*Log_A*B: An Algebraic Logic of Belief

Haythem O. Ismail

German University in Cairo
Department of Computer Science
haythem.ismail@guc.edu.eg

Abstract. Log_A**B** is a family of logics of belief. It holds a middle ground between the expressive, but prone to paradox, syntactical first-order theories and the often inconvenient, but safe, modal approaches. In this report, the syntax and semantics of Log_A**B** are presented. Log_A**B** is algebraic in the sense that it is a language of only terms; there is no notion of a formula, only proposition-denoting terms. The domain of propositions is taken to be a Boolean lattice, which renders classical truth conditions and definitions of consequence and validity theorems about Log_A**B** structures. Log_A**B** is shown to be sufficiently expressive to accommodate complex patterns of reasoning about belief while remaining paradox-free. A number of results are proved regarding paradoxical self-reference. They are shown to strengthen previous results, and to point to possible new approaches to circumventing paradoxes in syntactical theories of belief.

1 Introduction

Belief is usually viewed as a relation between a believing agent and a believed entity, typically a proposition or a sentence. Logics of belief come in two main flavors: the modal and the syntactical. Modal approaches [1–4, for instance] represent belief by a modal operator and employ some version of possible-worlds semantics. Syntactical theories [5–9, for instance] employ self-referential first-order languages, where belief is represented by a (typically) dyadic predicate of agents and sentences of the language. The semantics is standard Tarskian semantics, but complications arise due to the need to employ theories of arithmetic or string manipulation. On one hand, first-order logics are more expressive and more well-understood than their modal rivals. On the other hand, a result by Thomason [10] (following a similar result by Montague for the case of knowledge [11]) shows that, assuming some desirable properties of belief, first-order doxastic theories are paradoxical, whereas modal ones are not.

In this paper, I present Log_A**B**, a family of algebraic logics of belief. Log_A**B** is algebraic in the sense that it only contains terms, algebraically constructed from function symbols. No sentences are included in a Log_A**B** language. Instead, there are terms of a distinguished syntactic type that are taken to denote propositions. The inclusion of propositions in the ontology, though non-standard, has been suggested by a few authors [6, 12, for instance]. I refer the reader to Shapiro's article [12] for arguments in favor of adopting this approach in the representation of propositional attitudes in artificial intelligence. It turns out that, in addition to Shapiro's arguments, recognizing propositions as first-class inhabitants of our ontology has the additional benefit of avoiding the doxastic paradoxes referred to above. In particular, Log_A**B** holds a middle ground between modal and first-order syntactical theories of belief. On one hand, it is almost as expressive as the first-order theories; on the other hand, it is weak just enough to avoid the paradoxes to which those theories are susceptible.

Chalupsky and Shapiro [13] present a logic of belief, **SL**, based on Shapiro's proposal. Log_A**B** and **SL** differ in several important respects. Chiefly among these is that **SL** is fully-intensional; it adopts an excessively fine-grained representation of propositions [13, p. 168] and has no room for notions of truth, logical consequence, and validity.[1] Log_A**B** is much closer in spirit to standard extensional first-order theories. In the Log_A**B** ontology, propositions are structured in a Boolean lattice. This gives us, almost for free, all standard truth conditions, standard notions of consequence and validity, and an individuation of propositions that is neither too fine-grained, nor too coarse-grained, for a doxastic logic.[2] Moreover, Chalupsky and Shapiro are primarily concerned with simulative belief ascription, and include no mention of doxastic paradoxes of self-reference.

The paper is organized as follows. In Section 2, the syntax and semantics of Log_A**B** are presented. Section 3 shows how an account of truth may be associated with the otherwise truth-independent semantics of Log_A**B**. Proof theory is briefly discussed in Section 4. Section 5 analyzes the notion of belief in terms of properties of Log_A**B** semantic structures. In Section 6, the expressivity of Log_A**B** is demonstrated by showing how notions of common and distributed belief (cf. [2]) may be accounted for. Results pertaining to paradoxes of self-reference are presented in Section 7: We (i) show that Log_A**B** is not susceptible to paradox, (ii) strengthen a previous result of Bolander's [9], and (iii) point out possible new approaches to circumventing paradox in syntactical theories. For completeness, an appendix includes relevant background on Boolean algebra.

2 Log_AB Languages

Log_A**B** is a class of languages that share a common core of logical symbols and differ in a signature of non-logical symbols. A Log_A**B** language is a set of terms partitioned into two base syntactic types, σ_P and σ_I. Intuitively, σ_P is the set of terms denoting propositions and σ_I is the set of terms denoting anything else. A distinguished subset σ_A of σ_I comprises agent-denoting terms. In more specialized uses of Log_A**B**, the set σ_I may be further partitioned into more fine-grained syntactic types. For example, in a temporal setting, we can have a type for time-, state-, or event-denoting terms.

[1] Which is just fine for the purposes of Chalupsky and Shapiro in [13].

[2] The use of Boolean lattices may be seen as an application of the Boolean-valued models of set theory [14], or an extension of the mereology-based algebraic semantics of Link to the domain of propositions [15].

2.1 Syntax

As is customary in type-theoretical treatments, an alphabet of $Log_A\mathbf{B}$ is made up of a set of syncategorematic punctuation symbols and a set of denoting symbols each from a set σ of syntactic types. The set σ is the smallest set containing all of the following types.

1. σ_P.
2. σ_I.
3. $\tau_1 \longrightarrow \tau_2$, for $\tau_1 \in \{\sigma_P, \sigma_I\}$ and $\tau_2 \in \sigma$.

Intuitively, $\tau_1 \longrightarrow \tau_2$ is the syntactic type of function symbols that take a single argument of type σ_P or σ_I and produce a functional term of type τ_2. Given the restriction of the first argument of function symbols to base types, $Log_A\mathbf{B}$ is, in a sense, a first-order language.

A $Log_A\mathbf{B}$ alphabet is a union of four disjoint sets: $\Omega \cup \Xi \cup \Sigma \cup \Lambda$. The set Ω, the *signature* of the language, is a non-empty set of constant and function symbols. Each symbol in the signature has a designated syntactic type from σ and a designated adicity. (As usual, constants may be viewed as 0-adic function symbols.). Ω is what distinguishes one $Log_A\mathbf{B}$ language from another.

The set $\Xi = \{x_i, a_i, p_i\}_{i\in\mathbb{N}}$ is a countably infinite set of variables, where $x_i \in \sigma_I$, $a_i \in \sigma_A$, and $p_i \in \sigma_P$, for $i \in \mathbb{N}$. Σ is a set of syncategorematic symbols, including the comma, various matching pairs of brackets and parentheses, and the symbol $\forall$. The set Λ is the set of logical symbols of $Log_A\mathbf{B}$, defined as the union of the following sets.

1. $\{\neg\} \subseteq \sigma_P \longrightarrow \sigma_P$
2. $\{\wedge, \vee\} \subseteq \sigma_P \longrightarrow \sigma_P \longrightarrow \sigma_P$
3. $\{\mathbf{B}\} \subseteq \sigma_A \longrightarrow \sigma_P \longrightarrow \sigma_P$

A $Log_A\mathbf{B}$ language with signature Ω is denoted by L_Ω. It is the smallest set of terms formed according to the following rules, where t and t_i ($i \in \mathbb{N}$) are terms in L_Ω.

- $\Xi \subset L_\Omega$
- $c \in L_\Omega$, where $c \in \Omega$ is a constant symbol.
- $f(t_1, \ldots, t_n) \in L_\Omega$, where $f \in \Omega$ is of type $\tau_1 \longrightarrow \ldots \longrightarrow \tau_n \longrightarrow \tau$ and t_i is of type τ_i.
- $\neg t \in L_\Omega$, where $t \in \sigma_P$.
- $(t_1 \otimes t_2) \in L_\Omega$, where $\otimes \in \{\wedge, \vee\}$ and $t_1, t_2 \in \sigma_P$.
- $\forall x(t) \in L_\Omega$, where $x \in \Xi$ and $t \in \sigma_P$.
- $\mathbf{B}(t_1, t_2) \in L_\Omega$, where $t_1 \in \sigma_A$ and $t_2 \in \sigma_P$.

As usual, terms involving $\Rightarrow$, $\Leftrightarrow$, and $\exists$ may be introduced as abbreviations in the standard way.

2.2 Semantics

The basic ingredient of the $Log_A\mathbf{B}$ semantic apparatus is the notion of a $Log_A\mathbf{B}$ structure.

Definition 1 *A $Log_A\mathbf{B}$* structure *is a triple $\mathfrak{S} = \langle \mathcal{D}, \mathfrak{A}, \mathfrak{b} \rangle$, where*

- *$\mathcal{D}$, the domain of discourse, is a set with two disjoint, non-empty, countable subsets $\mathcal{P}$ and $\mathcal{A}$.*
- *$\mathfrak{A} = \langle \mathcal{P}, +, \cdot, -, \bot, \top \rangle$ is a complete, non-degenerate Boolean algebra.*
- *$\mathfrak{b} : \mathcal{A} \times \mathcal{P} \longrightarrow \mathcal{P}$.*

Intuitively, the domain $\mathcal{D}$ is partitioned by a set of propositions $\mathcal{P}$, structured as a Boolean lattice, and a set of individuals $\overline{\mathcal{P}}$, among which at least one agent in the set $\mathcal{A}$ of agents.[3] These stand in correspondence to the syntactic sorts of $Log_A\mathbf{B}$. In what follows, we let $\mathcal{D}_{\sigma_P} = \mathcal{P}$, $\mathcal{D}_{\sigma_I} = \overline{\mathcal{P}}$, and $\mathcal{D}_{\sigma_A} = \mathcal{A}$.

Definition 2 *Let L_Ω be a $Log_A\mathbf{B}$ language. A* valuation *$\mathcal{V}$ of L_Ω is a pair $\langle \mathfrak{S}, v_\Omega \rangle$, where $\mathfrak{S}$ is a $Log_A\mathbf{B}$ structure; and v_Ω is a function that assigns to each constant of sort τ in Ω an element of $\mathcal{D}_\tau$, and to each n-adic ($n \geq 1$) function symbol $f \in \Omega$ of sort $\tau_1 \longrightarrow \ldots \longrightarrow \tau_n \longrightarrow \tau$ an n-adic function $v_\Omega(f) : \bigtimes_{i=1}^{n} \mathcal{D}_{\tau_i} \longrightarrow \mathcal{D}_\tau$.*

Definition 3 *Let L_Ω be a $Log_A\mathbf{B}$ language and let $\mathcal{V}$ be a valuation of L_Ω. For a variable assignment $v_\Xi : \Xi \longrightarrow \mathcal{D}$, where, for every $i \in \mathbb{N}$, $v_\Xi(x_i) \in \overline{\mathcal{P}}$, $v_\Xi(a_i) \in \mathcal{A}$, and $v_\Xi(p_i) \in \mathcal{P}$, an* interpretation *of the terms of L_Ω is given by a function $[\![\cdot]\!]^{\mathcal{V},v_\Xi}$:*

- *$[\![x]\!]^{\mathcal{V},v_\Xi} = v_\Xi(x)$, for $x \in \Xi$*
- *$[\![c]\!]^{\mathcal{V},v_\Xi} = v_\Omega(c)$, for a constant $c \in \Omega$*
- *$[\![f(t_1, \ldots, t_n)]\!]^{\mathcal{V},v_\Xi} = v_\Omega(f)([\![t_1]\!]^{\mathcal{V},v_\Xi}, \ldots, [\![t_n]\!]^{\mathcal{V},v_\Xi})$, for an n-adic ($n \geq 1$) function symbol $f \in \Omega$*
- *$[\![(t_1 \wedge t_2)]\!]^{\mathcal{V},v_\Xi} = [\![t_1]\!]^{\mathcal{V},v_\Xi} \cdot [\![t_2]\!]^{\mathcal{V},v_\Xi}$*
- *$[\![(t_1 \vee t_2)]\!]^{\mathcal{V},v_\Xi} = [\![t_1]\!]^{\mathcal{V},v_\Xi} + [\![t_2]\!]^{\mathcal{V},v_\Xi}$*
- *$[\![\neg t]\!]^{\mathcal{V},v_\Xi} = -[\![t]\!]^{\mathcal{V},v_\Xi}$*
- *$[\![\forall x(t)]\!]^{\mathcal{V},v_\Xi} = \prod_{a \in \mathcal{D}_\tau} [\![t]\!]^{\mathcal{V},v_\Xi[a/x]}$, where x is of sort τ, $v_\Xi[a/x](x) = a$, and $v_\Xi[a/x](y) = v_\Xi[a/x](y)$ for every $y \neq x$*
- *$[\![\mathbf{B}(t_1, t_2)]\!]^{\mathcal{V},v_\Xi} = \mathfrak{b}([\![t_1]\!]^{\mathcal{V},v_\Xi}, [\![t_2]\!]^{\mathcal{V},v_\Xi})$*

In $Log_A\mathbf{B}$, logical consequence is defined in pure algebraic terms without alluding to the notion of truth. This is achieved using the natural partial order $\leq$ associated with $\mathfrak{A}$. (See the appendix for details.)

[3] I will have nothing much to say about the contentious issue of what propositions really are. I take propositions at least to be abstract particulars that are distinct from sentences or terms denoting them.

Definition 4 *Let L_Ω be a $Log_A\mathbf{B}$ language. For every $\phi \in \sigma_P$ and $\Gamma \subseteq \sigma_P$, ϕ is a logical consequence of Γ, denoted $\Gamma \models \phi$, if, for every L_Ω valuation $\mathcal{V}$ and $Log_A\mathbf{B}$ variable assignment v_Ξ, $\prod_{\gamma \in \Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi} \leq [\![\phi]\!]^{\mathcal{V},v_\Xi}$.*

By the above definition, and the algebraic properties of $\mathfrak{S}$, we can easily verify the validity of the following typical examples of logical consequence:

$$\{\phi \wedge \psi\} \models \phi, \{\phi\} \models \phi \vee \psi, \{\phi \Rightarrow \psi, \phi\} \models \psi, \{\bot\} \models \phi, \{\phi\} \models \top$$

In the appendix it is shown that $\models$ has the distinctive properties of classical Tarskian logical consequence.

Proposition 1 *Let L_Ω be a $Log_A\mathbf{B}$ language with $\phi \in L_\Omega$ and $\Gamma, \Delta \subseteq L_\Omega$.*

1. *If $\phi \in \Gamma$, then $\Gamma \models \phi$.*
2. *If $\Gamma \models \phi$ and $\Gamma \subseteq \Delta$, then $\Delta \models \phi$.*
3. *If $\Gamma \models \psi$ and $\Gamma \cup \{\psi\} \models \phi$, then $\Gamma \models \phi$.*

Definition 5 *Let L_Ω be a $Log_A\mathbf{B}$ language. For every $\phi, \psi \in \sigma_P$, ϕ is logically equivalent to ψ, denoted $\phi \equiv \psi$, if, for every L_Ω valuation $\mathcal{V}$ and $Log_A\mathbf{B}$ variable assignment v_Ξ, $[\![\phi]\!]^{\mathcal{V},v_\Xi} = [\![\psi]\!]^{\mathcal{V},v_\Xi}$. ϕ is logically valid if $[\![\phi]\!]^{\mathcal{V},v_\Xi} = \top$, for every L_Ω valuation $\mathcal{V}$ and $Log_A\mathbf{B}$ variable assignment v_Ξ.*

Again, all the standard logical equivalences are valid in our system as a direct corollary to the properties of Boolean algebras. Thus, for example, $\phi \wedge \psi$ and $\psi \wedge \phi$ are two different terms denoting the same proposition, and, hence, are logically equivalent.

3 Truth

As is clear from the previous section, the semantics of $Log_A\mathbf{B}$ has no place for a notion of truth. While we can happily accommodate the standard semantic relations of consequence and equivalence and the property of logical validity, our semantic apparatus has nothing to say about truth. But perhaps this is fine; for truth in the world and the language we use to describe that world and to carry out reasoning about it are not necessarily dependent.

However, it seems that we should at least provide truth conditions for σ_P terms of a $Log_A\mathbf{B}$ language. In standard Tarskian semantics, truth conditions of propositions are part of the definition of the interpretation of the language (Definition 3, in our case). We, however, seem to need more. What we need is what I shall call a *world structure*—a structure describing exactly which propositions in a $Log_A\mathbf{B}$ structure are true in the world.

Definition 6 *For every $Log_A\mathbf{B}$ structure $\mathfrak{S} = \langle \mathcal{D}, \mathfrak{A}, \mathfrak{b} \rangle$, a bivalent world structure $\mathfrak{W}_{\mathbf{2}}(\mathfrak{S})$ is a countably-complete ultrafilter of $\mathfrak{A}$.*[4]

[4] For the definition of ultrafilters, check the appendix. In future work, n-valent world structures are to be considered.

Intuitively, the world structure $\mathfrak{W}_{\mathbf{2}}(\mathfrak{S})$ comprises the set of propositions that are true. Members of the corresponding maximal ideal $\overline{\mathfrak{W}_{\mathbf{2}}(\mathfrak{S})}$ are the false propositions.

In what follows, a bivalent model of a $Log_A\mathbf{B}$ language L_Ω is a triple $\mathcal{M}_{\mathbf{2}} = \langle \mathfrak{W}_{\mathbf{2}}(\mathfrak{S}), \mathcal{V}, v_\Xi \rangle$, where $\mathfrak{W}_{\mathbf{2}}(\mathfrak{S})$ is a bivalent world structure, $\mathcal{V} = \langle \mathfrak{S}, v_\Omega \rangle$ is an L_Ω valuation, and v_Ξ is a $Log_A\mathbf{B}$ variable assignment.

Definition 7 *Let L_Ω be a $Log_A\mathbf{B}$ language. A σ_P-term $\phi \in L_\Omega$ is* true *in a bivalent model $\mathcal{M}_{\mathbf{2}} = \langle \mathfrak{W}_{\mathbf{2}}(\mathfrak{S}), \mathcal{V}, v_\Xi \rangle$, denoted $True_{\mathcal{M}_{\mathbf{2}}}(\phi)$, if $[\![\phi]\!]^{\mathcal{V}, v_\Xi} \in \mathfrak{W}_{\mathbf{2}}(\mathfrak{S})$. Otherwise, ϕ is* false *in $\mathcal{M}_{\mathbf{2}}$, denoted $False_{\mathcal{M}_{\mathbf{2}}}(\phi)$.*

With a bivalent world structure, a $Log_A\mathbf{B}$ logic satisfies the laws of bivalence, excluded-middle, and non-contradiction.

Proposition 2 *Let L_Ω be a $Log_A\mathbf{B}$ language with a bivalent model $\mathcal{M}_{\mathbf{2}}$. For every $\phi \in \sigma_P$ the following is true.*

1. *$True_{\mathcal{M}_{\mathbf{2}}}(\phi)$ or $False_{\mathcal{M}_{\mathbf{2}}}(\phi)$*
2. *$True_{\mathcal{M}_{\mathbf{2}}}(\phi)$ or $True_{\mathcal{M}_{\mathbf{2}}}(\neg\phi)$*
3. *It is not the case that both $True_{\mathcal{M}_{\mathbf{2}}}(\phi)$ and $True_{\mathcal{M}_{\mathbf{2}}}(\neg\phi)$*

The classical truth conditions for compound propositions follow from the above definition.

Proposition 3 *Let L_Ω be a $Log_A\mathbf{B}$ language with a bivalent model $\mathcal{M}_{\mathbf{2}} = \langle \mathfrak{W}_{\mathbf{2}}(\mathfrak{S}), \mathcal{V}, v_\Xi \rangle$ and let $\phi, \psi \in \sigma_P$ and $x \in \tau$.*

- *$True_{\mathcal{M}_{\mathbf{2}}}(\neg\phi)$ if and only if $False_{\mathcal{M}_{\mathbf{2}}}(\phi)$.*
- *$True_{\mathcal{M}_{\mathbf{2}}}(\phi \wedge \psi)$ if and only if $True_{\mathcal{M}_{\mathbf{2}}}(\phi)$ and $True_{\mathcal{M}_{\mathbf{2}}}(\psi)$.*
- *$True_{\mathcal{M}_{\mathbf{2}}}(\phi \vee \psi)$ if and only if $True_{\mathcal{M}_{\mathbf{2}}}(\phi)$ or $True_{\mathcal{M}_{\mathbf{2}}}(\psi)$.*
- *$True_{\mathcal{M}_{\mathbf{2}}}(\forall x(\phi))$ if and only if $True_{\mathcal{M}^b_{\mathbf{2}}}(\phi)$, for all $b \in \mathcal{D}_\tau$, where $\mathcal{M}^b_{\mathbf{2}}(\phi)$ is identical to $\mathcal{M}_{\mathbf{2}}(\phi)$ with v_Ξ replaced by $v_\Xi[b/x]$.*

Typically, Proposition 3 is given as the definition of truth conditions. In our system, the definition is given by membership in some ultrafilter of the underlying Boolean algebra of propositions. Now, one might suspect that there is something unsatisfying about the current state of affairs. For, whereas Proposition 3 provides the classical truth conditions for compound propositions, it is silent about atomic ones. The only thing that we have to say about the truth conditions of atomic propositions is said in Definition 7. But, according to this definition, an atomic propositional term such as $Dog(fido)$ is true if the proposition it denotes is true—something that is determined by fiat. This does not seem to explain much if compared to the classical assignment of truth based on Fido's membership in the extension of the predicate *Dog*.

Nevertheless, the membership test for atomic propositions features, albeit in a slightly different guise, in our semantics. Given a structure $\mathfrak{S}$, the semantics of a $Log_A\mathbf{B}$ symbol like *Dog* is a function from individuals to propositions that those individuals are dogs. However, a model $\mathcal{M}$ (including a world structure for $\mathfrak{S}$) gives rise to a derived function from individuals to truth values, roughly $True_{\mathcal{M}} \circ v_{\Omega}(Dog)$. But this is clearly the characteristic function of the classical set of dogs provided by a Tarskian model. Thus, whatever notion of meaning is provided by classical semantics is also inherent in our algebraic semantics. In addition, we seem to provide a level of meaning (the proposition), independent of a world structure, which is not explicitly available in classical theories.

4 Proof Theory

Our proof theory assumes a (possibly empty) finite knowledge base $\mathbb{K} \subset \sigma_P$ and an inference canon. I will take the inference canon to be a set of Fitch-style natural deduction rules of inference. Such rules come in two forms:

$$\frac{\Gamma}{\phi} \text{ and } \frac{\Gamma, \Delta}{\phi}$$

where $\phi \in \sigma_P$, Γ is a finite subset of σ_P-terms, and Δ is a finite set of items of the form $\Gamma_i \vdash \psi_i$, $\Gamma_i \cup \{\psi_i\} \subset \sigma_P$. As usual, the first form is interpreted as follows: If $\Gamma \subseteq \mathbb{K}$, then ϕ may be added to $\mathbb{K}$. For the second form, if $\Gamma \subseteq \mathbb{K}$ and ψ_i is derivable by the rules of inference with Γ_i as the knowledge base, for every $\Gamma_i \vdash \psi_i \in \Delta$, then ϕ may be added to $\mathbb{K}$. The notion of derivation is given the standard definition in terms of a finite sequence of justified σ_p-terms that ends with the derived expression. I will have nothing more to say about the proof theory here, but any system of Fitch-style natural deduction that is sound and complete for first-order logic will also be sound and complete for $Log_A\mathbf{B}$. I will also not commit myself to any particular set of rules or axiom schema for belief at this point, but Sections 5 and 7 present a thorough discussion of what the possibilities are.

5 Properties of Belief

Given the $Log_A\mathbf{B}$ semantics presented so far, our notion of belief, beside being a relation between agents and propositions, is otherwise totally unconstrained. Although flexibility is a virtue, we may still want our notion of belief to have certain reasonable properties. In what follows, I list some of these.

Definition 8 *Let* $\mathfrak{S} = \langle \mathcal{D}, \mathfrak{A}, \mathfrak{b} \rangle$ *be a* $Log_A\mathbf{B}$ *structure.*

1. $\mathfrak{S}$ *is* injective *if* $\mathfrak{b}$ *is injective.*
2. $\mathfrak{S}$ *is* non-trivial *if* $Range(\mathfrak{b}) \cap \mathfrak{W}_{\mathbf{2}}(\mathfrak{S}) \nsubseteq \{\top\}$, *for every bivalent world structure* $\mathfrak{W}_{\mathbf{2}}(\mathfrak{S})$.
3. $\mathfrak{S}$ *is* meet-distributive *if* $\mathfrak{b}(a, p \cdot q) = \mathfrak{b}(a, p) \cdot \mathfrak{b}(a, q)$, *for every* $p, q \in \mathcal{P}$, $a \in \mathcal{A}$.
4. $\mathfrak{S}$ *is* join-distributive *if* $\mathfrak{b}(a, p) + \mathfrak{b}(a, q) = \mathfrak{b}(a, p + q)$, *for every* $p, q \in \mathcal{P}$, $a \in \mathcal{A}$.
5. $\mathfrak{S}$ *is* consistent *if for every* $(a, p) \in \mathcal{A} \times \mathcal{P}$, $\mathfrak{b}(a, -p) \leq -\mathfrak{b}(a, p)$.

6. *A pair* $(a,p) \in \mathcal{A} \times \mathcal{P}$ *is an* autoepistemic pair *if* $-\mathfrak{b}(a,p) \leq \mathfrak{b}(a,-p)$. $\mathfrak{S}$ *is* autoepistemic *if every* $(a,p) \in \mathcal{A} \times \mathcal{P}$ *is an autoepistemic pair.*
7. $\mathfrak{S}$ *is* positively- (negatively-) introspective *if, for every* $(a,p) \in \mathcal{A} \times \mathcal{P}$, $q \leq \mathfrak{b}(a,q)$, *for* $q = \mathfrak{b}(a,p)$ *(respectively,* $-\mathfrak{b}(a,p)$*).*
8. $\mathfrak{S}$ *is* faithful *if, for every* $(a,p) \in \mathcal{A} \times \mathcal{P}$, $\mathfrak{b}(a, \mathfrak{b}(a,p)) \leq \mathfrak{b}(a,p)$.[5]

For injection, the intuition is that the proposition that a believes p is different from the proposition that b believes q, unless $a = b$ and $p = q$. This property is, in general, absent for other $\mathcal{D}$-valued functions. For example, the father of John may be identical to the father of Mary, and the proposition that John is a sibling of Mary may be identical to the proposition that Mary is a sibling of John. Note, however, that full injection may lead to awkward structures in the presence of other properties. For example, if $\mathfrak{S}$ is both positively-introspective and faithful, then $\mathfrak{b}(a,p) = \mathfrak{b}(a, \mathfrak{b}(a,p))$, for any $(a,p) \in \mathcal{A} \times \mathcal{P}$. Injection implies that $p = \mathfrak{b}(a,p)$, which trivializes the whole notion of belief. Careful definition of $\mathfrak{b}$ is, thus, recommended to avoid such anomalies.[6]

A trivial structure is one for which some world structure only admits either non-believing agents (practically, automata), or agents that believe only what they are bound to believe. While it might not seem reasonable to assume that $\top \in Range(\mathfrak{b})$, we do not rule out this possibility. For example, someone might argue that $\mathfrak{b}(a, \top) = \top$, mirroring the rule of necessity in modal doxastic logic: It is logically valid to believe what is logically valid. Also, sometimes the condition of conceit, $\mathfrak{b}(a, -\mathfrak{b}(a,p) + p) = \top$, is advisable (cf. [1, 10]).[7] Nevertheless, this will not be very useful in the absence of non-trivial beliefs.

Meet-distributivity is a strong condition that implies logical omniscience (cf. Observation 1 below). In general, this property should not be tolerated if we would like to account for realistic agents. The same applies to join-distributivity. Typically, only one direction of join-distributivity is desirable, namely $\mathfrak{b}(p) + \mathfrak{b}(q) \leq \mathfrak{b}(p+q)$. Unfortunately, this direction is *equivalent* to logical omniscience. Better than meet- and join-distributivity, a syntactic version involving $\wedge$ and $\vee$ instead of $\cdot$ and $+$ is preferred.[8]

The above properties of belief are not independent. The following observation lists some dependencies that will turn out to be important in Section 7.

Observation 1 *Let* $\mathfrak{S} = \langle \mathcal{D}, \mathfrak{A}, \mathfrak{b} \rangle$ *be a* $Log_A\mathbf{B}$ *structure.*

1. *If* $\mathfrak{S}$ *is meet-distributive (join-distributive) and* $p \leq q$, *for* $p, q \in \mathcal{P}$, *then* $\mathfrak{b}(a,p) \leq \mathfrak{b}(a,q)$, *for any* $a \in \mathcal{A}$.
2. *If* $\mathfrak{S}$ *is autoepistemic, consistent, and meet-distributive, then it is join-distributive.*

[5] The label "faithful" is inspired by [16].

[6] In their fully-intensional logic, Chalupsky and Shapiro [13] avoid these anomalies by refraining from making any assumptions about belief, including positive introspection and faithfulness.

[7] The label "conceit" is due to [17].

[8] In that case, our $\mathbf{B}$ will behave similar to Levesque's modality of explicit belief [3, p. 201].

3. *If $\mathfrak{S}$ is consistent and negatively-introspective, then it is faithful.*
4. *If $\mathfrak{S}$ is autoepistemic, consistent, positively-introspective, and meet-distributive then it is negatively-introspective.*
5. *If $\mathfrak{S}$ is autoepistemic and faithful, then it is negatively-introspective.*

6 Expressivity

How expressive is $Log_A\mathbf{B}$? As expected, it is more expressive than modal theories of belief. In particular, we can quantify over beliefs, which allows us, for example, to limit properties like introspection to any intensionally characterized set of agents. To demonstrate the expressivity of $Log_A\mathbf{B}$, consider the following example.

Example 1. In [2], the authors describe a proposition to be common knowledge for a group of agents if all agents in the group know it, all agents in the group know that all agents in the group know it, etc. In the modal framework adopted in [2], the authors had to introduce a new modal operator (actually, a class thereof, one operator for each possible group) to capture this notion of common knowledge. In $Log_A\mathbf{B}$, we can make use of our ability to quantify over propositions in order to represent the corresponding notion of *common belief.* If our group of agents is intensionally characterized by G, we first introduce a non-logical function symbol $\mathbf{B}^*_G$, which is akin to the reflexive transitive closure of $\mathbf{B}$ (viewed relationally):

$$\forall p\ [\mathbf{B}^*_G(p, p)]$$

$$\forall a, p\ [\mathbf{B}^*_G(p, \mathbf{B}(a, p)) \Leftrightarrow G(a)]$$

$$\forall p, q, r\ [\mathbf{B}^*_G(p, q) \wedge \mathbf{B}^*_G(q, r) \Rightarrow \mathbf{B}^*_G(p, r)]$$

Common belief (**CB**) can be defined as follows.

$$\mathbf{CB}_G(p) =_{\text{def}} \forall a\ [G(a) \Rightarrow \forall q\ [\mathbf{B}^*_G(p, q) \Rightarrow \mathbf{B}(a, q)]]$$

In [2], the related notion of distributed knowledge characterizes those propositions that are implied by the collective knowledge of a group. Again, the authors introduced a new modal operator to model distributed knowledge. In $Log_A\mathbf{B}$, we need a non-logical function symbol capturing the collective beliefs of the group:

$$\mathbf{B}^{\cup}_G(p) \Leftrightarrow \exists a[G(a) \wedge \mathbf{B}(a, p)]\ \vee\ \exists q, r[\mathbf{B}^{\cup}_G(q) \wedge \mathbf{B}^{\cup}_G(r) \wedge (p \Leftrightarrow (q \wedge r))]$$

We may now define distributive belief as follows.

$$\mathbf{DB}_G(p) =_{\text{def}} \exists q\ [\mathbf{B}^{\cup}_G(q) \wedge (q \Rightarrow p)]$$

□

Though more expressive than modal theories, $Log_A\mathbf{B}$ is, in a certain sense, less expressive than *syntactical* first-order theories. In particular, $Log_A\mathbf{B}$ languages are not self referential.

7 Self-Reference

Results of Montague [11] and Thomason [10] show that a first-order treatment of epistemic modalities (respectively, knowledge and belief) yield inconsistent systems. Of course, it is a particular mix of assumptions about the modalities that gives rise to inconsistencies. The inconsistencies appear as paradoxes of self-reference, akin to the famous Liar paradox. How does our system fair in this regard? Interestingly, our system is immune to such paradoxes for the same reason why it is not a classical first-order logic. Let me explain. First-order doxastic theories are *syntactical*, they include a dyadic belief predicate (akin to our functional **B**) whose first argument is an agent-denoting term and whose second argument is a term that denotes a *formula* of the same language. It is this ability of the language to refer to its own syntactic structures that makes such theories syntactical. However, such ability does not come for free; there are, in general, two methods to achieve syntacticity. The first is to equip the language with an axiomatization of arithmetic and denote formulas of the language by their Gödel numbers [18, 9, for instance]. The second is to provide the language with systematic means to manipulate strings, together with devices for substitution, quotation, and un-quotation [7, for instance].

In syntactical theories, we can have formulas that refer to themselves. For example, a formula $P(\ulcorner 123 \urcorner)$ may have as its Gödel number the very same 123, encoded by the string $\ulcorner 123 \urcorner$. In fact, the diagonalization lemma (see [19, for example]) states that, in a syntactical first-order theory, there is a formula ϕ such that $\phi \Leftrightarrow p(\ulcorner \phi \urcorner)$ is a theorem, for any (possibly complex) monadic predicate p. Note that $\phi \Leftrightarrow p(\ulcorner \phi \urcorner)$ is a *theorem*—we have no way of avoiding it—not just a sentence generated by the grammar of the language. It is this result that leads to doxastic paradoxes, when p is $\lambda x.\neg B(\alpha, x)$ and B is the belief predicate.

As demonstrated by several authors [20, 8, 21], it is the syntacticity of a system that is the catalyst for paradox, not whether it is first-order or modal. Interestingly, $Log_A\mathbf{B}$, which has all the advantages that a first-order doxastic theory has over a modal one (see Section 6), is not syntactical. There is no way for a $Log_A\mathbf{B}$ language to refer to its own terms. In particular, no σ_P term can refer to itself, since, *tout court*, the proper-substring relation is irreflexive. Granted, we can write expressions such as $\phi \Leftrightarrow \neg\mathbf{B}(\alpha, \phi)$ (or even $\phi = \neg\mathbf{B}(\alpha, \phi)$); we can have such expressions in our knowledge base; and, with a certain notorious suite of assumptions on **B**, we shall get a contradiction. But this inconsistency is an inconsistency of the knowledge base, not a natural product of our proof theory: paradoxical, self-referential expressions are *not* theorems of our logic.

Now, the immunity of $Log_A\mathbf{B}$ to paradox is clearly rooted in its relative expressive weakness, compared to syntactical theories, when it comes to representing its own syntax. However, syntactical theories have almost exclusively been employed to account for the propositional attitudes, which a $Log_A\mathbf{B}$ approach seems to effectively accommodate. Perlis [7, 8] argues that languages with self-reference are essential for commonsense reasoning in general. I might have something to say about this, but that is a story for another day.

Notwithstanding $Log_A\mathbf{B}$'s immunity to paradox, we can certainly construct $Log_A\mathbf{B}$ structures in which, for some $p \in \mathcal{P}$, $p = -\mathfrak{b}(a, p)$. Such structures will be incompatible with some of the properties in Definition 8. Syntactically, this means that (knowledge-base) inconsistency looms given $\phi \Leftrightarrow \mathbf{B}(\alpha, \phi)$ and specific axiomatizations of belief. In what follows, we prove several results, indicating exactly when that happens. We start with the following two useful lemmas.[9]

Lemma 1. *If $\mathfrak{S}$ is a consistent, positively-introspective $Log_A\mathbf{B}$ structure, then $p = -\mathfrak{b}(a, p)$ implies $p = \top$ for every $(a, p) \in \mathcal{A} \times \mathcal{P}$.*
Proof. Suppose that $p = -\mathfrak{b}(a, p)$. By consistency, $\mathfrak{b}(a, -\mathfrak{b}(a, p)) = \mathfrak{b}(a, p) \leq -\mathfrak{b}(a, \mathfrak{b}(a, p))$. On the other hand, by positive introspection, $\mathfrak{b}(a, p) \leq \mathfrak{b}(a, \mathfrak{b}(a, p))$. It follows that, $\mathfrak{b}(a, p) \leq -\mathfrak{b}(a, \mathfrak{b}(a, p)) \cdot \mathfrak{b}(a, \mathfrak{b}(a, p)) = \bot$. Thus, $\mathfrak{b}(a, p) = \bot$ and, hence, $p = \top$ (by the definition of $\leq$ and B7.1 in the appendix). □

Lemma 2. *If $\mathfrak{S}$ is a negatively-introspective $Log_A\mathbf{B}$ structure, then $p = -\mathfrak{b}(a, p)$ implies $p = \bot$ for every $(a, p) \in \mathcal{A} \times \mathcal{P}$.*
Proof. Assume that $p = -\mathfrak{b}(a, p)$. It follows from negative introspection that $-\mathfrak{b}(a, p) \leq \mathfrak{b}(a, -\mathfrak{b}(a, p))$. Then $-\mathfrak{b}(a, p) \leq \mathfrak{b}(a, p)$, and, consequently, $-\mathfrak{b}(a, p) = p = \bot$ (by the definition of $\leq$ and B5.2 in the appendix). □

Our first theorem is a variant of Theorem 4.7 in [9, p. 76], where the inconsistency result of Thomason [10] is regenerated by trading conceit ($\mathfrak{b}(a, -\mathfrak{b}(a, p)+p) = \top$) for the more subjective negative introspection.

Theorem 1. *If $\mathfrak{S}$ is a consistent, positively-, and negatively-introspective $Log_A\mathbf{B}$ structure, then there is no $(a, p) \in \mathcal{A} \times \mathcal{P}$ such that $p = -\mathfrak{b}(a, p)$.*
Proof. Assume $(a, p) \in \mathcal{A} \times \mathcal{P}$ such that $p = -\mathfrak{b}(a, p)$. By Lemma 1, $p = \top$. By Lemma 2, $p = \bot$. Consequently, $\bot = \top$, which is impossible since the algebra $\mathfrak{A}$ is non-degenerate. □

The following theorem shows that positive introspection is not responsible, after all, for the inconsistency.

Theorem 2. *If $\mathfrak{S}$ is a consistent, negatively-introspective, and meet-distributive $Log_A\mathbf{B}$ structure, then there is no $(a, p) \in \mathcal{A} \times \mathcal{P}$ such that $p = -\mathfrak{b}(a, p)$.*
Proof. Assume $(a, p) \in \mathcal{A} \times \mathcal{P}$ such that $p = -\mathfrak{b}(a, p)$. By negative introspection and Lemma 2, $p = \bot$. Consequently, $\mathfrak{b}(a, p) = \mathfrak{b}(a, \bot) = \top$. Now, let $q \in \mathcal{P}$ be an arbitrary proposition. Since $\bot = q \cdot -q$, it follows that $\mathfrak{b}(a, q \cdot -q) = \top$. By meet-distributivity, $\mathfrak{b}(a, q \cdot -q) = \mathfrak{b}(a, q) \cdot \mathfrak{b}(a, -q)$. By consistency, $\mathfrak{b}(a, q \cdot -q) \leq \mathfrak{b}(a, q) \cdot -\mathfrak{b}(a, q) = \bot$, which is impossible since $\mathfrak{A}$ is non-degenerate. □

Even though we require $\mathfrak{S}$ to be meet-distributive, the direction of meet-distributivity actually used in the proof ($\mathfrak{b}(a, p \cdot q) \leq \mathfrak{b}(a, p) \cdot \mathfrak{b}(a, q)$) is provably equivalent to logical omniscience ($p \leq q$ implies $\mathfrak{b}(a, p) \leq \mathfrak{b}(a, q)$). Logical omniscience is already a property of the systems of both Thomason [10] and Bolander

[9] Strictly speaking, standard results pertain to sentences, not propositions. But see [22] for a discussion of how such results extend to propositions.

[9]. Hence, the above result is a strengthening of Bolander's in that it shows that negative-introspection, consistency, and omniscience are sufficient—without positive introspection—to regenerate Thomason's paradox. In addition, the result also shows that negative introspection is problematic enough to induce an inconsistency if it replaces Thomason's conceit and positive introspection.

But, while negative introspection is rightly incriminated by the above result, it certainly is not necessary for the inconsistency. The following theorem shows that replacing negative-introspection with the less controversial property of faithfulness gives rise to inconsistency when a non-pervasive version of autoepistemology is enforced.[10]

Theorem 3. *If $\mathfrak{S}$ is a consistent, positively-introspective, and faithful $Log_A\mathbf{B}$ structure, then there is no autoepistemic pair (a, p) such that $p = -\mathfrak{b}(a, p)$.*
Proof. Assume there is an autoepistemic pair $(a, p) \in \mathcal{A} \times \mathcal{P}$ such that $p = -\mathfrak{b}(a, p)$. By Lemma 1, $p = -\mathfrak{b}(a, p) = \top$. Since (a, p) is an autoepistemic pair, $-\mathfrak{b}(a, p) \leq \mathfrak{b}(a, -p) = \mathfrak{b}(a, \mathfrak{b}(a, p))$. By faithfulness and transitivity of $\leq$, $\top = -\mathfrak{b}(a, p) \leq \mathfrak{b}(a, p) = \bot$. Consequently, $\bot = \top$, which is impossible since the algebra $\mathfrak{A}$ is non-degenerate. □

In the spirit of [23], we present the following non-theorem, demonstrating the necessity of autoepistemology for the above result.

Non-Theorem 1 *If $\mathfrak{S}$ is a consistent, positively-introspective, and faithful $Log_A\mathbf{B}$ structure, then there is no $(a, p) \in \mathcal{A} \times \mathcal{P}$ such that $p = -\mathfrak{b}(a, p)$.*
Counterexample 1. Let $\mathfrak{S}_{TWO} = \langle \mathcal{D}_2, \mathfrak{A}_2, \mathfrak{b}_2 \rangle$. Take $\mathcal{D}_2 = \{a, \bot, \top\}$, $\mathfrak{A}_2 = \langle \{\bot, \top\}, +, \cdot, -, \bot, \top \rangle$, and $Range(\mathfrak{b}_2) = \{\bot\}$. $\mathfrak{S}_{TWO}$ is trivially consistent. In addition, it is both positively-introspective and faithful, since $\mathfrak{b}_2(a, p) = \mathfrak{b}_2(a, \mathfrak{b}_2(a, p)) = \bot$ for $p \in \mathcal{P}$. In this structure, $\top = -\bot = -\mathfrak{b}_2(a, \top)$.

Our first counterexample, though falsifies the non-theorem, constructs a rather trivial structure. Our second example is more general.
Counterexample 2. Consider two disjoint sets $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_n\}$ of propositions. Let $\mathfrak{A}_P = \langle P', +, \cdot, -, \bot, \top \rangle$ and $\mathfrak{A}_Q = \langle Q', +, \cdot, -, \bot, \top \rangle$ be the Boolean algebras generated by P and Q, respectively, and let $\mathfrak{A}_\cup = \langle (P \cup Q)', +, \cdot, -, \bot, \top \rangle$ be the Boolean algebra generated by $P \cup Q$. Take $\mathfrak{i}$ to be an isomorphism from $\mathfrak{A}_P$ to $\mathfrak{A}_Q$, where $\mathfrak{i}(p_i) = q_i$ for $p_i \in P$. Now, define $\mathfrak{S}_{ISO} = \langle \{a\} \cup (P \cup Q)', \mathfrak{A}_\cup, \mathfrak{b}_\mathfrak{i} \rangle$, where $\mathfrak{b}_\mathfrak{i}$ is defined as follows.

$$\mathfrak{b}_\mathfrak{i}(a, x) = \begin{cases} \bot & \text{if } x \in \{\bot, \top\} \\ \mathfrak{i}(x) & \text{if } x \in P' \setminus \{\bot, \top\} \\ x & \text{if } x \in Q' \setminus \{\bot, \top\} \\ z & \text{if } x = y \odot z, \text{ where } \odot \in \{\cdot, +\}, y \in P' \setminus \{\bot, \top\}, \\ & \text{and } z \in Q' \setminus \{\bot, \top\} \end{cases}$$

It could be shown, by induction on the structure of x, that $\mathfrak{b}_\mathfrak{i}$ is well-defined. We now show that it satisfies the conditions stated in the non-theorem. First,

[10] Faithfulness is a theorem of Thomason's system [10]. Hence, the following result is, in a sense, a strengthening of Thomason's.

consider consistency. The case of $\bot$ and $\top$ is similar to Counterexample 1. For $p \in P' \setminus \{\bot, \top\}$, $\mathfrak{b}_{\mathfrak{i}}(a, -p) = \mathfrak{i}(-p) = -\mathfrak{i}(p) = -\mathfrak{b}_{\mathfrak{i}}(a, p)$. For $q \in Q' \setminus \{\bot, \top\}$, $\mathfrak{b}_{\mathfrak{i}}(a, -q) = -q = -\mathfrak{b}_{\mathfrak{i}}(a, q)$. For $x = y \cdot z$, with $y \in P' \setminus \{\bot, \top\}$ and $z \in Q' \setminus \{\bot, \top\}$, $\mathfrak{b}_{\mathfrak{i}}(a, -x) = \mathfrak{b}_{\mathfrak{i}}(a, -y + -z) = -z = -\mathfrak{b}_{\mathfrak{i}}(a, y \cdot z)$. Similarly for $x = y + z$. Note that, not only is $\mathfrak{S}$ consistent, but it is also the case that every pair $(a, x) \in \mathcal{A} \times (P \cup Q)' \setminus \{\bot, \top\}$ is autoepistemic.

Similar to Counterexample 1, $\mathfrak{b}_{\mathfrak{i}}(a, x) = \mathfrak{b}_{\mathfrak{i}}(a, \mathfrak{b}_{\mathfrak{i}}(a, x))$, for $x \in \{\bot, \top\}$. Otherwise, $\mathfrak{b}_{\mathfrak{i}}(a, x) \in Q' \setminus \{\top\}$ and is, hence, identical to $\mathfrak{b}_{\mathfrak{i}}(a, \mathfrak{b}_{\mathfrak{i}}(a, x))$. Thus, $\mathfrak{S}$ is both positively-introspective and faithful. Finally, similar to Counterexample 1, note that $\top = -\bot = -\mathfrak{b}_{\mathfrak{i}}(a, \top)$. □

In the $Log_A\mathbf{B}$ structure $\mathfrak{S}_{ISO}$ of Counterexample 2, all propositions, except the paradoxical one, are autoepistemic and satisfy the negative introspection schema. In addition, $\mathfrak{S}_{ISO}$ is *almost* meet-distributive (and join-distributive), which means that it almost satisfies logical omniscience (cf. Observation 1). Thus, not only have we shown that consistency, positive-introspection, and faithfulness are tolerant to the $p = -\mathfrak{b}(a, p)$ possibility, but we have also shown that the tolerance persists even in the presence of a high degree of logical omniscience. We are pretty close to Thomason's system.

It should be clear that $\mathfrak{S}_{ISO}$ could be varied along different dimensions. We may allow multiple agents, where instead of the single set Q, we have a family of sets indexed by $\mathcal{A}$. Omniscience may also be avoided by constructing the isomorphism differently. For example, instead of standing in 1-1 correspondence to the set P, Q can be defined to correspond 1-1 to the elements of the *algebra* generated by P. We may also construct a structure in which we are more conservative about which propositions are autoepistemic. One way to achieve this is to change the definition of $\mathfrak{b}$ such that $\mathfrak{b}(a, -p) = -\mathfrak{b}(a, p) \cdot AE(p)$, where $AE(p) = \top$ only for some $p \in \mathcal{P}$ (those that are intuitively autoepistemic). Note that this definition of $\mathfrak{b}$ maintains the property of consistency.

Theorem 3 and Non-theorem 1 tell us the following: For consistent, positively-introspective, and faithful (and almost omniscient) structures, a pair (a, p) will satisfy $p = -\mathfrak{b}(a, p)$ if and only if it is not autoepistemic. What is interesting here is that the offensive property—autoepistemology—is one that naturally applies only to select propositions (and agents) by fiat. For example, whereas my having a brother is plausibly autoepistemic, my first-grade teacher's being asleep right now is clearly not. Thus, if pressed, we may deem a proposition p, such that $p = -\mathfrak{b}(a, p)$, non-autoepistemic. Now, while it might be easy to semantically implement this decree (and to perhaps philosophically justify it), it is not immediately clear how we can syntactically enforce it. But the results obtained in [23, 18, 9] give us some hope. Based on purely syntactic properties, we may be able to quarantine some σ_P terms which are believed to give rise to paradoxical self-reference. Unlike [23, 18, 9], where the recommendation is to suspend the application of *all* doxastic schema on the quarantined expressions, we only need to make sure that we do not label any of them as autoepistemic. We, thus, get the full force of rational belief (for example, consistency, positive introspection, faithfulness), and the limited application of negative introspection to the autoepistemic agent-proposition pairs (cf. Observation 1). The exact ramifications of this result for syntactical theories is to be explored in future work.

8 Conclusions

By admitting propositions as first-class individuals in our ontology, we achieve two things: (i) the expressivity and semantic simplicity of first-order doxastic theories and (ii) the consistency and syntactic simplicity of rival modal theories. I hope I have managed to convince the reader of the above claim through the presentation of $Log_A\mathbf{B}$. No paradoxical self-referential propositional term is a theorem of $Log_A\mathbf{B}$, but results of Thomason's and Bolander's feature as conditions of incompatibility of certain properties of $Log_A\mathbf{B}$'s semantics structures. This leads to one direction of future work: How may the results presented here (in particular those pertaining to the interrelations among autoepistemology, faithfulness, conceit, and negative introspection) be applied to syntactical theories in order to avoid paradox, while minimally sacrificing the pervasive adoption of desirable properties of belief?

Other directions for future research include developing similar algebraic accounts for other propositional attitudes, notably knowledge (within a $Log_A\mathbf{K}$ framework). Also, as pointed out earlier, non-bivalent world structures of $Log_A\mathbf{B}$ may be studied. In particular, a trivalent world structure, corresponding to a three-valued logic, may be defined as a filter (as opposed to an ultrafilter) of the underlying Boolean algebra. Similarly, a quad-valent world structure could be defined as a filter-ideal pair. Connections of such systems to existing many-valued logics (for example, [24, 25]) are then to be systematically studied.

References

1. Hintikka, J.: Knowledge and Belief: An Introduction to the Logic of the Two Notions. Cornell University Press, Ithaca, New York (1962)
2. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning about Knowledge. The MIT Press, Cambridge, Massachusetts (1995)
3. Levesque, H.: A logic of implicit and explicit belief. In: Proceedings of the Fourth National Conference on Artificial Intelligence (AAAI-84), Palo Alto, California, Morgan Kaufmann (1984) 198–202
4. Kaplan, A., Schubert, L.: A computational model of belief. Artificial Intelligence **120** (2000) 119–160
5. Kaplan, D.: Quantifying in. Synthese **19** (1968) 178–214
6. Bealer, G.: Theories of properties, relations, and propositions. The Journal of Philosophy **76**(11) (1979) 634–648
7. Perlis, D.: Languages with self-refernce I: Foundations. Artificial Intelligence **25** (1985) 301–322
8. Perlis, D.: Languages with self-reference II: Knowledge, belief, and modality. Artificial Intelligence **34**(2) (1988) 179–212
9. Bolander, T.: Logical Theories for Agent Introspection. PhD thesis, Department of Informatics and Mathematical Modelling, Technical University of Denmark (2003)

10. Thomason, R.: A note on syntactical treatments of modality. Synthese **44**(3) (1980) 391–395
11. Montague, R.: Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. Acta Philosophica Fennica **16** (1963) 153–167
12. Shapiro, S.C.: Belief spaces as sets of propositions. Journal of Experimental and Theoretical Artificial Intelligence **5** (1993) 225–235
13. Chalupsky, H., Shapiro, S.: SL: A subjective, intensional logic of belief. In: Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society, Hillsdale, New Jersey, Lawrence Erlbaum (1994) 165–170
14. Bell, J.: Set Theory: Boolean-Valued Models and Independence Proofs. 3rd edn. Oxford University Press (2005)
15. Link, G.: Algebraic Semantics in Language and Philosophy. CSLI Publications, Stanford, CA (1998)
16. Konolige, K.: A computational theory of belief introspection. In: Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85), Los Angeles, California (1985) 502–508
17. Smullyan, R.: Logicians who reason about themselves. In [26] 342–352
18. Morreau, M., Kraus, S.: Syntactical treatments of propositional attitudes. Artificial Intelligence **106** (1998) 161–177
19. Boolos, G., Burgess, J., Jeffrey, R.: Computability and Logic. 5th edn. Cambridge University Press (2007)
20. Asher, N., Kamp, H.: The knower's paradox and representational theories of attitudes. In [26] 131–148
21. Grim, P.: Operators in the paradox of the knower. Synthese **94**(3) (1993) 409–428
22. Mills, E.: A simple solution to the liar. Philosophical Studies **89**(2/3) (1998) 197–212
23. des Rivières, J., Levesque, H.: The consistency of syntactical treatments of knowledge. In [26] 115–130
24. Kleene, S.C.: On notation for ordinal numbers. The Journal of Symbolic Logic **3**(4) (1938) 150–155
25. Belnap, N.: A useful four-valued logic: How a computer should think. In Anderson, A., Belnap, N., Dunn, M., eds.: Entailment. Volume II. Princeton University Press, Princeton, NJ (1992) 506–541
26. Halpern, J., ed.: Proceedings of the First International Conference on Theoretical Aspects of Reasoning about Knowledge (TARK-86), Monterey, California, Morgan Kaufmann (1986)
27. Burris, S., Sankappanavar, H.P.: A Course in Universal Algebra. Springer-Verlag (1982)

Appendix: Boolean Algebra

For the sake of completeness, I hereafter present some basic results about Boolean algebra that are relevant to the development of Log_A**B**. The presentation is based primarily on [27]. All proofs are omitted for limitations of space; the interested reader may consult [27] or any standard text on the topic.

A Boolean algebra is a sextuple $\mathfrak{B} = \langle B, +, \cdot, -, \bot, \top \rangle$ where B is a non-empty set and $\{\bot, \top\} \subseteq B$. B is closed under the two binary operators $+$ and $\cdot$ and the unary operator $-$. The operators satisfy the following conditions.

B1.1: $a + b = b + a$ (Commutativity)
B1.2: $a \cdot b = b \cdot a$
B2.1: $a + (b + c) = (a + b) + c$ (Associativity)
B2.2: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
B3.1: $a + (a \cdot b) = a$ (Absorption)
B3.2: $a \cdot (a + b) = a$
B4.1: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ (Distribution)
B5.1: $a + -a = \top$ (Complements)
B5.2: $a \cdot -a = \bot$

The following properties of Boolean algebras immediately follow.

B4.2: $a + (b \cdot c) = (a + b) \cdot (a + c)$
B6.1: $a \cdot a = a$
B6.2: $a + a = a$
B7.1: $a \cdot \bot = \bot$
B7.2: $a + \top = \top$
B8: $a \cdot \top = a + \bot = a$
B9: $-(-a) = a$
B10.1: $-(a \cdot b) = (-a) + (-b)$
B10.2: $-(a + b) = (-a) \cdot (-b)$

A Boolean algebra $\mathfrak{B} = \langle B, +, \cdot, -, \bot, \top \rangle$ is *complete* if, for every $A \subseteq B$, $\sum_{a \in A} a \in B$ and $\prod_{a \in A} a \in B$. $\mathfrak{B}$ is *degenerate* if $\bot = \top$, otherwise, it is *non-degenerate*. Elements of B are partially-ordered by the relation $\leq$, where $a \leq b$ if and only if $a \cdot b = a$. By B3.1 and B3.2, it follows that $a \leq b$ if and only if $a + b = b$.

A *filter* of $\mathfrak{B}$ is a subset F of B such that

F1. $\top \in F$
F2. $a, b \in F$ implies $a \cdot b \in F$
F3. $a \in F$ and $a \leq b$ imply $b \in F$

F is an *ultrafilter* of $\mathfrak{B}$ if it is maximal with respect to not including $\bot$. The following properties of ultrafilters follow from the definitions.

F4. For every $a \in B$, exactly one of a and $-a$ belong to F.
F5. For every $a, b \in B$, $a + b \in F$ if and only if $a \in F$ or $b \in F$.

Moreover, F is *countably-complete* if it satisfies

F6. For every $A \subseteq F$, if A is countable, then $\prod_{a \in A} a \in F$.

To illustrate the relevance of the above properties of Boolean algebras to $Log_A\mathbf{B}$, I present proofs for Propositions 1 and 3. (Proposition 2 follows immediately from F4.)

Proof of Proposition 1.

1. Suppose that $\phi \in \Gamma$. Then, for every L_Ω valuation $\mathcal{V}$ and $Log_A\mathbf{B}$ variable assignment v_Ξ, $\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi} = (\prod_{\phi\neq\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\phi]\!]^{\mathcal{V},v_\Xi}$. By B2.2 and B6.1,

$$(\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\phi]\!]^{\mathcal{V},v_\Xi} = (\prod_{\phi\neq\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) \cdot ([\![\phi]\!]^{\mathcal{V},v_\Xi} \cdot [\![\phi]\!]^{\mathcal{V},v_\Xi}) = \prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}$$

   Hence, $\Gamma \models \phi$.
2. Suppose that $\Gamma \models \phi$ and $\Gamma \subseteq \Delta$. Thus, there is a set Γ' such that $\Delta = \Gamma \cup \Gamma'$ and $\Gamma \cap \Gamma' = \varnothing$. By B1.2 and B2.2,

$$(\prod_{\delta\in\Delta} [\![\delta]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\phi]\!]^{\mathcal{V},v_\Xi} = (\prod_{\gamma'\in\Gamma'} [\![\gamma']\!]^{\mathcal{V},v_\Xi}) \cdot [(\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\phi]\!]^{\mathcal{V},v_\Xi}]$$

   Since $\Gamma \models \phi$, it follows that

$$(\prod_{\delta\in\Delta} [\![\delta]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\phi]\!]^{\mathcal{V},v_\Xi} = (\prod_{\gamma'\in\Gamma'} [\![\gamma']\!]^{\mathcal{V},v_\Xi}) \cdot (\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) = \prod_{\delta\in\Delta} [\![\delta]\!]^{\mathcal{V},v_\Xi}$$

   Hence, $\Delta \models \phi$.
3. Suppose $\Gamma \models \psi$ and $\Gamma \cup \{\psi\} \models \phi$. By definition of $\models$,

$$(\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\psi]\!]^{\mathcal{V},v_\Xi} \leq [\![\phi]\!]^{\mathcal{V},v_\Xi}$$

   But, since $\Gamma \models \psi$,

$$\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi} = (\prod_{\gamma\in\Gamma} [\![\gamma]\!]^{\mathcal{V},v_\Xi}) \cdot [\![\psi]\!]^{\mathcal{V},v_\Xi} \leq [\![\phi]\!]^{\mathcal{V},v_\Xi}$$

   Hence, $\Gamma \models \phi$. □

Proof of Proposition 3.

1. $True_{\mathcal{M}_\mathbf{2}}(\neg\phi)$ iff $-[\![\phi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ iff $[\![\phi]\!]^{\mathcal{V},v_\Xi} \notin \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ (by F4) iff $False_{\mathcal{M}_\mathbf{2}}(\phi)$.
2. $True_{\mathcal{M}_\mathbf{2}}(\phi \wedge \psi)$ iff $[\![\phi]\!]^{\mathcal{V},v_\Xi} \cdot [\![\psi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ iff $[\![\phi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ and $[\![\psi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ (by F2 and F3) iff $True_{\mathcal{M}_\mathbf{2}}(\phi)$ and $True_{\mathcal{M}_\mathbf{2}}(\psi)$.
3. $True_{\mathcal{M}_\mathbf{2}}(\phi \vee \psi)$ iff $[\![\phi]\!]^{\mathcal{V},v_\Xi} + [\![\psi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ iff $[\![\phi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ or $[\![\psi]\!]^{\mathcal{V},v_\Xi} \in \mathfrak{W}_\mathbf{2}(\mathfrak{S})$ (by F5) iff $True_{\mathcal{M}_\mathbf{2}}(\phi)$ or $True_{\mathcal{M}_\mathbf{2}}(\psi)$.
4. Similar to the case of $\wedge$, using F6 instead of F2. □