Beliefs, Propositions, and Definite Descriptions

Wesley H. Holliday* Eric Pacuit†

December 4, 2016

DRAFT: Please do not circulate.

Abstract

In this paper, we introduce a doxastic logic with expressions that are intended to represent definite descriptions for propositions. Using these definite descriptions, we can formalize sentences such as:

- Ann believes that the strangest proposition that Bob believes is that neutrinos travel at twice the speed of light.
- Ann believes that the strangest proposition that Bob believes is false.

The first sentence is represented as $B_a(\gamma \text{ is } \varphi)$, where γ stands for "the strangest proposition that Bob believes" and φ stands for "that neutrinos travel at twice the speed of light". The second sentence has both de re and de dicto readings, which are distinguished in our logic. We motivate our logical system with a novel analysis of the Brandenburger-Keisler paradox. Our analysis of this paradox uncovers an interesting connection between it and the Kaplan-Montague Knower paradox.

*Department of Philosophy and Group in Logic and the Methodology of Science, University of California, Berkeley, *email*: wesholliday@berkeley.edu, *web*: wesholliday.net

†Department of Philosophy, University of Maryland, College Park, *email*: epacuit@umd.edu, *web*: pacuit.org

1 Introduction

An important feature of the formal models of belief found in the philosophical logic and game theory literature is that agents can think about each others' beliefs. The modus operandi is to assume that each agent thinks about the others' beliefs in terms of their propositional contents. For example, consider two agents, Ann and Bob, thinking about each other's beliefs. The standard approach is to assume that "Ann believes that Bob believes that it is raining" means that Ann believes that the proposition expressed by 'it is raining' is believed by Bob. There are, of course, other ways of attributing higher-order beliefs. For example, we can use propositional quantification, as in "Ann believes that there is some false proposition that Bob believes." In this paper, we are interested in higher-order belief attributions involving *definite descriptions* for propositions. For instance, propositions can be described in terms of their status within an agent's web of beliefs:

Ann believes that the *strongest* proposition that Bob believes is false. (1)

Alternatively, propositions can be described in evaluative terms:

Ann believes that the *strangest* proposition that Bob believes is false. (2)

Finally, propositions can be described by referencing a time and place:

Ann believes that what Bob was thinking *yesterday in class* is false. (3)

In each of (1)–(3), a definite description is used to describe the proposition that Ann believes. For example, in (1), the proposition that Ann believes is described in terms of its relationship to all of the other propositions that Bob believes. That is, the definite description 'the strongest proposition that Bob believes' is intended to denote the proposition that (i) is believed by Bob and (ii) entails each of the propositions believed by Bob.[1]

[1]Below we will discuss what happens if such a definite description fails to denote.

Due to the presence of the definite descriptions, the belief attributions in (1)–(3) admit both *de re* and *de dicto* readings. Suppose the strangest proposition that Bob believes is that the universe has 63 spatial dimensions. Further suppose that Ann believes that it is *false* that the universe has 63 spatial dimension—but she does not think of this in terms of Bob's beliefs. Then (2) has a true de re reading, but not a true de dicto reading. We might suggest the intended de re reading by saying "Ann believes *of* the strangest proposition that Bob believes that it is false." On the other hand, suppose a trusted source tells Ann, "The strangest proposition that Bob believes is false," and Ann accepts this, despite not having any idea of what that proposition is. Then (2) has a true de dicto reading. This may be so even if the strangest proposition that Bob believes is that the universe has 63 spatial dimensions, and Ann herself believes this proposition.

In addition to Ann having beliefs about the truth values of described propositions, she may have beliefs about what the described propositions are. For example, she might believe that the strangest proposition that Bob believes is that neutrinos travel at twice the speed of light. This will be rendered in our formal language by a formula of the form $B_a(\gamma \mathsf{\ is\ } \varphi)$ where B_a stands for 'Ann believes that', γ stands for 'the strangest proposition that Bob believes', and φ stands for 'that neutrinos travel at twice the speed of light'. In general, φ may itself contain belief operators and definite descriptions. For instance, $\gamma \mathsf{\ is\ } \varphi$ may stand for "The strangest proposition that Bob believes is that Ann believes that the strongest proposition that Bob believes is false." This opens the door to self-reference and, in turn, threatens to lead to paradoxes. For example, our language can express the following:

> The strangest proposition that Ann believes is that she does not believe that the strangest proposition that she believes is true. (4)

Given certain assumptions about Ann's beliefs (see §2), which are standardly made in the

game theory literature, the proposition expressed by (4) is a *blindspot* for Ann in the sense of Sorensen 1988.[2] That is, if the assumptions hold, then Ann cannot believe the proposition expressed by (4). The argument for this parallels the proof that the *Buridan-Burge* sentence is paradoxical [Burge, 1978, 1984, Caie, 2012, Conee, 1987, Sorensen, 1988]:

Ann does not believe this sentence is true.

In this paper, we will introduce a modal logic that can capture the reasoning about (4) that leads to inconsistency. We will argue that this inference pattern also underlies the reasoning in two other semantic paradoxes. It is not hard to see this for a version of Kaplan and Montague's Knower Paradox [1960] involving beliefs [Thomason, 1980]. It is less obvious that this reasoning plays a key role in the Brandenburger-Keisler paradox [Brandenburger and Keisler, 2006], a two-person version of Russell's paradox that plays an important role in the epistemic foundations of game theory [Pacuit and Roy, 2015]. Like the Buridan-Burge paradox and Kaplan and Montague's Knower paradox, the BK paradox shows how aspects of the semantic paradoxes can infect reasoning about knowledge or belief.

The paper is organized as follows. In §2, we discuss the Brandenburger-Keisler paradox. With this motivating example, in §3 we introduce our modal logic for reasonings about beliefs with definite descriptions for propositions. In §4, we show that this modal logic can formalize both the Knower and the Brandenburger-Keisler paradoxes, and we give a semantic perspective on the logic in §5. Finally, we conclude with a discussion of related work in §6.

2 The Brandenburger-Keisler Paradox

Brandenburger and Keisler [2006] identified a fascinating two-person variant of Russell's paradox. They used this paradox to prove a result about formal models in epistemic game

[2]We will carefully distinguish between de re and de dicto readings of (4) in §4.

theory, namely that so-called assumption-complete belief models may not exist,[3] a result which has important consequences for the epistemic characterizations of some game-theoretic solution concepts [Brandenburger et al., 2008, Battigalli and Siniscalchi, 2002]. Given the technical nature of this result, it may seem as if the Brandenburger-Keisler (BK) paradox lacks a broader significance outside of game theory. We will argue, to the contrary, that the essential idea of the BK paradox is more general and does not depend on certain features of Brandenburger and Keisler's original presentation.

The BK paradox arises for agents whose beliefs satisfy the following constraints:

(B1) *The set of propositions believed by an agent is consistent and deductively closed.*

(B2) *Everyone is correct about their own beliefs.* This means, for instances, that Ann cannot believe that she believes p while at the same time *not* believing p; and Ann cannot believe that she does *not* believe p while at the same time believing p.

(B3) *Everyone is perfectly introspective about their own beliefs.* This means, for instance, that if Ann believes p, then she believes that she believes p; and if Ann does not believe p, then she believes that she does not believe p.

Interpreted as postulates for *rational* beliefs, each of the above has been the subject of much philosophical discussion. Our general position is that even if none of the postulates hold for "normal" believers, or even "rational" believers, it should not be *impossible* to reason about believers who satisfy all of the above constraints.

The BK paradox involves two agents, Ann (a) and Bob (b), thinking about each other's beliefs. The main component of the paradox is statement (1) from §1:

Ann believes that the strongest proposition that Bob believes is false. (1)

[3]See Brandenburger and Keisler [2006], Brandenburger et al. [2008], Halpern and Pass [2009], Mariotti et al. [2005] for discussions.

Brandenburger and Keisler use the phrase 'Ann believes that Bob's assumption is false' rather than what is stated in (1). According to their semantics [Brandenburger and Keisler, 2006], 'Bob's assumption' refers to the strongest proposition that Bob believes (cf. §6).

A proper interpretation of (1) depends on the assumption that the definite description 'the strongest proposition that Bob believes' (γ) denotes a proposition. This, in turn, depends on assumptions in the underlying theory of belief content. Notably, both Stalnaker [1984] and Lewis's [1986] theories of belief content entail that there is a strongest proposition believed by an agent. Furthermore, the standard interpretation of a propositional modal language of multiagent beliefs in Kripke models presupposes the existence of such a proposition for each agent at each world. At least under these influential views, γ does denote a proposition.

Assuming that γ denotes a proposition, then (1) also expresses a proposition, which may very well be the strongest proposition that Bob believes:

> The strongest proposition that Bob believes is that
>
> Ann believes that the strongest proposition that Bob believes is false. (5)

There is nothing paradoxical about (5). One can imagine that Ann is a contrarian with respect to Bob's beliefs. She does not believe any proposition that Bob believes (except, of course, any tautologies or universal truths). If Bob believes *this* about Ann, then the strongest proposition that Bob believes may well be expressed by (1).[4]

The BK paradox arises when we imagine situations in which Ann believes the proposition

[4]More plausibly, this may well be the strongest proposition that Bob believes *about Ann*. Arguably, this reading is implicit in the framework from [Brandenburger and Keisler, 2006]. They work in a two-sorted first-order logic that distinguishes "Ann states" from "Bob states". Let us suppose that we are restricting attention to Bob's beliefs about Ann and vice versa.

expressed by (5):

> Ann believes that
>
> the strongest proposition that Bob believes is that
>
> Ann believes that the strongest proposition that Bob believes is false. (6)

Prima facie there does not seem to be anything paradoxical about (6). Indeed, Ann may believe the story above about Bob, i.e., that Bob believes that Ann is a contrarian with respect to his beliefs. The difficulty arises when we try to evaluate (1), which appears as the third line of (6). Suppose that (1) is true, so Ann believes that the strongest proposition that Bob believes is false. By (6), Ann believes that the strongest proposition that Bob believes is expressed by (1). Thus, Ann believes that it is false that Ann believes that the strongest proposition that Bob believes is false. Assuming that Ann is correct about her own beliefs (postulate B2), this means that it *is* false that Ann believes that the strongest proposition that Bob believes is false. This contradicts the assumption that (1) is true. So, (1) is false. Assuming that Ann believes that the strongest proposition that Bob believes is either true or false, it follows that Ann believes that this proposition is true. Again by (6), Ann believes that the strongest proposition that Bob believes is expressed by (1). Thus, Ann believes that she believes that the strongest proposition that Bob believes is false. Assuming Ann is correct about her beliefs (postulate B2), this means that Ann believes that the strongest proposition that Bob believes is false. This contradicts the assumption that (1) is false. Whether we assume that (1) is true or false, we deduce a contradiction.

The first observation to make is that there are variants of the BK paradox that do not make any reference to the strongest proposition that Bob believes. Indeed, other definite descriptions of propositions that Bob believes can be used in place of 'the strongest proposition

that Bob believes'. For instance, consider the following variant of (6):

> Ann believes that
>
> the strangest proposition that Bob believes is that
>
> Ann believes that the stranges proposition that Bob believes is false. (7)

The above argument can obviously be adapted to show that 'Ann believes that the strangest proposition that Bob believes is false' cannot be assigned a truth value. There are further variants that employ other definite descriptions of propositions that Bob believes. For instance, 'the boldest proposition that Bob believes', 'the most fundamental proposition that Bob believes', and 'the most interesting proposition that Bob believes' are additional examples. These variants are noteworthy because they show that the BK paradox does not rely on particular substantive (and possibly controversial) assumptions about the structure of Bob's beliefs. It may be tempting to dismiss the BK paradox on the grounds that there is no proposition that Bob believes that qualifies as the logically strongest, so the definite description fails to denote. Yet the BK paradox does not depend on the choice of this particular definite description—other definite descriptions would suffice for generating the paradox.

3 A Logical Framework

In this section, we define a propositional modal logic that captures reasoning about multiagent beliefs with definite descriptions for propositions.

Let At be a countably infinite set of atomic sentences and Agt a finite set of agents. Similar to the atomic sentences that are intended to express propositions, the language will include a set Des of definite descriptions that are intended to denote propositions. The

language $\mathcal{L}$ is the smallest set of formulas generated by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i^{dicto}\mathsf{T}(\gamma) \mid B_i^{dicto}\mathsf{F}(\gamma) \mid B_i^{re}\mathsf{T}(\gamma) \mid B_i^{re}\mathsf{F}(\gamma) \mid \gamma \ \mathsf{is} \ \varphi \mid B_i\varphi$$

where $p \in \mathsf{At}$, $i \in \mathsf{Agt}$, and $\gamma \in \mathsf{Des}$. The additional propositional connectives $(\vee, \rightarrow, \leftrightarrow)$ are defined as usual. The intended meaning of $B_i\varphi$ is that "agent i believes that φ". We use γ to denote an arbitrary element of Des. The intended meaning of $\gamma \ \mathsf{is} \ \varphi$ is "the definite description γ denotes the proposition expressed by φ". We will also say that "the γ-proposition is φ". The remaining formulas are intended to represent the different ways that i believes that the proposition denoted by γ is true/false. We will discuss this below.

A few remarks about the language are in order. First of all, the $\mathsf{T}(\cdot)$ and $\mathsf{F}(\cdot)$ notation is reminiscent of truth predicates. Indeed, one may be tempted to add the following as axiom schemes: $\mathsf{T}(\gamma) \leftrightarrow \gamma$ and $\mathsf{F}(\gamma) \leftrightarrow \neg\gamma$. However, these are not well-formed formulas in $\mathcal{L}$ for two reasons. First, by itself a definite description γ does not make a declarative statement and hence cannot be operated on by connectives. Second, both for simplicity and to avoid the issue of what the truth value of $\mathsf{T}(\gamma)$ and $\mathsf{F}(\gamma)$ should be when γ does not denote, we do not allow $\mathsf{T}(\gamma)$ and $\mathsf{F}(\gamma)$ to occur outside of belief contexts.

This explains the syntactic restriction that $\mathsf{T}(\gamma)$ and $\mathsf{F}(\gamma)$ are always preceeded by expressions B_i^{re} or B_i^{dicto}. These formulas are intended to express the following notions:

- *de re belief*: $B_i^{re}\mathsf{T}(\gamma)$ ($B_i^{re}\mathsf{F}(\gamma)$). The intended interpretation is that "i believes *of* the proposition actually denoted by γ that it is true (false)."

- *de dicto belief*: $B_i^{dicto}\mathsf{T}(\gamma)$ ($B_i^{dicto}\mathsf{F}(\gamma)$). The intended interpretation is that "i believes that γ denotes a true (false) proposition."

The final remark is that the syntax allows the is operator to be nested. For instance, $\gamma \ \mathsf{is} \ (\gamma \ \mathsf{is} \ p)$ means that, for instance, "The strangest proposition that i believes is that

the strangest proposition that i believes is p." Although liar-sentences (e.g., γ is $\neg\gamma$) are not well-formed, our syntax does allow for some self-reference. For instance, consider the following formulas:

F1. γ is $\neg B_i^{re}\mathsf{T}(\gamma)$: the γ-proposition is that i does not believe of the γ-proposition that it is true;

F2. γ is $B_i^{re}\mathsf{F}(\gamma)$: the γ-proposition is that i believes of the γ-proposition that it is false.

Note that γ need not be a definite description of a proposition that i believes (it may be referring to some proposition that another agent believes). Formula F2 is used in the BK paradox, so we discuss it in §4. Formula F1 is a formalization of (4) from §1 (let γ be the definite description "the strangest proposition that Ann believes"). As we shall see, both formulas lead to a contradiction, given standard assumptions about the logic of belief.

Remark 3.1 (The Buridan-Burge Sentence) The Buridan-Burge sentence,

Ann does not believe this sentence is true. (8)

is not directly expressible in our language. If p represents (8), then the most direct formalization is $\mathsf{T}(p) \leftrightarrow \neg B_a\mathsf{T}(p)$. However, this is not a well-formed formula since the $\mathsf{T}(\cdot)$ and $\mathsf{F}(\cdot)$ predicates can only be applied to elements of Des. However, note that $\mathsf{T}(\gamma) \leftrightarrow \neg B_a(\mathsf{T}(\gamma))$ is also not well-formed (even if we replace B_a with B_a^{re} or B_a^{dicto}). The problem is that $\mathsf{T}(\gamma)$ must be directly preceded by either B_a^{re} or B_a^{dicto}. Thus, our translation of (8) is essentially the same as the translation of (4): γ is $\neg B_a^{re}\mathsf{T}(\gamma)$ (or γ is $\neg B_a^{dicto}\mathsf{T}(\gamma)$).

We can now be more precise about the postulates governing the agents' beliefs mentioned in §2. Each group of postulates is listed below.

(B1) *The set of propositions believed by an agent is consistent and deductively closed.*

(D)	$B_i\varphi \to \neg B_i\neg\varphi$
(K)	$B_i(\varphi \to \psi) \to (B_i\varphi \to B_i\psi)$
(Nec)	if φ is a theorem, so is $B_i\varphi$

(B2) *Everyone is correct about their own beliefs.*

(CorP)	$B_iB_i\varphi \to B_i\varphi$
(CorN)	$B_i\neg B_i\varphi \to \neg B_i\varphi$

(B3) *Everyone is perfectly introspective about their own beliefs.*

(PI)	$B_i\varphi \to B_iB_i\varphi$
(NI)	$\neg B_i\varphi \to B_i\neg B_i\varphi$

We need to include special axioms to deal with the formulas expressing beliefs about γ-propositions. Note that these formulas ($B_i^{re}\mathsf{T}(\gamma)$, $B_i^{re}\mathsf{F}(\gamma)$, $B_i^{dicto}\mathsf{T}(\gamma)$, and $B_i^{dicto}\mathsf{F}(\gamma)$) describe beliefs of an agent i, so they should satisfy the correctness and introspection axioms. However, note that, for example, $B_iB_i^{re}\mathsf{T}(\gamma) \to B_i^{re}\mathsf{T}(\gamma)$ is *not* an instance of (CorP). Similarly, $B_i^{re}\mathsf{T}(\gamma) \to B_iB_i^{re}\mathsf{T}(\gamma)$ is *not* an instance of (PI). The problem is that $B_i^{re}\mathsf{T}(\gamma)$ is not a formula of the form $B_i\varphi$, which is needed to instantiate (CorP) and (PI). This means that we need to include special axioms to deal with these formulas:

(Cor_P)	$B_i\chi \to \chi$
(Cor_N)	$B_i\neg\chi \to \neg\chi$
	for each $\chi \in \{B_i^{dicto}\mathsf{T}(\gamma), B_i^{dicto}\mathsf{F}(\gamma), B_i^{re}\mathsf{T}(\gamma), B_i^{re}\mathsf{F}(\gamma)\}$

(I_P)	$\chi \to B_i\chi$
(I_N)	$\neg\chi \to B_i\neg\chi$
	for each $\chi \in \{B_i^{dicto}\mathsf{T}(\gamma), B_i^{dicto}\mathsf{F}(\gamma), B_i^{re}\mathsf{T}(\gamma), B_i^{re}\mathsf{F}(\gamma)\}$

Finally, we include axiom schemes that govern the interaction between is-expressions and formulas in the scope of belief operators. Suppose that $\gamma \in \mathsf{Des}$. If γ *in fact* denotes a proposition that is expressed by φ, then believing of the proposition denoted by γ that it is true should be equivalent to believing that φ. Similarly, believing of the proposition denoted by γ that it is false should be equivalent to believing that $\neg\varphi$. This is the analogue of the usual *replacement of equivalents* rule that is valid in all propositional modal logics. In the de dicto case, if the agent *believes* that γ denotes a proposition that is expressed by φ, then believing that γ denotes a true (resp. false) proposition should be equivalent to believing that φ (resp. $\neg\varphi$). Thus, we include the following axioms schemes for each $\gamma \in \mathsf{Des}$:

$(S1^{re})$	$(\gamma\ \mathsf{is}\ \varphi) \to (B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i\varphi)$
$(S2^{re})$	$(\gamma\ \mathsf{is}\ \varphi) \to (B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i\neg\varphi)$
$(S1^{dicto})$	$B_i(\gamma\ \mathsf{is}\ \varphi) \to (B_i^{dicto}\mathsf{T}(\gamma) \leftrightarrow B_i\varphi)$
$(S2^{dicto})$	$B_i(\gamma\ \mathsf{is}\ \varphi) \to (B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i\neg\varphi)$

4 Formalizing the Paradoxes

In this section, we demonstrate the formalizing power of the framework just introduced. Our first proposition is that the formula F1 from the previous section is inconsistent given the axioms listed above. In the derivations below, 'Prop Reasoning' means that the formula follows by propositional reasoning. This proposition is similar to a version of the Knower Paradox using beliefs (cf. Thomason, 1980; Koons, 2009; and Egré, 2005, Theorem 2.9).

Proposition 4.1 The formula γ is $\neg B_i^{re}\mathsf{T}(\gamma)$ is inconsistent in any propositional modal logic containing S1^{re}, Cor$_N$, and I$_N$.

Proof.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i\neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reasoning, 1, 2)
4.	$B_i\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor$_N$)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reasoning, 3, 4)
6.	$\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reasoning, 5)
7.	$\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i\neg B_i^{re}\mathsf{T}(\gamma)$	(I$_N$)
8.	$B_i\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reasoning, 6, 7)
9.	$B_i^{re}\mathsf{T}(\gamma)$	(Prop Reasoning, 8, 3)
10.	Contradiction	(6, 9)

QED

Remark 4.2 It is instructive to compare the derivation in the proof of Proposition 4.1 with the derivation provided for the proof of Theorem 2.9 in Egré 2005. In that proof, necessitation (Nec) is used instead of negative introspection (NI). However, (Nec) cannot be used here since γ is $\neg B_i\gamma$ is an assumption, not a theorem of the logic. For versions of the Knower paradox, the celebrated Carnap-Gödel Fixed-Point Lemma guarantees the existence of a sentence φ such that $\varphi \leftrightarrow \neg B_i\varphi$ is derivable (the trade-off is that the logic must contain a formal arithmetic strong enough to capture all primitive recursive functions).

As should be expected, essentially the same derivation shows that γ is $B_i^{re}\mathsf{F}(\gamma)$ is inconsistent. We give the derivation here to facilitate a comparison with the BK paradox.

Proposition 4.3 The formula γ is $B_i^{re}\mathsf{F}(\gamma)$ is inconsistent in any propositional modal logic containing $\mathrm{S2}^{re}$, Cor_N, and I_N.

Proof.

1.	γ is $B_i^{re}\mathsf{F}(\gamma)$	(assumption)
2.	$(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)) \rightarrow (B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$	($\mathrm{S2}^{re}$)
3.	$B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma))$	(Prop Reasoning, 1, 2)
4.	$B_i(\neg B_i^{re}\mathsf{F}(\gamma)) \rightarrow \neg B_i^{re}\mathsf{F}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{F}(\gamma) \rightarrow \neg B_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning, 3, 4)
6.	$\neg B_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning, 5)
7.	$\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow B_i \neg B_i^{re}\mathsf{F}(\gamma)$	(I_N)
8.	$B_i \neg B_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning, 6, 7)
9.	$B_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning, 2, 7)
10.	Contradiction	(6, 9)

QED

The BK paradox is related to Proposition 4.3. Suppose that $\gamma \in \mathsf{Des}$ is a definite description for Ann of a proposition that Bob believes. The starting point of the BK paradox is a statement of the following form: Ann believes that the γ-proposition is that Ann believes that the γ-proposition is false. This can be expressed in two ways in our language:

$$B_a(\gamma \text{ is } B_a^{dicto}\mathsf{F}(\gamma)) \text{ or } B_a(\gamma \text{ is } B_a^{re}\mathsf{F}(\gamma)).$$

The BK paradox does not follow from Proposition 4.3. The reason is that the formulas $B_a^{dicto}(\gamma \text{ is } B_a^{dicto}\mathsf{F}(\gamma))$ and γ is $B_a^{dicto}\mathsf{F}(\gamma)$ are logically independent (neither implies the other) in logics of belief that do not satisfy factivity, i.e., $B_i\varphi \rightarrow \varphi$, and similarly for the de re formulas. Nonetheless, we can show that both formulas $B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma))$ and $B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma))$ are inconsistent with the axioms given above.

Proposition 4.4 The formula $B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma))$ is inconsistent in any propositional modal logic containing S2^{dicto}, Cor_N, and I_N.

Proof.

1.	$B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma))$	(assumption)
2.	$B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma)) \rightarrow (B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{dicto}\mathsf{F}(\gamma)))$	(S2^{dicto})
3.	$B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{dicto}\mathsf{F}(\gamma))$	(MP, 1, 2)
4.	$B_i^{dicto}\mathsf{F}(\gamma) \rightarrow B_i(\neg B_i^{dicto}\mathsf{F}(\gamma))$	(Prop Reasoning, 3)
5.	$B_i(\neg B_i^{dicto}\mathsf{F}(\gamma)) \rightarrow B_i^{dicto}\mathsf{F}(\gamma)$	(Prop Reasoning, 3)
6.	$B_i(\neg B_i^{dicto}\mathsf{F}(\gamma)) \rightarrow \neg B_i^{dicto}\mathsf{F}(\gamma)$	(Cor_N)
7.	$B_i^{dicto}\mathsf{F}(\gamma) \rightarrow \neg B_i^{dicto}\mathsf{F}(\gamma)$	(Prop Reasoning, 4, 6)
8.	$\neg B_i^{dicto}\mathsf{F}(\gamma)$	(Prop Reasoning, 7)
9.	$\neg B_i^{dicto}\mathsf{F}(\gamma) \rightarrow B_i \neg B_i^{dicto}\mathsf{F}(\gamma)$	(I_N)
10.	$B_i \neg B_i^{dicto}\mathsf{F}(\gamma)$	(MP, 8, 9)
11.	$B_i^{dicto}\mathsf{F}(\gamma)$	(MP, 5, 10)
12.	Contradiction	(8, 11)

QED

In the informal explanation of the BK paradox in §2, we did not carefully distinguish between de dicto and de re readings of the relevant belief attributions. We can now do so. For the following derivation in the de re case, recall that the monotonicity rule (Mon) states that if $\varphi \rightarrow \psi$ is a theorem, so is $B_i\varphi \rightarrow B_i\psi$. This rule is admissible given (Nec) and (K).

Proposition 4.5 The formula $B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma))$ is inconsistent in any propositional modal logic closed under Nec and containing K, S2^{dicto}, CorP, PI, Cor_P, Cor_N, I_P, and I_N.

Proof.

1.	$B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma))$	(assumption)
2.	$(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)) \rightarrow (B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$	(S2^{re})
3.	$B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)) \rightarrow B_i(B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$	(Mon, 2)
4.	$B_i(B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$	(MP, 1, 3)
5.	$B_i(B_i^{re}\mathsf{F}(\gamma) \rightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$	(K, Nec, Prop Reasoning, 4)
6.	$B_i(B_i(\neg B_i^{re}\mathsf{F}(\gamma)) \rightarrow B_i^{re}\mathsf{F}(\gamma))$	(K, Nec, Prop Reasoning, 4)
7.	$B_iB_i^{re}\mathsf{F}(\gamma) \rightarrow B_iB_i(\neg B_i^{re}\mathsf{F}(\gamma)))$	(K, Prop Reasoning, 5)
8.	$B_iB_i(\neg B_i^{re}\mathsf{F}(\gamma))) \rightarrow B_i\neg B_i^{re}\mathsf{F}(\gamma)$	(CorP)
9.	$B_i\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow \neg B_i^{re}\mathsf{F}(\gamma)$	(Cor$_N$)
10.	$B_iB_i^{re}\mathsf{F}(\gamma) \rightarrow \neg B_i^{re}\mathsf{F}(\gamma)$	(Prop Reas, 7, 8, 9)
11.	$B_iB_i^{re}\mathsf{F}(\gamma) \rightarrow B_i^{re}\mathsf{F}(\gamma)$	(Cor$_P$)
12.	$\neg B_iB_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning 10, 11)
13.	$B_iB_i\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow B_iB_i^{re}\mathsf{F}(\gamma)$	(K, Prop Reasoning, 6)
14.	$B_i\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow B_iB_i\neg B_i^{re}\mathsf{F}(\gamma)$	(PI)
15.	$B_i\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow B_iB_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning 13, 14)
16.	$\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow B_i\neg B_i^{re}\mathsf{F}(\gamma)$	(I$_N$)
17.	$\neg B_i^{re}\mathsf{F}(\gamma) \rightarrow B_iB_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning 15, 16)
18.	$\neg B_iB_i^{re}\mathsf{F}(\gamma) \rightarrow B_i^{re}\mathsf{F}(\gamma)$	(Prop Reasoning, 17)
19.	$B_i^{re}\mathsf{F}(\gamma)$	(MP, 12, 18)
20.	$B_i^{re}\mathsf{F}(\gamma) \rightarrow B_iB_i^{re}\mathsf{F}(\gamma)$	(I$_P$)
21.	$B_iB_i^{re}\mathsf{F}(\gamma)$	(MP, 19, 20)
22.	Contradiction	(12, 21)

QED

5 A Semantic Perspective

In this section, we present a semantic perspective on the language $\mathcal{L}$ from §3 and on the formalization of the BK paradox in §4. To do so, we define a Kripke-style semantics for $\mathcal{L}$.

Definition 5.1 A **frame** for $\mathcal{L}$ is a tuple $\mathcal{F} = \langle W, \{R_i\}_{i\in\mathsf{Agt}}, \{D_\gamma\}_{\gamma\in\mathsf{Des}}\rangle$, where:

1. W is a nonempty set (the *set of worlds*);
2. for each $i \in \mathsf{Agt}$, R_i is a binary relation on W (the *accessibility relation for agent i*);
3. for each $\gamma \in \mathsf{Des}$, $D_\gamma \colon W \nrightarrow \wp(W)$ is a partial function (the *denotation function*).

A **model** based on $\mathcal{F}$ is a tuple $\langle \mathcal{F}, V\rangle$ where $V \colon \mathsf{At} \to \wp(W)$ is a valuation function. A frame (model) is called a **quasi-partition** when for each $i \in \mathsf{Agt}$, R_i is serial, transitive and Euclidean. For convenient notation, given $w \in W$, let $R_i(w) = \{v \in W \mid wR_iv\}$.

The truth of a formula $\varphi \in \mathcal{L}$ at a world w in a model $\mathcal{M} = \langle W, \{R_i\}_{i\in\mathsf{Agt}}, \{D_\gamma\}_{\gamma\in\mathsf{Des}}, V\rangle$, written $\mathcal{M}, w \models \varphi$, is defined by recursion. In what follows, let $[\![\varphi]\!]_\mathcal{M} = \{w \mid \mathcal{M}, w \models \varphi\}$. The definition of truth for the Boolean connectives and the belief modality is as usual:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$, for $p \in \mathsf{At}$;
- $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$;
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$;
- $\mathcal{M}, w \models B_i\varphi$ iff $R_i(w) \subseteq [\![\varphi]\!]_\mathcal{M}$.

For each $w \in W$ and $\gamma \in \mathsf{Des}$, when $D_\gamma(w)$ is defined, $D_\gamma(w)$ is the proposition denoted by γ. Thus, the definition of truth for formulas of the form $\gamma\ \mathsf{is}\ \varphi$ is:

- $\mathcal{M}, w \models \gamma\ \mathsf{is}\ \varphi$ iff $D_\gamma(w)$ is defined and $D_\gamma(w) = [\![\varphi]\!]_\mathcal{M}$.

This means that γ is φ is false a state w when either γ does not denote a proposition at w or the proposition denoted by γ at w is not expressed by φ. Finally, the truth definition for *de re* and *de dicto* belief formulas makes precise their distinction precise:

- $\mathcal{M}, w \models B_i^{re}\mathsf{T}(\gamma)$ iff $D_\gamma(w)$ is defined and $R_i(w) \subseteq D_\gamma(w)$;
- $\mathcal{M}, w \models B_i^{re}\mathsf{F}(\gamma)$ iff $D_\gamma(w)$ is defined and $R_i(w) \subseteq W \setminus D_\gamma(w)$;
- $\mathcal{M}, w \models B_i^{dicto}\mathsf{T}(\gamma)$ iff for all $v \in R_i(w)$, $D_\gamma(v)$ is defined and $R_i(w) \subseteq D_\gamma(v)$;
- $\mathcal{M}, w \models B_i^{dicto}\mathsf{F}(\gamma)$ iff for all $v \in R_i(w)$, $D_\gamma(v)$ is defined and $R_i(w) \subseteq W \setminus D_\gamma(v)$.

We will illustrate the above definitions by showing that the formulas from Propositions 4.4 and 4.5 are not satisfiable in quasi-partition models. That is, there is no state in a quasi-partition model that makes the formulas from Propositions 4.4 and 4.5 true. This provides a complementary semantic explanation on the BK paradox.

Proposition 5.2 If $\mathcal{M}$ is a quasi-partition model, then for all $w \in W$,

$$\mathcal{M}, w \not\models B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma)).$$

Proof. Suppose that $\mathcal{M}$ is a quasi-partition model and that for some $w \in W$, we have $\mathcal{M}, w \models B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma))$. Thus, for all $v \in W$, if wR_iv, then $\mathcal{M}, v \models \gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma)$. It follows that for all $v \in W$,

$$\text{if } wR_iv, \text{ then } D_\gamma(v) \text{ is defined and } D_\gamma(v) = [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}. \qquad (*)$$

There are two cases to consider.

1. $\mathcal{M}, w \models B_i^{dicto}\mathsf{F}(\gamma)$. Then for all $v \in W$, if wR_iv, then $D_\gamma(v)$ is defined and $R_i(w) \subseteq W - D_\gamma(v)$. Since R_i is serial, there is a v_0 such that wR_iv_0. Since R_i is transitive, we

have $R_i(v_0) \subseteq R_i(w)$. Suppose that $v \in W$ with $v_0 R_i v$. Then since wR_iv, by $(*)$ we have that $D_\gamma(v)$ is defined with $D_\gamma(v) = [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Furthermore, we have

$$R_i(v_0) \subseteq R_i(w) \subseteq W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}} = W - D_\gamma(v).$$

Thus, $\mathcal{M}, v_0 \models B_i^{dicto}\mathsf{F}(\gamma)$, i.e., $v_0 \in [\![B^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Since $v_0 \in R_i(w)$ and $R_i(w) \subseteq W - D_\gamma(v_0) = W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$, we have $v_0 \notin [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$, which is a contradiction.

2. $\mathcal{M}, w \not\models B_i^{dicto}\mathsf{F}(\gamma)$. Then there is a v_0 such that wR_iv_0 such that either $D_\gamma(v_0)$ is not defined or $R_i(w) \not\subseteq W - D_\gamma(v_0)$. By $(*)$, we have that $D_\gamma(v_0)$ is defined and $D_\gamma(v_0) = [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Thus,

$$R_i(w) \not\subseteq W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}},$$

so there is a $v_1 \in R_i(w)$ such that $v_1 \notin W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. That is, $v_1 \in [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Suppose that $v \in R_i(w)$. Then by $(*)$, $D_\gamma(v)$ is defined and $D_\gamma(v) = [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Since R_i is transitive and $v_1 \in R_i(w)$, we have that $R_i(v_1) \subseteq R_i(w)$; and so, by $(*)$, for all $v^* \in R_i(v_1)$, $D_\gamma(v^*) = [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Therefore, since $\mathcal{M}, v_1 \models B_i^{dicto}\mathsf{F}(\gamma)$, we have that $R_i(v_1) \subseteq W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Now, since R_i is Euclidean and wR_iv_1, we have $v_1 R_i v$. Hence $v \in W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Since v is an arbitrary element of $R_i(w)$, we have $R_i(w) \subseteq W - [\![B_i^{dicto}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$, which is a contradiction. QED

Proposition 5.3 If $\mathcal{M}$ is a quasi-partition model, then for all $w \in W$,

$$\mathcal{M}, w \not\models B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)).$$

Proof. Suppose that $\mathcal{M}$ is a quasi-partition model and that for some $w \in W$, we have $\mathcal{M}, w \models B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma))$. Then for all $v \in W$, if $w\ R_i\ v$, then $\mathcal{M}, v \models \gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)$. Thus, for all $v \in W$,

$$\text{if } wR_iv, \text{ then } D_\gamma(v) \text{ is defined and } D_\gamma(v) = [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}} \qquad (*)$$

There are two cases to consider.

1. $\mathcal{M}, w \models B_iB_i^{re}\mathsf{F}(\gamma)$. Then $R_i(w) \subseteq [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Since R_i is serial, there is a v_0 such that wR_iv_0, so $\mathcal{M}, v_0 \models B_i^{re}\mathsf{F}(\gamma)$. Let $v \in W$ with v_0R_iv (such a state exists since R_i is serial). Then $\mathcal{M}, v_0 \models B_i^{re}\mathsf{F}(\gamma)$ implies $v \notin D_\gamma(v_0)$. By $(*)$, we have $D_\gamma(v_0) = [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Thus, $\mathcal{M}, v \not\models B_i^{re}\mathsf{F}(\gamma)$. This implies that either $D_\gamma(v)$ is not defined or there is a $v' \in W$ such that $v' \in D_\gamma(v)$. Since R_i is transitive, we have wR_iv. By $(*)$, this implies that $D_\gamma(v)$ is defined. Thus, $v' \in D_\gamma(v) = [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Since R_i is transitive, v_0R_iv, and vR_iv', we have v_0R_iv' with $v' \in [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}} = D_\gamma(v_0)$. This contradicts the fact that $\mathcal{M}, v_0 \models B_i^{re}\mathsf{F}(\gamma)$.

2. $\mathcal{M}, w \not\models B_iB_i^{re}\mathsf{F}(\gamma)$. Then there is a $v_0 \in W$ such that wR_iv_0 and $\mathcal{M}, v_0 \not\models B_i^{re}\mathsf{F}(\gamma)$. Thus, either $D_\gamma(v_0)$ is not defined or there is a $v_1 \in W$ with $v_0R_iv_1$ and $v_1 \in D_\gamma(v_0)$. By $(*)$, $D_\gamma(v_0)$ is defined with $D_\gamma(v_0) = [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Thus, $\mathcal{M}, v_1 \models B_i^{re}\mathsf{F}(\gamma)$. Since R_i is transitive with wR_iv_0 and $v_0R_iv_1$, we have wR_iv_1. By $(*)$, $D_\gamma(v_1)$ is defined with $D_\gamma(v_1) = [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$. Since R_i is Euclidean and wR_iv_1, we have $v_1R_iv_1$. Since $\mathcal{M}, v_1 \models B_i^{re}\mathsf{F}(\gamma)$, we have $v_1 \notin D_\gamma(v_1)$. Thus, $v_1 \notin [\![B_i^{re}\mathsf{F}(\gamma)]\!]_{\mathcal{M}}$, a contradiction. QED

Standard results about our logical system (e.g., completeness and decidability) will be discussed in the full version of the paper. We conclude this section by highlighting some interesting features of the system.

The first observation is that the axiom schemes from §3 are all valid on any quasi-partition frame. Recall that a formula is **valid** over a frame $\mathcal{F} = \langle W, \{R_i\}_{i\in\mathsf{Agt}}, \{D_\gamma\}_{\gamma\in\mathsf{Des}}\rangle$ provided that for every model $\mathcal{M} = \langle \mathcal{F}, V\rangle$ and $w \in W$, $\mathcal{M}, w \models \varphi$. To simplify our notation, let $\hat{B}_i\varphi$ be shorthand for $\neg B_i\neg\varphi$. Then in any model $\mathcal{M} = \langle W, \{R_i\}_{i\in\mathsf{Agt}}, \{D_\gamma, V\}_{\gamma\in\mathsf{Des}}\rangle$, we

have that $\mathcal{M}, w \models \hat{B}_i\varphi$ iff $R_i(w) \cap [\![\varphi]\!]_{\mathcal{M}} \neq \emptyset$. Our second observation is that the following formulas are valid on any quasi-partition frame:

F3. $\hat{B}_i(\gamma \text{ is } p) \rightarrow (B_i^{dicto}\mathsf{T}(\gamma) \rightarrow B_i p)$;

F4. $\hat{B}_i(\gamma \text{ is } p) \rightarrow (B_i^{dicto}\mathsf{F}(\gamma) \rightarrow B_i \neg p)$.

We sketch the proof that F3 is valid (the proof for F4 is similar). Suppose that $\mathcal{M}$ is a quasi-partition models and that $\mathcal{M}, w \models \hat{B}_i(\gamma \text{ is } p)$. Then there is a $v \in R_i(w)$ such that $D_\gamma(v)$ is defined and $D_\gamma(v) = [\![p]\!]_{\mathcal{M}}$. If $\mathcal{M}, w \models B_i^{dicto}\mathsf{T}(\gamma)$, then for all $v \in R_i(w)$, $D_\gamma(v)$ is defined and $R_i(w) \subseteq D_\gamma(v)$. Thus, since $v \in R_i(w)$, we have $R_i(w) \subseteq [\![p]\!]_{\mathcal{M}}$. It is also instructive to note that the following formulas are *not* valid:

F5. $(\gamma \text{ is } p) \rightarrow (B_i^{dicto}\mathsf{T}(\gamma) \leftrightarrow B_i p)$;

F6. $(\gamma \text{ is } p) \rightarrow (B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i \neg p)$.

To see that F5 is not valid (the proof that F6 is not valid is similar), let $\mathcal{M}$ be a model where $W = \{w, v_1, v_2\}$; $R_i = \{(w, v_1), (w, v_2), (v_1, v_2)\}$; $D_\gamma(w) = \{v_1\}$, $D_\gamma(v_1) = \{v_1, v_2\}$, and $D_\gamma(v_2) = \{w, v_1, v_2\}$; and $V(p) = \{v_1\}$. We have (i) $\mathcal{M}, w \models \gamma \text{ is } p$ since $D_\gamma(w) = \{v_1\} = [\![p]\!]_{\mathcal{M}}$; (ii) $\mathcal{M}, w \not\models B_i p$ since $R_i(w) = \{v_1, v_2\} \not\subseteq [\![p]\!]_{\mathcal{M}}$; and (iii) $\mathcal{M}, w \models B_i^{dicto}\mathsf{T}(\gamma)$ since $R_i(w) \subseteq D_\gamma(v_1) = \{v_1, v_2\}$ and $R_i(w) \subseteq D_\gamma(v_2)$.

These observations illustrate some of the relationships between the belief operator, the *de dicto* belief operators and the is-operator. These and other relationships add interest to the question of axiomatization discussed in the full version of this paper.

6 Related Literature

The original formulation of the BK paradox from Brandenburger and Keisler [2006] is:

> Ann believes that Bob *assumes* that
>
> Ann believes that Bob's *assumption* is false. (9)

To formalize (9), they add a modality to a multi-agent propositional modal language that is intended to represent "Bob assumes that...". Formally, the language, $\mathcal{L}^{BK}$, is the smallest set of formulas generated by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i\varphi \mid \boxplus_i \varphi.$$

Formulas from $\mathcal{L}^{BK}$ are interpreted in quasi-partition Kripke models $\langle W, \{R_i\}_{i\in\mathsf{Agt}}, V\rangle$. Truth for the atomic propositions, Boolean connectives, and belief modalities are defined as in §5. Truth for the assumption modality is defined as follows:

- $\mathcal{M}, w \models \boxplus_i\varphi$ iff $R_i(w) = [\![\varphi]\!]_{\mathcal{M}}$.

Thus, $\boxplus_i\varphi$ is true at w if φ defines the set of states $R_i(w)$.[5] Note that in a multi-agent Kripke model, $R_i(w)$ is the strongest proposition that i believes at w.

Suppose that $\mathcal{M} = \langle W, \{R_a, R_b\}, V\rangle$ is a quasi-partition Kripke model for the two agents, Ann (a) and Bob (b). Let d be an atomic proposition with

$$V(d) = \{x \mid R_a(x) \subseteq \{y \mid x \notin R_b(y)\}\}.$$

[5]This modality has been investigated by a number of different authors. Most notably, Humberstone [1987] provides an axiomatization and Passy and Passy [1991] discuss this modality as part of a larger discussion of Boolean Modal Logic [cf. Blackburn et al., 2001, Section 7.1]. See also Halpern and Lakemeyer [2001] on the "all I know" operator.

The intended meaning of d is that "Ann believes that Bob's assumption is false." Then, $B_a \boxplus_b d$ expresses (9). The main result from Brandenburger and Keisler [2006] is that it is impossible satisfy both (9) and that Ann's beliefs are not inconsistent, i.e., $\neg B_a \bot$ is true. More formally, it is shown that in any quasi-partition model $\mathcal{M}$:

$$\text{If } \mathcal{M}, w \models B_a \boxplus_b d \wedge \neg B_a \bot, \text{ then } \mathcal{M}, w \models d \leftrightarrow \neg d.$$

The other studies of this paradox [Mariotti et al., 2005, Pacuit, 2007, Abramsky and Zvesper, 2015, Baskent, 2015] build on the above formalization. The take-home message from this literature is that the difficulty stems from the interaction between Ann's belief (the B_a modality) and Bob's assumption (the $\boxplus_b$ modality). Our analysis of the paradox is different.

We argue that the central issue raised by the BK paradox is the use of a definite description to denote the proposition that Ann believes. Indeed, our formalization of the paradox only involves Ann's beliefs (see the formulas in Propositions 4.4 and 4.5). In the original formulation of the paradox, the definite description γ does denote a proposition that is believed by Bob; however, nothing changes if we assume that γ denotes a proposition believed by Ann. Thus, our analysis de-emphasizes the multi-agent aspect of the paradox.

Our approach to formalizing the BK paradox raises a number of issues that are also discussed in Caie 2012. The most relevant to this paper is the normative paradox from Section 2 in Caie 2012. This paradox is derived from the assumption that there is a sentence β that *names* the sentence $\neg B_a \mathsf{T}(\beta)$, where T is a truth predicate (i.e., β names "Ann does not believe that this sentence is true").[6] Then an instance of the truth-schema is:[7]

$$(*) \qquad \mathsf{T}(\beta) \leftrightarrow \neg B_a \mathsf{T}(\beta).$$

[6] Caie *stipulates* that there exists such a sentence (following Kripke [1975]), but notes that one can construct such a sentence using Gödel numbering. See Gaifman [2006] for a nice perspective on this.

[7] As we mentioned in Remark 3.1, this formula is not well-formed in the logic defined in §3.

Suppose that Ann believes $(*)$. This means that the following two statements about Ann's beliefs are true:

- $B_a(B_a\mathsf{T}(\beta) \rightarrow \neg\mathsf{T}(\beta))$: Ann believes that if she believes that β is true, then β is not true.
- $B_a(\neg B_a\mathsf{T}(\beta) \rightarrow \mathsf{T}(\beta))$: Ann believes that if she does not believe that β is true, then β is true.

Now suppose that Ann believes that β is true ($B_a\mathsf{T}(\beta)$). Assuming that she is perfectly introspective and correct about her own beliefs (see §2) and that she has no other evidence that bears on the truth of β, then Ann is in a position in which her evidence makes her certain that β is not true. Similarly, we can argue that if Ann does not believe that β is true ($\neg B_a\mathsf{T}(\beta)$), then she is in a position in which her evidence makes her certain that β is true. Caie is led to a paradox by assuming that Ann's beliefs are consistent and assuming the following postulate about the relationship between evidence and beliefs:

> EVIDENCE: For any proposition X, if an agent's evidence makes X certain, then the agent is rationally required to believe X [Caie, 2012, p. 5].

Our analysis in this paper is related to a version of this paradox using the following propositional analogue of the sentence β [Caie, 2012, Section 2.2]:

X Ann does not believe the proposition expressed by X.

Caie derives a contradiction as follows. Let $\rho(X)$ be *the proposition expressed by* X.[8] Now, X and '$\neg B_a\rho(X)$' name the same sentence. So we must have

$$(**) \qquad \rho(X) = \rho\text{`}\neg B_a\rho(X)\text{'}.$$

[8]We use $[\![\varphi]\!]_{\mathcal{M}}$ to denote the set of worlds in $\mathcal{M}$ that make φ true (i.e., the *truth-set* of φ in $\mathcal{M}$). So for a formula φ and model $\mathcal{M}$, "the proposition expressed by φ", written as $\rho(\varphi)$, is $[\![\varphi]\!]_{\mathcal{M}}$.

Adapting the assumptions that Ann is perfectly introspective and correct about her own beliefs gives us the following:

P1. $B_a\rho(X) \leftrightarrow B_a\rho(\text{`}B_a\rho(X)\text{'})$; and

P2. $\neg B_a\rho(X) \leftrightarrow B_a\rho(\text{`}\neg B_a\rho(X)\text{'})$.

It is not hard to derive a contradiction from $(**)$, P1 and P2. For instance, suppose that $\neg B_a\rho(X)$ is true. By P2, this is equivalent to $B_a\rho(\text{`}\neg B_a\rho(X)\text{'})$ being true. By $(**)$ and replacement of equals, this means that $B_a\rho(X)$ is true, a contradiction. Note that this contradiction is derived without appeal to EVIDENCE. However, Caie argues that "[c]hanging a conditional linking the truth-values of propositions to an identity between propositions has the same effect as assuming conformity to EVIDENCE" [Caie, 2012, p. 12].

We work at an intermediate level between assuming the existence of a sentence that names another sentence as in $(*)$ and reasoning directly about propositions as in $(**)$. An interesting question for future work is whether Caie's solution to the above paradox can handle our version of the BK paradox.[9]

We conclude by briefly mentioning a logical system touching on issues raised in this paper. Halpern and Kets [2014] introduce an epistemic logic in which agents may disagree about their interpretation of atomic propositions. We also have the resources to represent disagreement about the interpretation of an atomic formula. In our logic, we can represent this type of disagreement using elements of Des. For instance, suppose that γ_a denotes "the proposition expressed by p, according to Ann" (i.e., "Ann's interpretation of p") and γ_b

[9]Caie's solution to the above paradox is to "restrict the law of excluded-middle for certain claims about whether or not [an agent] believes certain propositions." However, there still remains a quesiton about the relevance of Caie's solution to the original game-theoretic motivation for the BK paradox. We do not discuss applications to the epistemic foundations of game theory in this paper.

denotes "the proposition expressed by p, according to Bob" (i.e., "Bob's interpretation of p"). It is an interesting question for future research to compare Halpern and Kets's [2014] epistemic logic with ambiguity to our doxastic logic with definition descriptions.

References

Samson Abramsky and Jonathan A. Zvesper. From Lawvere to Brandenburger-Keisler: Interactive forms of diagonalization and self-reference. *Journal of Computer and System Sciences*, 81:799 – 812, 2015.

Can Baskent. Some non-classical approaches to the Brandenburger-Keisler paradox. *Logic Journal of the IGPL*, 23(4):533 – 552, 2015.

Pierpaolo Battigalli and Marciano Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356 – 391, 2002.

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

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

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

Tyler Burge. Buridan and epistemic paradox. *Philosophical Studies*, 34:21 – 35, 1978.

Tyler Burge. Epistemic paradox. *Journal of Philosophy*, 81(1):5 – 29, 1984.

Michael Caie. Belief and indeterminacy. *The Philosophical Review*, 121(1):1 – 54, 2012.

Earl Conee. Evident, but rationally unacceptable. *Australasian Journal of Philosophy*, 65: 316 – 326, 1987.

Paul Egré. The knower paradox in the light of provability interpretations of modal logic. *Journal of Logic, Language and Information*, 14:13 – 48, 2005.

Heim Gaifman. Naming and diagonalization, from Cantor to Gödel to Kleene. *Logic Journal of the IGPL*, 14(5):709 – 728, 2006.

Joseph Y. Halpern and Willemien Kets. A logic for reasoning about ambiguity. *Aritifical Intelligence*, 209(1 - 10), 2014.

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

Joseph Y. Halpern and Rafael Pass. A logical characterization of iterated admissibility. In *Proceedings of Twelfth Conference on Theoretical Aspects of Rationality and Knowledge*, pages 146 – 155, 2009.

I. Lloyd Humberstone. The modal logic of 'all and only'. *Notre Dame Journal of Formal Logic*, 1987.

David Kaplan and Richard Montague. A paradox regained. *Notre Dame Journal of Formal Logic*, 1(3):79–90, 1960.

Robert C. Koons. *Paradoxes of Belief and Strategic Rationality*. Cambridge University Press, 2009.

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 19:690 – 716, 1975.

David Lewis. *On the Plurality of Worlds*. Basil Blackwell, Oxford, 1986.

Thomas Mariotti, Martin Meier, and Michele Piccione. Hierarchies of beliefs for compact possibility models. *Journal of Mathematical Economics*, 41:303 – 324, 2005.

Eric Pacuit. Understanding the Brandenburger-Keisler paradox. *Studia Logica*, 86(3):435 – 454, 2007.

Eric Pacuit and Olivier Roy. Epistemic foundations of game theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2015 edition, 2015.

Solomon Passy and Tinko Tinchev Passy. An essay in combinatory dynamic logic. *Information and Computation*, 93:263 – 332, 1991.

Roy Sorensen. *Blindspots*. Oxford University Press, 1988.

Robert C. Stalnaker. *Inquiry*. MIT Press, Cambridge, Mass., 1984.

Richmond H. Thomason. A note on syntaactical treatments of modality. *Synthese*, 44(3): 391 – 395, 1980.