Inquiry
An Interdisciplinary Journal of Philosophy

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/sinq20

Question-relative knowledge for minimally rational agents

Francisca Silva

To cite this article: Francisca Silva (10 Apr 2024): Question-relative knowledge for minimally rational agents, Inquiry, DOI: 10.1080/0020174X.2024.2338794

To link to this article: https://doi.org/10.1080/0020174X.2024.2338794

Published online: 10 Apr 2024.

INQUIRY
https://doi.org/10.1080/0020174X.2024.2338794

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Question-relative knowledge for minimally rational agents

Francisca Silva

Department of Philosophy, University of St Andrews, St Andrews, UK

ABSTRACT
Agents know some but not all logical consequences of what they know. Agents seem to be neither logically omniscient nor logically incompetent. Yet finding an intermediate standard of minimal rationality has proven difficult. In this paper, I take suggestions found in the literature [Lewis, D. 1988. Relevant Implication. *Theoria* 54 (3): 161174. https://doi.org/10.1111/theo.1988.54.issue-3; Hawke, P., A. Özgün, and F. Berto. 2020. The Fundamental Problem of Logical Omniscience. *Journal of Philosophical Logic* 49 (4): 727766. https://doi.org/10.1007/s10992-019-09536-6; Plebani, M., and G. Spolaore. 2021. Subject Matter: A Modest Proposal. *The Philosophical Quarterly* 71 (3): 605622. https://doi.org/10.1093/pq/pqaa054] and join the forces of subject matter and impossible worlds approaches to devise a new solution to this quandary. I do so by combining a space of FDE worlds [Berto, F., and M. Jago. 2019. *Impossible Worlds*. Oxford University Press.] with a Lewisian (1988) understanding of subject matters as partitions. By doing so, I show how subject matters impose some order in the anarchic space of FDE worlds, while the worlds allow for distinctions between contents which would not otherwise be available. Combining the two approaches, then, brings us closer to the desired closure principles for knowledge of minimally rational agents.

ARTICLE HISTORY Received 29 April 2023; Accepted 28 March 2024

KEYWORDS Impossible worlds; subject matters; minimal rationality; logical omniscience

1. Introduction

The so-called problem of logical omniscience has been a central topic of discussion among epistemic logicians for several decades now (Berto and Jago 2019; Bjerring and Skipper 2019; Hawke, Özgün, and Berto 2020; Hintikka 1975; Jago 2007, 2014; Hoek, forthcoming) regarding how to more realistically model human-like cognitive agents. These are agents who, in spite of failing to uphold ideal standards of rationality, are still

CONTACT Francisca Silva fs97@st-andrews.ac.uk 5 The Scores, St Andrews KY16 9AL, UK

logically competent. This is known in the literature as the project of accounting for minimal rationality (Hoek, forthcoming).

Two influential lines of approach to this problem have been to extend modal space to include impossible worlds (Berto and Jago 2019; Bjerring and Skipper 2019; Jago 2014) and to introduce a second component of meaning – a sentence's topic or subject matter (Berto 2022; Hawke, Özgün, and Berto 2020). Theories in the former camp face a dilemma: either they end up validating closure principles that are too demanding, or they are based on a space of worlds that is too fine-grained if no conditions are imposed (Bjerring 2013; Bjerring and Skipper 2019; see also Berto and Jago 2019; Jago 2014 for further discussion). Theories in the latter camp, on the other hand, have been criticized for not being able to distinguish between all contents towards which agents might have differing epistemic attitudes. These include the case of an unknown necessity and the disjunction of it and its negation (Hawke, Özgün, and Berto 2020, 748–749).

These objections are further explored below, but I take them to be generally correct. There have been, however, illuminating remarks in the subject-matter literature (Fine 2020; Hawke, Özgün, and Berto 2020; Lewis 1988; Plebani and Spolaore 2021) regarding the possible benefits of introducing impossible worlds to various accounts of subject matter.[1]

In this paper, I take this suggestion head-on and argue that combining the strengths of impossible worlds and subject matter provides us with a promising solution to the problem of logical omniscience. The core idea of the paper is that while impossible worlds serve the purpose of distinguishing between contents, subject matter serves the purpose of delimiting what would otherwise be a too-anarchic space of worlds, while also introducing distinctions of its own accord.

In Section 2, I begin by presenting in greater detail the problem of logical omniscience. In Section 3, I motivate the impossible worlds solution developed by Berto and Jago (2019) and Jago (2014) as well as the dilemma it faces in terms of modelling logically competent but non-ideal agents. In Section 4, I motivate what I take to be the best view of subject matter in an intensional framework, and I present the issue it faces when it comes to differentiating between necessary propositions and the disjunctions of them and their negations. In Section 5, I combine the approaches previously considered and propose a new solution to the problem of logical omniscience. In Section 6, I highlight

[1]One could equally well here talk of impossible states, following Fine (2020): I am in agreement with Berto and Jago (2023) that apart from a difference in focus on the relation of exact entailment, truthmaker semantics and impossible world semantics are notational variants.

the benefits of the current proposal, notably the closure principles it validates and invalidates, as well as its limitations (that it validates adjunction and monotonicity). In Section 7, I compare my approach to other views in the literature, namely, to those of Bjerring and Skipper (2019) and of Berto (2022), finding much in common and presenting reasons in favour of my own. I also present a non-monotonic version of my account inspired by Berto's. Finally, in Section 8, I take stock and conclude.

2. Logical omniscience and necessary equivalence

One of the most prominent applications of possible worlds semantics, due to Hintikka (1962), has been to epistemic logic. Treating knowledge as a modal operator, K, Hintikka proposed to characterize knowledge as 'truth in all the epistemically accessible worlds', with modal epistemic logic then imposing different conditions on knowledge depending on the conditions imposed on the accessibility relation between the possible worlds. Regardless of the conditions imposed on the accessibility relation, however, some conditions are imposed on knowledge by default, simply by virtue of its being defined on a space of possible worlds. Two of these have to do with how entailment is defined and with how what is known – the propositions – are individuated in the space of possible worlds. In standard possible world semantics, a given proposition, $\boldsymbol{p}$, entails another, $\boldsymbol{q}$, if and only if $p \subseteq q$ in all models. Propositions are understood as entities without structure, as simply the set of possible circumstances in which they are true, that is, a set of possible worlds. Entailment is therefore defined in terms of the subset relation and propositions are understood as sets of possible worlds. Consequently, when it comes to the logic of knowledge agents know all the logical consequences of what they know and if agents (fail to) know a proposition, they thereby (fail to) know all propositions that are logically or necessarily equivalent to it. I designate these conditions respectively LOGICAL OMNISCIENCE and NECESSARY EQUIVALENCE for modal epistemic logic.

To appreciate how LOGICAL OMNISCIENCE arises, consider that in all the possible worlds where p is true, q is also true. So if p is just a set of possible worlds, and knowledge is just a relation between an agent and a set of possible worlds, then the agent will bear that relation to everything that is true in the worlds on which p is defined. So, if Kp is true at a given world, the agent will also know everything that the worlds accessible from that world represent as being the case. By definition, such worlds will represent all t such that $p \subseteq t$, for they are included in them. Agents fall very short of knowing all the logical consequences of what they know, however. To know all such consequences is plausibly *not* a requirement of rationality – quite the opposite, given how varied and irrelevant some logical consequences of what we know are. We might have strong normative reasons not to clutter up our minds in that way (Harman 1986). One could retort that logically omniscient agents, so described, nevertheless represent a normative ideal that cannot be achieved. However, Harman's conclusions raise doubts as to whether that is the case, for even if it were practically possible, it is doubtful that we should want to come to know all the logical consequences of what we know since only a very select few of them would bear any weight on our goals.

To appreciate how NECESSARY EQUIVALENCE arises, consider that, since propositions are just sets of possible worlds, when two propositions are true in the same possible worlds, they correspond to the same set. This has as a consequence that there is only one necessary proposition – the set of all possible worlds – and one impossible proposition – the empty set of worlds. However, it seems that agents can know some necessary truths – some truths of mathematics, for instance – while ignoring other necessary truths. Or suppose, with Kripke (1980), that identities such as *Water is* H_2O are necessary. Then according to this picture of content whenever an agent knows that they are drinking water, they know that they are drinking H_2O. However, it seems that before the chemical structure of water was discovered, this was clearly not the case: agents who knew the former proposition certainly did not know the latter.

Philosophers like Soames (2010) might retort that the agents *did know* that they were drinking H_2O, they simply did not know it under the guise of the sentence 'I am drinking H_2O', as opposed to under the guise of the sentence 'I am drinking water'. While we could appeal to a distinction between propositions and the linguistic guises under which they appear, in order to explain away hyperintensional distinctions, this strategy does not seem to be general enough. If someone was transported to twin-Earth from Earth before the discovery that water was H_2O had taken place, and they were confronted later on with the discovery that the liquid they thought was water was *XYZ*, they would think that water was *XYZ*. But someone who knows that water is H_2O would instead know that *XYZ* isn't water. And this seems to hang on the fact that the relevant knowledge is not merely of propositions which can appear under different guises, but of contents – and substances – themselves.

I proceed then by favouring the intuition that something extralinguistic is being appealed to in the cases where agents assent to knowing that they're drinking water but not to knowing that they're drinking H_2O: they fail to possess an item of knowledge *about* the chemical structure of what they're drinking. This allows for a uniform treatment of all such cases, avoiding the introduction of guises into the theory. Of course, much can be accomplished by introducing guises into a theory of knowledge, but my stance is that they should be saved as a last resource for when it is uncontroversial that what is grounding an agent's diverging assent to a pair of intensionally (or otherwise) equivalent propositions is a linguistic misgiving on the agent's part.

In the sections that follow I look into proposed ways of blocking these two unwelcome consequences for the logic of knowledge, starting with the idea, from Hintikka (1975), of enriching the space of possible worlds with impossible worlds – ways that the world could *not* have been. Impossible worlds can be understood as valuation points in which necessities can be false and/or necessarily equivalent propositions can be attributed to distinct truth-values.

3. Impossible worlds and Bjerring's problem

Taking seriously the option of extending the space of worlds with ways the world could not have been, one hasn't still been given enough guidance on how to proceed. For instance, should impossible worlds be maximal and assign a truth-value to all propositions, but non-exclusive, so that they might assign both *True* and *False* to the same proposition? Or should impossible worlds (also) be allowed to not assign any truth-value to some propositions, so that unlike possible worlds, they can also be non-maximal, aside from being inconsistent?

If we think of impossible worlds as representing *all* the ways the world could *not* have been, then impossible worlds should be allowed to be both inconsistent and non-maximal. If we think of worlds as ways for spatio-temporally isolated universes to be (here I'm following Lewis (1986) in taking universes to be what Van Inwagen (1986) calls 'Lewis-style worlds'), then any world is of necessity maximal, so if a world is non-maximal, even if consistent, it is not a way the universe could have been, so we may aptly describe it as an impossible world.[2] Similarly,

[2]Note that these are ways that the universe could not have been, not ways that *things* could not have been. Had we taken this latter interpretation, then partial/non-maximal but consistent states would be taken as possible states that are parts of worlds. How we refer to these states and whether we take them to be possible states or impossible worlds, is a terminological/classificatory point that nonetheless has great theoretical import, for impossible world theories look very different if they include only maximal states.

(including all possible alternative universes, or only our own, if only one such universe exists) is of necessity consistent, so every world that is inconsistent, even if maximal, is not a way the universe could have been.

I follow Berto and Jago (2019) and Nolan (1997) in accepting the following parallel clauses for possible and impossible propositions:

- If *p* is possible then there is a possible world in which *p*.
- If *p* is impossible then there is an impossible world in which *p*.

Importantly, these clauses tell us only what worlds must represent, and are therefore silent on whether impossible worlds should fail to represent anything. That is, however, a crucial question with respect to LOGICAL OMNISCIENCE. Arguably, the problem was that agents' knowledge in a setting with only possible worlds is closed under classical logical consequence. Possible worlds are such that for any classically valid inference $\Gamma \vdash_C \phi$, they represent ϕ whenever they represent all $\gamma \in \Gamma$. The parallel clauses above allow for an interpretation according to which impossible worlds are valuation points such that for any inference in a weaker logic, *W*, $\Gamma \vdash_W \phi$ and any world *w* representing all $\gamma \in \Gamma$, *w* also represents that ϕ. This would mean that the impossible worlds of the expanded modal space would be closed under a weaker relation of logical consequence, and therefore so too would agents' knowledge. But then the cluttering problem would remain.

In addition to failing to be logically omniscient with respect to classical logic, agents fail to know all the consequences of what they know with respect to weaker logics too (Berto and Jago 2019). So one would (seemingly) want it to be the case that for any set of propositions Γ and any proposition ϕ, there are impossible worlds that represent all propositions $\gamma \in \Gamma$ and fail to represent ϕ. We effectively end up with what Priest (2005) calls 'open worlds'.

3.1. A trivial logic for knowledge

The space of open worlds comes with many benefits, but it also has significant disadvantages. For instance, this space contains worlds that represent contradictions and contradictories, worlds that represent conjunctions but not their conjuncts, and also a world that does not represent anything at all.

I mention these particular cases as they seem to correspond directly to three outstanding issues for the logic of knowledge. To wit: that worlds representing contradictions as well as a proposition and its negation are not compatible with any agent's knowledge[3] (Jago 2014) – i.e. that blatantly contradictory scenarios are *a priori* excluded from the space of epistemically possible worlds; that to know a conjunction $K(p \wedge q)$ is to thereby already know each conjunct Kp and Kq (Williamson 2000); and that a world that does not represent anything is also incompatible with any agent's knowledge, as any agent must know that at least *some* proposition.

LOGICAL OMNISCIENCE prompted us to make the space of worlds more fine-grained, but now impossible worlds have taken us too far. The granularity of epistemic space must lie somewhere in-between.

In order to find the proper middle ground, we must do more than tackle the three issues just mentioned. In a space of open worlds, all propositions are treated as atomic. For this reason, if we land on a space of open worlds, we would be left with a trivial logic for knowledge, a logic whose only closure principle is identity – which does not fare better than syntactic approaches (Konolige 1986), as shown by Jago (2007, 2009). But we want our logic for knowledge to be informative and to be predictive of what agents know and how they will update their information when confronted with given scenarios. An open space of worlds will not do, then.

If we keep our impossible-worlds setting but start to introduce some logical inferences that knowledge is closed under, however, we run into what Jago (2014) calls BJERRING'S PROBLEM (Bjerring 2013). According to this dilemma, in an impossible worlds setting, agents can only be characterized as either logically omniscient or logically incompetent. Once knowledge is closed under certain logical inference rules – say, some easily implemented rules such as Conjunction Elimination or *Modus Ponens* – then knowledge is closed under full logical consequence. This is so as even the most complex of logical inferences can be expressed as sequences of trivial inferences. So if knowledge were closed under trivial logical consequence, it would be closed under logical consequence *simpliciter*.

Jago (2014) favours a solution to this problem according to which knowledge is not closed under trivial logical consequence, but it is never determinate which trivial logical consequence an agent fails to draw, the reason being that attributing such a failure to an agent

[3]Dialetheists like Priest, would, of course, disagree. Here I have nothing to add that would convince them.

would amount to revoking their status as a rational agent. Jago also considers all trivial inferences to be on a par, so that a failure to infer $A \vee B$ from A would be as grave an epistemic oversight as a failure to infer A from $A \wedge B$. I disagree with Jago on both features of his reply, for reasons to do with relevance, which will be further explored in the next section.

Attributing to an agent failures of abiding by the standards of rationality is not equivalent to revoking their status as a rational agent. The agent may not have been in optimal conditions to draw the inference (was not attentive, did not have enough time, was emotionally unstable…), or the inference may not have been relevant (in a sense to be specified below) in a given context. Crucially, it may be that rationality only requires of agents that, when optimal conditions are met, they reliably, but not infallibly, infer the given (relevant) trivial consequences of what they know. Furthermore, it does not seem that all trivial inferences are on a par. While both $A \vdash A \vee B$ and $A \wedge B \vdash A$ are trivial, one-step inferences, in the former the conclusion may introduce concepts (those needed to grasp B) that are not present in the premise, while the latter inference can never take the agent from something they grasp to something they do not grasp.

Having seen the limitations of the impossible worlds approach as well as Jago's proposed solutions for BJERRING'S PROBLEM, I now turn to subject matter, promises a completely different approach to avoiding LOGICAL OMNISCIENCE and NECESSARY EQUIVALENCE. Afterwards, I will combine the two approaches, and show that there is an intermediate space of worlds between the space of open worlds and the space of possible worlds that is best suited to capture distinctions of relevance and epistemic contents.

4. Subject matters: a second component of meaning

The approaches to content seen so far presuppose a view of propositions as mere sets of worlds: those in which they are true. For a proposition, p, we can call these the truth-conditions of p. I'll use $T_W p$ to refer to the set of worlds in which an atomic proposition p is true. Similarly, the set of worlds in which the propositions are false, their 'falsity-sets', will be referred to as $F_W p$. In a setting with only possible worlds, $T_W p \cup F_W p = PW$, with PW being the set of all possible worlds. Furthermore, $T_W p \cap F_W p = \emptyset$. A proposition is then identified with the set of worlds in which it is true.

Subject matter theorists take a different approach by considering one of the central roles of propositions – serving as the content of sentences. In addition to truth-conditions, they take the meaning of a sentence to be given by what it is about, its subject matter, or which question it gives an answer to.[4] . So in addition to $T_W p$, a proposition should have a second element, $\sigma(p)$, its subject matter.

To see the general approach in action, consider again the case of necessary propositions. Instead of differentiating between, say, $2+2=4$ and Water is H_2O by adding valuation points in which one is true and the other is false, which is not how we intuitively distinguish between them, we can make the distinction by saying that even though $T_W(2+2=4) = T_W$ (Water is H_2O), $\sigma(2+2=4) \neq \sigma$ (Water is H_2O). Or, to put it informally, '$2+2=4$' and 'Water is H_2O' are about different things: only one is about numbers, and only one is about water.

Subject matter theories can distinguish between necessarily equivalent propositions, dealing in this way with NECESSARY EQUIVALENCE. What about LOGICAL OMNISCIENCE? As seen earlier, some inferences are such that they introduce 'new content' that wasn't in the premises, such as the inference $A \vdash A \vee B$. With the help of the notion of a subject matter, we can now make sense of the idea of the conclusion introducing new content that wasn't present in the premises.

Just as truth-conditions are partially ordered by the subset relation, subject matters can be partially ordered by a relation of parthood (for instance, see Berto 2022), $\sqsubseteq$. The subject matter *mathematics* includes the subject matter *analysis*, the subject matter *Peter's favorite toys and Mary's favorite shows* includes *Peter's favorite toys*, and so on. Intuitively, for instance, the subject matter of $A \wedge B$ includes the subject matter of A, so that $\sigma(A) \sqsubseteq \sigma(A \wedge B)$. When I tell you 'It's rainy in Lisbon', and you say 'It's rainy in Lisbon and it's sunny in Rome', part of what you say has already been said by me, in the sense that part of what you talk *about* has already been talked *about* by me: the *weather in Lisbon*. But the subject matter of A does not include the subject matter of $A \vee B$, so that $\sigma(A \vee B) \not\sqsubseteq \sigma(A)$. $A \vee B$ contains the subject matter of B ($\sigma(B) \sqsubseteq \sigma(A \vee B)$), which is not necessarily included in the subject matter of A ($\sigma(B) \not\sqsubseteq \sigma(A)$). Intuitively, for instance, the subject matter of *It is rainy or sunny* includes the subject matter of *It is rainy*, whereas *It is sunny* does not.

[4]Here I assume, following philosophers such as Lewis (1988) and Yablo (2014), that for any given subject matter, such as **the number of stars**, one can identify a question under discussion, such as *What is the number of stars?* So, subject matters can be thought of *as* questions.

With this informal notion of subject matter inclusion, we can then restrict the acceptable logical inferences for closure principles for knowledge to those that do not add new subject matters in the conclusion. We can now present a global view of content inclusion. The thick proposition A (the set having as two members A's truth-set and A's subject matter) includes the thick proposition B, where $B \leq A$ if and only if Yablo (2014, 2018):

(i) $T_W A \subseteq T_W B$;
(ii) $\sigma(B) \sqsubseteq \sigma(A)$.

So, for instance, the proposition corresponding to the sentence 'The cat is on the mat' will correspond to a set having as members a set of possible worlds in which the cat is on the mat, and further a second element which is the topic *the cat being on the mat*. I can come to know this proposition directly in the usual ways, and also if it is a logical consequence of what I know, *as long as* its subject matter is also included in what I already know. The intuitive idea is that we know propositions when we are sensitive to the circumstances in which they are true, and also are 'on top of' what they are about.

There are various accounts in the literature on what subject matters are, for instance, Lewis (1988) takes them to be partitions of logical space, Yablo (2014) takes them to be divisions of logical space, Fine (2020) takes them to be fusions of states, Perry (1989) takes them to be sets of entities.[5] The thought I want to consider for now is simply that knowledge is closed under logical consequence from given premises Γ to ϕ whenever the content of ϕ is included in that of the propositions $\gamma \in \Gamma$. So before giving my own account, I want to remain neutral on these various understandings.

Like extant subject matter approaches (Berto 2022; Fine 2020; Hawke 2018), I accept that the logical connectives are subject matter transparent, that is, that a logical connective does not contribute to the subject matter of the proposition in which it figures, and that the subject matter of complex propositions is just the fusion (+) of the subject matter of the atomic propositions that compose it.[6]

[5]See Hawke (2018) for a more comprehensive overview of subject matter approaches.

[6]The fusion of given subject matters is just the least upper bound in the partial order on the subject matters imposed by the relation of parthood. The fusion of the subject matters of A and B, $\sigma(A) + \sigma(B)$ is just the $\sigma(C)$ such that $\sigma(A) \sqsubseteq \sigma(C)$ and $\sigma(B) \sqsubseteq \sigma(C)$ and for all other $\sigma(D)$ such that $\sigma(A) \sqsubseteq \sigma(D)$ and $\sigma(B) \sqsubseteq \sigma(D)$, $\sigma(C) \sqsubseteq \sigma(D)$ as well.

We can then give clauses for what the subject matter of a proposition is in a simple propositional language where the only connectives are ¬, ∧ and ∨:

$$\sigma(\neg A) = \sigma(A)$$
$$\sigma(A \wedge B) = \sigma(A \vee B) = \sigma(A) + \sigma(B)$$

4.1. Remaining conflations

While various necessities are indeed distinguished in any theory of subject matter like the one outlined above, the theory will still not distinguish *all* necessarily equivalent propositions which agents might have different epistemic attitudes towards. That is, it won't be able to if only possible worlds are accepted.

Consider the example of *Goldbach's Conjecture* (hereafter *GC*), according to which *Every even number greater than two can be expressed as the sum of two primes*.[7] As the label suggests, it remains a *conjecture*, which has not been proven or disproven by mathematicians. It seems, then, that mathematicians do not know *GC* (assuming that only a proof of *GC* will count as justification for it and that knowledge requires justification). However, let us suppose that *GC* is true. Being a proposition from number theory, it is consequently not a contingent truth, but rather a *necessary truth*, so that $T_W GC = PW$. Furthermore, it does not seem to be an indeterminate or vague matter whether the conjecture is true: it is the case that the conjecture is true or that its negation is true. This, too, is a necessary truth, so that $T_W(GC \vee \neg GC) = T_W GC = PW$. Finally, it seems that there are agents who know this last disjunction – they know that the conjecture is either true or false.

But if mathematicians are trying to prove or disprove the conjecture, then while they plausibly know the disjunction, they don't know *GC*. However, according to any plausible subject matter proposal, they should know *GC* if *GC* is true whenever they know $GC \vee \neg GC$, for the two propositions have the same truth conditions and the logical connectives do not add subject matter, so that $GC \vee \neg GC$ has the same subject matter as *GC*. This, however, is the wrong result. Plenty of mathematicians know the disjunction but not the conjecture.

The problem seems simple and the solution trivial: the restriction to possible worlds is to blame, and we should move past them to accept other valuation points. In what follows, I take on the suggestion, found across the subject matters literature (Fine 2020; Hawke, Özgün, and

[7]This case is discussed in Hawke, Özgün, and Berto (2020) and in Berto (2022), however, how to solve it is in both places left as an open question.

Berto 2020; Lewis 1988; Plebani and Spolaore 2021), that to make the hyperintensional distinctions one wants in a theory of subject matters, one should add impossible worlds (or in Fine's 2020 preferred terminology, states (possible and impossible)).[8]

5. Combining impossible worlds and subject matters

In trying to overcome both approaches' shortcomings, I begin by returning to the issues that came with the expansion of modal space beyond possible worlds. As seen, if extended all the way so as to include all open worlds, epistemic space would be too anarchic. The problem, however, was not due to the inconsistency and non-maximality of worlds, but rather due to the way each proposition was represented as true, false, both, or neither at worlds. In short, the problem seemed to be that in open worlds, all propositions are treated as atomic, a threat to the very compositionality of this version of impossible worlds semantics (Berto and Jago 2019, 2023; Fine, 2021).

Even if *qua* impossibilities this is the right result, when it comes to the characterization of epistemic space, some logical structure needs to be imposed, so that, for instance, worlds representing $A \wedge B$ come out as representing A and representing B.

Following Berto and Jago (2019, 115), in what follows I consider instead the space of so-called *FDE worlds*, i.e. worlds closed under the logic of First-Degree Entailment (FDE) of the relevant logic E of Anderson and Belnap (1975).

These worlds treat atomic sentences in the same way that open worlds do: they can be related to neither, one of, or both of the truth-values *True* and *False* at any given world. What changes is that the valuation relation from worlds to truth-values is defined recursively. So if ρ_w is a valuation relation at a world w,[9] and A and B are arbitrary formulae, we have:

$$\rho_w(\neg A, True) \quad \text{iff } \rho_w(A, False)$$
$$\rho_w(\neg A, False) \quad \text{iff } \rho_w(A, True)$$
$$\rho_w(A \wedge B, True) \quad \text{iff } \rho_w(A, True) \quad \text{and} \quad \rho_w(B, True)$$
$$\rho_w(A \wedge B, False) \quad \text{iff } \rho_w(A, False) \quad \text{or } \rho_w(B, False)$$
$$\rho_w(A \vee B, True) \quad \text{iff } \rho_w(A, True) \quad \text{or } \rho_w(B, True)$$
$$\rho_w(A \vee B, False) \quad \text{iff } \rho_w(A, False) \quad \text{and} \quad \rho_w(B, False)$$

[8]I understand this as a mere terminological dispute since Fine thinks of impossible worlds as maximal states. However, as highlighted earlier, I think there are reasons for taking even non-maximal consistent states as impossible worlds, since they are not complete alternatives to the way the world is.

[9]The use of a valuation relation instead of a valuation function comes from Priest (1998), as a response to Everett's (1993) challenge to the dialethitian solution to the Liar's Paradox.

The space of FDE worlds is still not the epistemic space I wish to characterize. Let me recap some of the worries one might have, briefly gestured at above. All of these will suppose that knowledge is truth at all the epistemically possible worlds for a given agent.[10]

Most crucially, in FDE worlds (Berto and Jago 2019), we have that if a world represents A as *True*, then it represents $A \vee B$ as *True* for arbitrary B. Agents, however, as considered earlier, might fail to have the concepts necessary to grasp B even though they know that A. Furthermore, a world may represent both A and $\neg A$ as *True*, which would allow for agents to have blatantly inconsistent worlds in their epistemic space.

Still, I believe FDE worlds are a good starting place. They represent a way of getting gaps and gluts from a space of possible worlds while minimally changing the clauses for how formulae are evaluated at worlds: conjunction, disjunction and negation behave as usual, but we now have to account for truth and falsity separately, as formulae might get both or neither of them instead of only one. So the main motivation for appealing to a space of FDE worlds is simply this: they allow us to move away from possible worlds while not getting too anarchic – we're simply adding both gaps and gluts.

Why add gluts? If agents' epistemic space, as I have been stressing, should not include worlds representing contradictions, why should we include worlds that have gluts? The answer to this is that while agents don't seem to consider worlds with blatant contradictions as epistemically possible, we nonetheless very often have inconsistent implicit beliefs. We do not, for that reason, believe anything and everything. It is for this reason that a space that allows for truth-value gluts, and on top of which we can build a paraconsistent logic is preferred.

Following Berto and Jago (2019) and Jago (2014), I impose a restriction on epistemic space to the effect that worlds that represent blatant inconsistencies are excluded. These are worlds representing contradictions like $A \wedge \neg A$. Unlike these authors, however, I believe this to be more perspicuously done with the help of subject matters instead of through a stipulation on what goes into the underlying space of worlds. I also impose my own restriction such that the empty world – the world that represents no propositions at all as being the case or not the case – is excluded on the grounds of being 'blatantly incomplete'. The motivation for this exclusion is that it should be *a priori* the case for every agent that *something* is true

[10]Here I am ignoring considerations stemming from multi-agent scenarios and focusing only on single-agent (and therefore not agent-relative) models.

or false. Put another way: every minimally rational agent has at least one item of *a priori* knowledge, something that holds in all epistemically possible worlds.

Let us see, then, how these worlds are excluded from epistemic space. For that, we move on to the subject-matter side of the view. In a broadly intensional framework, subject matters only introduce further distinctions than those initially made by sets of possible worlds. In a setting with impossible worlds, however, they might both introduce new distinctions and erase some already introduced – after all, impossible world theories are already quite fine-grained.[11]

If one accepts a partition-based account of subject matters, then there will be a set closely associated with a given subject matter. This is the set of worlds w such that w is in one of the cells of the partition – i.e. the set that the partition is a partition of. For a given subject matter $\sigma(p)$, I call this set $EW_{\sigma(p)}$.

The core thought motivating the view here presented is that if one thinks of subject matters as questions-under-discussion, and these as naturally associated with a partition view where different cells correspond to different answers, then one can define, for each question an agent is sensitive to in their inquiry, the set of epistemically possible worlds (which might be impossible) that answer the question.

These worlds will then always 'speak to the question' and, at least with respect to the question, be consistent. They can, however, be inconsistent with regard to other questions. Someone who knows that Socrates is a philosopher might have inconsistent attitudes toward some tricky logical and mathematical puzzles, say.

These are the highlights. Let's see how it works more formally.

On this view of subject matters, there is not one unique topic or question that a sentence is about or an answer to – 'Jane is a lawyer' could be just as aptly an answer to the question *Is Jane a lawyer?* as to *What is Jane's profession?*. Lewis (1988) points out, and Yablo (2014) seems to agree, however, that for a given thin atomic proposition (the set of possible and impossible worlds in which it is true), p, there is a minimal subject matter, $\sigma_{\min}(p)$, such that all other subject matters of p include $\sigma_{\min}(p)$.

$\sigma_{\min}(p)$, for p an atom, is going to be *Is* p *the case?* and it will be a two-celled partition of the set of worlds representing p (as *True* or *False*). If one favours an ontology of worlds in line with Jago's (2014) and Berto and

[11]However, see Gioulatou (2016) for a system in which subject matters are used only to add further distinctions to a space with impossible worlds.

Jago's (2019), as I do, then this will be a partition of the set of worlds containing the Lagadonian sentence that is the translation of *p* to the lagadonian language.[12] In what follows, I will refer to the minimal subject matter of a given proposition as *the* subject matter of the proposition, and ignore the subscript (in this usage I follow Yablo (2014) and others).

Following Yablo (2014), we can in general identify the minimal subject matter of a given sentence *A* with the question *What needs to be true for* A *to be true/false?*[13] It's just that in the case of atomic sentences the answer to the question *What needs to be true for* p *to be true/false?* is just *p* itself. So we just need to ascertain whether *p* is *True* or *False*.

This is not so for complex propositions. It is plausibly the case that the truth or falsity of a conjunction depends on the truth or falsity of its conjuncts, and similarly for a disjunction on its disjuncts, and a negation on what is negated.

Using again $T_W A$ for the set of worlds where *A* is true, letting *A* and *B* be arbitrary sentences, and S_1 and S_2 be variables for sets of FDE worlds, then we can define subject matters recursively from the atomic case[14] :

$$\sigma(p) = \{T_W p, T_W \neg p\}$$
$$\sigma(\neg A) = \sigma(A)$$
$$\sigma(A \vee B) = \sigma(A \wedge B) = \{S_1 \cap S_2 : S_1 \in \sigma(A) \wedge S_2 \in \sigma(B)\}$$

If we take as an assumption that $\sigma(A \wedge B)$ is the same as $\sigma(A) + \sigma(B)$, for + a binary fusion operator, then the last clause also serves as a definition of fusion.

In an intensional setting, one subject matter being part of another is just defined as the latter being a refinement of the former, where both are partitions of the same set of worlds, namely the set of all possible worlds. In the present setting, however, different questions/subject matters correspond to partitions of different subsets of the space of FDE worlds. So if *P* is the partition that is $\sigma(A)$ and P^* is the partition that is $\sigma(B)$, then there will be at least one world *w* such that $w \in S_1$ and $S_1 \in P$ and there is no S_2 for which $w \in S_2$ and $S_2 \in P^*$. We would not get the intuitive results we want (for instance that $\sigma(A)$ is part ($\sqsubseteq$)

[12] A Lagadonian language, according to Lewis (1986), is a language in which every object is a designator for itself and every property and relation is a predicate designating itself.

[13] In the sense that it explains in a non-causal way the truth/falsity of the sentence.

[14] In the case where both *A* and *B* are atomic, say respectively *p* and *q*, we have four possible combinations: $\sigma(p \wedge q) = \sigma(p \vee q) = \{T_W p \cap T_W q, T_W p \cap T_W \neg q, T_W \neg p \cap T_W q, T_W \neg p \cap T_W \neg q\}$. There is, then, a close association between the number of members of a sentence's subject matter and the number of rows in its truth-table.

of $\sigma(A \wedge B)$) if parthood between subject matters were defined in the same way.

We can instead define parthood in the usual way through the notion of fusion, where as usual:

$$\sigma(A) \sqsubseteq \sigma(B) := \sigma(A) + \sigma(B) = \sigma(B)$$

If we combine this with the definition of fusion given above, we get that $\sigma(A)$ is part of $\sigma(B)$ if the intersection of each set in $\sigma(B)$ with each set in $\sigma(A)$ is the same as each set in $\sigma(A)$. Here, it might help to take a look at two cases: one where there isn't parthood and one where there is parthood among subject matters.

Let us consider the example of the subject matters of the atoms p, q, as well as of their conjunction $p \wedge q$. We know that *every* world that represents that p is the case is going to be in $\sigma(p)$, as well as every world that represents that p is not the case, i.e. every world that represents that $\neg p$ is the case. But given the clauses above for how FDE worlds represent, some worlds that represent that p is the case or that p is not the case will be worlds that represent q as neither being the case nor not being the case. Take the sets $T_W p$ and $T_W q$, for instance, which are both sets that belong respectively to $\sigma(p)$ and $\sigma(q)$. Clearly, $T_W p \cap T_W q$ is neither $T_W p$ nor $T_W \neg p$ (which are the only sets in $\sigma(p)$). But then clearly $\sigma(p)$ is not part of $\sigma(q)$. Consider, in turn, the relation between $\sigma(p)$ and $\sigma(p \wedge q)$. For $\sigma(p)$ to be part of $\sigma(p \wedge q)$, then intersecting a member of the latter with a member of the former should yield a member of the former. We can do this by exhaustion. Remember that $\sigma(p \wedge q)$ just is $\{T_W p \cap T_W q, T_W p \cap T_W \neg q, T_W \neg p \cap T_W q, T_W \neg p \cap T_W \neg q\}$. And clearly, the first two members intersect with $T_W p$ to yield $T_W p$, and the last two members with $T_W \neg p$ to yield $T_W \neg p$. This results in the fact that $\sigma(p) \sqsubseteq \sigma(p \wedge q)$, as desired.

I now define a question-relative notion of knowledge – $K_{\sigma(A)} B$ – which will correspond to the truth of B in all the epistemically-accessible worlds in a given $EW_{\sigma(A)}$, for some $\sigma(A)$. I also define an absolute notion of knowledge, which is given by the following condition[15] :

$$KB \text{ if and only if for some } \sigma(A), \quad K_{\sigma(A)}\ B$$

Having presented the view at a semi-formal level, I am now in a position to give a precise model for knowledge.

[15] Of course, in the models below, these sentences will be evaluated relative to worlds. In [REDACTED] I expand these models to the first-order case and also evaluate them relative to variable-assignments.

5.1. The model

I start by providing the epistemic language $\mathscr{EL}$. This is given by the following BNF:

$$A := p \mid \neg p \mid p \vee q \mid p \wedge q \mid p \mid K_{\sigma(p)}q \mid Kp$$

I use capital letters from the beginning of the alphabet as metavariables for formulae. I use the capital letter S with numbers as subscripts for variables for sets of FDE worlds.

A question-relative epistemic model $\mathscr{QEM}$ will be a tuple $\langle W, N, R, \rho\rangle$ where W is a set of worlds, N is a subset of W, intuitively the normal (or possible) worlds, and R is a dyadic epistemic accessibility relation between worlds that is assumed to be reflexive (given the factivity of knowledge). Finally, ρ is as above, with the added information that at worlds in N, ρ behaves like a classical valuation function – i.e. it assigns one and only one truth-value to any formula, and it follows the two clauses for question-relative knowledge and the two clauses for the absolute notion:

- $\rho_w(K_{\sigma(A)}B, True)$ iff $\forall w'(wRw' \wedge w' \in EW_{\sigma(A)}) \rightarrow \rho_{w'}(B, True)$
- $\rho_w(K_{\sigma(A)}B, False)$ iff $\neg\exists w'((wRw' \wedge w' \in EW_{\sigma(A)}) \wedge \rho_{w'}(B, True))$
- $\rho_w(KB, True)$ iff for some $\sigma(A)$, $\rho_w(K_{\sigma(A)}B, True)$
- $\rho_w(KB, False)$ iff for no $\sigma(A)$, $\rho_w(K_{\sigma(A)}B, True)$

As is standard, logical consequence is defined as necessary truth preservation in all models, i.e. for any model, if all formulae A in a set of formulae Γ are true in a world $w \in N$, then B is true in w if and only if Γ entails B. Logical truth is necessary truth in all models.

6. Meeting the desiderata and the limits of the proposal

We now take a look at the results of the proposal and how it meets the desiderata it was set out to accomplish.

Subject matters help with all of the first three puzzlements. With respect to worlds representing contradictions, when agents are considering a specific question, they consider distinct direct answers to it that are indeed exclusive, so that relative to that subject matter, the propositions that stand for different possible answers are either *True* or *False*, but neither both nor neither. However, other propositions that are irrelevant to the question at hand are still permitted to be both *True* and *False*. This goes in line with the intuition that agents can have implicitly, but not explicitly inconsistent systems of belief and information.

This line of thought also leads to a stronger version of the earlier claim that the world that represents all propositions as neither *True* nor *False* is not epistemically possible for any agent. Underlying the intuition that such a world is epistemically impossible is the fact that the question *What is a priori true?* always has a non-empty set of propositions as an answer for any minimally rational agent. More generally, as can be read from the model just presented, relative to a subject matter, agents consider worlds that 'speak to the question', that is, that represent propositions that are direct answers to it. So, relative to a given question, such as the simple case *Is* p *the case?*, all the worlds that are in the cells of the partition, should represent p as *True* or as *False*.

As for the third puzzlement – of agents knowing $A \wedge B$ but failing to know A and/or failing to know B – we need to consider the Yablovian notion of content inclusion appealed to above. Consider the case of the subject matter of $A \wedge B$. As seen, $\sigma(A \wedge B)$ includes by fiat $\sigma(A)$. So the subject-matter part of content inclusion is accounted for: $\sigma(A) \sqsubseteq \sigma(A \wedge B)$. As for truth-conditions, as highlighted above, we only care about what goes on in normal worlds, and we want it to be the case that the normal worlds where $A \wedge B$ is true are a subset of the worlds where A (or B) is true (though it is generally true that $T_W(A \wedge B) \subseteq T_W A$ in the full space of FDE worlds).

We move on to consider how this approach deals with the particular problems usually faced by *FDE*-worlds, as well the earlier problem for subject-matter views of distinguishing between the contents of necessities that employ the same concepts (bar the logical connectives).

Let's start with the latter. Again, I'll use GC and $GC \vee \neg GC$ as examples and I'll assume that one should allow for agents who know the latter but fail to know the former. Again, we want to know if $GC \geq (GC \vee \neg GC)$, since we have accepted that knowledge is closed under content inclusion, and we are testing whether merely knowing that $GC \vee \neg GC$ is enough to come to know GC.

Let's present, then, a countermodel for an agent who knows $GC \vee \neg GC$, in which GC is true at all possible worlds (i.e. all worlds in N), and yet, one does not know that GC. Clearly $\sigma(GC) \sqsubseteq \sigma(GC \vee \neg GC)$, so we will only check for truth conditions. As above, a model is a tuple $\langle W, N, R, \rho \rangle$. Let $W = \{w_1, w_2\}$, $N = \{w_1\}$, $R = \{\langle w_1, w_1 \rangle, \langle w_1, w_2 \rangle, \langle w_2, w_2 \rangle\}$. Finally, we let $\rho_{w_1}(GC, True)$,[16] $\rho_{w_2}(\neg GC, True)$, and $\rho_{w_2}(GC, False)$.

[16]Here note that since w_1 is a possible world, then $\neg GC$ is also *False* at w_1.

It follows that $\rho_{w_1}(GC \vee \neg GC, True)$, and similarly $\rho_{w_2}(GC \vee \neg GC, True)$. Supposing, to simplify, that GC is atomic and that $T_W GC = w_1$ and $T_W \neg GC = w_2$, then $\sigma(GC) = \{\{w_1\}, \{w_2\}\}$. Clearly, then, $EW_{\sigma(GC)} = W$. In this model, $\rho_{w_1}(K_{\sigma(GC)}(GC \vee \neg GC), True)$, since $w_1 R w_1$, $w_1 \in EW_{\sigma(GC)}$, and $\rho_{w_1}(GC \vee \neg GC, True)$. Similarly, $w_1 R w_2$, $w_2 \in EW_{\sigma(GC)}$, and $\rho_{w_2}(GC \vee \neg GC, True)$. Therefore, $\rho_{w_1}(K(GC \vee \neg GC), True)$. Yet, it is not the case that $\rho_{w_1}(K_{\sigma(GC)} GC, True)$, because $\rho_{w_2}(GC, False)$, and $w_2 \in EW_{\sigma(GC)}$. And clearly, if the agent doesn't know that *GC* relative to the question associated with *GC*'s subject matter, then they don't know it relative to any other subject matter.[17]

Even recent accounts employing subject matters within a possible worlds framework, like Berto's (2022), face the problem of distinguishing between some necessarily equivalent contents. Combining impossible worlds and subject matters in the way I have presented here is, thus, a major advantage of the present account.[18]

Let us now take a look at the other problems mentioned for strategies employing FDE worlds without subject matters. One of them, I believe, merits a distinct answer from the others: the problem that worlds are closed under Double Negation Introduction, and therefore agents who know a given thick proposition know that same proposition is preceded by a very high number of '¬'-signs. As the reader might surmise, I am not sympathetic to a general metalinguistic approach[19] to puzzles concerning representational hyperintensionality. However, I believe such an approach is appropriate for this case.

As mentioned earlier, it is a common position in the literature on subject matters to take the logical connectives to be transparent. If that is so, then subject matters will also not help with the problematic case arising from closure under Double Negation Introduction. However, there is a sense in which saying A, $\neg\neg A$, $\neg\neg\neg\neg A$, and so forth is saying the *same* thing. No new complexity at the level of *what is being said* is added, merely syntactic complexity. And where only syntactic complexity is added, it is plausible that what agents really fail to know is not $\neg \cdots \neg A$, but what proposition the sentence is expressing (Stalnaker 1984). They fail to parse the syntax, for various potential reasons – time constraints, lack of attention, failure to count the number of negation signs, etc.

[17] I leave this to be proven, for reasons of space.

[18] Thanks to an anonymous referee for recommending I stress and clarify this point.

[19] That is, the approach according to which, in general, there is no difference in content between necessarily equivalent propositions. Instead, it is claimed, agents just are ignorant of which contents are being expressed by certain linguistic vehicles, a view championed by Stalnaker (1984).

This move is not ad-hoc: even though the proposed account distinguishes contents in a more fine-grained than as sets of possible worlds, such distinctions must stop somewhere before the level of syntax, and they do so here. I believe it is a natural place to stop since we just seem to increase the number of symbols without introducing any complexity in terms of what is conveyed or talked about.

I now turn to the main worry facing the simple FDE worlds approach: that all worlds that represent A represent $A \vee B$ for arbitrary B. Agents' knowledge should not be closed in this way, given that they might not possess the concepts that feature in B. What seems to go wrong in this case is precisely that the subject matter of $A \vee B$ goes beyond the subject matter of A. Of course, $\sigma(A \vee B) \not\sqsubseteq \sigma(A)$ on the account given above, and therefore we can easily construct a countermodel where an agent has knowledge of A without having knowledge of $A \vee B$.

Finally, consider the issue of the logic for knowledge being adjunctive – that is, that for any normal world w, whenever $\rho_w(KA, True)$ and $\rho_w(KB, True)$, $\rho_w(K(A \wedge B), True)$. This is the case in the models presented so far. Indeed, suppose that $\rho_w(K_\sigma(A)A, True)$, so that relative to the question (roughly put) *What does it take to make* A *true?*, the agent knows that A. Similarly, $\rho_w(K_\sigma(B)B, True)$. This, of course, still leaves open the possibility that $\rho_w(K_\sigma(A)A \wedge B, False)$. But clearly, one would get $\rho_w(K_\sigma(A \wedge B)A \wedge B, True)$, and this would be enough for $\rho_w(KA \wedge B, True)$, proving that knowledge is adjunctive.

Lewis (1982) presents an influential case against belief being adjunctive, as do other fragmentists like Yalcin (2018). Since knowledge implies belief, such cases also impact on any attempts at developing an epistemic logic. A further worry related to knowledge being adjunctive has to do with the lottery paradox. An agent might have good reasons to believe of each ticket in a lottery that it will be a losing ticket (as there are many of them), and yet the agent does not have good reason to believe *Ticket 1 will lose* $\wedge$*Ticket 2 will lose* $\wedge \ldots$ for all tickets. Of course, here I have been presenting a model for *knowledge*, so the translation of the paradox is not direct: since knowledge is factive, the agent does not know for each ticket that it will be a losing ticket, for one of them is a winning ticket. Yet it can be easily adapted. Let there be 1000 tickets, and assume that the winning ticket is T_{901}. Then for each T_i in $T_1 \ldots T_{900}$, the agent has (fallible) knowledge that T_i is a losing ticket. However, the agent might not have even fallible knowledge that $T_1 \wedge \ldots \wedge T_{900}$ are losing tickets, for they amount to a very large proportion of the tickets. So there is indeed a problem regarding adjunctive reasoning associated with the lottery paradox.

Here, there are various approaches one could take. The one I prefer would be to block the move from KA and KB to $K(A \wedge B)$ and so reject that knowledge is adjunctive. A possible way of doing this is to follow Lewis and to say that the agent fails to know anything with respect to the subject matter $\sigma(A \wedge B)$. What would fail when adjunction fails is that while the agent knows A relative to $\sigma(A)$ and B relative to $\sigma(B)$, they fail to know anything relative to $\sigma(A \wedge B)$, as they fail to integrate the information available to them and to combine the information they have on both questions. This is the same strategy as has been employed in Hawke, Özgün, and Berto (2020) as well for other forms of logical non-omniscience.

A closely related worry is that, in this model, agents' body of knowledge is represented as monotonic – that is, upon acquiring new items of knowledge, agents retain the knowledge that they previously had. As Hawke, Özgün, and Berto (2020) perspicuously show, however, there are scenarios in which this is not the case. Suppose that Jane is taking a logic course and she has been erroneously told by her very reliable and brilliant professor that 'Theorem' 1 is valid in a given logic. During one of her tutorial sessions, Jane is told by a tutor that a different proposition, Theorem 2, is valid. When diligently doing her homework, Jane stumbles upon the result that 'Theorem' 1 and Theorem 2 are incompatible with one another. She therefore gains the knowledge of the incompatibility of the corresponding propositions. Trusting her (more reliable) professor over the tutor, she abandons her belief in the validity of Theorem 2 and proceeds to make use of the proposition the professor conveyed in class instead. As it turns out, the professor made a mistake, a very rare occurrence for them, and 'Theorem' 1 is in fact no theorem of the given logic at all. By abandoning the belief in Theorem 2, however – which was in fact true – she therefore lost the knowledge she had of Theorem 2. Acquiring new knowledge might lead to loss of knowledge one already had.

This is a recognized limitation of the present proposal, but it is by no means a knock-down objection against it: it can be easily modified to yield non-monotonic results. Below, I show exactly how to do so by comparing my approach to that of Berto's (2022). In short, what makes my approach monotonic is that the accessibility relation on worlds stays the same regardless of what subject matter the agent is sensitive to. But very naturally, we should think that what is epistemically possible for agents varies as they consider different questions in a dynamic way so that also the accessibility relation varies. I consider this change to my account in the next section.

To conclude this point on fragmentation, I will note that even if the dynamic processes of considering questions and updating one's knowledge is not developed here in full, it wouldn't in principle be hard to modify my account in ways similary to Hawke et al.'s (2020) proposal, which delivers both fragmentation and various forms of non-monotonicity. My proposal works with an extended space of worlds (which they also consider doing) by allowing the valuation of the atomic propositions to contain gaps and gluts, which it combines with subject matters considered as partitions and linked with the notion of a question, as in inquisitive semantics (Ciardelli, Groenendijk, and Roelofsen 2019; Hoek 2022). I see no reason why these differences between our respective approaches would prevent me from adopting their solution to the problem of adjunction. Still, this is a limitation of the current proposal and a satisfactory response is left for future work.

Finally, I would like to address the opposite worry: that on this account, agents might be too logically incompetent.[20] While I don't have a general argument for why agents are not too logically incompetent on this account, I believe this can be established on a case-by-case basis. For instance, the reader should notice that while closure under material implication ($KA,\ A \rightarrow B \vDash KB$) fails, closure under *known* material implication holds ($KA,\ K(A \rightarrow B) \vDash KB$), and that it is the latter instead of the former that generates consensus among epistemic logicians. Further, one should note that in the models presented, as long as the agent, to borrow terminology from Berto (2022), is 'on top of' the relevant subject matters, their knowledge is closed under content inclusion. This guarantees that agents know many propositions as long as they meet some relevancy constraints (which is also ensured by the gaps of FDE worlds in the case of necessary truths).

My proposal further allows for agents to know conjunctions of what they know but not come out as believing everything once they have inconsistent belief systems, and to meet some rational consistency standards as they address different questions.

Agents also come out as knowing sentences of a high degree of syntactic complexity once they know equivalent sentences of a low degree of syntactic complexity, even if they don't know them under the 'guise'

[20]Thanks to an anonymous reviewer for calling my attention to this worry.

of the more syntactic complex expressions. Here, I disagree with (forthcoming) that we should make a distinction between the attitudes of agents towards such sentences and I agree with Berto (2022) that the best answer is to take a metalinguistic stance (Stalnaker 1984). Nothing changes, for instance, in terms of what is said between $p \wedge q$ and $\neg(\neg p \vee \neg q)$ in the context of a DeMorgan negation.

Having said all that, it seems, then, that the current proposal is well-suited to provide an interesting account of the standards of minimal rationality – one which doesn't make agents logically incompetent on the one hand, nor logically omniscient, on the other.

7. The account in the current landscape

Having presented my own question-sensitive account of knowledge, which combines impossible worlds and subject matters and attempts to model the epistemic state of minimally rational agents, I now compare the approach to other proposals that have closely inspired it and boast similar results.

In particular, I consider more closely the dynamic approach of Bjerring and Skipper (2019), which makes use of impossible worlds but not of subject matters, as well as the approach of Berto (2022), which makes use of possible worlds and subject matters, but not of impossible worlds.[21]

7.1. Bjerring and Skipper's (2019) dynamic epistemic logic

Following Bjerring and Skipper (2019) and Jago (2014) attempt to characterize the distinction between agents who are and aren't logically competent in terms of agents who are able to perform given *trivial* inferences. Like Jago, they recognize that simply stating that one's knowledge is closed under trivial logical consequence would entail admitting that one's knowledge is closed under full logical consequence (Bjerring and Skipper 2019). Unlike Jago, however, they don't claim that it is, therefore, vague which trivial logical consequence the agent fails to draw, but rather note that agents must spend resources and carry out a deduction when reasoning from premises to a conclusion, even when the deduction is trivial. Their language contains the sentence $\langle n \rangle p$, which states that

[21]Thanks to an anonymous referee who called my attention to the similarities between my approach and Bjerring and Skipper's (2019) and Berto's (2022) and encouraged me to compare them to my own in a more detailed way.

after some chain of n steps of reasoning according to a given set of rules of inference $\mathscr{R}$, one can derive p.

Importantly, the agent need not carry the specific chain of inference steps that leads one to p. However, as I understand the authors, if p is a trivial consequence of what the agent knows in n steps of reasoning and the agent has the resources to perform a deduction of n steps using the rules of $\mathscr{R}$, then when asked whether p is the case, the agent should be primed to perform this chain of inferences and come to believe that p.

I believe the authors and I are largely in agreement, and I'd like to start by pointing out the various features that our respective approaches have in common. The first is that we both recognize that, when confronted with the axioms of Peano Arithmetic, and assuming that *GC* is a theorem of the axioms of Peano Arithmetic, agents might fail to believe *GC*, and this is best modelled as agents having *impossible* worlds among their epistemically possible worlds. The second is that we recognize the crucial importance of modelling agents who are neither logically omniscient nor logically incompetent, setting a normative standard for minimally rational agents. The third and most important is that we recognize the importance of agents considering questions for updating their doxastic states and, therefore, their epistemic states.

Bjerring and Skipper's (2019) formal account of what agents come to know given n steps of reasoning does not include yet considerations from subject matter semantics. To a large extent, my contribution comes from noticing that it would be natural to join subject matters to their proposal. After all, as Berto and Jago (2019) note, Bjerring and Skipper are only able to characterize an interesting notion of what agents *may* or *can* come to believe after n steps of reasoning ($\langle n \rangle Bp$), whereas we wish to characterize an interesting notion of what agents *will* come to know (and therefore believe) after n steps of reasoning ($[n]Bp$), even if only when primed by a question. The reason for this is that we're interested in capturing what agents will come to know given what they know. As Berto and Jago (2019) point out, however, given minimal assumptions about the inferential rules in $\mathscr{R}$, $[n]Bp$ collapses into Bp (i.e. what the agent already believes, taking 0 steps of reasoning from where they start). Bjerring and Skipper (2019) seem to accept this as a natural result.

I believe, therefore, that the addition of subject matters makes explicit in a specific way some of the key insights of Bjerring and Skipper's – namely, the insight that we should make knowledge question-relative.

We can find considerations going in this direction also in Hoek (2022, forthcoming) and Yalcin (2018).

Still, it is important to highlight some differences between our accounts. One of them is that Bjerring and Skipper allow for worlds in which contradictions are realized to feature in an agent's epistemic space. According to my account, and in agreement with Berto and Jago (2019), such worlds are ruled out due to their being blatantly inconsistent. Perhaps the most important divergence, however, is that my account does not appeal to a distinction between trivial and non-trivial inferences, as well as to a dynamic process of deduction. Instead, I appeal to the questions to which an agent is sensitive, and I take it that an agent's knowledge is closed under a relation of content inclusion that preserves not only truth but also subject matter. On the contrary, the inference rules that Bjerring and Skipper appeal to are topic-neutral and may include rules such as Disjunction Introduction, which, even if trivial, will easily take one to propositions that include content that bears no relation to the content one started from. I take it, then, that instead of focusing on inference rules which may add new topics and then constrain those inference rules so that they don't 'go wild', we should opt for a more elegant solution according to which knowledge is preserved via content inclusion, and the questions an agent attends to have a more central, formal, role.

7.2. Berto's (2022) topics of thought

Following the thought of according to subject matters more of an active role, Berto's (2022) *Topics of Thought* represents a *locus* point in treating subject matters as a core component of meaning, including in epistemic logic. The author arrives at many of the same results when it comes to the logic of knowledge as I have here.

Still, some differences are important to highlight. (forthcoming) present a helpful critical summary of Berto's (2022) main stances, which has helped me to build the present comparison between his views and my own.

Just like in Berto (2022) and Hawke, Özgün, and Berto (2020) make use of a space of possible worlds with a set of subject matters that are taken as primitive or unanalysed entities. This is a departure from my own account already in two ways: I work with a space of states that includes impossible worlds, and I define subject matters as partitions of different subsets of the set of possible and impossible worlds.

This need not be interpreted as a deep disagreement, however. My own account of subject matters can be interpreted as a way-based account (Hawke 2018), which is the preferred view of subject matters of one of the co-authors of *Topics of Thought*. Secondly, both in Hawke, Özgün, and Berto (2020) and in Berto (2022), it is suggested that to avoid the problem highlighted earlier of knowledge of GC and $GC \vee \neg GC$, adding impossible worlds to one's models is a natural way to go.

So, let's ignore these initial differences, and explore what other distinctions and similarities we may find in the resulting epistemic logics.

For starters, Berto (2022) does not attempt to define, as is usual, a unary knowledge operator. Rather, in the overarching framework, Berto and his collaborators characterize what they call 'Topic-Sensitive Intentional Modals' (TSIMs), which take the general form of $X^{\varphi}\psi$, standing for 'given input φ, the agent X's that ψ'. Two conditions are imposed for an agent X-ing that ψ on the input that φ at a given world w: (i) that for all worlds w' such that $wR_{\varphi}w'$, ψ holds; and (ii) that $\sigma(\psi) \sqsubseteq \sigma(\varphi)$. The biggest novelty in these conditions is that R_{φ} is an accessibility relation between worlds that is indexed to φ. As (forthcoming) put it '[i]ntuitively, we can think of the index [φ] as determining which class of worlds the agent "attends to" given input [φ]'. This feature of Berto's (2022) TSIMs, so considered, approximates his models from the intuitions driving my account and that of Bjerring and Skipper (2019) just considered.

Depending on which TSIM the general semantics for TSIMs is applied to, there are different constraints on R_{φ}. Importantly, none of the TSIMs developed by Berto (2022) attempt to capture a notion of knowledge appropriate for minimally rational agents. The closest we get instead is, based on a paper co-authored with Hawke (Berto and Hawke 2021), the notion of 'knowability relative to information' (KRI). This TSIM is represented by the operator $K^{\varphi}\psi$, which stands for 'relative to information φ, the agent is in a position to know ψ'. As you would expect, $K^{\varphi}\psi$ holds at a given possible world w if and only if ψ holds at all worlds that are R_{φ} accessible from w, and further if $\sigma(\psi) \sqsubseteq \sigma(\varphi)$. Yet, it is also imposed that all the worlds that make φ true are included among the worlds w' such that $wR_{\varphi}w'$. That is all the worlds in which the information that one uses a basis for knowledge is factual are included among the worlds one is sensitive to.

An important difference between my account and Berto's (2022) is that his account considers the attitude of *being in a position to know* relative to *information* (which might not be factive), whereas I consider the attitude of knowledge as a normative ideal that is suitable for minimally rational agents, relative not to information but to questions or subject matters. In principle, this doesn't make the projects incompatible, but it should be noted that they focus on different phenomena and try to explain different attitudes.

While it's a great advantage of Berto's (2022) account as it stands that it is able to model how agents are in a position to know in a non-monotonic way, this is achieved through an appeal to a notion of information that is syntactic and which allows accessible worlds to include worlds in which one's input is not factual. This non-monotonicity can in fact easily be replicated in my account in a natural way.

If we use the conception proposed above of subject matters, then since Berto works with only a space of possible worlds, in his framework, at any world w in which one is given input φ and is in a position to know ψ, all the φ-accessible worlds are trivially included in what I have called $EW_{\sigma(\varphi)}$, i.e. the set of worlds in the subject matter of φ, which in this case is the set of all possible worlds.

Could we say the same more generally when we move on to the space of FDE worlds and rewrite Berto's condition for $K^{\varphi}\psi$ in terms of an operator $K^{\sigma(\varphi)}\psi$, such that: (i) all the worlds in what would be $R_{\sigma(\varphi)}$ are in $EW_{\sigma(\varphi)}$ and (ii) $\sigma(\psi) \sqsubseteq \sigma(\psi)$? Yes! This would further have the benefit that we would then not need to appeal to a syntactic notion of information (so that, say, we could take p and $p \wedge p$ to be the same input as they have the same subject matter), and we would not need to appeal to the condition that all the φ-worlds are among the accessible worlds, for that is automatically the case as it's always the case that $T_W\varphi \subseteq EW_{\sigma(\varphi)}$.

I don't believe that this does any injustice to Berto's account, in the sense that it does not allow us to derive any consequences that we were not already able to derive from the original position. Even though we force all worlds that are $R_{\sigma(\varphi)}$-accessible to represent φ or $\neg\varphi$, this was already the case in Berto's original account, as it relied on possible worlds, and it is nonetheless natural as the agent is updating on the information that φ.

Our models, which we may call non-monotonic question-relative epistemic models, now look like tuples where we replace R with a set of $R_{\sigma(\varphi)}$ accessibility relations, and everything else stays the same: $\langle W, N, R_{\sigma(\varphi)}, \rho \rangle$. We change the clause for knowledge relative to the subject matter such that, naturally, one now has to know a proposition relative to the accessibility relation of its subject matter:

$$\rho_w(K_{\sigma(A)}B, True) \quad \text{iff } \forall w'(wR_{\sigma(A)}w' \wedge w' \in EW_{\sigma(A)}) \rightarrow \rho_{w'}(B, True)$$
$$\rho_w(K_{\sigma(A)}B, False) \quad \text{iff } \neg\exists w'((wR_{\sigma(A)}w' \wedge w' \in EW_{\sigma(A)}) \wedge \rho_{w'}(B, True))$$

We can now look at a simple countermodel for monotonicity. Let $W = N = \{w_1, w_2\}$, $R_{\sigma(p)} = R_{\sigma(q)} = \{\langle w_1, w_1\rangle, \langle w_2, w_2\rangle\}$, $R_{\sigma(p\wedge q)} = \{\langle w_1, w_1\rangle, \langle w_1, w_2\rangle, \langle w_2, w_2\rangle\}$, and finally $\rho_{w_1}(p, True)$, $\rho_{w_1}(q, True)$, $\rho_{w_2}(p, False)$, and $\rho_{w_2}(q, False)$. We then have that $\rho_{w_1}(K_{\sigma(p)}p \wedge q, True)$, since all worlds accessible from w_1 relative to the subject matter $\sigma(p)$ that are in $EW_{\sigma(p)}$ make $p \wedge q$ true, but we don't have $\rho_{w_1}(K_{\sigma(p\wedge q)}p \wedge q)$, since relative to the subject matter $\sigma(p \wedge q)$, we have that w_2 is accessible from w_1, and in w_2 both p and q are false and w_2 is in $EW_{\sigma(p\wedge q)}$. So there is a normal world w in a model where $\rho_w(K_{\sigma(A)}B, True)$ but where it's not the case that $\rho_w(K_{\sigma(A\wedge C)}B, True)$. So monotonicity does not hold.

It seems to me, then, that Berto's main advantage – non-monotonicity – can be easily regained in my proposal in a way that is more economical (subject matters are simply understood as sets of sets of worlds), does not rely on a syntactic notion of information, and, more importantly, does not conflate between some logical necessities and the disjunctions of them and their negations.

8. Conclusion

In this paper, I have attempted to provide closure principles for knowledge well-suited for minimally rational but non-omniscient agents. In doing so, I have noted symmetric issues found in impossible-world and subject-matter approaches: the former allows too many worlds into their models, and the latter cannot manage the filtering effected by possible worlds.

The idea to go beyond possible worlds in subject matter theory is not original, but as far as I am aware it has so far been kept as an off-hand remark.[22]

Given that I wanted to characterize *minimally rational* agents, I took as desiderata for any plausible theory of knowledge that contradictions should not be compatible with any agent's epistemic states, nor should

[22]Gioulatou (2016), again, is here an exception. However, Gioulatou considers that Jago's (2014) model is not sufficiently fine-grained and, therefore, needs subject matters to better characterize certain hyper-intensional distinctions. This paper is still the first developed attempt to combine the two approaches for the case of closure principles for knowledge with subject matters *delimiting* a space of impossible worlds.

a world that does not represent any proposition as being the case or as not being the case, and that any agent who knows that $A \wedge B$ *ipso facto* knows that A and that B. Simple impossible-world approaches cannot meet these three desiderata without accepting various other undesirable closure principles. One such principle is that an agent would come out as knowing $A \vee B$ whenever they know A, where B might bear no relevant connection to A.

Combining impossible worlds and subject matters allowed me to satisfy several desiderata: the three just presented; avoiding Disjunction Introduction; and avoiding conflating necessary truths with the disjunction of them and their negations. In order to do so, I accepted one specific view of impossible worlds and of subject matters: FDE worlds, subject matters as questions, and these as partitions of subsets of the space of FDE worlds.

This combination of views led to the right results on all listed desiderata. Along the way, reasons were provided for why this is a natural combination of tools and not just a gerrymandered artificial construction. In fact, the resulting view ends up being theoretically simple in the sense that the subject matters are themselves constructible from the space of impossible worlds, *viz.* as partitions of it, and are not taken to be extra primitive entities (as they are, for instance, in Berto 2022).

While my account can be expanded to include fragmentation and an account of the dynamic processes of knowledge update as in Hawke, Özgün and Berto (2022) so as to avoid adjunction, the current proposal represents a natural way of joining two traditions in epistemic logic. It is striking that in doing so, it bears close ties to views like those of Bjerring and Skipper (2019), Hoek (2022), and Yalcin (2018) in that it gives central focus to the questions which agents consider and try to answer. In its best, non-monotonic version, my account also borrows heavily from Berto (2022). If nothing else, this highlights the importance of collaborative projects and the interplay of different traditions in philosophy and logic.

Acknowledgments

I would like to thank Franz Berto and Greg Restall for comments on drafts of this paper, as well as the audience at the 2022 Summer School on Truthmaker Semantics of the University of Italian Switzerland, and three anonymous referees for this journal. The paper is much improved thanks to their detailed feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is funded by the FCT doctoral research [grant number 2021.06938.BDC.]

References

Anderson, A., and N. Belnap. 1975. *Entailment: The Logic of Relevance and Necessity I*. Princeton: Princeton University Press.

Berto, F. 2022. *Topics of Thought*. Oxford: Oxford University Press.

Berto, F., and P. Hawke. 2021. "Knowability Relative to Information." *Mind: A Quarterly Review of Psychology and Philosophy* 130 (517): 1–33. https://doi.org/10.1093/mind/fzy045.

Berto, F., and M. Jago. 2019. *Impossible Worlds*. Oxford: Oxford University Press.

Berto, F., and M. Jago. 2023. "Impossible Worlds." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/impossible-worlds/.

Bjerring, J. C. 2013. "Impossible Worlds and Logical Omniscience: An Impossibility Result." *Synthese* 190 (13): 2505–2524. https://doi.org/10.1007/s11229-011-0038-y.

Bjerring, J. C., and M. Skipper. 2019. "A Dynamic Solution to the Problem of Logical Omniscience." *Journal of Philosophical Logic* 48 (3): 501–521. https://doi.org/10.1007/s10992-018-9473-2.

Bjerring, J. C., and M. Skipper. Forthcoming. "Hyperintensionality and Topicality: Remarks on Berto's Topics of Thought." *Analysis*. https://doi.org/10.1093/analys/anad077.

Ciardelli, I., J. Groenendijk, and F. Roelofsen. 2019. *Inquisitive Semantics*. Oxford: Oxford University Press.

Everett, A. 1993. "A Note on Priest's 'Hypercontradictions'." *Logique et Analyse* 36:39–43.

Fine, K. 2020. "Yablo on Subject-matter." *Philosophical Studies* 177 (1): 129–171. https://doi.org/10.1007/s11098-018-1183-7.

Fine, Kit. 2021. "Constructing the impossible." In *Conditionals, Paradox, and Probability: Themes from the Philosophy of Dorothy Edgington*, edited by Lee Walters and John Hawthorne. Oxford: Oxford University press.

Gioulatou, I. 2016. "Hyperintensionality." Master's thesis, Master of Logic at the University of Amsterdam.

Harman, G. 1986. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.

Hawke, P. 2018. "Theories of Aboutness." *Australasian Journal of Philosophy* 96 (4): 697–723. https://doi.org/10.1080/00048402.2017.1388826.

Hawke, P., A. Özgün, and F. Berto. 2020. "The Fundamental Problem of Logical Omniscience." *Journal of Philosophical Logic* 49 (4): 727–766. https://doi.org/10.1007/s10992-019-09536-6.

Hintikka, J. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.

Hintikka, J. 1975. "Impossible Possible Worlds Vindicated." *Journal of Philosophical Logic* 4 (4): 475–484. https://doi.org/10.1007/BF00558761.

Hoek, D. 2022. "Questions in Action." *Journal of Philosophy* 119 (3): 113–143. https://doi.org/10.5840/jphil202211938.

Hoek, D. Forthcoming. "Minimal Rationality and the Web of Questions." In *Unstructured Content*, edited by D. Kindermann, P. van Elswyk, A. Egan, and C. D. Kirk-Giannini. Oxford: Oxford University Press.

Jago, M. 2007. "Hintikka and Cresswell on Logical Omniscience." *Logic and Logical Philosophy* 15:325–354. https://doi.org/10.12775/LLP.2006.019.

Jago, M. 2009. "Logical Information and Epistemic Space." *Synthese* 167:327–341. https://doi.org/10.1007/s11229-008-9411-x.

Jago, M. 2014. *The Impossible: An Essay in Hyperintensionality*. Oxford: Oxford University Press.

Konolige, K. 1986. *A Deduction Model of Belief*. San Francisco: Morgan Kaufman.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. 1982. "Logic for Equivocators." *Noûs* 16 (3): 431–441.

Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. 1988. "Relevant Implication." *Theoria* 54 (3): 161–174. https://doi.org/10.1111/theo.1988.54.issue-3.

Nolan, D. 1997. "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic* 38 (4): 535–572. https://doi.org/10.1305/ndjfl/1039540769.

Plebani, M., and G. Spolaore. 2021. "Subject Matter: A Modest Proposal." *The Philosophical Quarterly* 71 (3): 605–622. https://doi.org/10.1093/pq/pqaa054.

Priest, G. 1998. "What is so Bad About Contradictions." *Journal of Philosophy* 95 (8): 410–426. https://doi.org/10.2307/2564636.

Priest, G. 2005. *Towards Non-being: The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.

Soames, S. 2010. *What is Meaning*. Princeton: Princeton University Press.

Stalnaker, R. 1984. *Inquiry*. Cambridge, MA: MIT Press.

Van Inwagen, P. 1986. "Two Concepts of Possible Worlds." In *Midwest Studies in Philosophy, XI*, edited by T. Uehling, and H. Wettstein, 185–213. Minneapolis, Minnesota: University of Minnesota Press.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Yablo, S. 2014. *Aboutness*. Princeton: Princeton University Press.

Yablo, S. 2018. "Reply to Fine on Aboutness." *Philosophical Studies* 175 (6): 1495–1512. https://doi.org/10.1007/s11098-017-0922-5.

Yalcin, S. 2018. "Belief as Question-Sensitive." *Philosophy and Phenomenological Research* 97 (1): 23–47. https://doi.org/10.1111/phpr.2018.97.issue-1.