a set of strategies and a set of types, each strategy-type pair is a state for a player, and each type for one player has beliefs about the states for the other player. It is intended that the strategy sets (the sets S^a, S^b below) are part of an underlying game with payoff functions (maps π^a, π^b from $S^a \times S^b$ to the reals), but the payoff functions are not needed in the formal treatment here.

Definition 8.1. Given a pair of nonempty sets (S^a, S^b), an (S^a, S^b)-based **strategic model** is a belief model $\mathcal{M} = (U^a, U^b, P^a, P^b, \ldots)$ where:

(a) U^a and U^b are Cartesian products $U^a = S^a \times T^a$, $U^b = S^b \times T^b$; members of S^a and S^b are called **strategies**, and members of T^a and T^b are called **types**.

(b) $P^a((s^a, t^a), y)$ depends only on (t^a, y), and $P^b((s^b, t^b), x)$ depends only on (t^b, x).

(c) The vocabulary of $\mathcal{M}$ must include the following additional relations, which capture the sets S^a, T^a, S^b, T^b: The binary relation τ^a on U^a which says that two states in U^a have the same type, for each $s^a \in S^a$ the unary relation $s^a(x)$ on U^a which holds when x has strategy s^a, and the analogous relations with b in place of a.

Thus a strategic model has the form

$$\mathcal{M} = (U^a, U^b, P^a, P^b, \tau^a, \tau^b, s^a, s^b, \ldots : s^a \in S^a, s^b \in S^b).$$

In view of condition (b), each type t^a for Ann assumes a nonempty set of states for Bob, and vice versa. Also, the extra relations in the vocabulary of a strategic model give us the following useful fact. (Recall that an elementary submodel of $\mathcal{M}$ is a submodel $\mathcal{N}$ such that each tuple of elements of $\mathcal{N}$ satisfies the same first order formulas in $\mathcal{N}$ as in $\mathcal{M}$.)

Proposition 8.2. *If $\mathcal{M}$ is an (S^a, S^b)-based strategic model and $\mathcal{N}$ is an elementary submodel $\mathcal{M}$, then $\mathcal{N}$ is an (S^a, S^b)-based strategic model.*

Proof. Let $\mathcal{N} = (V^a, V^b, \ldots)$. We must find sets of types T_0^a, T_0^b such that $V^a = S^a \times T_0^a$ and $V^b = S^b \times T_0^b$. We first observe that for each $s^a \in S^a$, the sentence $\exists x s^a(x)$ holds in $\mathcal{N}$ because it holds in $\mathcal{M}$. Since $\mathcal{N}$ is a submodel of $\mathcal{M}$, every $x \in V^a$ satisfies $s^a(x)$ for some $s^a \in S^a$. Moreover, for each $r^a, s^a \in S^a$, $\mathcal{N}$ satisfies the sentence

$$\forall x[s^a(x) \rightarrow \exists u[r^a(u) \wedge \tau^a(x, u)]],$$

It follows that $V^a = S^a \times T_0^a$ where

$$T_0^a = \{t^a \in T^a : (s^a, t^a) \in V^a\}.$$

We can define T_0^b in a similar way and get $V^b = S^a \times T_0^b$. It is clear that the relations P^a, P^b, τ^a, τ^b have the required properties in $\mathcal{N}$. Therefore $\mathcal{N}$ is an (S^a, S^b)-based strategic model. ■

$$\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{S^b} \\
 & & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R} \\ \cline{3-4}
 & U & 1,-1 & -1,1 \\ \cline{3-4}
S^a & D & -1,1 & 1,-1 \\ \cline{3-4}
\end{array}$$

Figure 8.1

$$\begin{array}{cc|c|c|}
v^a & & \multicolumn{2}{c}{S^b} \\
 & & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R} \\ \cline{3-4}
 & v^b & + & \\ \cline{3-4}
T^b & w^b & & \\ \cline{3-4}
\end{array}
\qquad
\begin{array}{cc|c|c|c}
 & & \multicolumn{2}{c}{T^a} & w^b \\
 & & \multicolumn{1}{c}{w^a} & \multicolumn{1}{c}{v^a} & \\ \cline{3-4}
 & U & & + & \\ \cline{3-4}
S^a & D & & & \\ \cline{3-4}
\end{array}$$

$$\begin{array}{cc|c|c|}
v^b & & \multicolumn{2}{c}{T^a} \\
 & & \multicolumn{1}{c}{w^a} & \multicolumn{1}{c}{v^a} \\ \cline{3-4}
 & U & & \\ \cline{3-4}
S^a & D & + & \\ \cline{3-4}
\end{array}
\qquad
\begin{array}{cc|c|c|c}
 & & \multicolumn{2}{c}{S^b} & w^a \\
 & & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R} & \\ \cline{3-4}
 & v^b & & & \\ \cline{3-4}
T^b & w^b & & + & \\ \cline{3-4}
\end{array}$$

Figure 8.2

Here is a simple example of a strategic model.

Example 8.3. Consider Matching Pennies (Figure 8.1) and the associated strategic model in Figure 8.2. Here, Ann's type is either v^a or w^a; Bob's type is either v^b or w^b. Type v^a of Ann assumes the singleton set $\{(L, v^b)\}$, as depicted by the plus sign in the upper left of Figure 8.2. In words, this type of Ann assumes that Bob chooses the strategy L and is of type v^b. The interpretation of the rest of Figure 8.2 is similar.

Fix the state (U, v^a, R, w^b). At this state, Ann chooses strategy U, and assumes that Bob chooses L and is of type v^b. Bob in fact chooses R (contrary to what Ann assumes), and assumes that Ann chooses U and is of type

v^a (which is indeed the situation). This is an example of the kind of game scenario that strategic models are designed to describe.

9. Weakly Complete and Semi-Complete Models

In this section we give two positive results. First, we introduce weakly complete models and use them to show that the notion "assumes" is an essential ingredient in our main impossibility result. Second, we introduce semi-complete models and use them to show that the existence problem for complete models is intrinsically multi-player, i.e., game-theoretic, in nature. We will get stronger positive results by building strategic models with strategy sets given in advance, rather than plain belief models.

To show that the notion "assumes" is essential, we will find strategic models which are complete in the weaker sense that every statement which is possible can be believed (instead of assumed) by the player. Thus a paradox like the one in the Introduction does not arise in the ordinary modal logics which can express "believes" but cannot express "assumes".

Definition 9.1. Fix sets S^a and S^b. An (S^a, S^b)-based strategic model $\mathcal{M}$ will be called **weakly complete** if each nonempty set $Y \subseteq S^b \times T^b$ is believed by some $t^a \in T^a$, and each nonempty set $X \subseteq S^a \times T^a$ is believed by some $t^b \in T^b$.

Note that a weakly complete model $\mathcal{M}$ is "weakly complete for *every* language $\mathcal{L}$", even when extra relations are added to the vocabulary. That is, each nonempty set $Y \in \mathcal{L}^b$ is believed by some $t^a \in T^a$ and each nonempty set $X \in \mathcal{L}^a$ is believed by some $t^b \in T^b$.

Proposition 9.2. *Given sets S^a and S^b, there is a weakly complete (S^a, S^b)-based strategic model.*

Proof. Let T^a and T^b be sets of cardinality $\aleph_0 + |S^a| + |S^b|$. Then $U^a = S^a \times T^a$ and $U^b = S^b \times T^b$ have the same cardinality as T^a and T^b. Let $f : T^a \to U^b$ and $g : T^b \to U^a$ be bijections. Let $P^a((s^a, t^a), y)$ hold if and only if $f(t^a) = y$, and define P^b in the analogous way from g. Then $\mathcal{M} = (S^a \times T^a, S^b \times T^b, P^a, P^b, \ldots)$ is weakly complete; if $y \in Y \subseteq U^b$ then $f^{-1}(y)$ believes Y, and similarly with a, b reversed. ■

We next show that the existence problem for complete models is intrinsically multi-player, i.e., game-theoretic, in nature. We do this by showing that there is no difficulty if the condition is that every nonempty set of Bob's states is assumed by one of Ann's states, or vice versa. The impossibility arises when we want both conditions simultaneously.

Definition 9.3. A belief model $\mathcal{M}$ is **semi-complete** (for player a) if every nonempty subset of U^b is assumed by some element of U^a. $\mathcal{M}$ is **first order semi-complete** (for a) if $\mathcal{M}$ is complete for the language $\mathcal{H}$ consisting of all first order formulas $\varphi(y)$ where y has sort U^b.

Note that a semi-complete model $\mathcal{M}$ is first order semi-complete, and remains first order semi-complete even when extra relations are added to the vocabulary. Moreover, semi-completeness for a depends only on the relation P^a, not on P^b.

Proposition 9.4. *Given sets S^a, S^b, T^b, there is an (S^a, S^b)-based strategic model $\mathcal{M}$ with the given type set T^b which is semi-complete for a. If the sets S^a, S^b, T^b are finite, then $\mathcal{M}$ may be taken to be finite.*

Proof. Let T^a be the set of all nonempty subsets of $S^b \times T^b$, and define P^a by $P^a(s^a, t^a, s^b, t^b)$ if and only if $(s^b, t^b) \in t^a$. Then for any P^b, the strategic model $\mathcal{M} = (S^a \times T^a, S^b \times T^b, P^a, P^b, \ldots)$ is semi-complete for a. ■

We now show that if the sets S^a, S^b are at most countable, there is a countable (S^a, S^b)-based strategic model which is first order semi-complete.

Proposition 9.5. *Given finite or countable strategy sets S^a, S^b, there is a countable (S^a, S^b)-based strategic model $\mathcal{N}$ which is first order semi-complete.*

Proof. By Proposition 9.4 there is an infinite (S^a, S^b)-based strategic model $\mathcal{M}$ which is semi-complete for a. Then $\mathcal{M}$ is first order semi-complete for a. By the Downward Löwenheim-Skolem-Tarski Theorem, $\mathcal{M}$ has a countable elementary submodel $\mathcal{N}$. By Proposition 8.2, $\mathcal{N}$ is an (S^a, S^b)-based strategic model. It follows that $\mathcal{N}$ is first order semi-complete for a. ■

The above argument actually gives the following more general fact.

Proposition 9.6. *Suppose $\mathcal{M}$ is an (S^a, S^b)-based strategic model which is complete for a language $\mathcal{L}$. Then any elementary submodel of $\mathcal{M}$ is an (S^a, S^b)-based strategic model which is complete for $\mathcal{L}$. If the vocabulary of $\mathcal{M}$ is countable, then $\mathcal{M}$ has a countable elementary submodel, giving a countable (S^a, S^b)-based strategic model which is complete for $\mathcal{L}$.*

Proof. By the downward Löwenheim-Skolem-Tarski Theorem and Propositions 4.3 and 8.2. ■

Our next example shows that Theorem 5.7, which pinpointed the location of the holes in a belief model, cannot be improved by shortening the list of formulas (i)—(v) to (i)—(iv). The example produces a belief model which is complete for U^a, U^b and formulas (i)—(iv), and as a bonus, is semi-complete for b.

Example 9.7. Let S^a, S^b be nonempty sets. There is an (S^a, S^b)-based strategic model which is semi-complete for b and is complete for a language containing U^a, U^b and the formulas (i)—(iv) from Theorem 5.7.

Moreover, if S^a, S^b are finite or countable, there is a countable (S^a, S^b)-based strategic model which is first order semi-complete for b and complete for the above language.

Proof. Let T^a be any set with at least 3 distinct elements c, d, e. Put $U^a = S^a \times T^a$, let T^b be the set of all nonempty subsets of U^a, and put $U^b = S^b \times T^b$. Let P^b be the relation $P^b((s^b, t^b), x)$ iff $x \in t^b$. Then choose any relation P^a such that c assumes U^b, d assumes $S^b \times \{U^a\}$, e assumes $S^b \times \{\{d\}\}$, and every other element of T^a assumes a subset of U^b which is not contained in $S^b \times \{U^a\}$.

As in Proposition 9.4, the resulting model $\mathcal{M}$ is a semi-complete strategic model for b. One can readily check that in $\mathcal{M}$, formula (i) defines the set $S^b \times \{U^a\}$ which is assumed by d, and formula (iii) defines the set $S^b \times \{\{d\}\}$ which is assumed by e. Since $\mathcal{M}$ is semi-complete for b, it is complete for every subset of U^a. Thus $\mathcal{M}$ is complete for U^a, U^b, and the formulas (i)—(iv).

The moreover clause now follows from Proposition 9.6. ■

10. Positively and Topologically Complete Models

We now give some more positive results on the existence of complete models. (We already had Example 4.4 in Section 4, and also the semi-complete models of the previous section.) We show that complete models exist for the fragment of first order logic which has the positive connectives $\wedge, \vee$, the quantifiers $\forall, \exists$, and the belief and assumption operators, but does not have the negation symbol. We will work with strategic models. As usual, it is understood that x is a variable of sort U^a and y is a variable of sort U^b.

Definition 10.1. Let $\mathcal{M}$ be a strategic model. A **positive formula** is a first order formula for $\mathcal{M}$ which is built according to the following rules:

- Every atomic formula is positive.

- If φ, ψ are positive formulas, then so are $\varphi \wedge \psi, \varphi \vee \psi, \forall x \varphi, \forall y \varphi, \exists x \varphi, \exists y \varphi$.
- If $\varphi(y)$ is a positive formula, then so are $[x$ *believes* $\varphi]$ and $[x$ *assumes* $\varphi]$.
- If $\varphi(x)$ is a positive formula, then so are $[y$ *believes* $\varphi]$ and $[y$ *assumes* $\varphi]$.

The **positive language for** $\mathcal{M}$ is the set $\mathcal{P}$ of positive formulas in the first order language for $\mathcal{M}$. Thus $\mathcal{P}^a$ is the set of all subsets of U^a which are definable by a positive formula $\varphi(x)$, and similarly for $\mathcal{P}^b$.

Theorem 10.2. *For every pair of finite or countable strategy sets S^a, S^b, there is a countable (S^a, S^b)-based strategic model $\mathcal{M}$ which is complete for its positive language.*

Before giving the proof, which uses topological methods, let us consider what our main impossibility theorem tells us about models which are complete for the positive language. Since $\mathcal{M}$ cannot be complete for its first order language, there must be a set which is definable in the first order language $\mathcal{L}$ but is not definable in the positive language $\mathcal{P}$. That is, if the players use the positive language, there must be a first order property that the players cannot express. Using Theorem 5.7, we can pinpoint exactly where this happens.

Proposition 10.3. *Let $\mathcal{M}$ be a strategic model which is complete for its positive language. Then the set D defined by the diagonal formula*

$$\forall y[P^a(x,y) \to \neg P^b(y,x)]$$

does not belong to $\mathcal{P}^a$, but its negation does belong to $\mathcal{P}^a$.

Proof. The sets U^a, U^b are definable by the positive formulas $x = x$ and $y = y$, and the formulas (i)—(iii) of Theorem 5.7 are positive formulas. Moreover, the complement $U^a \setminus D$ is defined by the positive formula $\exists y[P^a(x,y) \wedge P^b(y,x)]$.

Assume that D is definable in $\mathcal{M}$ by a positive formula. Since the positive formulas are closed under the belief and assumption operators, the set defined by the formula (v) is also defined by a positive formula. But then, since $\mathcal{M}$ is complete for its positive language, there cannot be a hole at any of the formulas (i)—(v), contradicting Theorem 5.7. We conclude that the set D does not belong to $\mathcal{P}^a$. ■

We now construct strategic models which are complete in a topological sense. We will then show that such models are also complete for the positive language. Given a topological space X, let $K(X)$ be the space of all nonempty compact subsets of X endowed with the Vietoris topology. If X is compact metrizable, then so is $K(X)$ (Kechris [22, 1995, Theorem 4.26 and Exercise 4.20i)]).[††]

THEOREM 10.4. *Let S^a and S^b be compact metrizable spaces. There are compact metrizable spaces T^a, T^b and an (S^a, S^b)-based strategic model*

$$\mathcal{M} = (U^a, U^b, P^a, P^b, \tau^a, \tau^b, s^a, s^b, \ldots : s^a \in S^a, s^b \in S^b),$$

such that

(a) Each $t^a \in T^a$ assumes a compact set $\kappa^a(t^a) \in K(U^b)$, and each $t^b \in T^b$ assumes a compact set $\kappa^b(t^b) \in K(U^a)$.

(b) The mappings $\kappa^a : T^a \to K(U^b)$ and $\kappa^b : T^b \to K(U^a)$ are continuous surjections.

PROOF. Let $\mathcal{C}$ be the Cantor space, i.e., the space $\{0,1\}^{\mathbb{N}}$. There is a continuous surjection from $\mathcal{C}$ to any compact metrizable space (Kechris [22, 1995, Theorem 4.18]). Set $T^a = T^b = \mathcal{C}$, $U^a = S^a \times T^a, U^b = S^b \times T^b$. The space $K(U^b)$ is compact metrizable, so there is a continuous surjection κ^a from T^a to $K(U^b)$, and similarly with a, b reversed. The model $\mathcal{M}$ is obtained by setting $P^a((s^a, t^a), y)$ if and only if $y \in \kappa^a(t^a)$, and similarly for P^b. Then conditions (a) and (b) hold. The relation P^a is a proper subset of $U^a \times U^b$ because the space T^b has compact nonempty proper subsets. Therefore $\mathcal{M}$ is a strategic belief model. ■

DEFINITION 10.5. We will call a strategic model with the properties (a)—(b) of Theorem 10.4 an (S^a, S^b)-based **topologically complete model**.

Note that a topologically complete model $\mathcal{M}$ is complete for any language $\mathcal{K}$ such that $\mathcal{K}^a \subseteq K(U^a)$ and $\mathcal{K}^b \subseteq K(U^b)$.

We now make the connection between topologically complete models and the positive language.

LEMMA 10.6. *Let*

$$\mathcal{M} = (U^a, U^b, P^a, P^b, \tau^a, \tau^b, s^a, s^b, \ldots : s^a \in S^a, s^b \in S^b),$$

be a topologically complete model such that each of the extra relations in the list ... is compact. Then every positive formula $\varphi(x)$ defines a compact set in U^a and similarly for $\varphi(y)$ and U^b. Hence $\mathcal{M}$ is complete for its positive language.

[††] All topological spaces are understood to be nonempty.

Proof. It is clear that the unary relations are compact sets, that is, for each strategy $s^a \in S^a$ the set of states x for Ann with strategy s^a is compact, and similarly for S^b. It is also clear that the relations τ^a, τ^b, which hold for pairs of states with the same type, are compact. We next show that the relation P^a is compact. By Kechris [22, 1995, Exercise 4.29.i], the relation $C = \{(y, Y) : y \in Y\}$ is compact in $U^b \times K(U^b)$. Since the function $f^a : (s^a, t^a) \mapsto \kappa^a(t^a)$ is continuous from U^a to $K(U^b)$, the set $P^a = \{(x, y) : (y, f^a(x)) \in C\}$ is compact. Similarly, the relation P^b is compact. Thus every atomic formula defines a compact set. Since the spaces are Hausdorff, compact sets are closed, finite unions and intersections of compact sets are compact, and projections of compact relations by universal or existential quantifiers are compact.

To complete the proof, it suffices to show that for each compact set $Y \subseteq U^b$, the sets

$$A = \{x \in U^a : x \textit{ assumes } Y\}, \quad B = \{x \in U^a : x \textit{ believes } Y\}$$

are compact (and similarly with a and b reversed). We have $A = (f^a)^{-1}(\{Y\})$. Since the finite set $\{Y\}$ is closed in $K(U^b)$, A is compact in U^a. Similarly, $B = (f^a)^{-1}(\{Z : Z \subseteq Y\})$ and the set $\{Z : Z \subseteq Y\}$ is closed in $K(U^b)$, so B is compact in U^a. ■

We remark that the positive language $\mathcal{P}$ depends only on the relations of the model $\mathcal{M}$, while the sets $K(U^a)$ and $K(U^b)$ depend on the topology of $\mathcal{M}$. If the vocabulary of $\mathcal{M}$ is countable, then the positive language $\mathcal{P}$ will be countable, and since $K(U^a)$ is uncountable, $\mathcal{P}^a$ will be a proper subset of $K(U^a)$. In fact, the sets U^a and U^b are uncountable and each finite set is compact, so $K(U^a) \setminus \mathcal{P}$ will even contain finite sets. At the other extreme, in the above lemma we can take the vocabulary of $\mathcal{M}$ to be uncountable and even to contain all compact relations, and in this case we will have $\mathcal{P}^a = K(U^a)$ and $\mathcal{P}^b = K(U^b)$.

Proof of Theorem 10.2. Since S^a and S^b are finite or countable, there exist compact metrizable topologies on S^a and S^b. By Theorem 10.4, there is an (S^a, S^b)-based topologically complete model

$$\mathcal{M} = (U^a, U^b, P^a, P^b, \tau^a, \tau^b, s^a, s^b, \ldots : s^a \in S^a, s^b \in S^b).$$

It is clear from the definition that $\mathcal{M}$ without the extra relations indicated by the three dots is still topologically complete, so we may take $\mathcal{M}$ to have no extra relations. Then $\mathcal{M}$ has a countable vocabulary, and $\mathcal{M}$ is complete for its positive language by Lemma 10.6. In general, $\mathcal{M}$ is uncountable, but it follows from Proposition 9.6 that there is a countable (S^a, S^b)-based strategic model $\mathcal{N}$ (a countable elementary submodel of $\mathcal{M}$) which is complete for $\mathcal{P}$. ■

11. Other Models in Game Theory

The purpose of this section is to explain briefly how complete belief models are related to other models in the game theory literature. Here is an attempt to classify models of all possible beliefs that have been considered.

i. Universal models Start with a space of underlying uncertainty. (This could be the players' strategy sets, or sets of possible payoff functions for the players, etc.) The players then form beliefs over this space (their zeroth-order beliefs), beliefs over this space and the spaces of zeroth-order beliefs for the other players (their first-order beliefs), and so on inductively through the ordinals. The question is whether this process 'ends' at some level? More precisely, do the beliefs at some ordinal level α determine the beliefs at all subsequent levels? If so, we get a universal model. If not, i.e., if the set of α-level beliefs increases through all ordinals α, we get a non-existence result.

There are many papers on universal models. Existence results are given by Armbruster and Boge [1, 1979], Boge and Eisele [5, 1979], Mertens and Zamir [27, 1985], Brandenburger and Dekel [9, 1993], Heifetz [15, 1993], Epstein and Wang [11, 1996], Battigalli and Siniscalchi [3, 1999], [4, 2002], Mariotti, Meier, and Piccione [24, 2005], and Pintér [30, 2005], among others. Fagin, Halpern, and Vardi [14, 1991], Fagin [12, 1994], Fagin, Geanakoplos, Halpern, and Vardi [13, 1999], and Heifetz and Samet [19, 1999] give non-existence results. The positive results are obtained by making various topological or measure-theoretic hypotheses (as we did in the previous section). We should also note that many of these papers formalize belief as probability and not possibility as we have done. But—again in line with our results in Section 10—the key point is the topological assumptions. (Epstein-Wang consider preferences, with a topological structure.) Aumann [2, 1999] treats knowledge rather than belief, and uses S5 logic to get a positive result. (On this last result, see also Heifetz [16, 1999].)

ii. Complete models These are defined as in our Definition 4.2 or Definition 10.5—i.e., in terms of 'two-way surjectivity.' One way to obtain complete models is to construct a universal model. For example, Mariotti, Meier, and Piccione [24, 2005] get a complete model for compact Hausdorff spaces, as a corollary of the existence of a universal model. Our proof of Theorem 10.4 gives a simple direct construction of a complete model for compact metrizable spaces. Battigalli and Siniscalchi [3, 1999], [4, 2002] get a complete model for spaces of conditional probability systems, again as a corollary of the existence of a universal model. Brandenburger, Friedenberg, and Keisler [10, 2006] has a direct construction of a complete model for spaces of lexicographic probability systems. Salonen [31, 1999] gives a variety of existence results on completeness, under a variety of assumptions.

iii. Terminal models Given a category $\mathbf{C}$ of models of beliefs, call a model $\mathcal{M}$ in $\mathbf{C}$ terminal if for any other model $\mathcal{N}$ in $\mathbf{C}$, there is a unique belief-preserving morphism from $\mathcal{N}$ to $\mathcal{M}$.[*] Heifetz and Samet [17, 1998a] show existence of a terminal model for probabilities, without topological assumptions. Meier [26, 2006] shows existence of a terminal model with finitely additive measures and κ-measurability (for a fixed regular cardinal κ), and non-existence if all subsets are required to be measurable. Heifetz and Samet [18, 1998b] show non-existence for the case of knowledge models, and Meier [25, 2005] extends this non-existence result to Kripke frames.

We end by noting that, to the best of our knowledge, no general treatment exists of the relationship between universal, complete, and terminal models (absent specific structure). Such a treatment would be very useful.

References

[1] Armbruster, W., and W. Boge, "Bayesian Game Theory," in O. Moeschlin and D. Pallaschke (eds.), *Game Theory and Related Topics*, North-Holland, Amsterdam, 1979.

[2] Aumann, R., "Interactive Epistemology I: Knowledge," *International Journal of Game Theory*, 28, 1999, 263-300.

[3] Battigalli, P., and M. Siniscalchi, "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory*, 88, 1999, 188-230.

[4] Battigalli, P., and M. Siniscalchi, "Strong Belief and Forward-Induction Reasoning," *Journal of Economic Theory*, 106, 2002, 356-391.

[5] Boge, W., and T. Eisele, "On Solutions of Bayesian Games," *International Journal of Game Theory*, 8, 1979, 193-215.

[*] Meier [26, 2006] proposes the same terminology.

[6] Bonanno, G., "A Simple Modal Logic for Belief Revision," *Synthese (Knowledge, Rationality and Action)* 147, 2005, 193-228.

[7] Boolos, G., *The Logic of Provability*, Cambridge University Press, 1993.

[8] Brandenburger, A., "On the Existence of a 'Complete' Possibility Structure," in Basili, M., N. Dimitri, and I. Gilboa, eds., *Cognitive Processes and Economic Behavior*, Routledge, 2003, 30-34.

[9] Brandenburger, A., and E. Dekel, "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 1993, 189-198.

[10] Brandenburger, A., A. Friedenberg, and H.J. Keisler, "Admissibility in Games," 2006, at www.stern.nyu.edu/∼abranden.

[11] Epstein, L., and T. Wang, "Beliefs about Beliefs Without Probabilities," *Econometrica*, 64, 1996, 1343-1373.

[12] Fagin, R., "A Quantitative Analysis of Modal Logic," *Journal of Symbolic Logic*, 59, 1994, 209-252.

[13] Fagin, R., J. Geanakoplos, J. Halpern, and M. Vardi, "The Hierarchical Approach to Modeling Knowledge and Common Knowledge," *International Journal of Game Theory*, 28, 1999, 331-365.

[14] Fagin, R., J. Halpern, and M. Vardi, "A Model-Theoretic Analysis of Knowledge," *Journal of the Association of Computing Machinery*, 38, 1991, 382-428.

[15] Heifetz, A., "The Bayesian Formulation of Incomplete Information—The Non-Compact Case," *International Journal of Game Theory*, 21, 1993, 329-338.

[16] Heifetz, A., "How Canonical is the Canonical Model? A Comment on Aumann's Interactive Epistemology," *International Journal of Game Theory*, 28, 1999, 435-442.

[17] Heifetz, A., and D. Samet, "Topology-free Typology of Beliefs," *Journal of Economic Theory*, 82, 1998a, 324-381.

[18] Heifetz, A., and D. Samet, "Knowledge Spaces with Arbitrarily High Rank," *Games and Economic Behavior*, 22, 1998b, 260-273.

[19] Heifetz, A., and D. Samet, "Coherent Beliefs are Not Always Types," *Journal of Mathematical Economics*, 32, 1999, 475-488.

[20] Humberstone, I.L. "The Modal Logic of All and Only,"*Notre Dame Journal of Formal Logic*, 28, 1987, 177-188.

[21] Huynh, H.L., and B. Szentes, "Believing the Unbelievable: The Dilemma of Self-Belief," 1999, at http://home.uchicago.edu/∼szentes.

[22] Kechris, A., *Classical Descriptive Set Theory*, Springer-Verlag, 1995.

[23] Lomuscio, A., *Knowledge Sharing Among Ideal Agents*, Ph.D. Thesis, University of Birmingham, 1999.

[24] Mariotti, T., M. Meier, and M. Piccione, "Hierarchies of Beliefs for Compact Possibility Models," *Journal of Mathematical Economics*, 41, 2005, 303-324.

[25] Meier, M., "On the Nonexistence of Universal Information Structures," *Journal of Economic Theory*, 122, 2005, 132-139.

[26] Meier, M., "Finitely Additive Beliefs and Universal Type Spaces," *The Annals of Probability*, 34, 2006, 386-422.

[27] Mertens, J-F., and S. Zamir, "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1985, 1-29.

[28] Pacuit, E., "Understanding the Brandenburger-Keisler Paradox," to appear, *Studia Logica*, 2006.

[29] Pearce, D., "Rational Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1984, 1029-1050.

[30] Pintér, M., "Type Space on a Purely Measurable Parameter Space," *Economic Theory*, 26, 2005, 129-139.

[31] Salonen, H., "Beliefs, Filters, and Measurability," 1999, University of Turku.

[32] Thomason, R., "A Note on Syntactical Treatments of Modality," *Synthese*, 44, 1980, 391-395.

[33] Yanofsky, N., "A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points," *The Bulletin of Symbolic Logic*, 9, 2003, 362-386.

Adam Brandenburger
Stern School of Business
New York University
New York, NY 10012
`adam.brandenburger@stern.nyu.edu`
`www.stern.nyu.edu/∼abranden`

H. Jerome Keisler
Department of Mathematics
University of Wisconsin-Madison
Madison, WI 53706
`keisler@math.wisc.edu`
`www.math.wisc.edu/∼keisler`