'I believe' in a ranking-theoretic analysis of 'believe'*

Sven Lauer

University of Konstanz, Konstanz, Germany
sven.lauer@uni-konstanz.de

Abstract

There is a *prima facie* tension between two well-known observations about sentences of the form *I (don't) believe p*. The first is *Moore's paradox*, i.e., the fact that sentences of the form *p, but I don't believe p* and $\neg p$ *but I believe p* sound 'contradictory' or 'incoherent'. The second observation is that *I believe* often functions as a hedge: A speaker who asserts *I believe p* often (but not always) conveys that she is not certain that *p*, or that she does not want to commit entirely to *p* being true.

I argue that (natural explantations of) these two observations are in conflict with respect to the following question: Does saying *I believe p* commit the speaker to taking *p* to be true? Moore's paradox says *Yes!*, while the fact that *I believe* functions as a hedge says *No!* I argue that resolving this tension (along with other desiderata) requires a theory of *graded belief*. I reject a probabilistic threshold analysis for familiar reasons (lack of closure under conjunction), and show that an alternative based on the *ranking theory* of Spohn (1988, 1990, 2012) can compositionally deliver the desired results when embedded in a commitment-based theory of declarative force.

1 Two observations and a dilemma

1.1 Moore's paradox

Moore's paradox is the observation that (1) and (2) sound 'incoherent' or 'contradictory'.

(1) It is raining, but I don't believe it (is raining). $p \wedge \neg \mathsf{Bel}_{\mathrm{Sp}}(p)$ or $p \wedge \mathsf{Bel}_{\mathrm{Sp}}(\neg p)$

(2) It is not raining, but I believe it (is raining). $\neg p \wedge \mathsf{Bel}_{\mathrm{Sp}}(p)$

(1) has two readings, depending on the relative scope of negation and *believe*. For simplicity, in the following, I will talk mostly about (2), which is equally Moore-paradoxical. The account offered in the second half of this paper, however, also extends to the two readings of (1).

These sentences sound contradictory, even though they should, given standard assumptions about their meanings, express perfectly consistent propositions. One way to see is this to consider the sentences in (3).

(3) a. It is not raining, but John believes/thinks it is (raining).
b. It was not raining, but I believed/thought it was (raining).

These sentences are consistent because it is possible that John has false beliefs, and it is possible that the speaker used to have false beliefs. But, of course, it is also possible that the speaker presently has false beliefs, so (2) should have a consistent content, as well. And yet, (2) sounds contradictory.

Here is a sketch of a natural explanation of these observations: With uttering $\neg p$, a speaker commits to taking *p* to be false, but with uttering *I believe p*, she commits to taking *p* to be true. Thus, while (2) has a consistent content, it gives rise to incompatible commitments for the speaker. Hence, it sounds contradictory.

*I'd like to thank Cleo Condoravdi, Ciyang Qing, Eric Raidl, Wolfgang Spohn, as well as audiences at the Konstanz *What if?*-group and the *Questioning Speech Acts* workshop, for helpful comments and discussion.

1.2 Hedging with 'I believe'

When a speaker utters *I believe p*, she often conveys that she is not entirely sure that p is true.

(4) I believe it is raining. $\rightsquigarrow$ Speaker is not sure that it is raining.

However, this is by no means always the case. The intuitive implication in (4) can be coherently denied (5a), it can be suspended by inserting an adverb like *firmly* (5b) and even without such an indication, the implication of uncertainty can be absent if the context is right.

(5) a. I believe that the president should be impeached. I am positive/absolutely certain.
b. I firmly believe that the president should be impeached.

A natural explanation of these facts is the following: The inference in (4) is a *conversational implicature*, which roughly is derived at follows. The speaker could have asserted p instead of uttering *I believe p*, which is longer and more complex. She must have had a reason to opt for *I believe p*. A plausible reason (in the right context) is that she did not want to commit to p, because she is not sure that p is true. Hence she opted for only committing to *I believe p* instead of p.

1.3 A dilemma

The two 'natural explanations' just sketched are very intuitive individually, but they are problematic in so far as they appear to presuppose opposite answers to the question 'Does uttering *I believe p* commit the speaker to taking p to be true?'

To bring this out more clearly, let $\mathsf{Ass}_{\mathrm{Sp}}(\cdot)$ be an operator that represents the normative consequences of assertion ('doxastic commitment' / 'commitment to believe' Condoravdi and Lauer 2011, Lauer 2013; 'assertoric commitment' Krifka 2014; 'truth commitment' Searle 1969, Krifka 2015; ...) and let $\mathsf{Bel}_{\mathrm{Sp}}(\cdot)$ be an operator representing the content of *I believe* statements. The question before us is whether the two operators should support the following principle ($\rightarrow$ is material implication):[1]

(6) **Mixed extraspection**: $\mathsf{Ass}_{\mathrm{Sp}}(\mathsf{Bel}_{\mathrm{Sp}}(\phi)) \rightarrow \mathsf{Ass}_{\mathrm{Sp}}(\phi)$

Note that (6) is independent, in principle, from both (7) and (8):

(7) **Extraspection for belief**: $\mathsf{Bel}_{\mathrm{Sp}}(\mathsf{Bel}_{\mathrm{Sp}}(\phi)) \rightarrow \mathsf{Bel}_{\mathrm{Sp}}(\phi)$

(8) **Extraspection for commitment**: $\mathsf{Ass}_{\mathrm{Sp}}(\mathsf{Ass}_{\mathrm{Sp}}(\phi)) \rightarrow \mathsf{Ass}_{\mathrm{Sp}}(\phi)$

(7) and (8) are intuitively plausible. (7) says that an agent cannot mistakenly think he has a belief. In a Kripke-model for $\mathsf{Bel}_{\mathrm{Sp}}(\cdot)$, (7) corresponds to density of the accessibility relation ($\forall w_1, w_2 : w_1 R w_2 \rightarrow \exists v : w_1 R v \wedge v R w_2$), which follows from Euclideanity ($\forall w, v_1, v_2 : wRv_1 \wedge wRv_2 \rightarrow v_1 R v_2$). Euclideanity, in turn, correponds to the following principle, which is standardly assumed for belief:

(9) **Negative introspection**: $\neg\mathsf{Bel}_{\mathrm{Sp}}(\phi) \rightarrow \mathsf{Bel}_{\mathrm{Sp}}(\neg\mathsf{Bel}_{\mathrm{Sp}}(\phi))$

(8) says that an agent who is committed to being committed to p is automatically committed to p. This seems also very plausible, and the principle plays a crucial role in Condoravdi and Lauer's (2011) analysis of explicit performatives.

[1] I borrow the term 'extraspection' from van der Hoek (1993). Rieger (2015) calls the principle 'positive belief infallibility'.

But what about the mixed extraspection principle in (6)? Here we are in a bind. The 'natural explanation' for the fact that *I believe* is used as a hedge presupposes that (6) is *not valid* (else, saying *I believe p* is not a way of avoiding commitment to *p*). On the other hand, the 'natural explanation' for Moore's paradox apparently presupposes that (6) *is valid* (else, Moore-paradoxical sentences do not give rise to contradicting commitments).

In the rest of this paper, I will work towards a compositional analysis of *believe*-sentences that avoids this dilemma while maintaining the intuitive core of the 'natural explanations' sketched above. In the present context, the desideratum of compositionality amounts to the following two requirements: First, the sentences in (10) should get the same kind of content, modulo the belief subject and tense.

(10) a. I believe it is raining.
b. John believes it is raining.
c. I believed that it was raining.

Second, *'p'* and *'I believe p'* should be assigned a uniform (declarative) force, with the different implications of the two kinds of sentences tracing to their (different) contents.

2 Diagnosis: Four desiderata

Intuitively, if we want to account for hedging-effects as an implicature, then **mixed extraspection** (6) must fail because, in the titular slogan of Hawthorne et al. (2015), 'belief is weak'. Their slogan (and central thesis) gives rise to a first desideratum for a theory of belief (self-)ascriptions:

(11) **Weakness.** Assertion of '*I believe p*' induces a weaker commitment than assertion of '*p*'.

At the same time, however, the commitment induced by belief self-ascriptions should not be too weak. It must be strong enough to deliver on the following two desiderata:

(12) **Moore's paradox.**
Assertion of '*p, but I don't believe p*' gives rise to inconsistent commitments.

(13) **Consistency.**
Assertions of '*I believe p*' and '*I believe* $\neg p$' jointly give rise to inconsistent commitments.

Consistency in particular requires that the commitment induced by *I believe p* must be stronger than the one induced by *might p*, as witnessed by the non-contradictoriness of (14).

(14) It might be raining, and/but it might also not be raining.

So we want *I believe p* to induce a stronger commitment than *might p*, but a weaker commitment than *p*. This motivates employing a representation of belief that allows for more than two grades of belief. A popular theory of this kind is Bayesian probability. Setting aside the issue of compositionality for a moment, suppose that assertoric commitments are represented as constraints on the speaker's subjective probability distribution P_{Sp} and suppose that assertion of '*p*' and 'I believe *p*' induce the following commitments, for some $\theta_{\mathsf{a}}, \theta_{\mathsf{b}} > 0.5$ (cf. Swanson 2006, Lassiter 2017 on epistemic *must*):

(15) '*p*' induces the commitment $P_{\mathrm{Sp}}(p) \geq \theta_{\mathsf{a}}$.
(16) 'I believe *p*' induces the commitment $P_{\mathrm{Sp}}(p) \geq \theta_{\mathsf{b}}$.

Such a theory meets the desiderata **Moore's paradox** (12) and **consistency** (13), since P_{Sp} cannot assign probability > 0.5 to both p and $\neg p$. However, as it stands, a probabilistic threshold analysis can account for at most one of **weakness** (11) and the following desideratum:[2]

(17) **Closure.**
Assertions of '*I believe* p' and '*I believe* q' commit the speaker to '*I believe* $p \wedge q$'.

To account for **weakness**, it must be that $\theta_{\mathsf{b}} < \theta_{\mathsf{a}} \leq 1$. But then $\theta_{\mathsf{b}} < 1$, and hence **closure** is unaccounted for, because it is always possible to assign probabilities to p and q such that they are larger than θ_{b}, but their conjunction has a probability smaller than θ_{b}.[3]

3 Ranking Theory

We want a theory of graded belief that, together with a theory of assertion, meets all four desiderata: **weakness** (11), **Moore's paradox** (12), **consistency** (13) and **closure** (17).

In the following, I will spell out such a theory in a compositional fashion, employing the *ranking theory* of Spohn (1988, 1990, 2012). The basic construct for the representation of beliefs in the version of ranking theory I am going to use here is that of a *ranking mass function*.[4]

Definition 1 (Ranking mass function, after Spohn 2012, p. 70)**.** *Given a set of worlds W, a* ranking mass function *is any function $k : W \to (\mathbb{N} \cup \{\infty\})$ such that $k^{-1}(0) \neq \emptyset$.*

The set of all ranking mass functions over W is denoted by $\mathbb{K}_W$.

Raidl (t.a.) articulates well how ranking mass functions should be interpreted:

> "One may think of a ranking mass [function] as a doxastic ordering source. The zero worlds are the closest worlds, or the best candidates for the actual world. The greater the rank of a world, the less plausible that world is as a candidate for the actual world, the more it is disbelieved or the more the agent has doubts about it. Worlds ranked with $n < \infty$ are within the doxastic modal horizon (or modal base). Worlds with rank ∞ are crazy worlds outside the modal horizon. Although the agent acknowledges their eventual possibility, she disregards them for matters of actual judgements." (Raidl t.a.)

Even though all our definitions could be stated in terms of ranking mass functions, it is convenient to define the derived notion of a *ranked belief function*.

Definition 2 (Ranked belief functions, after Spohn 2012, p. 75)**.** *Given a ranking mass function k, the* corresponding ranked belief function *β_k is that function $\wp(W) \to (\mathbb{N} \cup \{\infty\})$ such that*

[2] Whether rational belief must satisfy a constraint analogous to **closure** is of some debate in philosophical epistemology, most notably in discussions of the 'lottery paradox' (Kyburg 1961) and 'preface paradox' (Makinson 1965). For a recent discussion of these issues, and a vote for **closure** see Leitgeb (2014).

Here I confine myself to noting that, as an observation about sincere assertion (rather than rational belief), (17) appears to me unassailable: Someone who sincerely asserts *I believe p* and *I believe q*, but then (immediately) goes on to deny or withhold assent from *I believe* $p \wedge q$ clearly has failed up to live up to the commitments he has taken on with his two assertions.

[3] There are more articulated probabilistic analyses of belief that may well satisfy all our desiderata when combined with a suitable analysis of declarative force. Of particular relevance are the proposals in Lin and Kelly (2012) and Leitgeb (2015), which feature probabilistic proposals that can deliver **closure**.

[4] Here and throughout, the notation and some of the terminology departs from previous presentations of this work (including the submitted abstract) to harmonize with other recent presentations of ranking theory, especially Raidl (t.a., 2017). What Raidl (and I) call a 'ranking mass (function)' is called a 'complete pointwise ranking function' in Spohn's work.

(a) $\beta_k(W) = \infty$

(b) $\beta_k(\emptyset) = 0$

(c) *for all non-emtpy* $A \subseteq W$: $\beta_k(A) = \min\{k(w) \mid w \notin A\}$

A ranked belief function assigns to each proposition a measure of its belief (rather than disbelief): The higher the rank of a proposition is, the more it is believed. A crucial property of ranked belief functions is the following, which follows directly from Definition 2:

Fact 3. *For any ranked belief function* β_k *and any proposition* A*: At most one of* $\beta_k(A)$ *and* $\beta_k(W - A)$ *can be larger than* 0.

Ranked belief functions have a number of further notable properties:

Fact 4 (Properties of ranked belief functions)**.** *Any ranked belief function* β_k *is a 'positively minimizing ranking function' in the sense of Spohn (2012). That is:*

(a) $\beta_k(W) = \infty$
The tautology is always absolutely believed.

(b) $\beta_k(\emptyset) = 0$
The contradictory proposition is never believed to a positive degree.)

(c) *For all propositions* A, B: $\beta_k(A \cap B) = \min(\beta_k(A), \beta_k(B))$.
The rank of an intersection of two propositions is the smaller of the rank assigned to the individual propositions.

4 The object language: Syntax, semantics and pragmatics

For simplicity, I work with a simple propositional language, enriched with a family of modal operators for belief. This section spells out its syntax, semantics, and pragmatics.

4.1 Syntax

Definition 5 (Language)**.** *Let* P *and* I *be disjoint sets (of proposition letters and individuals, resp.). Then* $\mathcal{L}_{P,I}$ *is the smallest set such that*

1. $P \subseteq \mathcal{L}_{P,I}$ *(proposition letters)*
2. *If* $\phi \in \mathcal{L}_{P,I}$*, then* $\neg\phi \in \mathcal{L}_{P,I}$. *(negation of formulas)*
3. *If* $\phi, \psi \in \mathcal{L}_{P,I}$*, then* $(\phi \wedge \psi) \in \mathcal{L}_{P,I}$. *(conjunction of formulas)*
4. *If* $\phi \in \mathcal{L}_{P,I}$ *and* $i \in I$*, then* $(\mathsf{Bel}_i(\phi)) \in \mathcal{L}_{P,I}$. *(belief formulas)*

Other connectives are introduced as the usual abbreviations.

4.2 Semantics

Models are standard possible-worlds ones, and the interpretation of non-belief formulas is a standard one. In order to interpret the belief operator, we add two elements to the models: A function K that specifies the believe state for each agent at each world, and a threshold b.[5]

[5] In a system for English, it might be more appropriate to have b provided by the context instead of fixing it lexically. It would also likely be desirable to give the predicate *believe* a degree argument, to account for the fact that *believe* is compatible with what looks like degree modification, as in *I firmly believe p*.

Definition 6 (Models). *A **model** for $\mathcal{L}_{P,I}$ is a qadruple $M = \langle W, V, K, \mathsf{b}\rangle$, such that*

1. *W is a set of possible worlds,*
2. *$V : P \to \wp(W)$ is a valuation for the proposition letters.*
3. *$K : I \times W \to \mathbb{K}_W$ is a function that assigns to each individual-world pair a ranking mass function.*
4. *$0 < \mathsf{b} \in \mathbb{N}$ is the threshold for belief ascriptions.*

Definition 7 (Denotation function). *Given a model $M = \langle W, V, K, \mathsf{b}\rangle$, the **denotation function** $[\![\cdot]\!]^M : \mathcal{L}_{P,I} \to \wp(W)$ is as follows:*

1. *$[\![p]\!]^M = V(p)$ for all $p \in P$.*
2. *$[\![\neg\phi]\!]^M = W - [\![\phi]\!]^M$.*
3. *$[\![\phi \wedge \psi]\!]^M = [\![\phi]\!]^M \cap [\![\psi]\!]^M$.*
4. *$[\![\mathsf{Bel}_i(\phi)]\!]^M = \{w \in W \mid \beta_{K(i,w)}(\phi) \geq \mathsf{b}\}$*

In the following, I will generally omit the model parameter from $[\![\cdot]\!]$.

4.2.1 Admissibility

In order to guarantee some desirable properties, I introduce a notion of *admissibility* for ranking mass functions and models:[6]

Definition 8 (Admissibility of ranking mass functions and models). *Given a language $\mathcal{L}_{P,I}$, model $M = \langle W, V, I, \mathsf{b}\rangle$, a pointwise ranking mass function k is* admissible for $i \in I$ in M *iff* $\forall v \in k^{-1}(0) : K(i, v) = k$
A model is admissible iff $\forall w \in W, i \in I : K(i, w)$ is admissible for i in M.

Admissibility will play a crucial role in the pragmatics, defined in the next subsection. A semantic consequence of requiring admissibility is that it ensures extraspection for belief.

Fact 9 (Extraspection for belief). *For any admissible model M, for all $i \in I, \phi \in \mathcal{L}_{P,I}$:* $[\![\mathsf{Bel}_i(\mathsf{Bel}_i(\phi))]\!] \subseteq [\![\mathsf{Bel}_i(\phi)]\!]$

4.3 Pragmatics

4.3.1 Commitment frames and states

The commitments an agent has are represented as *constraints on ranking functions.*

Definition 10 (Commitment frames). *A* commitment frame *for $\mathcal{L}_{P,I}$ is any $\mathfrak{C}_\mathsf{i} = \langle M, i, \mathsf{a}\rangle$:*

1. *$M = \langle W, V, K, \mathsf{b}\rangle$ is a model for $\mathcal{L}_{P,I}$.*
2. *$i \in I$ is an individual.*
3. *$\mathsf{b} < \mathsf{a} \in \mathbb{N}$ is the* assertion threshold.

[6] For a systematic investigation of a 'ranked semantics' of the type used here, including correspondence results, see Raidl t.a., 2017).

Definition 11 (Commitment states)**.** *Given a commitment frame* $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$*, a* commitment state C_i *is partial truth function of ranking mass functions such that* $C_i(k)$ *is defined iff* k *is admissible for* i *in* M*. There are two distinguished commitment states:*

(a) $\bot = \lambda k.0$ *(the contradictory state)*

(b) $\top = \lambda k.1$ *(the uncommitted state)*

4.3.2 Updates for commitment states

The commitment dynamics is essentially the same as in Veltman's (1996) *Update Semantics*.[7]

Definition 12 (Declarative update)**.** *Given a commitment frame* $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$*, the* update function $+$ *is the following function from commitment states and formulas to commitment states:*

$$C + \phi = \lambda k.C(k) \;\&\; \beta_k([\![\phi]\!]) \geq \mathsf{a}$$

Definition 13 (Support)**.** *For any commitment state* C *and formula* ϕ*:*

$$C \vDash \phi \textit{ iff } C + \phi = C$$

We immediately obtain the following minimal requirement for a notion of commitment. If an agent is committed, in C_i to a proposition ϕ, then update with any proposition ψ that is incompatible with ϕ results in the inconsistent state.

Fact 14 (Inconsistency of commitments)**.** *Let* C_i *be a commitment state, and for some* ϕ, ψ *such that* $[\![\phi]\!] \cap [\![\psi]\!] = \emptyset$ *and* $C_i \vDash \phi$*. Then* $C_i + \psi = \bot$*.*

Proof. Follows immediately from the following Fact 15. □

Further, we can show that update with a conjunction is equivalent to subsequent update with the two conjuncts:

Fact 15 (Conjunction)**.** *For any commitment state* C_i *and formulas* ϕ, psi *:* $C_i + (\phi \wedge \psi) = C_i + \phi + \psi$*.*

Proof. First, note that

(*) For any k : $\beta_k([\![\phi \wedge \psi]\!]) = \beta_k([\![\phi]\!] \cap [\![\psi]\!]) = \min(\beta_k([\![\phi]\!]), \beta_k([\![\psi]\!]))$. (by Fact 4c)

We have to show that for any commitment state C_i and ranking mass function k: $[C_i + \phi \wedge \psi](k) = 1 \Leftrightarrow [C_i + \phi + \psi](k) = 1$. ($\Rightarrow$) Suppose that $[C_i + \phi \wedge \psi](k) = 1$, then $\beta_k([\![\phi \wedge \psi]\!]) \geq \mathsf{a}$, hence by (*) $\min(\beta_k([\![\phi]\!]), \beta_k([\![\psi]\!])) \geq \mathsf{a}$ and so $k([\![\phi]\!]) \geq \mathsf{a}$ and $k([\![\psi]\!]) \geq \mathsf{a}$. Further, since $[C_i + \phi \wedge \psi](k) = 1$, it must be that that $C_i(k) = 1$. But then $[C_i + \phi \wedge \psi](k) = 1$. ($\Leftarrow$) Suppose $[C_i + \phi + \psi](k) = 1$. Then $C_i(k) = 1$ and $k([\![\phi]\!] \geq \mathsf{a}$ and $k([\![\psi]\!]) \geq \mathsf{a}$. But then $\min(\beta_k([\![\phi]\!]), \beta_k([\![\psi]\!])) \geq \mathsf{a}$, hence by (*) $\beta_k([\![\phi \wedge \psi]\!]) \geq \mathsf{a}$. But then $[C_i + \phi \wedge \psi](k) = 1$. □

[7] I could have brought this out even more clearly by letting commitment states be sets of ranking functions. Then the contradictory state would be $\emptyset$, the uncommitted state would be $\{k \mid k \text{ is admissible for } i\}$, and the update operation $+\phi$ would be intersection with $\{k \mid \beta_k(\phi) > \mathsf{a}\}$.

For present purposes, such a set-based representation would have been sufficient. I opted for the truth-function representation instead because it generalizes better to additional phenomena (not treated here) and because it highlights the idea that commitment states are to be thought of as constraints on ranking functions.

4.4 Predictions

4.4.1 Desideratum 1: Closure

Fact 16 (Closure explained)**.** *For any commitment state C_i:* $C + (\mathsf{Bel}_i(\phi)) + (\mathsf{Bel}_i(\psi)) \models \mathsf{Bel}_a(\phi \wedge \psi)$.

Proof. Direct corollary of Fact 15. □

4.4.2 Desideratum 2: Consistency

Fact 17 (Consistency)**.** *For any commitment state C_i and any formula ϕ:* $C_i + \mathsf{Bel}_i(\phi) + (\mathsf{Bel}_i(\neg\phi)) = \bot$

Proof. Obviously, $C_i + \mathsf{Bel}_i(\phi) \models \mathsf{Bel}_i(\phi)$ and $[\![\mathsf{Bel}_i(\phi)]\!] \cap [\![\neg\mathsf{Bel}_i(\phi)]\!]$. But then, by Fact 14, $C_i + \mathsf{Bel}_i(\phi) + (\mathsf{Bel}_i(\neg\phi)) = \bot$. □

4.4.3 Desideratum 3: Weakness

Fact 18 (Weakness)**.** $C_i + \mathsf{Bel}_i(\phi) \nvDash \phi$.

Proof. Let $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$ with $M = \langle W, V, K, \mathsf{b}\rangle$ such that there are $w, v \in W$ such that $\beta_{K(i,w)}(V(p)) = \mathsf{b}$ and $\beta_{K(i,v)}(V(p)) = \mathsf{a}$. Recall that it must be that $\mathsf{a} > \mathsf{b}$. Then $\top + \mathsf{Bel}_i(p)$ is true of $K(i,w)$ and $K(i,v)$, but $\top + \mathsf{Bel}_i(p) + p$ is true of $K(i,v)$, but not of $K(i,w)$. But then $\top + \mathsf{Bel}_i(p) \neq \top + \mathsf{Bel}_i(p) + p$ and hence $\top + \mathsf{Bel}_i(p) \nvDash p$. □

Note that even though we require that $\mathsf{b} < \mathsf{a}$, of course nothing prevents an agent who assigns rank a to ϕ from asserting $\mathsf{Bel}_i(\phi)$. Formally: A ranking mass function k such that $\beta_k([\![\phi]\!]) \geq \mathsf{a}$ is compatible with commitment state $C_i + \mathsf{Bel}_i(\phi)$ (provided that k is compatible with the original C_i). This is as it should be: Recall that the 'natural explanation' for hedging with *I believe* that we started out treats the hedging-inference as a conversational implicature. Thus the update with $\mathsf{Bel}_i(\phi)$ should not preclude that the agent is absolutely certain that ϕ, it should only be compatible with her not being absolutely certain.

4.4.4 Desideratum 4: Moore's paradox

Fact 18 assures us that, in our system **mixed extraspection** fails. At the same time, Lemma 19 assures us that belief is not *too* weak: Declarative update with *I believe that* ϕ does not induce full assertoric commitment to ϕ (as **mixed extraspection** would have it), but it *does* induce a commitment to ϕ that goes over and above the commitment to $\mathsf{Bel}_i(\phi)$.

Lemma 19 (Belief update)**.** *Let $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$ with $M = \langle W, V, K, \mathsf{b}\rangle$. Then for any commitment state C_i and formula ϕ:*

$$\textit{For any } k\textit{: if } [C_i + \mathsf{Bel}_i(\phi)](k) = 1 \textit{ then } k([\![\phi]\!]) \geq \mathsf{b}$$

Proof. Suppose that for some $k : [C_i + \mathsf{Bel}_i(\phi)](k) = 1$. Then it must be that $\beta_k([\![\mathsf{Bel}_i(\phi)]\!] \geq \mathsf{a} > 0$, which implies that $k^{-1}(0) \subseteq [\![\mathsf{Bel}_i(\phi)]\!]$. Let $w \in k^{-1}(0)$ (such w must exist by Definition 1). Since $w \in [\![\mathsf{Bel}_i(\phi)]\!]$, $K(i,w)([\![\phi]\!] \geq \mathsf{b}$. But, by the admissibility requirement on commitment states, $k = K(i,w)$. So $k([\![\phi]\!]) \geq \mathsf{b}$. □

In addition, we have the converse of Fact 18: Declarative update with φ ensures that assertoric commitment to $\mathsf{Bel}_i(\varphi)$, which follows directly from the requirement that $\mathsf{a} > \mathsf{b}$:

Fact 20 (Assertoric commitment implies belief commitment)**.** *For all* $C_i, \varphi : C_i + \varphi \vDash \mathsf{Bel}_i(\varphi)$.

With these two results in place, we can show that our system accounts for **Moore's paradox**.

Corollary 21 (Moore's paradox explained)**.** *For any commitment state* C_i*:*

(a) $C_i + (\neg\phi \wedge \mathsf{Bel}_i(\phi)) = \bot$ *(b)* $C_i + (\phi \wedge \mathsf{Bel}_i(\neg\phi)) = \bot$ *(c)* $C_i + (\phi \wedge \neg\mathsf{Bel}_i(\phi)) = \bot$

Proof. (a) By Fact 15, it is sufficient to show that $C_i + \neg\phi + \mathsf{Bel}_i(\phi) = \bot$. Suppose otherwise, i.e. that there is k such that $[C_i + \neg\phi + \mathsf{Bel}_i(\phi)](k) = 1$. Clearly, it must be that $\beta_k([\![\neg\phi]\!]) = \geq \mathsf{a} > 0$ and, by Lemma 19, it must be that $\beta_k([\![\phi]\!]) \geq \mathsf{b} > 0$. But (Fact 3), a belief function can assign positive rank to at most one of $[\![\phi]\!]$ and $[\![\neg\phi]\!]$. Contradiction. (b) Analogous. (c) Again, it is sufficient to show that $C_i + \phi + \neg\mathsf{Bel}_i(\phi) = \bot$. By Fact 20, $C_i + \phi \vDash \mathsf{Bel}_i(\phi)$. Since $[\![\mathsf{Bel}_i(\phi)]\!] \cap [\![\neg\mathsf{Bel}_i(\phi)]\!] = \emptyset$, by Fact 14, $C_i + \phi + \neg\mathsf{Bel}_i(\phi) = \bot$. □

The cases (a-c) in this corollary correspond to the Moorean sentence we have been working with (2) and the two readings of the classical Moore-sentence, (18) and (19):

(2)	It is not raining but I believe it is.	Corollary (21a)
(18)	It is raining but I believe it is not raining.	Corollary (21b)
(19)	It is raining but it is not the case that I believe it is raining.	Corollary (21c)

5 Conclusion

In this paper, I have combined a ranking-theoretic semantics for *believe* with an update-semantics style commitment-based account of declarative force. I have shown that the resulting system compositionally accounts for Moore's paradox, and at the same time can account for the fact that *I believe* functions as a hedge, while predicting a number of other desiderata.

It is noteworthy that the factor that motivated me to employ ranking theory is the desideratum **closure**. I still think that **closure** is desirable, and so it is encouraging to see that the two main phenomena of interest can be jointly accounted for with a theory of ranked belief that delivers on **closure**. However, committed Bayesians may be willing to jettison the principle.

The good news for such a Bayesian is that the commitment-based account as developed here can be combined with a probabilistic threshold account, in a way that predicts all of our desiderate except **closure**. Here is how: Replace the function K in our models with a function $\mathbb{P}$ that assigns to each agent-world pair a probability distribution, and replace b by a threshold $\theta_{\mathsf{b}} > 0.5$. Admissibility then amounts to the following: A probability distribution P is admissible for i iff for all worlds w in the *support* of $P : \mathbb{P}(i, w) = P$, and a model is admissible iff $\mathbb{P}$ only assigns admissible probability distributions. Commitment states become truth functions of admissible probability distributions, the assertion threshold is $\theta_{\mathsf{a}} \in \mathbb{R}$ such that $\theta_{\mathsf{a}} > \theta_{\mathsf{b}}$. The update operation is adjusted in the obvious way: $C_i + \phi = \lambda P.C_i(P)\ \&\ P(\phi) \geq \theta_{\mathsf{a}}$.

What this shows is that the choice of a theory of graded belief can be made on independent grounds. People like me, who find **closure** compelling will find the version of the account proposed in the main text attractive and comforting. Committed Bayesians, being Bayesians, may prefer the account sketched in the previous paragraph. The contribution of the present paper is not an argument for one representation of graded belief over another, but rather in that it shows that Moore's paradox and hedging with *I believe* can be explained in the 'natural ways' sketched at the beginning of the paper, that both can be explained together, and that this can be done compositionally, by combining a graded theory of belief with a commitment-based understanding of declarative force.

References

Condoravdi, C. and Lauer, S.: 2011, Performative verbs and performative acts, *in* I. Reich, E. Horch and D. Pauly (eds), *Sinn and Bedeutung 15*, Universaar, Saarbrücken, pp. 149–164.

Hawthorne, J., Rothschild, D. and Spectre, L.: 2015, Belief is weak, *Philsophical Studies* **173**(5).

Krifka, M.: 2014, Embedding illocutionary acts, *in* T. Roeper and P. Speas (eds), *Recursion, Complexity in Cognition*, Springer, Berlin, pp. 125–155.

Krifka, M.: 2015, Bias in commitment space semantics: Declarative questions, negated quetions, and question tags, *Semantics and Linguistic Theory* **25**, 328–345.

Kyburg, H. E.: 1961, *Probability and the Logic of Rational Belief*, Wesleyan University Press.

Lassiter, D.: 2017, *Graded Modality: Qualitative and Quantitative Perspectives*, Oxford University Press.

Lauer, S.: 2013, *Towards a dynamic pragmatics*, PhD thesis, Stanford University.

Leitgeb, H.: 2014, The review paradox: On the diachronic costs of not closing rational belief under conjunction, *Noûs* **48**(4), 781–793.

Leitgeb, H.: 2015, The Humean thesis on belief, *Aristotelian Society Suppl. Vol.* **89**(1), 143–185.

Lin, H. and Kelly, K. T.: 2012, A geo-logical solution to the lottery paradox, *Synthese* **186**, 531–575.

Makinson, D.: 1965, The paradox of the preface, *Analysis* **25**(6), 205–207.

Raidl, E.: 2017, Completeness for counter-doxa conditionals–using ranked semantics. ms., University of Konstanz.

Raidl, E.: t.a., Ranking semantics for doxastic necessities and conditionals, *The Logica Yearbook 2017*.

Rieger, A.: 2015, Moore's paradox, introspection and doxastic logic, *Thought* **4**, 215–227.

Searle, J. R.: 1969, *Speech Acts: An essay in the philosophy of language*, Cambridge University Press.

Spohn, W.: 1988, Ordinal conditional functions: A dynamic theory of epistemic states, *in* W. Harper and B. Skyrms (eds), *Causation in Decision, Belief Change, and Statistics*, Kluwer, pp. 105–134.

Spohn, W.: 1990, A general non-probabilistic theory of inductive reasoning, *in* R. Shachter, T. Levitt, J. Lemmer and L. Kanal (eds), *Uncertainty in Artificial Intelligence (Volume 4)*, pp. 149–158.

Spohn, W.: 2012, *The laws of belief*, Oxford University Press, Oxford, UK.

Swanson, E.: 2006, *Interactions with context*, PhD thesis, MIT, Cambridge, MA.

van der Hoek, W.: 1993, Systems for knowledge and belief, *Journal of Logic and Computation* **3**(2), 173–195.

Veltman, F.: 1996, Defaults in update semantics, *Journal of Philosophical Logic* **25**, 221–261.