A Note on Assumption-completeness in Modal Logic

Jonathan A. Zvesper* Eric Pacuit†

Introduction

The literature on the epistemic foundations of game theory uses a variety of mathematical models to formalise talk about the players' beliefs about the game, beliefs about the rationality of the other players, beliefs about the beliefs of the other players, beliefs about the beliefs about the beliefs of the other players, and so on (see [Bra07] for a recent survey). Examples include Harsanyi's type spaces ([Har67]), interactive belief structures ([Bra03]), knowledge structures ([Aum76]) plus a variety of logic-based frameworks (see, for example, [Ben01, HM06, Bon02, Boa02, BSZ08]). A recurring issue involves defining a space of *all possible* beliefs of the players and whether such a space exists. In this paper, we study one such definition: the notion of *assumption-complete* models. This notion was introduced in [Bra03], where it is formulated in terms of "*interactive belief models*" (which are essentially qualitative versions of type spaces). Assumption-completeness is also explored in [BK06], where a number of significant results are found, and connections to modal logic are mentioned. A discussion of that paper, and a syntactic proof of its central result, are to be found in [Pac07].

Within and between these different mathematic models, different epistemic notions can be formalised, one of which is the notion of an "assumption", which is closely related to the only-knowing operator studied by Levesque [Lev90] (cf. [HL01]). Roughly speaking, a player's *assumption* is defined as her strongest belief: the conjunction of all her beliefs (equivalently, a belief that implies all her other beliefs).[1] Call the (two) players "Ann" and "Bob". An *interactive belief model* (we will use the shorter *belief model* in this paper) consists of *states* for Ann and for Bob. It specifies the beliefs, and some other information like the strategy chosen, of a player at any of his/her states. Each player's beliefs are defined over the other player's states: that is, beliefs of the players are given in terms of a set of the opponent's states. Thus to each Ann state is associated a set of Bob states that Ann considers possible, and to each Bob state a set of Ann states that Bob considers possible. Bob's *assumption* at a state is the set of Ann

*ILLC and CWI, Amsterdam, jonathan@illc.uva.nl

†Stanford University, epacuit@stanford.edu

[1] The formal details will follow in Section 1. This definition of "assumption" might seem strange, and we certainly do not claim that it, nor the formal definition that will follow, capture the common-sense meaning of the English word "assumption". However, when we present the semantics of belief models we will see that, formally speaking, *assumption-completeness* will be a natural enough property, albeit poorly named. Note that, while the term "assumption" is used in [BK06], there the notion that we will study is called simply "completeness". We prefer the more specific, if less wieldy, term "assumption-completeness".

states to which that state is related. A belief model is *assumption-complete* for a language of Ann-states just when for every sentence of the language, there is some Bob-state (in the model) where Bob assumes (the set of states satisfying) that sentence. The idea of assumption-completeness is that the language of Ann-states should be accessible to Bob. And if it is accessible to Bob then he should be able to *assume* (in this artificial sense) any member of it.

Brandenburger and Keisler prove the following impossibility result:

Theorem 4 [BK06, Theorem 5.4]**.** There are *no* assumption-complete models for the first-order language.

This was taken to be a limitative result, and one that should be of significance for game theory:

> *[O]ur 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 [assumption-]incomplete. This appears to be a kind of basic limitation in the analysis of games.* [BK06]

As Brandenburger and Keisler point out, the existence of assumption-complete models is not only of theoretical interest [BK06]. It turns out to be relevant for the "epistemic program" in game theory. The goal of this program is to provide epistemic conditions on the players (e.g., common belief in rationality) that lead[2] to various solution concepts (e.g. Nash equilibrium, iterated dominance, backwards induction). The assumption that the belief model is assumption-complete has been non-trivially used in two analyses: in Battigalli and Siniscalchi's analysis of extensive-form rationalisability [BS02] and Brandenburger, Friedenberg and Keisler's analysis of iterated admissibility [BFK08a]. We return to the role that assumption-completeness plays in epistemic analyses of games in Section 4.

Given the above interpretation of Theorem 4, a natural question[3] is: can one consider instead a restricted set of "tools" which *can* be "available" to the players, and which are also useful for the analyst? Theorem 4 shows that the first-order language is too powerful a tool to be available; what about weaker languages?

We will address this question from the perspective of modal logic, defining a modal language for belief models, giving a complete axiomatisation (Theorem 9), and obtaining, as a corollary of the completeness proof, the following:

Theorem 8. There are assumption-complete models for the basic modal language.

What about strengthening our possibility result? In Section 3 we will look briefly at perspectives for doing exactly this. We will conjecture that the bounded fragment of first-order logic has assumption-complete models. The bounded fragment is expressive enough to express some concepts that are very

[2]More precisely *representation* theorems are proved stating that players satisfying such-and-such epistemic condition will play a particular solution concept and conversely if the players play according to some solution concept then there is an epistemic model where the players satisfy the epistemic conditions. See [dB04] for a critical survey of this line of reasoning.

[3]This is also raised in [BK06, Section 2].

important for game theory. Notably, in the appropriate framework, it can express the proposition that a player is *rational*[4]

1 Preliminaries

Except in Appendix A, we will work in the framework from [BK06]. Belief models are two-sorted first-order structures.[5]

Definition 1 ([BK06], Definition 3.1)**.** *A* **belief model** *is a structure*

$$(U^a, U^b, R^a, R^b, \{P^\alpha\}_{\alpha\in\mathbb{N}}), \text{ where}$$

1. $U^a \neq \emptyset \neq U^b$*,* $U^a \cap U^b = \emptyset$*;*
2. $R^a \subsetneq U^a \times U^b$*,* $R^b \subsetneq U^b \times U^a$*;*
3. *for every* $u \in U^a$ *there is a* $v \in U^b$ *such that* $uR^a v$*, and similarly for every* $v \in U_b$*;*

The elements of the domain $U = U^a \cup U^b$ are called "***states***". Specifically those in U^a are called "*Ann states*" and those in U^b "*Bob states*". We might also call them "*types*", because each one specifies an epistemic type (in the sense of an Harsanyi type space) of the relevant player. The relations R^a and R^b specify those states considered possible by Ann and Bob respectively: $uR^a v$ means that the Ann-state u considers the Bob-state v to be possible. We write $R^a(u)$ to mean $\{v \in U^b \mid uR^a v\}$, and we use similar terminology with b switched for a. We say that for $u \in U^a$ and $E \subseteq U^b$, u ***believes*** E just when $R^a(u) \subseteq E$, and (stronger) that u ***assumes*** E just when $R^a(u) = E$. We will write R for $R^a \cup R^b$.

The predicates P^α are there to carry additional information about the states. For example, when considering models for *games*, each P^α may represent which *strategy* is chosen by each player state.

The conditions imposed by Definition 1 are natural: (1) says that there are Ann states and Bob states, (2) that for both Ann and Bob there is at least one non-trivial belief state, i.e. at which some Bob or Ann state is ruled out as a possibility, and (3) that at every Ann or Bob state, some Bob or Ann state is taken to be possible.

State-based models, and a definition of belief like the one given, are familiar from epistemic logic since the work of Hintikka [Hin62]. In those single-sorted models that are standard in epistemic logic, states (sometimes called "possible worlds") specify a type for each player, i.e. in the two-player case a state would specify an Ann-type *and* a Bob-type. (The connection between qualitative type spaces and single-sorted models is discussed in [BB99, Chapter 3].) We discuss

[4]Here *rationality* is interpreted in the standard way: optimising given the agent's *current* beliefs. Making the statement that a particular logic can/cannot express this notion of rationality precise is still ongoing work. This is also briefly discussed in [BK06, Section 2].

[5]In [BK06], the definition of belief models is more general, allowing for almost arbitrary signatures for belief models. With condition (3), we restrict our attention to monadic predicates (the P^α's) because it is natural to do so in the stated field of application of the belief models, viz. to games, in which the P^α's represent choices made by the players. Furthermore, they allow arbitrary strategy sets, where we restrict our attention, for simplicity's sake, to countable strategy sets.

in Appendix A how to define assumption-completeness for standard models. Modulo this different perspective given by the two-sortedness of the models, the notion of an assumption is essentially the same as that of "only knowing", which was introduced by Levesque [Lev90], and axiomatised by Halpern and Lakemeyer ([HL01], cf. [Ben79, Hum87]).

Fix a belief model $(U^a, U^b, R^a, R^b, \{P^\alpha\}_{\alpha \in \mathbb{N}})$. Following [BK06], a **language for Ann** (based on the fixed belief model), denoted $\mathcal{L}^a$, is any subset of U^a (similarly for Bob). For example, the **powerset language** for Ann (Bob) is $\mathcal{L}^a = \mathfrak{p}(U^a)$ ($\mathcal{L}^b = \mathfrak{p}(U^b)$), where we write $\mathfrak{p}(X)$ to denote the power set of X. Another natural example is the ***first-order language*** defined to be the sets definable by sentence of first-order logic. Formally, a first-order formula is defined by the following recursion schema:

$$\varphi ::= \mathbf{U^a} \mid \mathbf{P}^\alpha x \mid x\mathbf{R^a}y \mid x\mathbf{R^b}y \mid \neg\varphi \mid \varphi \wedge \varphi \mid \exists x_a \varphi \mid \exists x_b \varphi$$

As usual, sentences are closed formulae (i.e., those without free occurrences of variables). Given a belief model, each first-order sentence defines a subset of that model in the standard way (with a caveat about two-sorted quantification: $\exists x_a$ quantifies only over Ann states). Then the ***first-order language*** $\mathcal{L}_1^a$ is the set of subsets of U^a that are definable by a first-order sentence (similarly for Bob). We are now ready to define the notion that will be central to our concerns in this paper.

Definition 2 ([BK06], Definition 4.2)**.** *A belief model is* **assumption-complete** *for the language $\mathcal{L} = \mathcal{L}^a \cup \mathcal{L}^b$ just if*

(Ca) for every $\emptyset \neq E_b \in \mathcal{L}^b$, there is an $u \in U^a$ such that $R^a(u) = E_b$, and for every $\emptyset \neq E_a \in \mathcal{L}^a$, there is a $v \in U^b$ such that $R^b(v) = E_a$.

That is, (Ca) ensures that every definable set of Ann (or Bob) states can be assumed by Bob (resp. Ann). Formally speaking, this is a natural condition to impose. We will say of a *language* that it is assumption-complete just when there is *some* model that is assumption-complete for it, and that it is ***assumption-incomplete*** just when there is *no* such model.

2 Main Results

Intuitively, if a language is assumption-complete then it has a "big" model, which means, roughly, that any property of Bob states (expressible in the language) can be assumed by Ann (and vice versa for Bob). The following theorem demonstrates that some restriction on the language is needed for it to be assumption-complete:

Theorem 3 ([Bra03])**.** *The powerset language (where $\mathcal{L}^a = \mathfrak{p}(U^a)$ and $\mathcal{L}^b = \mathfrak{p}(U^b)$) is assumption-incomplete.*

The proof of Theorem 3 uses Cantor's Theorem that there is no surjection from a set onto its power set. But what about slightly weaker languages that still are stronger than the first-order language? (For example, first-order logic with fixed-points, or some second-order logic.) The following theorem states that the first-order language is already too expressive to be assumption-complete:

Theorem 4 ([BK06, Theorem 5.4])**.** *$\mathcal{L}_1$ is assumption-incomplete.*

Thus the standard tool first-order logic is too strong to have this formally natural property.[6] The proof of Theorem 4 uses a generalised version of Russell's paradox, and essentially relies on the same kind of diagonalisation argument used to prove Cantor's Theorem. Although we refer to [BK06] for the proof, we note for later reference that it is important that the sentence (BK) is expressible in the first-order language $\mathcal{L}_1$:

(BK) Ann believes that Bob's assumption is that Ann believes that Bob's assumption is wrong.

Theorem 4 rules out the first-order language; what then might we use in its stead? – What are our alternatives? We now mention very briefly one positive result, for the so-called "*positive fragment*" (see [BK06, Definition 10.1]), which is essentially a mixture of first-order and modal languages without negation.

Theorem 5 ([BK06, Theorem 10.4])**.** *The positive fragment is assumption-complete.*

Related to this Theorem, Mariotti, Meier and Piccione show that there exists a "universal possibility structure" [MMP05]. We now begin an investigation of *other* fragments of first-order logic that may be assumption-complete. We start with the basic *modal language* $\mathcal{ML}$. The *basic modal formulae* are those defined by the following schema:

$$\varphi ::= ♀ \mid p_\alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi$$

We write $\Diamond$ to abbreviate $\neg\Box\neg$, $\varphi \supset \psi$ for $\neg(\varphi \wedge \neg\psi)$, and ♂ for $\neg$♀. The ***basic modal language*** $\mathcal{ML}$ is the set of subsets that are definable by some basic modal formula, where ♀ defines the Ann states U^a; p_α defines the set P^α; negation and conjunction work as usual; and $\Box\varphi$ defines the set where the state-owner believes φ. That is, where $[\![\varphi]\!]$ is the set defined by φ, $\Box\varphi$ defines the following set:

$$\{u \in U \mid R(u) \subseteq [\![\varphi]\!]\},$$

It was shown in [BK06, Section 9] that, since the basic modal language cannot express the assumption operator, there are belief models "complete in a weaker sense that every statement which is possible can be believed (instead of assumed) by the player." What about the stronger statement, that if a language cannot express the assumption operator, then there are assumption-complete models? Certainly, being able to talk about Bob's assumptions *is* essential in the particular proof of Theorem 4; however, the converse is open: that if one *cannot* talk about the players' assumptions in the language then there is an assumption-complete model (we are not making any claims concerning the truth of this statement, just that it has not been proved). We do have the following theorem:

Theorem 6. *$\mathcal{ML}$ is assumption-complete.*

In fact we can strengthen Theorem 6, by adding a property which is also in effect present in Brandenburger and Keisler's positive result.

[6] Of course, there may be some other language which is expressively *incomparable* with $\mathcal{L}_1$, but we will not pursue this line of reasoning here.

Definition 7. *Say that a belief model is* **total** *just when*

(Cb) for every P^α, $U^a \cap P^\alpha \neq \emptyset$ and $U^b \cap P^\alpha \neq \emptyset$.

Condition (Cb) means that every possible 'basic configuration' is present, so in the case where the P^α's represent choice of strategy, it means that for each of player i's strategies s_i, there is a state at which i chooses that s_i. (This is assuming the same strategy sets for both players; the case of disjoint strategy sets does not add any mathematical complication.) We will therefore prove the following:

Theorem 8. *There are total assumption complete belief models for the basic modal language.*

Although the totality condition is never made explicit out in [BK06], it is implicit there. (Not only does their Theorem 5 actually show the existence of total complete belief models, but also if the requirement that the model be total is dropped, then [BK06, Problem 7.7] is trivially answerable [Zve07].)

We give now a complete axiomatisation for $\mathcal{ML}$ over belief models. The axioms consist of an axiomatisation of the propositional connectives $\neg$ and $\wedge$, plus the following:

$\vdash \Box(p \supset q) \supset (\Box p \supset \Box q)$	K
$\vdash \Box p \supset \Diamond p$	D
$\vdash ♂ \supset \Box ♀$	$U1$
$\vdash \Diamond ♀ \supset ♂$	$U2$

The following are the rules of inference:

$$\frac{\vdash \varphi \qquad \vdash \varphi \supset \psi}{\vdash \psi}\ MP \qquad \frac{\vdash \varphi}{\vdash \Box\varphi}\ Nec \qquad \frac{\vdash \varphi}{\vdash \varphi[p \mapsto \psi]}\ Sub$$

If there is an inference (using only these axioms and rules) of φ then we write $\vdash \varphi$. We write $\models \varphi$ just when φ is *valid*, that is: when in *every* belief model, φ defines the whole set U. To show that a logic is (weakly) *complete*[7] is to show that $\models \varphi \Rightarrow \vdash \varphi$.

Theorem 9. $\models \varphi \ \Leftrightarrow \ \vdash \varphi$

Theorem 9 is proved in a standard way by building a "canonical model" (cf. [BdRV01]). The states in the canonical model are maximally-consistent sets, and the relation is defined as follows:

$$R(\Gamma) = \{\Delta \mid \forall\psi \in \Delta, \Diamond\psi \in \Gamma\}$$

Theorem 8 can then be proved by observing that this canonical model is appropriately assumption-complete. We sketch now the proof of Theorem 8. Take any definable subset of the canonical model $E \in \mathcal{ML}^a$ (without loss of generality: the same will hold with a switched for b). Then by definition of $\mathcal{ML}^a$, there is some modal formula φ such that $[\![\varphi]\!] = E$. Furthermore, we must have $\vdash \varphi \supset ♀$, because otherwise there would be some Bob-state $\Gamma \in E$. Then it remains to show that there is a state Γ_φ that assumes φ in the canonical model. We use the following lemma:

[7] This notion of completeness, familiar from formal logic, has (usually!) nothing to do with assumption-completeness.

Lemma 10. *The set* Γ'_φ *is consistent:*

$$\Gamma'_\varphi := \{\Diamond\gamma \mid \nvdash \neg(\varphi \wedge \gamma)\} \cup \{\Box\varphi\}$$

Proof. We appeal to completeness and invariance of basic modal formulae under disjoint unions and generated submodels ([BdRV01, Propositions 2.3 and 2.6]):

Let $\Pi_\varphi = \{\gamma \mid \nvdash \neg(\varphi \wedge \gamma)\}$ be the set of formulae consistent with φ. For each such $\gamma \in \Pi_\varphi$, by completeness there is a pointed model $\mathfrak{M}_\gamma, \omega_\gamma$ such that $\mathfrak{M}_\gamma, \omega_\gamma \vDash \varphi \wedge \gamma$. Since the language is preserved under generated submodels and disjoint unions, we will still have that each $\omega_\gamma \vDash \varphi$ when we take the disjoint union of all of the submodels generated by the ω_γ's for every $\gamma \in \Pi_\varphi$. Now define a new model $\mathfrak{M}$ by taking the disjoint union of the submodels generated by each ω_γ, adding one new state ω_0, and stipulating that $R(\omega_0) = \{\omega_\gamma | \gamma \in \Pi_\varphi\}$. Notice that since by hypothesis $\varphi \vdash ♀$, this is indeed a model, since every $\omega_\gamma \vDash ♀$ (and so $\omega_0 \vDash ♂$). By construction we have $\omega_0 \vDash \Box\varphi$, because for each $\gamma \in \Pi_\varphi$, $\omega_\gamma \vDash \varphi$. Furthermore, for each $\gamma \in \Pi_\varphi$, we also have $\omega_0 \vDash \Diamond\gamma$, since $\omega_0 R \omega_\gamma$ and $\omega_\gamma \vDash \gamma$. This simple construction is illustrated in figure 1.

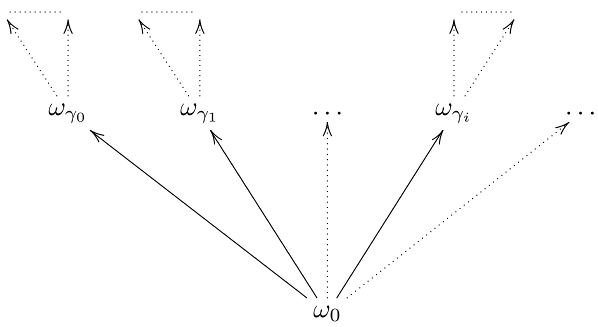

Figure 1: The construction described in the proof of lemma 10

■

From Lemma 10, it follows that Γ'_φ can be extended to form a state Γ_φ in the canonical model. We must show that $R(\Gamma_\varphi) = [\![\varphi]\!]$:

- Since $\Box\varphi \in \Gamma_\varphi$, $R(\Gamma_\varphi) \subseteq [\![\varphi]\!]$;
- Take any $\Delta \in [\![\varphi]\!]$. Then take any $\psi \in \Delta$; we know that $\nvdash \neg(\varphi \wedge \psi)$, since Δ is by hypothesis consistent. So by definition of Γ'_φ, $\Diamond\psi \in \Gamma'_\varphi \subseteq \Gamma_\varphi$. So by definition of the canonical model, $\Delta \in R(\Gamma_\varphi)$.

This concludes the proof of Theorem 8. The argument can in fact be made more general, to the effect that any *saturated* canonical model [BdRV01] is in this sense assumption-complete – a suggestion made to the second author by van Benthem (p.c.).

3 Beyond the Basic Modal Language

The basic modal language is natural, but is arguably not expressive enough a "tool" for the analyst. It is not possible to give general results about the expressibility of game-theoretical concepts in terms of a language, since there are different ways of representing the various elements involved in a game. However, we get some indications from the fact that in [BOR05] the authors deem it necessary to extend the basic modal language (with so-called "nominals") in order to express *Nash equilibrium*, and in [AZ07] the authors express *rationality* in a natural modal way using *bounded quantification*.[8] (Again, in these papers the framework is one of *single-sorted models*, which we will discuss below in Appendix A.) It is therefore natural to search for more expressive extensions of the basic modal language that nonetheless are assumption complete; given Theorem 4, it is natural to look for fragments of first-order logic that have the property.[9] In this section we briefly discuss perspectives for finding such fragments. Since we take the vocabulary of the basic modal language to be intuitively appealing, and of use to the analyst, we will look at *extensions* of the basic modal language.

[Cat05] studies a number of extensions of the basic modal language. Of particular interest to us now is $\mathcal{L}_\downarrow$, known as "hybrid language with binder". Let the ***binder formulae*** be those of the following form:

$$\varphi ::= ♀ \mid p_\alpha \mid x \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid \downarrow x.\varphi$$

Binder formulae ($\downarrow$ is know as the "binder") define sets relative to a *variable assignment*, that is: a function $\sigma : Var \to U$.[10] We write $[\![\varphi]\!]_\sigma$ for the subset of U defined by the formula φ relative to σ. And we write $\sigma[x \mapsto u]$ for the assignment that agrees everywhere with σ except that it maps x to u. The semantics of the new connectives are as follows:

- $[\![x]\!]_\sigma = \{\sigma(x)\}$;
- $[\![\downarrow x.\varphi]\!]_\sigma = \{u \in U \mid u \in [\![\varphi]\!]_{\sigma[x \mapsto u]}\}$

For a binder *sentence* φ, we write just $[\![\varphi]\!]$ (since it makes no difference which assignment we choose). The language $\mathcal{L}_\downarrow$ of subsets defined by a binder sentence is indeed a language in our strict sense, because $\mathcal{L}_\downarrow$ is a fragment of the first-order language. Indeed, it is expressively equivalent to the "bounded fragment" of the first-order language [ABM99]. Since Feferman has studied the bounded fragment [Fef68], we know that it is *invariant under generated submodels*. In the details of the proof of Theorem 8, the only 'modal' behaviour we exploit, in showing that the set Γ'_φ is consistent, is that sentences are preserved under disjoint unions and generated submodels. Thus – although there is no canonical model construction for $\mathcal{L}_\downarrow$– we are still lead to suspect that this language also has assumption-complete models:

Conjecture 11. *There are (total) assumption-complete belief models for the bounded fragment.*

[8] Indeed, (instrumental) rationality essentially says that an agents choice '*now*' is optimal given her beliefs '*now*', something which is typically not expressible in a basic modal language, but which calls for the kind of "hybrid" language, i.e. using nominals.

[9] Though fixpoint logics for common knowledge go beyond first-order logic, and are obviously of relevance to the epistemi analysis of games; we do not examine those here.

[10] The same is true of first-order formulae, but we skipped the details because they are more standard.

An inspection of the proof of Theorem 4 reveals three conditions that *together are sufficient* to show that a language $\mathcal{L}$ is not assumption-complete. The first condition, satisfied by all of these modal languages, is that $\mathcal{L}^a$ be closed for the belief operator:

(C0) $X \in \mathcal{L}^b \Rightarrow \{u_a \in U_a \mid R^a(u_a) \subseteq X\} \in \mathcal{L}^a$

The second condition is that $\mathcal{L}$ contain the following set:

$$D_A := \{\omega \in U_A \mid \forall \omega' \in (R(\omega) \cap U_B),\ \omega \notin R(\omega')\}$$

(C1) $D_A \in \mathcal{L}^a$.

C1 says that Bob has "available" (i.e. in $\mathcal{L}^a$, the language for defining sets of Ann states) the sentence "Ann believes that Bob's assumption is wrong".

The third condition is that $\mathcal{L}^a$ be closed under the assumption operator, i.e.:

(C2) $X \in \mathcal{L}^a \Rightarrow \{u_b \in U_b \mid R^b(u_b) = X\} \in \mathcal{L}^b$

If all of (C0)–(C2) hold, then $\mathcal{L}$ is assumption-incomplete. In particular, $\mathcal{L}$ will allow sentence (BK) to be expressed. Note that D_A *is* expressible by a binder sentence, so is in $\mathcal{L}_\downarrow$:

Fact 12. $D_A = [\![♀ \wedge \downarrow x.\Box\Box\neg x]\!] \in \mathcal{L}_\downarrow$

However, since the language is closed under generated submodels then, importantly, the assumption operator is *not* expressible:

Fact 13. $\mathcal{L}_\downarrow$ *does not have condition C2.*

Any extension of the modal language has C0; we've looked at a language with C1; so what about C2? Clearly, adding an assumption operator to $\mathcal{L}_\downarrow$ will make the language assumption-incomplete. (So this would be a strictly weaker language than first-order logic, that is nonetheless assumption-incomplete.) Nonetheless, while we do not investigate the matter further here, we conjecture that adding an assumption operator into $\mathcal{ML}$ would *not* leave the happy realm of assumption-completeness.

4 Discussion

We plan to explore a number of issues in the future. The most pressing issue is finding interesting languages, primarily fragments of first-order logic, that are assumption-complete, especially languages that can express concepts that are of interest to game-theorists, and that have a natural appeal in terms of being languages that we would want to attribute to the agents to capture their ability to think about the situation they are in (e.g., the bounded fragment). We also think that it is important to explicate the notion of assumption-completeness in terms of some of the other models that are used in the game-theory literature, for example those that we mentioned in the introduction. – If assumption-completeness is an important epistemic notion, then it is important to understand it also in terms of other epistemic models. We address this issue, for the case of single-sorted epistemic models that are familiar from the epistemic logic literature since [Hin62], in the Appendix A.

It is also natural to question the quoted interpretation of the Brandenburger and Keisler impossibility result: does Theorem 4, which is essentially a generalisation of Russell's paradox, really point to "a kind of basic limitation in the analysis of games"?

Why work in assumption-complete models? A natural reaction to the Brandenburger and Keisler impossibility result (Theorem 4) is to wonder what all the fuss is about. Assumption-completeness is a nice abstract property of a logical language, but what is the harm in working with languages without this property? One answer from the literature on the epistemic foundations of game theory is that assumption-complete belief models are needed to provide an epistemic analysis of certain solution concepts. Indeed, Battigalli and Siniscalchi argue convincingly that "*analysing an extensive-form game in the framework of an incomplete type space*[11] *introduces implicit and potentially undesirable restrictions on forward-induction reasoning*" ([BS02, pg. 368], original italics). More broadly, Brandenburger, Friedenberg and Keisler point out that

> We think of a particular incomplete structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory [...], who the players are in the given game can be seen as a shorthand for their experiences before the game. The players' possible characteristics — including their possible types — then reflect the prior history or context. (Seen in this light, complete structures represent a special "context-free" case, in which there has been no narrowing down of types.) [BFK08b, pg. 319]

So, we have two examples of epistemic analyses of solution concepts where it is crucial that the analysis takes place in assumption-complete models. In fact, [BFK08a] introduce a new form of *irrationality* where a player optimises, but does not consider *all possibilities*. Making this notion of irrationality precise requires defining a formal model of "all possibilities", and assumption-completness is one approach to rigorously defining such a model. Finally, we note that there are other "epistemic" analyses of iterated admissibility, or removal of weakly dominated strategies, where assumption-completeness plays less of a role (see [Ben07, Bon02, AZ07]). Of course, comparing these different analyses to [BFK08b] and judging the need for assumption-complete models is difficult without a precise set of criteria (cf. [dB04]).

Acknowledgements

The first author was supported by a Marie Curie Early Stage Research fellowship in the project GLoRiClass (MEST-CT-2005-020841). He would also like to thank his supervisor Krzysztof Apt for tireless encouragement and inspiration, and the participants in Amsterdam's Dynamic Logic seminar for their feedback, especially Alexandru Baltag, Johan van Benthem and Daisuke Ikegami. Definition 15 was clarified in discussion with Cédric Dégremont.

[11] The notion of incomplete type from [BS02] is analogous to our notion of a belief model that is assumption-incomplete (with respect to some language). For the purposes of this paper, the differences are not important.

References

[ABM99] Carlos Areces, Patrick Blackburn, and Maarten Marx. Hybrid logic is the bounded fragment of first order logic. In R. de Queiroz and W. Carnielli, editors, *Proceedings of 6th Workshop on Logic, Language, Information and Computation, WOLLIC99*, pages 33–50, Rio de Janeiro, Brazil, 1999.

[Aum76] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[AZ07] Krzysztof R. Apt and Jonathan A. Zvesper. Common beliefs and public announcements in strategic games with arbitrary strategy sets. Manuscript, 2007.

[BB99] P. Battigalli and G. Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53:149–225, 1999.

[BdRV01] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, 2001.

[Ben79] Johan van Benthem. Minimal deontic logics. *Bulletin of the Section of Logic*, 8(1):36–42, 1979.

[Ben01] Johan van Benthem. Games in dynamic-epistemic logic. volume 53, pages 219–248. 2001. Proceedings LOFT-4, Torino.

[Ben07] Johan van Benthem. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9(1):13–45, 2007. (Erratum reprint, 9(2), 377–409).

[BFK08a] Adam Brandenburger, Amanda Friedenberg, and H. Jerome Keisler. Admissibility in games. *Econometrica*, 76(2):307 – 352, 2008.

[BFK08b] Adam Brandenburger, Amanda Friedenberg, and H. Jerome Keisler. Admissibility in games. *Econometrica*, 76:307–352, 2008.

[BK06] Adam Brandenburger and H. Jerome Keisler. An impossibility theorem on beliefs in games. *Studia Logica*, 84(2):211–240, November 2006.

[Boa02] Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49:49–80, 2002.

[Bon02] Giacomo Bonanno. Modal logic and game theory: Two alternative approaches. *Risk, Decision and Policy*, 7(3):309–324, November 2002.

[BOR05] Johan van Benthem, Sieuwert van Otterloo, and Olivier Roy. Preference logic, conditionals, and solution concepts in games. PP-2005(28), 2005.

[Bra03] Adam Brandenburger. *On the Existence of a "Complete" Possibility Structure*, pages 30–34. 2003.

[Bra07] Adam Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35(4):465–492, April 2007.

[BS02] Peirpaolo Battigalli and Marciano Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106:356 – 391, 2002.

[BSZ08] Alexandru Baltag, Sonja Smets, and Jonathan Zvesper. When all is done but not (yet) said: Dynamic rationality in extensive games. To be presented at Logic and Intelligent Interaction (ESSLLI workshop), 2008.

[Cat05] Balder ten Cate. *Model theory for extended modal languages*. PhD thesis, University of Amsterdam, 2005. ILLC Dissertation Series DS-2005-01.

[dB04] B. de Bruin. *Explaining Games: On the Logic of Game Theoretic Explanations*. PhD thesis, University of Amsterdam, 2004.

[Fef68] Solomon Feferman. Persistent and invariant formulas for outer extensions. *Compositio Mathematica*, 20:29–52, 1968.

[Har67] J. C. Harsanyi. Games with incompletete information played by bayesian players parts I-III. *Management Sciences*, 14, 1967.

[Hin62] Jaakko Hintikka. *Knowledge and Belief*. Cornell University Press, 1962.

[HL01] Joseph Halpern and G Lakemeyer. Multi-agent only knowing. *Journal of Logic and Computation*, 11(1):41–70, 2001.

[HM06] Joe Halpern and Yoram Moses. Characterizing solution concepts in games using knoweldge-based programs. arXiv:cs/0610098v1, 2006.

[Hum87] I. L. Humberstone. The modal logic of all and only. *Notre Dame Journal of Formal Logic*, 28:177–188, 1987.

[Lev90] Hector J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42(2-3):263–309, 1990.

[MMP05] Thomas Mariotti, Martin Meier, and Michele Piccione. Hierarchies of beliefs for compact possibility models. *Journal of Mathematical Economics*, 41(303 - 324), 2005.

[Pac07] Eric Pacuit. Understanding the Brandenburger-Keisler paradox. *Studia Logica*, 86(3), 2007.

[Zve07] Jonathan A. Zvesper. The Brandenburger-Keisler paradox in normal modal logics. Manuscript, 2007.

A Single-sorted Models

The "belief models" of [BK06] are not standard in the epistemic logic literature. There is another more standard class of models used in epistemic logic, which are mentioned in [BK06] as "[a]nother approach". The obvious way of defining assumption-completeness turns out to be much too strong: not even extremely un-expressive languages would have assumption-complete models in this sense, on the condition that the players are "introspective".[12] We will point out that, on the other hand, if the players *are not* introspective, then the sentence (BK) can be satisfiable. So, since it is central to the proof of Theorem 4 that (BK) be unsatisfiable on the relevant class of models, the impossibility theorem *might* fail for non-introspective models.

In any case, it is usually assumed that players *are* introspective, so we will also look at a different definition of assumption-completeness for single-sorted models. This second definition is more faithful to the original definition of assumption-completeness for belief models: When we translate a belief model into a single-sorted model, the resulting single-sorted model is assumption-complete in the second sense if and only if the original belief model was assumption-complete.

Single-sorted models consist of a non-empty set of "possible worlds" Ω and a relation $R_i \subseteq \Omega \times \Omega$ for each player. In the two-player case, a "*frame*" is a list (Ω, R_a, R_b). In belief models there was additionally some information (in the P^{α}'s) about strategies and so forth; since we are now in a modal framework, we will encode this using a set $\Phi = \Phi^a \cup \Phi^b$ of proposition letters, and adding a *valuation* $V : \Phi \to \mathfrak{p}(\Omega)$: a ***single-sorted model*** is a frame equipped with a valuation. (We will henceforth sometimes refer to single-sorted models simply as "models".) The idea behind dividing the proposition letters into two sets is that some pertain to Ann (for example, those saying which strategy she plays), and some to Bob. So in the case when we want a model for an interactive model, we will want a proposition letter for each player i and each strategy P^{α} That is:

$$\Phi^i = \{p_i^{\alpha} \mid \alpha \in \mathbb{N}\}$$

We say that player i is *consistent* and ***introspective*** just if the following conditions hold:

(D_i) $R_i(s) \neq \emptyset$

(4_i) sRt & $tRu \Rightarrow sRu$.

(5_i) sRt & $sRu \Rightarrow tRu$.

If a model (or frame) satisfies all of these properties for all players, then we say that it is "KD45". It is straightforward (though not *entirely* trivial) to translate in a meaningful way between belief models and single-sorted models. That is, to give a pair of functions (ρ, τ) with ρ taking a single-sorted model and returning an "equivalent" belief model, and τ taking a belief model and returning an equivalent single-sorted model. We give the details of such a translation below:

[12] A player i is *introspective* just if when i believes φ she believes that she believes it, and when she does not believe it she believes that she does not believe it. Introspection is usually taken for granted (often tacitly) in formulating epistemic models used in game theory

Definition 14 (Translation Belief Models to Single-Sorted Models)**.**

$$\tau(U^a, U^b, R^a, R^b, \{P^\alpha\}_{\alpha\in\mathbb{N}}) = (U^a \times U^b, R_a, R_b, V),$$

where:

$$(u_a, u_b)R_i(u'_a, u'_b) \Leftrightarrow (u_i = u'_i \ \& \ u_i R^i u'_j),$$

$$V(p^\alpha_a) = (P^\alpha \cap U^a) \times U^b,$$

&

$$V(p^\alpha_b) = U^a \times (P^\alpha \cap U^b).$$

Definition 15 (Translation from Single-Sorted Models to Belief Models)**.**

$$\rho(\Omega, R_a, R_b, V) = (U^a, U^b, R^a, R^b, \{P^\alpha\}_{\alpha\in\mathbb{N}}),$$

where:

First define for $i \in \{a, b\}$ the equivalence relation $\sim_i$:

$$s \sim_i t \Leftrightarrow (R_i(s) = R_i(t) \& s \sim_{V_i} t)$$

Then let $U^i = \Omega_{/\sim_i}$, and say:

$$[s]_i R^i [u]_j \Leftrightarrow \exists v \in [u] : sR_i v.$$

Finally, let:

$$P^\alpha = \{[s]_i \mid s \in V(p^\alpha_i)\}$$

How are we to define assumption-completeness for models? The naïve approach would be to say that a model $\mathcal{S}$ is assumption-complete for $\mathcal{L} \subseteq \mathfrak{p}(\Omega)$ just if for any $X \subseteq \mathcal{L}$, there is $\omega \in \Omega$ such that $R_a(\omega) = X$, and similarly for b. However, it is not difficult to see that this is not an innocent approach. For then even very simple languages are not assumption-complete:

Fact 16. *For the definition of assumption-completeness just proposed, any language that is closed under unions would be assumption-incomplete with respect to KD45 models.*

Furthermore, if we were to take this as a sign that we should not be working in KD45 models, here is another fact:

Fact 17 ([Zve07])**.** *The formal translation of the sentence (BK) is satisfiable if D_a or 4_a or 5_a does not hold.*

That is, if introspection fails then the sentence (BK) is consistent. It might seem puzzling that an informal argument is given in [BK06, Section 1] to the effect that (BK) is not satisfiable, an argument where the word "introspection" is never used, nor is any concept like it employed. It turns out that corners were cut in the informal argument. eliminating needed talk introspection; the threads of the argument are unpicked in [Zve07].

(BK) cannot hold in belief models because in belief models there is an implicit assumption of introspection:

Fact 18. *For any belief model $\mathcal{M}$, $\tau(\mathcal{M})$ is KD45.*

We will now give a proper definition of assumption-completeness for models. We write $s \sim_{V_i} t$ to mean that s and t have the same propositional valuation with respect to Φ^i, i.e.:

$$\forall p \in \Phi^i,\ \ s \in V(p) \Leftrightarrow t \in V(p)$$

Definition 19. *A model $\mathcal{S} = (\Omega, R_a, R_b, V)$ is* **assumption-complete** *for a language $\mathcal{L} \subseteq \mathfrak{p}(\Omega)$ just if for any $X \in \mathcal{L}$, for $\{i,j\} = \{a,b\}$, there exists $y \in \Omega$ such that the following two conditions hold:*

- $\forall x \in X, \exists v \in R_i(y) : R_j(v) = R_j(x)$ *&* $v \sim_{V_j} x$*;*
- $\forall v \in R_i(y), \exists x \in X : R_j(v) = R_b(x)$ *&* $v \sim_{V_j} x$*.*

We say that $\mathcal{S}$ is assumption complete *tout court* when it is assumption complete for a and for b.

Definition 19 might seem more long-winded, but it is equivalent to the definition for belief models, in the following sense:

Theorem 20. *Any belief model $\mathcal{M}$ is assumption-complete iff $\tau(\mathcal{M})$ is assumption-complete. And any model $\mathcal{S}$ is assumption-complete iff $\rho(\mathcal{S})$ is assumption-complete.*

Thus we have found the "correct" definition of assumption-completeness for single-sorted models.