Responses to Review Comments

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

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: We agree with this comment, and have changed the informal descriptions of soundness in the abstract and introduction. Reviewer 2 also noted this issue and put forward a counterexample to this principle. We refer to the response to reviewer 2 for the extra assumptions one needs in order for this principle to hold, and the changes we have made to the text for clarification.

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: *In general, I think this introduction could do a better job at motivating the paper, and could perhaps benefit from being a few lines longer*

Response: In addition to changes described above, we have added an extra paragraph with some example settings in which reasoning about expertise is relevant, in an effort to better motivate the problem we study.

> Problems associated with expertise have been exacerbated recently by the COVID-19 pandemic, in which false information from non-experts has been shared widely on social media (Van Dijck & Alinejad, 2020; Llewellyn, 2020). There have also been high-profile instances of experts going beyond their area of expertise to comment on issues of public health (Xaudiera & Cardenal, 2020), highlighting the importance of *domain-specific* notions of expertise. Identifying experts is also an important task for *liquid democracy* (Blum & Zuber, 2016), in which voters may delegate their votes to expertise on a given policy issue.

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 Alexandrov 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}.

> P_J^{dist} also has a topological interpretation. As in Section 4, each P_j gives rise to an Alexandrov topology τ_j (where P_j are the closed sets) if it is closed under unions and intersections. By the aforementioned properties, τ_J^{dist} corresponds to the coarsest Alexandrov topology finer than each τ_j. On the other hand, since the join (in the lattice of topologies on X) of finitely many Alexandrov topologies is again Alexandrov (Steiner, 1966, Theorems 2.4, 2.5), it follows that τ_J^{dist} is equal to the join $\bigvee_{j \in J} \tau_j$.

Note that we have taken a slightly different interpretation to the reviewer's original suggestion, by viewing each P_j as the *closed* sets of a topology, i.e. $\tau_j = \{A \subseteq X \mid X \setminus A \in P_j\}$ (since P_j is closed under unions and intersections, so too is τ_j). This is to match our earlier topological interpretations in sections 2, 4 and 5, e.g. so that $\|\mathsf{S}\varphi\|$ is the closure of $\|\varphi\|$.

Reviewer 2

Comment: *Also, I think that the title "Expertise and knowledge: a modal logic perspective" is confusing in some sense. It may make one think of a logic on both expertise and knowledge, but the main contribution of the paper is logics just on expertise.*

Response: We have now changed the title to "Expertise and information: an epistemic logic perspective". We hope this eliminates the suggestion that we introduce new logics of knowledge, and instead refer to the existing literature on epistemic logic. We also mention "information" in the title, since the core notion of *soundness* in the paper is centred around non-expert information.

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. It seems the reviewer's model does not quite constitute a countermodel of $\mathsf{E}\varphi \wedge \neg\mathsf{E}\psi \wedge \varphi \to \mathsf{S}(\varphi \wedge \psi)$, since it shows $M, 2 \models \mathsf{E}p \wedge \neg\mathsf{E}q \wedge p$ – so that φ is p and ψ is q – but $M, 2 \not\models \mathsf{S}(p \wedge \neg q)$. Note that q is negated, so this corresponds to $\mathsf{S}(\varphi \wedge \neg\psi)$ instead of $\mathsf{S}(\varphi \wedge \psi)$. However, a minor modification yields a counterexample: taking $P = \{\{1, 2\}, \{2\}\}$ instead, we have $M, 1 \models \mathsf{E}p \wedge \neg\mathsf{E}q \wedge p \wedge \neg\mathsf{S}(p \wedge q)$.

As suggested, there are some further restrictions upon which the principle from the abstract 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 we go on to study.