There are total assumption-complete belief models for the basic modal language (DRAFT)

Jonathan A. Zvesper

March 2007
revised May 2008

1 Introduction

The notion of an *assumption* in epistemic logic is used by Brandenburger and Keisler [BK06] to reveal "a kind of basic limitation in the analysis of games" (op. cit.). An agent's assumption is her strongest belief: a belief that implies all her other beliefs. The formal translation of the sentence (BK)

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

is used in [BK06] to demonstrate limits on the cognitive powers of the agents being modeled (Ann and Bob). Specifically [BK06, Theorem 5.4] states that the agents cannot have access to a full first-order language description of the belief model.

The question arises: to which fragments of the full first-order language *can* the players have access? Brandenburger and Keisler define what we will call *assumption-completeness* of a belief model for a language $\mathcal{L}$ in order to formalise the notion that players have access to $\mathcal{L}$. Furthermore, [BK06, Theorem 10.4] states that there *are* complete belief models for a positive fragment $\mathcal{L}_1^+$ of first-order logic, i.e. one not closed under complementation.

The more expressive a language, the more "likely" it is that it will be able to express a problematic sentence like (BK), and therefore not have complete belief models, i.e. be beyond the access of the modeled agents. How expressive is the positive fragment $\mathcal{L}_1^+$? It cannot express even for example a sentence like "Ann considers it possible that Bob will choose strategy s", or "Ann does not believe that Bob will choose strategy s".[1]

[1]Note however that is *is* possible to express some forms of *rationality* using positive first-order formulae, cf. [AZ07]. However, many forms of rationality (notably those in-

What about languages that *do* have this sort of expressivity? This is the question that we address here. Specifically, we will show (Theorem 6 below) that there *are* complete belief models for the *basic modal language*, which is built by adding belief modalities for Ann and Bob to propositional logic. We also briefly discuss some expressive extensions of the basic modal language, situated between it and the full first-order language.

2 Preliminaries

We recall some definitions from [BK06].

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

$$\mathfrak{U} = (U, U^a, U^b, R, R^a, R^b, \{P^\alpha\}_{\alpha\in\lambda}), \text{where}$$

1. $U^a \neq \emptyset \neq U^b$;
2. $U^a \cup U^b = U$, $U^a \cap U^b = \emptyset$;
3. $R = R^a \cup R^b$;
4. $R^a \subsetneq U^a \times U^b$, $R^b \subsetneq U^b \times U^a$;
5. *for every* $u \in U^a$ *there is a* $v \in U^b$ *such that* $uR^a v$, *and similarly for every* $v \in U_b$;
6. λ *is an arbitrary ordinal and each* $P^\alpha \subseteq U$.

Brandenburger and Keisler allow the more general case of almost arbitrary signatures for belief models. With condition (6), we restrict our attention to monadic predicates (the P^α's) partly 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; and partly because this will lead to a more natural modal language.

The elements of the domain U are called 'states', specifically those in U^a are called 'Ann states' and those in U^b 'Bob states'. The relations R^a and R^b give those states which are considered possible by Ann and by Bob respectively: $xR^a y$ means that if Ann's state is x, then she considers it possible that Bob's state might be y. Thus we say that for $x \in U^a$ and $E \subseteq U^b$, x ***believes*** E, just when $R^a(x) \subseteq E$, and stronger, that x ***assumes*** E just when $R^a(x) = E$. (We write $R^a(x)$ to mean $\{y \in U^b \mid xR^a y\}$, and we use similar terminology with b switched for a.)

The remaining conditions imposed by Definition 1 are natural: (1) says that there are Ann and Bob states, (2) that there are only Ann and Bob

volving weak dominance) are *not* expressible in the positive fragment. And perhaps more importantly, negation seems prima facie like a natural operator that should not be beyond the representational capacities of the players being modeled.

states, and that Ann states are distinct from Bob states; (3) we have included to make the later analogy with modal logic more transparent; (4) says that for both Ann and Bob there is at least one non-trivial belief state, i.e. at which some Bob or Ann state has been excluded from consideration, and (5) that at every Ann or Bob state, some Bob or Ann states are taken to be possible.

State-based models, and a defition of belief like the one given, are familiar from epistemic logic since the work of Hintikka [Hin62]. In Hintikka's work, states are called "possible worlds"; there the term is more apt than it would be in this setting, since "possible world" suggests that a total description of all relevant features of the world is given, i.e. a specification of the Ann and of the Bob state. As Brandenburger and Keisler note, the definition of assumption is essentially the *only know* operator introduced in epistemic logic by Levesque [Lev90], transposed to these models in which a state is associated with a single player.

Let the ***first-order formulae*** be those given by the following recursive definition:

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

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*** $\mathcal{L}_1$ is the set of subsets of U that are definable by a first-order sentence (similarly for Bob).

In general a ***language*** $\mathcal{L}$ (given some model) is any subset of $\mathcal{L}_1$. Given a language $\mathcal{L}$, we write $\mathcal{L}^a$ to mean $\{E \cap U^a \mid E \in \mathcal{L}\}$ and $\mathcal{L}^b$ to mean $\{E \cap U^b \mid E \in \mathcal{L}\}$. (Note that $\mathcal{L}^a \subseteq \mathcal{L} \supseteq \mathcal{L}^b$.)

Definition 2 ([BK06], definition 4.2)**.** *The belief model*

$$(U, U^a, U^b, R, R^a, R^b, \{P^\alpha\}_{\alpha\in\lambda})$$

*is **assumption-complete** for the language $\mathcal{L}$ just if*

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

That is, (Ca) ensures that every definable set of Ann-states can be assumed by Bob, and vice-versa. Described in this way, (Ca) seems like a natural condition to impose. If a language $\mathcal{L}$ has an assumption-complete belief model then we say that $\mathcal{L}$ is assumption-complete. If it does not, then we say that $\mathcal{L}$ is ***assumption-incomplete***. Brandenburger and Keisler show that for sufficiently strong (expressive) languages, (Ca) is not satisfiable. Specifically, they exploit the formal translation of the sentence (BK) in order to prove Theorem 3:

Theorem 3 ([BK06], Theorem 5.4)**.** *$\mathcal{L}_1$ is assumption-incomplete.*

A natural question then is: for what languages *are* there assumption-complete belief models? The same paper also gives a positive result (Theorem 10.4) for a *positive language* (i.e. without negation). That is, the authors show that a language that does not include negation *is* assumption-complete. This is certainly an interesting result. However, the absence of negation is clearly quite a strong restriction to make on the language. Furthermore, given that Brandenburger and Keisler claim that they are investigation what languages are "available to the players" [BK06], it would be interesting to look at languages *with* negation, which (at least seems like it) *should* be available to the players. Therefore our aim here is the investigate other fragments of first-order logic. Our main result is another positive one, specifically we will prove Theorem 4.

Theorem 4. *The* basic modal language $\mathcal{ML}$ *is assumption-complete.*

The basic modal language $\mathcal{ML}$ is the bisimulation-invariant fragment of first-order logic $\mathcal{L}_1$ (the van Benthem Characterisation Theorem [Ben76]). Since Hintikka [Hin62] it has been used to reason about knowledge and beliefs. 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:

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

In fact we will prove a stronger result than Theorem 4, adding a strengthening which is also in effect present in Brandenburger and Keisler's positive result.

Definition 5. *A belief model $(U, U^a, U^b, R, R^a, R^b, \{P^\alpha\}_{\alpha\in\lambda})$ 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 that Brandenburger and Keisler have in mind, where the P^α's represent choice of strategy, it means that for each player and each strategy, there is a state at which that player chooses that strategy. If a language has a total assumption-complete model, then we say that it is ***totally assumption-complete***.

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

Although the totality condition is never made explicit out in [BK06], it is implicit there. The proof of Theorem 10.4 actually entails the existence of *total* complete belief models for the positive fragment.[2]

3 Proof of Theorem 6

We will show how syntatic methods familiar to modal logicians [BdRV01] allow us to obtain a positive result in Brandburger and Keisler's framework. Specifically we will define a normal modal logic KBK, and then show that *its "canonical model" is a total complete belief model* for the basic modal language.

We now give a Hilbert-style proof system for KBK, i.e. using axioms and rules of inference. Along with all propositional tautologies, the following are axioms:

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

And the following are rules of inference:

$$\frac{\varphi}{\Box\varphi}\ Nec \qquad \frac{\varphi \quad \varphi \supset \psi}{\psi}\ MP \qquad \frac{\varphi(p)}{\varphi(\psi)}\ Sub$$

Now we can define the standard notion of logical ***consistency***: we say that a set of formulae Γ is consistent just when it is not possible to derive a contradiction (i.e. a formula of the form $\varphi \wedge \neg\varphi$) using only axioms, formulae from Γ and one of the rules.

The logic KBK is the smallest set of formulae containing the axioms and closed under the rules of inference. We write $\vdash \varphi$ to mean $\varphi \in \mathsf{K}BK$. An alternative characterisation of consistency for single formulae is now available: a formula φ is consistent just when $\neg\varphi$ is not a member of KBK, i.e. just when $\nvdash \neg\varphi$.

Notice that belief models form a natural semantics for these formulae; we can interpret them on a belief model as follows:

- $[\![$♀$]\!] = U^a$
- $[\![p_\alpha]\!] = P^\alpha$

[2] Furthermore, if the requirement that the model be total is dropped, then [BK06, Problem 7.7] is answerable (in particular a finite model can be constructed that is assumption-complete for the bounded fragment $\mathcal{L}_\downarrow$, introduced below).

- $[\![\neg\varphi]\!] = U - [\![\varphi]\!]$
- $[\![\varphi \wedge \psi]\!] = [\![\varphi]\!] \cap [\![\psi]\!]$
- $[\![\Box\varphi]\!] = \{\omega \in U \mid R(\omega) \subseteq [\![\varphi]\!]\}$

Inspection shows that we have just given an alternative characterisation of $\mathcal{ML}$ (defined in the previous section):

Fact 7. $\mathcal{ML} = \{[\![\varphi]\!] \mid \varphi \in \mathsf{K}BK\}$

Completeness theorems are theorems that show that a given syntactic and semantic characterisation of a set of formulae coincide. We will define the semantic turnstyle $\Vdash$ and then obtain a completeness result.[3] This is incidental to our main aim, but we will find that a familiar tool for modal logicians standardly used in completeness proofs [BdRV01] – the canonical model – is enough for us to establish Theorem 6.

Definition 8. *The* ***canonical model*** *for* K *is* $(\Omega^C, R^C, U_a^C, V^C)$*, where*

- Ω^C *is the set of all maximally consistent sets of formulae,*
- $\begin{array}{lll} \Gamma R^C \Delta & \text{iff} & \forall\psi \in \Delta, \Diamond\psi \in \Gamma \\ & \text{iff} & \forall\Box\psi \in \Gamma, \psi \in \Delta \end{array}$ *(cf. [Che80], Theorem 5.10),*
- U_a^C *is the set of all maximally consistent sets of formula containing* $\male$*.*
- $V(p) = \{\Theta \in \Omega^C \mid p \in \Theta\}$.

The canonical model is a rich structure, in the sense that it contains every possible configuration that KBK allows. Recall that the states of the canonical model are all the maximally consistent sets of formulae. Then a corollary of the following lemma (the so-called "Truth Lemma") is that ever possible (i.e. consistent) configuration of formulae is present in the model.

Lemma 9 (Truth Lemma)**.** $\varphi \in \Gamma \Leftrightarrow \Gamma \Vdash \varphi$

Importantly for our interests, we can exploit this in order to show that the canonical model is assumption-complete in the sense of Proposition 12.

First we state the completeness result that follows in a standard way from Lemma 9. We write $\Vdash \varphi$ to mean that for *any* model $\mathcal{M}$ and state u in it, $\mathcal{M}, u \Vdash \varphi$. Then we obtain:

Corollary 10 (Completeness of $\mathsf{K}BK$)**.** $\Vdash \varphi \Leftrightarrow \vdash \varphi$

[3]Completeness in this Logical sense should not be confused with assumption-completeness.

Now let us define ***typed formulae*** to be those that prove ♀ or ¬♀. Typed formulae are those that "specify" a player, that is that "say" whether the states at which they are true are Ann states or Bob states. We can get the following lemma as a corollary of Fact 7:

Lemma 11. *For every* $E \subseteq \mathcal{ML}^a$*, there is a* typed φ *such that* $E = [\![\varphi]\!]$.

Proposition 12. *Every consistent typed formula is assumed in the canonical model. I.e. for every consistent typed formula* φ*, there is some point* τ_φ *in the canonical model such that* $R(\tau_\varphi) = [\![\varphi]\!]^C$.

Proof. Take a consistent typed formula φ. Without loss of generality, suppose that $\varphi \vdash ♂$.

Consider the set $\Pi_\varphi = \{\Diamond\gamma | \nvdash \neg(\varphi \wedge \gamma)\}$ of 'possibilities' for φ.

Lemma 13. $\Pi_\varphi \cup \{\Box\varphi\}$ *is consistent.*

Proof. We appeal to completeness (Corollary 10) and to the invariance of basic modal formulae under disjoint unions and generated submodels ([BdRV01, Propositions 2.3 and 2.6]):

Since for each $\gamma \in \Pi_\varphi$, φ and γ are pairwise consistent, then by completeness there is a pointed model $\mathcal{M}_\gamma, \omega_\gamma \vDash \varphi \wedge \gamma$. So we define a new model $\mathcal{M}$ by taking a new point ω_0, and the disjoint union of the submodels generated by each ω_γ, and stipulating that $R(\omega_0) = U_a = \{\omega_\gamma | \gamma \in \Pi_\varphi\}$. (Note that the fact that φ is *typed* ensures that the resulting structure is indeed a $\mathsf{K}BK$ model.) By construction we have $\omega_0 \vDash \Box\varphi$, because each $\omega_\gamma \vDash \varphi$. Furthermore, for each $\gamma \in \Pi_\varphi$, we also have $\omega_0 \vDash \Diamond\gamma$ because $\omega_0 R \omega_\gamma$ and $\omega_\gamma \vDash \gamma$. This simple construction is illustrated in Figure 1. ■

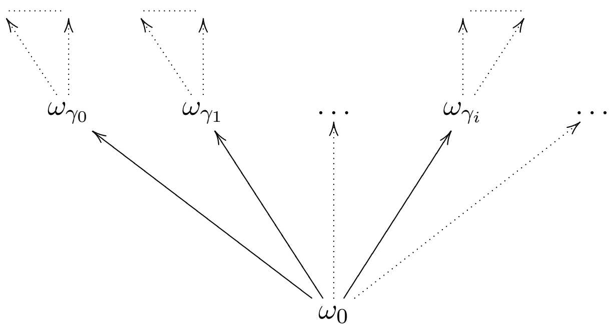

Figure 1: The construction described in Lemma 13

By Lemma 13 and Lindenbaum's lemma, we can extend $\Pi_\varphi \cup \{\Box\varphi\}$ to a maximally consistent set $\tau_\varphi \in U_a^C$. (Or we just take $\tau_\varphi = \{\psi | \omega_0 \vDash \psi\}$.)

Lemma 14. $\tau_\varphi R^C \Delta \Leftrightarrow \varphi \in \Delta$.

Proof. Immediate from construction. We give the details just to be sure:

⇒: Suppose that $\tau_\varphi R^C \Delta$. Then since $\Box\varphi \in \tau_\varphi$, by Definition 8, $\varphi \in \Delta$.

⇐: Right to left: Suppose that $\varphi \in \Delta$. Take any $\psi \in \Delta$; by Definition 8, it will suffice to show that $\Diamond\psi \in \tau_\varphi$: well, $\psi \wedge \varphi$ must be consistent, so $\Diamond\psi \in \tau_\varphi$.

■

But by Lemma 9 $[\![\varphi]\!]^C = \{\Delta \in \Omega^C \mid \varphi \in \Delta\}$, so we have shown all we need to show.

■

We then obtain Theorem 6 as a corollary of Proposition 12:

Proof. Consider the canonical model for KBK

$$(\Omega^C, R^C, U_a^C, V^C)$$

The canonical model in effect is a complete belief model: Define the belief model

$$\mathfrak{U}^C = (U, R^a, R^b, U^a, U^b, (P^\alpha)_{\alpha\in\lambda})$$

as follows:

- $A = \Omega^C$
- $U^a = U_a^C$
- $U^b = U - U^a$
- $R^a = R^C \cap (U^a \times U^b)$
- $R^b = R^C \cap (U^b \times U^a)$
- $R = R^b \cup R^c$
- $P^\alpha = V^C(p_\alpha)$

The structure $\mathfrak{U}^C$ is an assumption-complete and total belief model with respect to the basic modal language. (For each property we will only treat the case for Ann, the case for Bob being analogous can be treated the same.)

To show that $\mathfrak{U}^C$ is a belief model, the only conditions that we will consider are (4) and (5); the others are even more immediate.

For (4), we must show that $R^a \neq U^a \times U^b$. Well, we know that $p_A \wedge \Box p_0$ is consistent, and therefore that there is a point Γ in the canonical model such that $\Gamma \models p_A \wedge \Box p_0$. But also, $\neg p_A \wedge \neg p_0$ is consistent, so there is a point $\Delta \models p_A \wedge \neg p_0$. In that case, $(\Gamma, \Delta) \in U^a \times U^b$, but $(\Gamma, \Delta) \notin R^a$.

(5) is 'seriality', i.e. that for every point $\Gamma \in U^a$ there is some $\Delta \in U^b$ such that $\Gamma R^b \Delta$. This is assured by axioms D and $U1$. For suppose that there were no such Δ. Then $\Gamma \Vdash \Box a$. But by (U1), $\Gamma \Vdash \Box\neg a$, so propositional logic and normality $\Gamma \Vdash \Box\bot$, but then by D, $\Gamma \Vdash \Diamond\bot$, which is not possible.

Totality of $\mathfrak{U}^C$ is immediate: take any p_α, then it is consistent with U^a, so can be extended to a maximally consistent set $\Gamma \in U \cap P^\alpha$.

Finally for assumption-completeness. Take some $E_a \in \mathcal{ML} \cap U^a$ such that $E_a \neq \emptyset$; by Corollary 11 it is $[\![\varphi]\!]$ for some consistent typed formula φ. So by Proposition 12, there is a point Γ such that $R^C(\Gamma) = E_a$.

■

4 Beyond the basic modal language

What about other, larger fragments of first-order logic than the basic modal language?

Remark 15. *Since the only 'modal' thing we need here is preservation under disjoint unions and generated submodels, there may be something to my earlier conjecture that a similar result holds for the hybrid language with binder, $\mathcal{L}_\downarrow$ (without constants (i.e. non-variable nominals)).*[4]

Note that the 'diagonal' proposition D_A ('Ann believes that Bob's assumption is wrong') that Brandenburger and Keisler discuss is expressible in $\mathcal{L}_\downarrow$:

Define $[\![D_A]\!] = \{\omega \in U_A \mid \forall\omega' \in (R(\omega) \cap U_B),\ \omega \notin R(\omega')\}$ (D_B can also be defined symmetrically). D_A is to be read 'at this Ann state, Ann believes that Bob's assumption is wrong'.

Proposition 16. $\omega \vDash D_A \Leftrightarrow \omega \vDash p_A \wedge \downarrow x.\Box(\Diamond x \supset p_A)$

Fact 17. *$\mathcal{L}_\downarrow$ enriched with a 'global modality' U is exactly FOL*

Fact 18. *$\mathcal{ML} + D$ is strictly weaker than $\mathcal{L}_\downarrow$*

Fact 19. *$\mathcal{ML}_\downarrow$ cannot express the formal translation of (BK).*

Remark 15 and Fact 19 suggest the following:

Conjecture 20. *$\mathcal{L}_\downarrow$ is totally assumption-complete.*

As a final remark: we should also be interested in languages expressively *incomparable* with $\mathcal{L}_1$, We have focused in this note on sublanguages of $\mathcal{L}_1$, but it is commonly known for example that fixpoint operators, like those for *common knowledge*, are not first-order definable. Common knowledge is an important notion for epistemic analysis of games, so one that should presumably be taken into account in any concerns about assumption-completeness.

[4]A good presentation of hybrid modal languages is contained in [Cat05].

Acknowledgements

This paper has benefited from conversations with Eric Pacuit, who introduced the author to the phenomenon of assumption-completeness (and studied it in [Pac07]). Also the participants in Amsterdam's Dynamic Logic working sessions provided useful feedback that should feed back into a future version of this paper, particularly Johan van Benthem, Alexandru Baltag and Daisuke Ikegami.

References

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

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

[Ben76] Johan van Benthem. *Modal Correspondence Theory*. PhD thesis, Mathematisch Instituut & Instituut voor Grondslagenonderzoek, Universiteit van Amsterdam, 1976.

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

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

[Che80] Brian F. Chellas. *Modal Logic: an introduction*. Cambridge University Press, 1980.

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

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

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