Responses to reviewer comments for “Expertise and knowledge: a modal logic perspective”

Dear editor and reviewers,

We thank you for your detailed comments and suggestions for improving the paper. The typos, minor points on presentation and bibliographic errors raised by both reviewers have been fixed in the revised paper. Please find more detailed responses to each reviewer below.

Reviewer 1

[**TODO:** General intro responses?]
Comment: *In order to fully understand the intuitive idea of “soundness”, I had to go to Singleton (2021). The way it’s phrased here, one would think that everything which entails a proposition about which the source is an expert would be sound. Of course, this would mean that reporting a contradiction like $p \wedge \neg p$ would be vacuously sound, which we see is not the case when the semantics is introduced.*

Response: [**TODO:** Response.]

Comment: *I know what this sentence is driving at, but I don’t quite agree with it on an intuitive level. Say A has a PhD on how the weather affects mental health (but she’s not an expert on mental health broadly speaking), and B is a weatherman; between the two of them they do not make an expert on mental health, unless A was an expert already. I guess there is a subtle* *but important distinction between being an expert on the assertion "φ implies ψ' and being an expert on the way in which φ affects ψ.*

Response: [**TODO:** Response.]

Comment: *Page 13: In 1-agent S5 models one can assume without loss of generality that the "knowledge" modality is universal, for the equivalence class partitions the model into the generated submodels given by the equivalence classes, each of them with a universal relation. Of course, if we attempt to do this in this framework we obtain that the* A *and the* K *modality have the same semantics; I would have loved to see some discussion on why this does not work in the present framework.*

Response: The difference here is that we work with the language $\mathcal{L}_{\mathsf{KA}}$, which includes not just the knowledge modality K but also the universal modality A. Since in the relational semantics over this language $\mathsf{A}\varphi$ is always interpreted with the universal relation $X \times X$, the notation of a generated submodel trivialises: the whole space X itself is the only set closed under the universal relation. For this reason, the submodels corresponding to the equivalence classes do not preserve the truth value of modal formulas, as would usually be the case with just the K modality, and thus the assumption that knowledge is universal is *not* without loss of generality.

Comment: *Page 15: if I'm not mistaken, Corollary 1 follows not from the immediately previous Lemma 8 but from Lemma 7 plus the (EA) axiom: if* $\mathsf{E}\varphi \in \Sigma$ *then (by (EA))* $\mathsf{AE}\varphi \in \Sigma$ *which implies (by the definition of R)* $\mathsf{E}\varphi \in \Gamma$.

Response: This is absolutely correct for the second part of Corollary 1. One can also show the first part ($\mathsf{A}\varphi \in \Gamma$ iff $\mathsf{A}\varphi \in \Delta$) without Lemma 8 by additionally showing that **4** ($\mathsf{A}\varphi \to \mathsf{AA}\varphi$) is provable in **KT5**. While this derivation is fairly short, we do need Lemma 8 elsewhere in the paper (e.g. in the truth lemma, Lemma 9). For this reason (and to save limited space), we have kept Corollary 1 *after* Lemma 8.

Comment: *Page 21: since the P_j's in this section are Alevandrov topologies, it would be nice to bring up the notion of topological join when defining P_J^{dist}.*

Response: This is a good point, and we have added a paragraph to point out this connection after the introduction of P_J^{dist}.

Reviewer 2

[**TODO:** Journal extension of Singleton (2021)? What can we say here?]
[**TODO:** Title change?]

Comment: *My third worry is whether there is any inconsistency between the logical designs and the example used in Abstract, which states that "if a source has expertise on φ but not ψ, when the conjunction $\varphi \wedge \psi$ is sound whenever φ holds, since we can ignore ψ (on which the source has no expertise)". It looks that the example in Abstract aims to argue that the principle "$\mathsf{E}\varphi \wedge \neg\mathsf{E}\psi \wedge \varphi \to \mathsf{S}(\varphi \wedge \psi)$" is valid, but it is not hard to construct a counterexample to it. This makes me doubt whether or not the logical design developed in the paper really captures the example in its motivation. Definitely, it may become valid if we restrict the class of models, but what the desired conditions should be?*

Response: This is a good point, and indeed the principle extrapolated from the abstract is not in general valid. As suggested, there are some further restrictions upon which it becomes valid, e.g. if one additionally assumes expertise is closed under (finite) intersections, that the source does not have expertise on any non-empty set strictly stronger than φ (in the sense that $\emptyset \subset A \subseteq \|\varphi\|$ implies $A = \|\varphi\|$ for all $A \in P$) and that $\varphi \wedge \psi$ is consistent. This is the case in Example 2, for instance.

Since these additional assumptions are not suitable to introduce the intuitive concept of soundness before the semantics are defined, we have removed the example from the abstract, replacing it with the following:

> Closely connected with expertise is a notion of *soundness of information*: φ is said to be "sound" if it is true *up to lack of expertise* of the source. **That is, any statement logically weaker than φ on which the source has expertise must in fact be true. This is relevant for modelling situations in which sources make claims beyond their domain of expertise.**

We have also reworked the same example in the introduction:

> This formalises the idea of "filtering out" parts of a statement within a source's expertise. For example, suppose $\varphi = p \wedge q$, and the source has expertise on p but not q. Supposing p is true but q is false, φ is false. However, if we discard information by ignoring q (on which the source has no expertise), we obtain the weaker formula p, on which the source *does* have expertise, and which is true. **If this holds for all possible ways to weaken $p \wedge q$ (this is the case, for instance, if the source does not have expertise on any statement strictly stronger than p), then $p \wedge q$ is *false* but *sound* for the source to report.**

We believe the revised versions better match the logical framework which we go on to introduce.

[**TODO:** Check these quotes match the final versions.]

[**TODO:** Shall we point out that their proposed counterexample is not actually a counterexample?]