Journal of Philosophical Logic
https://doi.org/10.1007/s10992-022-09694-0

The Logic of Framing Effects

Francesco Berto[1,2] · **Aybüke Özgün**[2]

Received: 2 April 2021 / Accepted: 8 December 2022

Abstract
Framing effects concern the having of different attitudes towards logically or necessarily equivalent contents. Framing is of crucial importance for cognitive science, behavioral economics, decision theory, and the social sciences at large. We model a typical kind of framing, grounded in (i) the structural distinction between beliefs activated in working memory and beliefs left inactive in long term memory, and (ii) the topic- or subject matter-sensitivity of belief: a feature of propositional attitudes which is attracting growing research attention. We introduce a class of models featuring (i) and (ii) to represent, and reason about, agents whose belief states can be subject to framing effects. We axiomatize a logic which we prove to be sound and complete with respect to the class.

Keywords Aboutness · Subject matter · Framing · Hyperintensionality · Logical omniscience · Knowledge representation · Working memory · Long term memory

1 Framed Believers

Physicians tend to believe that some lung cancer patients should get surgery with a 90% one-month survival rate. Physicians tend not to believe that such patients should get surgery with a 10% first month mortality [35, 367]. People often have different attitudes towards differently presented logically or necessarily equivalent contents. This is called the *framing effect* [36]. A good deal of behavioral economics takes its cue from it. Unlike Econs, the fully consistent agents of classical economic

✉ Francesco Berto
fb96@st-andrews.ac.uk; f.berto@uva.nl

Aybüke Özgün
a.ozgun@uva.nl

[1] Arché, University of St. Andrews, St Andrews, UK

[2] ILLC, University of Amsterdam, Amsterdam, Netherlands

Published online: 17 January 2023

theory who well-order their preferences and maximize expected utility, humans can be framed: nudged into believing different things depending on how equivalent options are presented to them [58]. People will believe more in a certain economic policy when its employment rate is given, than when the corresponding unemployment rate is [20]. Early student registration is boosted by threatening lateness penalty more than by promising early bird discount [25]. Framing has momentous social consequences [14, 19, 40, 50]. We need a logic to represent and reason about framed believers.

Standard epistemic logic in the tradition of [32] won't do. It models agents closer to Econs than to humans: perfect thinkers whose belief or knowledge states are fully closed under logical consequence. Hintikkan agents cannot have different attitudes towards logically or necessarily equivalent propositional contents. Such contents are just sets of possible worlds: their identity is given by co-intensionality.

It is often granted, however, that human thought is hyperintensional: one can have an attitude towards some content, without having it towards contents implied by or necessarily equivalent to it. One is not informed that one's neighbor John is Jack the Ripper although one has no doubts on John's self-identity. One can know that 7 + 5 = 12, without knowing that $x^n + y^n = z^n$ has no solutions in positive integers for $n > 2$; or that $p \rightarrow p$ is a logical truth, without knowing that a long and complicated propositional tautology is. One can desire that one's headache goes away, but one doesn't desire that one has a headache. More controversially, one can know that one has hands without being in a position to know that one is no brain in a vat. One can believe that φ without believing that $\varphi \wedge (\varphi \vee \psi)$ although the two are equivalent in classical and in various non-classical logics, for one lacks some concept needed to grasp ψ's content.

What *kind* of logical non-omniscience is involved in typical framing effects like the ones of the aforementioned examples? It cannot be tied to the a priori/a posteriori distinction, as in the 'John is John' vs. 'John is Jack the Ripper' case: that the survival rate is 90% is neither more nor less a priori than that mortality is 10%. Nor can it be due to computational limitations (it's easy to compute 7 + 5, whereas the proof of Fermat's Last Theorem has baffled mathematicians for two centuries), or difficulties in parsing long and syntactically complex sentences (the $p \rightarrow p$ versus complicated tautology case): the sentences 'The survival rate is 90%' and 'The mortality (rate) is 10%' are just as easy to parse as each other.

One may not want to assume at the outset that the problem is with the nature of the attitude itself. Idealized, perfect reasoners can desire that their headache goes away without desiring that they have a headache to begin with. Perhaps, due to looming skeptical paradoxes, they don't know they are not handless recently envatted brains although they know they have hands and (they even know that) the former follows from the latter: perhaps logical closure fails even for their knowledge states. (The jury is out on this: see e.g. [29, 39, 46, 55] among the anti-closure, [31, 38, 61] among the pro-closure.) What we are after now, however, is belief. When the case for knowledge not being closed under logical consequence even for perfect reasoners is presented, their being 'ideally astute' is usually *defined* in terms of belief: they believe all the logical consequences of what they know, and therefore believe [18, 33, 46]. The open issue is whether that's sufficient for the closure of their knowledge states.

Could the kind of logical non-omniscience displayed by agents with framed beliefs be due to lack of concepts, as when one believes φ but not an entailed ψ because one doesn't have some notion required to grasp ψ? To adapt [57], 88: William III believed that England could avoid war without believing that England could avoid nuclear war. That's because he had no idea what a nuclear weapon might be. He could have no such attitude towards propositions whose grasp involved a concept he lacked.

This gets us closer to the phenomenon we're after, but not close enough. Surely human thinkers have a limited repertoire of concepts, but that's not what is involved in typical framing effects. Framed physicians have all the concepts needed to fully grasp both the proposition that the survival rate is 90% and the one that the mortality is 10%. In particular, they are fully on top both of the concept *survival* and of the concept *mortality* by any conceptual or semantic competence test. Still, only the former proposition gets them to believe that the patients should take surgery.

What is going on, framing theorists say [35, 36], is that 'The mortality is 10%', but not 'The survival rate is 90%', makes people think about mortality. The thought that the survival rate is 90% is not *about* that: on the face of it, it's about survival. Survival and death are deeply connected in anyone's mind. But, cognitively limited as we are, we may not think about mortality – and much of what comes with it – when we think about survival rates, even if we have the concept *mortality* firmly in our repertoire. We leave it asleep. In order to think that the mortality is 10%, instead, we have to think about mortality, for that's what the proposition is about. As [62], Ch. 7, has it: an epistemic rift can open up between logically or necessarily equivalent propositions when they differ in *subject matter* – what they are about – even for thinkers who have all the relevant concepts.

The framing effects we aim to model, thus, involve the having of different attitudes towards equivalent propositions one perfectly grasps, due to differences in what those propositions are about. We grant that this is not the only way agents can be framed: qua psychological phenomenon, framing may involve all sorts of subtle pragmatic cues and mental associations triggered by word order, emphasis, etc. – we will see that approaches in doxastic logic different from ours may be even more suitable to model kinds of framing tied to the syntax-sensitivity of agents. But aboutness-based framing is a typical kind of framing, we conjecture, because it has deep roots, on the one hand, in the structure of our belief system, and on the other, in the nature of its contents. Our logic of framing will represent both roots. We introduce them in the following Section.

2 Working Memory, Long Term Memory, Aboutness

To model the structural features of our belief system responsible for aboutness-based framing, we should look at a widely accepted acquisition of cognitive psychology: the distinction between working and long term memory [22, Part II]. Researchers disagree on the nature of both. Qua logical modelers, we don't want our account to be held hostage to the next empirical discovery, or consensus switch in psychological research. Luckily, we can be neutral on the more controversial issues and take on board the less controversial ones. For instance, working memory (WM), which deals

with the processing and short-term storage of information, is at times understood as encompassing a buffer of data at hand for the performing of cognitive tasks, plus a central executive unit: the locus of attention and cognitive control [2, 3]; at times, as a plurality of modules or structures [6]. For our purposes, we only need to consider its most agreed-upon feature: it has limited capacity. Only few chunks of information can be retained in WM, and only for a limited amount of time (see the views compared in [44]).

Instead, long-term memory (LTM) (or, the declarative part of it: [52, 56]) is that vast knowledge base where cognitive agents store, or encode, their beliefs and knowledge about specific events (the so-called episodic memory) as well as general laws and principles (the so-called semantic memory). There's a divide in cognitive psychology, on whether WM and LTM are separate (contents are stored in LTM and retrieved from it for use in WM), or the former is just the activated part of the latter [1, 15, 44]. We can be neutral on this as well.

Now our framed agents, we propose, can have the belief that patients should get surgery with a 90% one-month survival rate activated in their *working* memory, without having the intensionally equivalent belief that patients should get surgery with a 10% first month mortality there. However, our agents can have all the relevant information and, in particular, the concept *mortality*, in their (declarative) LTM. Let's call beliefs activated in WM, *active*, and beliefs left asleep in LTM, *passive*. A belief is active when it is available in WM to perform cognitive tasks with it. It is passive when it is stored, or encoded, in the agent's LTM, and left inactive there.

As for the contents of active and passive beliefs: we propose to embed topics or subject matters in the very notion of proposition. Our starting point is the venerable idea of intentionality: mental states such as belief bear *aboutness* – as Yablo has it, 'the relation that meaningful items bear to whatever it is that they are *on* or *of* or that they *address* or *concern*' [62, p. 1]. This is their topic, or subject matter. The aboutness of an intentional state, such as believing that φ, must be in line with that of the proposition, P, which makes for the content of φ. P should not be understood, then, just as a set of possible worlds, but also in terms of its topic or subject matter. When one believes that John is handsome, one's belief is about *John's looks* insofar as that's what the believed proposition is about.

What is a topic or subject matter? We are familiar with truth conditions and truth sets (typically, sets of possible worlds) as specifying propositional contents. One may be less familiar with topics, although the literature is burgeoning. One way of understanding them links to questions. Lewis [41] interpreted subject matters as partitions of the modal space (see also [34]). Take *the number of stars*. The associated question is, 'How many stars are there?'. This splits the total set of worlds into equivalence classes. Two worlds are in the same cell when they agree on the answer: all zero-star worlds, all one-star worlds, etc. [62] goes for divisions (admitting overlap), rather than partitions. Others [24] understand topics in terms of truthmakers, interpreted as constructions out of states (think of something like the situations of [7]).

For our logical modeling purposes, however, we don't need to take a stance on what topics are. We just take on board three constraints needed for our logic of framing, following also [9–11, 30]. We will not defend the constraints, because

researchers on subject matter generally agree on them, and we piggyback on their agreed upon results:

1. Logically or necessarily equivalent sentences φ and ψ can differ in their propositional content because of differences in what they are about. E.g., in [62]'s version of subject matter theory, propositional content is hyperintensional because it's not given only by truth set, but also by aboutness. 'Equilateral triangles are equiangular' and 'Either John passes the exam or not' differ in content in spite of sameness of truth set, because only one is about *equilateral triangles*, and made true by how such triangles are. Yablo calls mere truth sets, 'thin propositions'; truth sets enriched with subject matters, 'thick propositions'.
2. The space of topics has a mereological structure [24, 34, 62]. Topics can have proper parts; distinct topics may have common parts. *Mathematics* includes *arithmetic*. *Mathematics* and *philosophy* share subject matter, having (certain parts of) *logic* in common. Correspondingly, what a sentence is about can overlap with, or be properly included in, what another one is about.
3. The Boolean operators are subject-matter-transparent: they add no topic of their own [24, 49, 62]. 'John isn't handsome' is exactly about what 'John is handsome' is about, say, *John's looks*. It certainly does not speak about *negation*. The topic of $\neg\varphi$ should be the same as that of φ. Conjunction and disjunction merge topics: 'John is tall and handsome' and 'John is tall or handsome' are both about *the height and looks of John*. The topic of $\varphi \wedge \psi$ and that of $\varphi \vee \psi$ are the same: the fusion of the topic of φ and that of ψ.

Supposing this is enough qua introduction to WM, LTM, and topics, here are some desiderata our logic of framing should comply with. First, an interpreted sentence expresses a thick proposition. Thick propositions are the contents of both active and passive belief. Belief as such, then, is, as [47, 63] have it, topic-sensitive. We evaluate ascriptions of active belief with respect to the agents' WM, and ascriptions of passive belief with respect to their LTM. In this sense, both WM and LTM are topic-sensitive.

Next, a realistic framed believer should be non-omniscient with respect to *both* WM *and* LTM. Psychologists contrast the limited capacity of the former with the breadth of the latter. However, neither should host all the logical consequences of what it hosts, or display an omni-inclusive conceptual repertoire. In particular, both passive and active belief must be hyperintensional: framed agents are not logically closed with respect to either.

Next, whether WM is separate from LTM, or just the activated part of the latter, no information or concept can be in WM unless it is in LTM to begin with: agents cannot have any attitude on subject matters whose concepts they simply lack. They are as blind to them as William III was to the topic of nuclear war avoidance.

To get an idea of how such desiderata cooperate, consider the following two triplets of group-wise intensionally equivalent sentences:

A. $7 + 5 = 12$.
B. $x^n + y^n = z^n$ has no solutions in positive integers for $n > 2$.
C. Extremally disconnectedness is not a hereditary property of topological spaces.

D. Triangles have three sides.

E. Bachelors are unmarried.
F. Baryons are hadrons with odd numbers of valence quarks.

A.–C. express contents which are necessary, of the same kind of necessity (say, mathematical necessity). Ditto for D.–F. (say, definitional necessity). Our framed believers could find themselves in the following situation with respect to each triplet. (i) They passively believe the content expressed by the first item, A., or D.: they have the relevant information and they are on top of the basic arithmetical or geometrical subject matter involved, so it's all stored or encoded in LTM. They are just not thinking about arithmetic, or about triangles, at the moment. (ii) They actively believe the content expressed by the second item, B. or E.: they have the relevant proposition in their WM because they are currently engaged in thoughts about diophantine equations, or John's marital status. (iii) They neither actively nor passively believe the content expressed by the third item, C. or F.: they just have no idea what topological spaces are and what features they have, or they have never heard about exotic notions from particle physics. They are as blind to them as William III was to nuclear war avoidance.

Before we get to our own proposal to model agents of this kind, we briefly discuss some hyperintensional epistemic logics for non-logically omniscient agents already on the market, to see to what extent they could be used to represent framing.

3 Some Ideas on the Market

As far as we know, few hyperintensional epistemic logics have aimed at directly representing the difference between WM and LTM. One distinction which may look similar is the one between *explicit* and *implicit* knowledge and belief, found in awareness logics developed with an eye on the logical omniscience problem [8, 23, 59]. Unawareness is lack of conception, rather than information: being unaware of φ is understood as not having φ present in the mind, or not thinking about φ [53, 79-80]. Thus, it seems especially suitable to model framing.

In [23]'s approach, awareness is represented *syntactically*. One is aware of φ when φ belongs to a set of formulas, thus linguistic items: the agent's awareness set $\mathcal{A}$. Implicit knowledge or belief get the usual Hintikkan characterization, whereas the corresponding explicit attitudes are defined as the combination of the implicit ones with awareness: an agent has the explicit attitude towards φ when the agent has the implicit one and $\varphi \in \mathcal{A}$.

The view has been claimed to mix syntax and semantics, essentially imposing a syntactic filter over a standard Hintikkan semantics [37]. Resorting to syntax, however, allows very fine-grained distinctions: as any bunch of sentences can serve as the awareness set $\mathcal{A}$, explicit attitudes obey no non-trivial logical closure properties. This is all good and well if one has a syntactic or sentential conception of belief. However, this is not the conception of belief qua topic-sensitive intentional state we have endorsed above. Such a conception puts a limit to the amount of fine-grainedness one can plausibly assign to belief contents. In our approach, a framed agent who actively believes $\varphi \wedge \psi$ should actively believe $\psi \wedge \varphi$, and should actively believe φ, provided the agent has parsed the syntax of the sentences that express the relevant contents.

Although the sentences we may use in our logical language to ascribe beliefs to the agent are different, the propositional content that John is tall and handsome and the one that John is handsome and tall are intensionally equivalent, and the agent who actively believes either is already thinking about the other's topic – because it is the same topic, say, *John's height and looks*. That John is tall and handsome entails that John is tall, and one who actively believes the former is already thinking about the topic of the latter, as it is part of that of the former. As Williamson – a subscriber to the semantic conception of belief – has it:

> If a positive propositional attitude is closed under at least some forms of logical consequence [...], we may expect it to be closed under a very intimate one such as the $\wedge$-elimination inference from $p \wedge q$ to p and to q. [...] There is no obstacle to the idea that knowing a conjunction *constitutes* knowing its conjuncts, just as, in mathematics, we may *count* a proof of a conjunction as a proof of its conjuncts, so that if $p \wedge q$ is proved then p is proved, not just provable. [61, 277, 283]

To be fair (and to follow a useful suggestion by a reviewer of our paper), syntactic approaches can easily mimic closure properties of belief, including conjunction elimination and inversion: see e.g. the recent cutting-edge work of [42, 43] on belief bases (classic works on belief bases include [28, 51]). Besides, works in the syntactic-sentential tradition can be useful in modeling some specific kinds of framing, e.g., presentation order effects. We get a long list of search results on Amazon and we stop when we find an article we judge satisfactory. Had the same items in the list been arranged differently, with the article further down, we may never have bought it. If we take the list as a long conjunction of sentences, φ & ψ & χ & ..., order matters [53, 83].

We have already seen, however, that this is not the kind of non-omniscience going on in our typical cases of framing above. In 'The survival rate is 90%' vs. 'The mortality (rate) is 10%', or 'You get early bird discount' vs. 'You get late registration surcharge', the sentences are no conjunct lists and the syntax of either in the pair is just as easy to parse as that of the other. The framed agents whose modelling we're after have correctly parsed the syntax of the relevant sentences and are fully on top of the expressed *contents*.

One variant of the awareness approach, 'propositionally determined awareness' (see [27, 327], focusing on knowledge, and [53, 84]), puts a constraint on awareness sets: one is aware of φ just in case one is aware of all of its atomic constituents taken together. This automatically delivers closure under conjunction elimination and other closure properties, taking it closer to the approach we present below – but still features a mixture of syntax and semantics to achieve the result. A properly semantic account of topics or subject matters as non-linguistic items, like the one we are pursuing, should allow for different (atomic) sentences to be assigned, on occasion, the same subject matter, just as they can be assigned, on occasion, the same truth set. Overall, explicit attitudes in the awareness setting do not map very neatly to our active belief as a topic-sensitive attitude implemented in the agent's WM.

Nor do implicit attitudes neatly map to our passive belief as implemented in LTM. Logics featuring the explicit/implicit distinction usually take the implicit attitude as

a normal Hintikkan modality. The attitude displays, thus, full logical omniscience: the agent implicitly knows or believes all logical truths, and all logical consequences of what it knows or believes. The agent has no awareness or conceptual limitations there: it is simply on top of all the relevant propositions. But, as we have remarked, LTM is not like that. Realistic agents don't possess all concepts, and don't have all the logical consequences of their passive beliefs stored or encoded in LTM: passive belief should be hyperintensional, too.

Balbiani et al. [4] present one of the few logical works that explicitly aim at modelling the WM/LTM distinction. It's a powerful framework in the tradition of dynamic epistemic logic [17], modelling the processes through which a non-omniscient agent forms its beliefs via operations of perception and inference in WM, and can store and retrieve them from LTM. Their language has an operator for explicit belief, tied to WM, and one expressing background knowledge, tied to LTM. The latter is a normal modality, and so faces the same issue as implicit knowledge in the awareness setting: the agent is logically omniscient with respect to its background knowledge.

What's more worrying for the prospects of applying the logic to framing, is that explicit belief gets a Scott-Montague neighborhood semantics [48, 54]: one explicitly believes that φ at world w when φ's truth set is in the neighborhood set assigned to w. Famously, neighborhood semantics gives weak non-normal modal logics capable of breaking a number of logical closure features for their operators. In particular, one can explicitly believe a conjunction without explicitly believing the conjuncts. If one does want to enforce conjunction elimination for explicit belief, one can, of course, add conditions (specifically, one could close the neighborhoods under supersets for $\wedge$-elimination: see [48, 81]).

But even in the basic neighborhood setting, when φ and ψ are logically or necessarily equivalent, they will be in the same set of neighborhoods. Thus, explicit belief in either will automatically entail explicitly belief in the other. This is exactly what shouldn't happen if we want to capture framing for explicit beliefs. Neighborhoods alone don't deliver the right kind of hyperintensionality and non-omniscience.

We now move on to our own proposal and start making things precise.

4 Logic for Framed Believers

4.1 Language and Models

Our logic of belief for framed agents is based on a modal language $\mathcal{L}$ with a countable set of propositional variables $\mathsf{Prop} = \{p_1, p_2, \dots\}$. $\mathcal{L}$ is defined recursively by the grammar:

$$\varphi := p_i \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi \mid B_A\varphi \mid B_P\varphi$$

where $p_i \in \mathsf{Prop}$. We read $B_A\varphi$ as 'the agent actively believes that φ', $B_P\varphi$ as 'the agent passively believes that φ', and $\Box\varphi$ as a normal epistemic-a priori modality, 'it is a priori that φ'. (Thus, the box does not represent what *our* modeled agents know a priori; rather, what logically omniscient Hintikkan agents may know a priori. We will use it in some logical invalidities below, to mark how our agents differ from the logically omnisicient ones.) When we talk about both B_A and B_P, we simply

write $B_\star$. We often use $p, q, r, ...$ for propositional variables and employ the usual abbreviations for propositional connectives $\vee, \rightarrow, \leftrightarrow$ as $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi := \neg\varphi \vee \psi$, and $\varphi \leftrightarrow \psi := (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$; and for the duals $\Diamond\varphi := \neg\Box\neg\varphi$ and $\widehat{B_\star}\varphi := \neg B_\star\neg\varphi$. We define $\top := p_1 \vee \neg p_1$ and $\bot := \neg\top$. We follow the usual rules for the elimination of the parentheses. For any $\varphi \in \mathcal{L}$, $Var(\varphi)$ denotes the set of propositional variables occurring in φ.

We interpret this language on *topic-sensitive subset space models* (inspired by the original subset space semantics of [45], see below for a comparison):

Definition 1 (Topic-sensitive Subset Space Model) A *topic-sensitive subset space model* (*model*) $\mathcal{M}$ is a tuple $\langle W, \mathcal{O}, T, \oplus, t, v\rangle$ where

1. W is a non-empty set of possible worlds;
2. $\mathcal{O}$ is a non-empty, finite subset of $\mathcal{P}(W)$ such that $\mathcal{O} \neq \{\emptyset\}$: each non-empty $O \in \mathcal{O}$ represents the informational content of a *memory cell*;
3. T is a non-empty set of topics;
4. $\oplus : T \times T \rightarrow T$ is a binary idempotent, commutative, associative operation: *topic fusion*. We assume unrestricted fusion, that is, $\oplus$ is always defined on T: $\forall a, b \in T\ \exists c \in T(c = a \oplus b)$;
5. $t : \mathsf{Prop} \cup \mathcal{O} \rightarrow T \cup \mathcal{P}(T)$ is a *topic function* assigning a topic to each element in Prop and a non-empty, finite set of topics to each element in $\mathcal{O}$: $t(p) \in T$ for all $p \in \mathsf{Prop}$, and $t(O) \in \mathcal{P}(T)$ is non-empty and finite for all $O \in \mathcal{O}$. Topic function t extends to the whole language $\mathcal{L}$ by taking the topic of a sentence φ as the fusion of the elements in $Var(\varphi)$:

$$t(\varphi) = \oplus Var(\varphi) = t(p_1) \oplus \cdots \oplus t(p_k)$$

   where $Var(\varphi) = \{p_1, \ldots, p_k\}$;
6. $v : \mathsf{Prop} \rightarrow \mathcal{P}(W)$ is a valuation function that maps every propositional variable in Prop to a set of worlds in W.

In the metalanguage we use variables a, b, c $(a_1, a_2, \ldots)$ ranging over possible topics. We define *topic parthood*, $\sqsubseteq$, out of topic fusion as

$$\forall a, b(a \sqsubseteq b \text{ iff } a \oplus b = b).$$

Thus, $(T, \oplus)$ is a join semilattice and $(T, \sqsubseteq)$ a partially ordered set. The *strict topic parthood*, denoted by $\sqsubset$, is defined as usual as $a \sqsubset b$ iff $a \sqsubseteq b$ and $b \not\sqsubseteq a$.

Here's what the model represents: the agent's belief system is composed of *memory cells*. These are chunks of LTM which can be put into (or, if one prefers, activated as) WM, that is, made available for actions of cognitive processing. A memory cell is represented by an indexed set O_a, where $\emptyset \neq O \in \mathcal{O}$ and $a \in t(O)$: O_a is made of informational content O and topic a.

Memory cells are topic-sensitive: when one is in (or activated as) WM, the agent is actively thinking about its subject matter, and has its informational content available for processing. $t(O)$ and $\mathcal{O}$ are assumed to be finite in order to make the framework realistic: our cognitive agents can only have finitely many memory cells.

Every $O \in \mathcal{O}$ is assigned a set of topics, rather than a single topic (Definition 1.5), to capture the key idea that the same informational content can be associated with different topics. Take our triplet of intensionally equivalent, topic-diverging sentences A., B. and C. in Section 2. Intensional equivalence means that they have the same bunch of worlds as their truth set. Call it X. Let the topics be a, b, and c, respectively. Each of X_a, X_b, and X_c can make for a distinct memory cell, differing from the others in topic but not in informational content.

The agent's LTM, then, is easily defined:

$$LTM := \left(\bigcap \mathcal{O}\right)_{\oplus(\bigcup_{O\in\mathcal{O}} t(O))}.$$

The information stored in LTM is the information available in all memory cells, taken together. The topic of LTM is the fusion of those of all memory cells: the total repertoire of topics or subject matters the agent has grasped. To simplify the notation, we set $\bigcap \mathcal{O} := O^{\cap}$ and $\oplus(\bigcup_{O\in\mathcal{O}} t(O)) := \mathfrak{b}$. Then the LTM of the agent is $O^{\cap}_{\mathfrak{b}}$. Notice that $\mathfrak{b}$ is guaranteed to be in T, since $\bigcup_{O\in\mathcal{O}} t(O)$ is finite.

LTM is at least as large as any single memory cell which can be activated as, or put into, WM, with respect to both information and topic. The agent passively believes, i.e., has in LTM, at least as much as it can actively believe, i.e., activate and process in WM: the latter has quite limited capacity compared to LTM, as cognitive psychology taught us.

This is made precise by spelling out the truth conditions for sentences of our language $\mathcal{L}$. We evaluate formulas with respect to *world-memory pairs* (w, O_a), with $w \in W$ representing the actual world, and O_a a memory cell. We denote the set of all world-memory pairs of a model $\mathcal{M}$ by $P(\mathcal{M})$ (P for *pair*). The *working memory* WM is the designated world-memory pair with respect to which we evaluate formulas:

Definition 2 ($\Vdash$-Semantics for $\mathcal{L}$) Given a model $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, v\rangle$ and a world-memory pair $(w, O_a) \in P(\mathcal{M})$, the $\Vdash$-semantics for $\mathcal{L}$ is defined recursively as:

$$\begin{array}{lll}
\mathcal{M}, (w, O_a) \Vdash p & \text{iff} & w \in v(p) \\
\mathcal{M}, (w, O_a) \Vdash \neg\varphi & \text{iff} & \text{not } \mathcal{M}, (w, O_a) \Vdash \varphi \\
\mathcal{M}, (w, O_a) \Vdash \varphi \wedge \psi & \text{iff} & \mathcal{M}, (w, O_a) \Vdash \varphi \text{ and } \mathcal{M}, (w, O_a) \Vdash \psi \\
\mathcal{M}, (w, O_a) \Vdash \Box\varphi & \text{iff} & W \subseteq [\![\varphi]\!]^{O_a}_{\mathcal{M}} \\
\mathcal{M}, (w, O_a) \Vdash B_A\varphi & \text{iff} & O \subseteq [\![\varphi]\!]^{O_a}_{\mathcal{M}} \text{ and } t(\varphi) \sqsubseteq a \\
\mathcal{M}, (w, O_a) \Vdash B_P\varphi & \text{iff} & O^{\cap} \subseteq [\![\varphi]\!]^{O_a}_{\mathcal{M}} \text{ and } t(\varphi) \sqsubseteq \mathfrak{b}
\end{array}$$

where $[\![\varphi]\!]^{O_a}_{\mathcal{M}} := \{w \in W : \mathcal{M}, (w, O_a) \Vdash \varphi\}$. We omit the subscript $\mathcal{M}$ for models and write $[\![\varphi]\!]^{O_a}$ when it is contextually clear. When it is not the case that $\mathcal{M}, (w, O_a) \Vdash \varphi$, we write $\mathcal{M}, (w, O_a) \nVdash \varphi$.

As the semantics has it, the agent actively believes whatever is entailed by their WM with respect to both informational content and topic. The agent passively believes whatever is entailed by their LTM (ditto).

As the following lemma shows, only the truth value of an ascription of *active* belief depends on the chosen O_a:

Lemma 1 *Given a model* $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$, $w \in W$, *two world-memory pairs* $(w, O_a), (w, U_c) \in P(\mathcal{M})$, *and* $\varphi \in \mathcal{L}$ *such that* φ *does not have any occurrences of* B_A, *we have*

$$\mathcal{M}, (w, O_a) \Vdash \varphi \text{ iff } \mathcal{M}, (w, U_c) \Vdash \varphi.$$

Proof See Appendix. □

However, the agent can actively believe φ with respect to one memory cell without actively believing the same content with respect to another one. That is, given a model $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$ and two world-memory pairs $(w, O_a), (w, U_c) \in P(\mathcal{M})$, it could be that $\mathcal{M}, (w, O_a) \Vdash B_A\varphi$ and $\mathcal{M}, (w, U_c) \nVdash B_A\varphi$ for some $\varphi \in \mathcal{L}$ as shown by the following example.

Example 1 Consider the model $\mathcal{M} = \langle \{w, w_1, w_2\}, \{O, U\}, \{a, \mathfrak{b}, c, d\}, \oplus, t, \nu \rangle$ such that $O = \{w, w_1\}$, $U = \{w, w_2\}$, and $(\{a, \mathfrak{b}, c, d\}, \oplus)$ constitutes the join-semilattice in Fig. 1b. For t and ν, we have $t(p) = a, t(q) = c$, $t(r) = d$, $t(O) = t(U) = \{a, c\}$, $\nu(p) = \nu(q) = \{w, w_1\}$, and $\nu(r) = \{w, w_2\}$. It is than easy to see that $(w, O_a) \Vdash B_A p$ since $O \subseteq \nu(p)$ and $t(p) \sqsubseteq a$. However, $(w, O_c) \nVdash B_A p$ since $t(p) \not\sqsubseteq c$, that is, the agent does not have the subject matter of p in working memory O_c. Similarly, we also have, e.g., $(w, U_c) \nVdash B_A p$ for two reasons: (1) $U \not\subseteq \nu(p)$ and (2) $t(p) \not\sqsubseteq c$, that is, the informational content of U_c does not eliminate all non-p possibilities and the subject matter of p is not part of the subject matter of working memory U_c, respectively (see Fig. 1).

Next comes the definition of *logical consequence (with respect to* $\Vdash$*):* with $\Sigma \subseteq \mathcal{L}$ and $\varphi, \psi \in \mathcal{L}$,

- $\Sigma \vDash \varphi$ iff for all models $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$ and all $(w, O_a) \in P(\mathcal{M})$: if $\mathcal{M}, (w, O_a) \Vdash \psi$ for all $\psi \in \Sigma$, then $\mathcal{M}, (w, O_a) \Vdash \varphi$.
- For single-premise entailment, we write $\psi \vDash \varphi$ for $\{\psi\} \vDash \varphi$.
- As a special case, *logical validity*, $\vDash \varphi$, truth at all world-memory pairs of all models, is $\emptyset \vDash \varphi$, entailment by the empty set of premises. φ is called

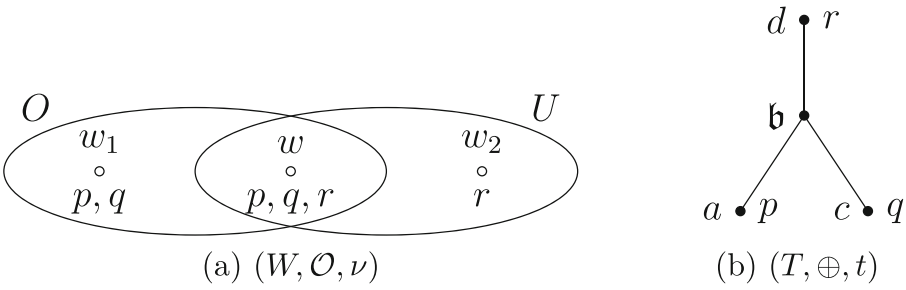

Fig. 1 Model $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$ in Example 1: White nodes represent possible worlds, black nodes represent possible topics, ellipses represent the informational contents of memory cells. Valuation and topic assignment are given by labelling each node with atomic formulas

invalid, denoted by $\nvDash \varphi$, if it is not a logical validity, that is, if there is a model $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$ and a world-memory pair $(w, O_a) \in P(\mathcal{M})$ such that $\mathcal{M}, (w, O_a) \nVdash \varphi$.

Soundness and *completeness* with respect to the proposed semantics are defined standardly (see, e.g., [13, Chapter 4.1]).

The abbreviation $\overline{\varphi} := \bigwedge_{x \in Var(\varphi)} (x \vee \neg x)$ will play a role in formalizing validities and invalidities.[1] Given a model $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$, it is easy to see that $\overline{\varphi}$ is true at every world-memory pair in $P(\mathcal{M})$ and $Var(\overline{\varphi}) = Var(\varphi)$ for any $\varphi \in \mathcal{L}$. Thus, we can talk in the language about what topics the agent is actively thinking about in WM, and what topics the agent has grasped and stored in LTM. Formulas of the form $B_A\overline{\varphi}$ ($\neg B_A\overline{\varphi}$) express within the language $\mathcal{L}$ statements such as 'The agent has (does not have) the subject matter of φ in WM':

$$\begin{aligned} \mathcal{M}, (w, O_a) \Vdash B_A\overline{\varphi} \quad & \text{iff } O \subseteq [\![\overline{\varphi}]\!]^{O_a} \text{ and } t(\overline{\varphi}) \sqsubseteq a \\ & \text{iff } O \subseteq W \text{ and } t(\varphi) \sqsubseteq a \\ & \qquad (t(\overline{\varphi}) = t(\varphi), \text{since} Var(\overline{\varphi}) = Var(\varphi)) \\ & \text{iff } t(\varphi) \sqsubseteq a. \end{aligned}$$

Similarly, formulas of the form $B_P\overline{\varphi}$ ($\neg B_P\overline{\varphi}$) express within the language $\mathcal{L}$ statements such as 'The agent has (does not have) the subject matter of φ in LTM' (the proof follows similarly).

Our semantics is structurally similar to the *subset space semantics* of [45] in that the component $\langle W, \mathcal{O} \rangle$ of a topic-sensitive subset space model $\langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$ constitutes a subset space and we evaluate sentences not at worlds but at world-set pairs. Subset space semantics was originally designed to model an evidence-based notion of absolutely certain knowledge and epistemic effort. The evaluation pairs of the form (w, O) within this framework obey the constraint $w \in O$ (for knowledge is veridical) and are often called 'epistemic scenarios'. O represents the agent's current truthful evidence.

Our framework comes with a distinct formalism, however, and a different interpretation of a subset space model's components. We focus on belief rather than knowledge, so the evaluation pairs are tailored accordingly: as belief is not factive, a memory cell (w, O_a) does not have to meet the constraint $w \in O$ (see also [12] for a different subset space semantics for belief without this constraint). More importantly, our subset spaces and the corresponding evaluation pairs are endowed with topics. This makes the resulting logic of belief hyperintensional, as opposed to the intensional epistemic logics of the traditional subset space semantics (see, e.g., [12, 16, 45, 60]).

[1] In order to have a unique definition of each $\overline{\varphi}$, we set the convention that elements of $Var(\varphi)$ occur in $\bigwedge_{x \in Var(\varphi)} (x \vee \neg x)$ from left-to-right in the order they are enumerated in $\mathsf{Prop} = \{p_1, p_2, \dots\}$. For example, for $\varphi := B_\star(p_{10} \rightarrow p_2) \vee \Box p_7$, $\overline{\varphi}$ is $(p_2 \vee \neg p_2) \wedge (p_7 \vee \neg p_7) \wedge (p_{10} \vee \neg p_{10})$, and not $(p_{10} \vee \neg p_{10}) \wedge (p_7 \vee \neg p_7) \wedge (p_2 \vee \neg p_2)$ or $(p_7 \vee \neg p_7) \wedge (p_{10} \vee \neg p_{10}) \wedge (p_2 \vee \neg p_2)$ etc. This convention will eventually not matter since our logic cannot differentiate two conjunctions of different order: $\varphi \wedge \psi$ provably and semantically equivalent to $\psi \wedge \varphi$.

Table 1 Axiomatization of the logic of framed belief L over $\mathcal{L}$

(CPL)	All classical propositional tautologies and Modus Ponens
($\mathsf{S5}_{\Box}$)	S5 axioms and rules for $\Box$
	(I) Axioms for $B_\star$, with $\star \in \{A, P\}$:
($\mathsf{C}_{B_\star}$)	$B_\star(\varphi \wedge \psi) \leftrightarrow (B_\star\varphi \wedge B_\star\psi)$
($\mathsf{Ax1}_{B_\star}$)	$B_\star\varphi \rightarrow B_\star\overline{\varphi}$
($\mathsf{Ax2}_{B_\star}$)	$(\Box(\varphi \rightarrow \psi) \wedge B_\star\varphi \wedge B_\star\overline{\psi}) \rightarrow B_\star\psi$
($\mathsf{Ax3}_{B_\star}$)	$B_\star\varphi \rightarrow \Box B_\star\varphi$
	(II) Axioms for B_A:
(D_{B_A})	$B_A\varphi \rightarrow \neg B_A\neg\varphi$
	(III) Axioms connecting B_A and B_P:
(Inc)	$B_A\varphi \rightarrow B_P\varphi$

4.2 Axiomatization, Soundness and Completeness

Table 1 gives a sound and complete axiomatization L of the *logic of framed belief* over $\mathcal{L}$.

The notion of derivation, denoted by $\vdash$, in L is defined as usual. Thus, $\vdash \varphi$ means φ is a theorem of L.

Theorem 2 L *is a sound and complete axiomatization of* $\mathcal{L}$ *with respect to the class of topic-sensitive subset space models: for every* $\varphi \in \mathcal{L}$, $\vdash \varphi$ *if and only if* $\vDash \varphi$.

Proof See Appendix. □

The axioms in Group I give general closure features of belief, both active and passive, for our framed agents. $\mathsf{C}_{B_\star}$ ensures that beliefs are fully conjunctive: one who believes that John is tall and handsome, believes both that John is tall and that John is handsome, and vice versa. $\mathsf{Ax1}_{B_\star}$ captures, as desired, the topic-sensitivity of belief: one can actively believe φ only if one is actively thinking about the relevant topic in WM; one can passively believe φ only if one has concepts for the relevant topic stored in LTM. $\mathsf{Ax2}_{B_\star}$ states a limited deductive closure principle for both active and passive belief: if ψ follows from φ a priori, and one believes φ, and one is on top of the subject matter of ψ, then one does believe ψ. $\mathsf{Ax3}_{B_\star}$ has it that beliefs are not world-relative.

In Group II, D_{B_A} states a consistency principle for active belief: one who has φ in WM will not also have $\neg\varphi$ there. This does not hold for passive belief: a realistic agent may have all sorts of inconsistent beliefs stored or encoded in its LTM. They can stay there insofar as one does not think about them all together, i.e., the inconsistencies are shielded from the focus of attention in WM.

As for Group III, the Inc principle bridging active and passive belief guarantees, as desired, that whatever is activated in WM be available in LTM to begin with.

Just as important as validities are *in*validities, as they display the extent to which our framed agents are non-omniscient. We discuss a few prominent ones:

1. **Omniscience Rule**: from φ infer $B_\star\varphi$
2. **A Priori Omniscience**: $\Box\varphi \to B_\star\varphi$
3. **Closure Under Strict Implication**: $(\Box(\varphi \to \psi) \wedge B_\star\varphi) \to B_\star\psi$
4. **Negative Introspection**: $\neg B_\star\varphi \to B_\star\neg B_\star\varphi$
5. **Framing a**: from $\varphi \leftrightarrow \psi$ infer $B_\star\varphi \leftrightarrow B_\star\psi$
6. **Framing b**: $(B_A\varphi \wedge B_P(\varphi \leftrightarrow \psi)) \to B_A\psi$
7. **Framing c**: from $\varphi \leftrightarrow \psi$ infer $(B_A\varphi \wedge B_P\overline{\psi}) \to B_A\psi$.

[*Countermodel*: Consider the model given in Example 1. We then have 1 and 2 invalid since $\models r \vee \neg r$ (therefore also $(w, O_a) \Vdash \Box(r \vee \neg r)$), but $(w, O_a) \nVdash B_A(r \vee \neg r)$ (since $t(r) \not\sqsubseteq a$) and $(w, O_a) \nVdash B_P(r \vee \neg r)$ (since $t(r) \not\sqsubseteq \mathfrak{b}$). For 3, take $\varphi := p$ and $\psi := r \vee \neg r$: $(w, O_a) \Vdash \Box(p \to (r \vee \neg r))$, $(w, O_a) \Vdash B_A p$, and $(w, O_a) \Vdash B_P p$, however, $(w, O_a) \nVdash B_A(r \vee \neg r)$, and $(w, O_a) \nVdash B_P(r \vee \neg r)$ as shown above. For 4, take $\varphi := r$: world-memory pair (w, O_a) falsifies it for B_A since $t(r) = t(\neg B_A r) \not\sqsubseteq a$ and falsifies it for B_P since $t(r) = t(\neg B_P r) \not\sqsubseteq \mathfrak{b}$. For 5, take $\varphi := p \vee \neg p$ and $\psi := r \vee \neg r$, and (w, O_a) falsifies the principle. For 6, take $\varphi := p$ and $\psi := q$: $(w, O_a) \Vdash B_A p$ and $(w, O_a) \Vdash B_P(p \leftrightarrow q)$, but $(w, O_a) \nVdash B_A q$ (since $t(q) \not\sqsubseteq a$). For 7, take $\varphi := p \vee \neg p$ and $\psi := q \vee \neg q$, and observe that (w, O_a) falsifies the principle.]

The failure of 1–3 tells us that our agents, unlike the logically omniscient Hintikkan agents, don't believe all (a priori) truths and that their beliefs are not closed under strict-a priori implication. 4 says that they also lack the wisdom of negative introspection: they can fail to believe that they don't believe something.

We've put the most important bit at the end: the last three invalidities, 5–7, crucially capture typical framing. **Framing a** guarantees that agents can have different attitudes towards equivalent formulas. **Framing b** says that one can have the belief that φ (e.g., patients should get surgery with a 90% one-month survival rate) activated in WM, without having the belief that ψ (patients should get surgery with a 10% first month mortality) there, even when one *does* have their equivalence in one's belief base: one is on top of all the relevant concepts and believes that either is true iff the other is. But all of this is left asleep in LTM: one is just not thinking about it. **Framing c** says that one's actively believing φ does not imply that one actively believes ψ, even when the two are equivalent *and* one has the subject matter of ψ in one's LTM. And so, we claim, our models capture precisely the phenomenon of framing we were after.

5 Conclusion and Further Work

We have presented a class of models and a logic, sound and complete with respect to the class, to reason about the beliefs of typical framed agents: non-logically

omniscient agents who can have different attitudes towards logically or necessarily equivalent contents they perfectly grasp. The two key ingredients we have adopted are (i) the topic-sensitivity of belief states, mirroring the one of the propositional contents of such states; and (ii) a distinction between WM and LTM, to model the idea that framed agents can actively think about φ in their WM, without thinking about an equivalent ψ which they, however, have in their LTM belief base.

Three directions of further investigation: first, both active and passive belief are plain, categorical forms of belief. It will be interesting to expand the language and formal semantics so that they include conditional, topic-sensitive belief.

Second, working memory is properly so-called in cognitive psychology because it is the locus of cognitive *activity*: beliefs are in there in order to be manipulated, expanded, revised via operations of combination, deduction, etc. And there's a tradition of *active logic* [21, see e.g.] which models the dynamic process of drawing inferences within the limitations of working memory. Active logic is a wide-ranging framework that can model the dynamics of commonsensical and episodic reasoning across time, whereas our approach to working memory in this paper has been rather static. However, one possible direction of expansion would then feature the addition to our language of dynamic operators in the style of Dynamic Epistemic Logic [5, 17], following a similar pattern as in [4]. This will allow to properly model, e.g., how agents operate on their active beliefs in the light of new incoming information, before storing the results in LTM.

Third, our subset space-style semantics suggests another natural dynamic extension: one could add the so-called *effort modality* of subset space logics as, e.g., a *working memory improvement* operator. The original effort modality, here denoted by $\circledast\varphi$, is intended to capture a notion of epistemic effort that leads to acquiring more evidence. In our topic-sensitive, hyperintensional logic for reasoning about framed believers, we can read $\circledast\varphi$ as 'φ is true in a stronger memory cell (with respect to both information and topic)' and interpret it as

$(w, O_a) \Vdash \circledast\varphi$ iff there is $U \in \mathcal{O}$ and $c \in T$ s.t. $U \subseteq O, a \sqsubseteq c \sqsubseteq \mathfrak{b}$ and $(w, U_c) \Vdash \varphi$.

So, $\circledast\varphi$ is modelled as a working memory transformation operator that takes a memory cell and gives us another that approximates better to the LTM.

Appendix A: Proofs

A.1 Proof of Lemma 1

The proof follows by induction on the structure of φ, where cases for the propositional variables, Boolean connectives, and $\varphi := \Box\psi$ are trivial. So assume inductively that the result holds for ψ and show that it holds also for $\varphi := B_P\psi$. Observe that the inductive hypothesis says that $[\![\psi]\!]^{O_a}_{\mathcal{M}} = [\![\psi]\!]^{U_c}_{\mathcal{M}}$. We then have

$$\begin{aligned}
\mathcal{M}, (w, O_a) \Vdash B_P\psi \quad & \text{iff } O^{\cap} \subseteq [\![\psi]\!]^{O_a}_{\mathcal{M}} \text{ and } t(\varphi) \sqsubseteq \mathfrak{b} && \text{(by the semantics of } B_P\psi) \\
& \text{iff } O^{\cap} \subseteq [\![\psi]\!]^{U_c}_{\mathcal{M}} \text{ and } t(\varphi) \sqsubseteq \mathfrak{b} && \text{(by the induction hypothesis)} \\
& \text{iff } \mathcal{M}, (w, U_c) \Vdash B_P\psi && \text{(by the semantics of } B_P\psi)
\end{aligned}$$

A.2 Proof of Theorem 2

A.2.1 Soundness of L

Soundness is a matter of routine validity check, so we spell out only the relatively tricky cases.

Proof Let $\mathcal{M} = \langle W, \mathcal{O}, T, \oplus, t, \nu \rangle$ be a model and $(w, O_a) \in P(\mathcal{M})$. Checking the soundness of the system S5 for $\Box$ is standard: recall that $\Box$ is interpreted as the global modality. Validity of D_{B_A} is guaranteed since $O \neq \emptyset$ by the definition of memory cells. Validity of $\mathsf{Ax1}_{B_\star}$ is an immediate consequence of the semantic clauses for $B_\star$ and the definition of $\overline{\varphi}$. $\mathsf{Ax3}_{B_\star}$ is valid since truth of a belief sentence $B_\star\varphi$ is state independent: it is easy to see that either $[\![B_\star\varphi]\!]^{O_a}_{\mathcal{M}} = W$ or $[\![B_\star\varphi]\!]^{O_a}_{\mathcal{M}} = \emptyset$, for any $\varphi \in \mathcal{L}$. Here we spell out the details only for C_{B_A}, $\mathsf{Ax2}_{B_P}$, and Inc.

C_{B_A}:

$$
\begin{aligned}
&\mathcal{M}, (w, O_a) \Vdash B_A(\varphi \wedge \psi)\\
&\text{iff } O \subseteq [\![\varphi \wedge \psi]\!]^{O_a} \text{ and } t(\varphi \wedge \psi) \sqsubseteq a\\
&\text{iff } O \subseteq [\![\varphi]\!]^{O_a} \cap [\![\psi]\!]^{O_a} \text{ and } t(\varphi) \oplus t(\psi) \sqsubseteq a\\
&\text{iff } (O \subseteq [\![\varphi]\!]^{O_a} \text{ and } O \subseteq [\![\psi]\!]^{O_a}) \text{ and } (t(\varphi) \sqsubseteq a \text{ and } t(\psi) \sqsubseteq a)\\
&\text{iff } (O \subseteq [\![\varphi]\!]^{O_a} \text{ and } t(\varphi) \sqsubseteq a) \text{ and } (O \subseteq [\![\psi]\!]^{O_a} \text{ and } t(\psi) \sqsubseteq a)\\
&\text{iff } \mathcal{M}, (w, O_a) \Vdash B_A\varphi \wedge B_A\psi
\end{aligned}
$$

Validity proof for C_{B_P} follows similarly: replace O by $O^\cap$ and a by $\mathfrak{b}$.

$\mathsf{Ax2}_{B_P}$:
Suppose that $\mathcal{M}, (w, O_a) \Vdash \Box(\varphi \to \psi) \wedge B_P\varphi \wedge B_P\overline{\psi}$, i.e., (1) $\mathcal{M}, (w, O_a) \Vdash \Box(\varphi \to \psi)$, (2) $\mathcal{M}, (w, O_a) \Vdash B_P\varphi$, and (3) $\mathcal{M}, (w, O_a) \Vdash B_P\overline{\psi}$. (1) means that $[\![\varphi]\!]^{O_a} \subseteq [\![\psi]\!]^{O_a}$, (2) implies that $O^\cap \subseteq [\![\varphi]\!]^O_a$. Therefore, (1) and (2) together implies that $O^\cap \subseteq [\![\psi]\!]^O_a$. Moreover, (3) is the case if and only if $t(\psi) \sqsubseteq \mathfrak{b}$. We therefore conclude that $\mathcal{M}, (w, O_a) \models B_P\psi$. Validity proof for $\mathsf{Ax2}_{B_A}$ follows similarly: replace $O^\cap$ by O and $\mathfrak{b}$ by a.

Inc:
Suppose that $\mathcal{M}, (w, O_a) \Vdash B_A\varphi$, i.e, that $O \subseteq [\![\varphi]\!]^O_a$ and $t(\varphi) \sqsubseteq a$. By the definitions of $O^\cap$ and $\mathfrak{b}$, we have that $O^\cap \subseteq O$ and $a \sqsubseteq \mathfrak{b}$. Therefore, $O^\cap \subseteq [\![\varphi]\!]^{O_a}$ and $t(\varphi) \sqsubseteq \mathfrak{b}$, i.e., $\mathcal{M}, (w, O_a) \Vdash B_P\varphi$. □

A.2.2 Completeness of L

We establish the completeness result via a canonical model construction. While the construction of memory cells uses methods presented by [26], the construction of canonical topics is inspired by the canonical model construction of awareness models (see, e.g., [27]).

Auxiliary Definitions and Lemmas:

The notion of derivation, denoted by $\vdash$, in L is defined as usual. Thus, $\vdash \varphi$ means φ is a theorem of L.

Lemma 3 *The following are derivable in* L*:*

1. $(\Box\varphi \wedge B_\star\overline{\varphi}) \rightarrow B_\star\varphi$ *2.* $B_\star\overline{\varphi} \rightarrow B_\star\overline{\psi}$, *if* $Var(\psi) \subseteq Var(\varphi)$

Proof

1. $\vdash (\Box\varphi \wedge B_\star\overline{\varphi}) \rightarrow B_\star\varphi$:

1. $\vdash (\Box(\overline{\varphi} \rightarrow \varphi) \wedge B_\star\overline{\varphi}) \rightarrow B_\star\varphi$	$\mathsf{Ax2}_{B_\star}$
2. $\vdash (\Box(\overline{\varphi} \rightarrow \varphi) \wedge B_\star\overline{\varphi}) \leftrightarrow (\Box\varphi \wedge B_\star\overline{\varphi})$	CPL, $\mathsf{S5}_\Box$
3. $\vdash (\Box\varphi \wedge B_\star\overline{\varphi}) \rightarrow B_\star\varphi$	1, 2, CPL

2. $\vdash B_\star\overline{\varphi} \rightarrow B_\star\overline{\psi}$, if $Var(\psi) \subseteq Var(\varphi)$
   Follows directly from $\mathsf{C}_{B_\star}$ and CPL.

□

For any set of formulas $\Gamma \subseteq \mathcal{L}$ and any $\varphi \in \mathcal{L}$, we write $\Gamma \vdash \varphi$ if there exists finitely many formulas $\varphi_1, \ldots, \varphi_n \in \Gamma$ such that $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \varphi$. We say that Γ is L-*consistent* if $\Gamma \nvdash \bot$, and L-inconsistent otherwise. A sentence φ is L-consistent with Γ if $\Gamma \cup \{\varphi\}$ is L-consistent (or, equivalently, if $\Gamma \nvdash \neg\varphi$). Finally, a set of formulas Γ is a *maximally* L*-consistent set* (or, in short, *mcs*) if it is L-consistent and any set of formulas properly containing Γ is L-inconsistent [13]. We drop mention of the logic L when it is clear from the context.

Lemma 4 *For every mcs* Γ *of* L *and* $\varphi, \psi \in \mathcal{L}$*, the following hold:*

1. $\Gamma \vdash \varphi$ *iff* $\varphi \in \Gamma$,
2. if $\varphi \in \Gamma$ *and* $\varphi \rightarrow \psi \in \Gamma$ *then* $\psi \in \Gamma$,
3. if $\vdash \varphi$ *then* $\varphi \in \Gamma$,
4. $\varphi \in \Gamma$ *and* $\psi \in \Gamma$ *iff* $\varphi \wedge \psi \in \Gamma$,
5. $\varphi \in \Gamma$ *iff* $\neg\varphi \notin \Gamma$.

Proof Standard. □

In the following proofs, we make repeated use of Lemma 4 in a standard way and often omit mention of it.

Lemma 5 (Lindenbaum's Lemma) *Every* L*-consistent set can be extended to a maximally* L*-consistent one.*

Proof Standard. □

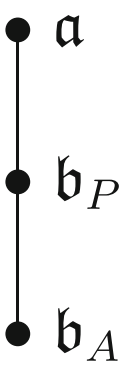

Fig. 2 $(T^c, \oplus^c)$ for Γ_0, where $\mathfrak{b}_A \sqsubset^c \mathfrak{b}_P \sqsubset^c \mathfrak{a}$

Canonical Model of $\mathcal{L}$ for a mcs Γ_0:

Let $\mathcal{X}^c$ be the set of all maximally L-consistent sets. For each $\Gamma \in \mathcal{X}^c$, define

$$\begin{aligned}
\Gamma[\Box] &:= \{\varphi \in \mathcal{L} : \Box\varphi \in \Gamma\},\\
\Gamma[B_\star] &:= \{\varphi \in \mathcal{L} : B_\star\psi \wedge \Box(\psi \to \varphi) \in \Gamma \text{ for some } \psi \in \mathcal{L}\}, \text{ and}\\
\Gamma[B_\star, \Box] &\coloneqq \{\varphi \in \mathcal{L} : B_\star\psi \wedge \Box(\psi \to \varphi) \in \Gamma \text{ for some } \psi \in \mathcal{L}\} \cup\\
&\qquad \{\varphi \in \mathcal{L} : \Box\varphi \in \Gamma\}.
\end{aligned}$$

In short, $\Gamma[B_\star, \Box] = \Gamma[B_\star] \cup \Gamma[\Box]$. By axiom Inc, we have $\Gamma[B_A] \subseteq \Gamma[B_P]$, therefore, also have $\Gamma[B_A, \Box] \subseteq \Gamma[B_P, \Box]$. Moreover, we define $\sim_\Box$ on $\mathcal{X}^c$ as

$$\Gamma \sim_\Box \Delta \text{ iff } \Gamma[\Box] \subseteq \Delta.$$

Since $\Box$ is an S5 modality, it is easy to see that $\sim_\Box$ is an equivalence relation. For any maximally L-consistent set Γ, we denote by $[\Gamma]_\Box$ the equivalence class of Γ induced by $\sim_\Box$, i.e., $[\Gamma]_\Box = \{\Delta \in \mathcal{X}^c : \Gamma \sim_\Box \Delta\}$. It is easy to see that if $\Gamma[B_\star, \Box] \subseteq \Delta$, then $\Delta \in [\Gamma]_\Box$.

Lemma 6 *For any two maximally consistent sets Γ and Δ such that $\Gamma \sim_\Box \Delta$, $\Gamma[\Box] = \Delta[\Box]$ and $\Gamma[B_\star] = \Delta[B_\star]$. Therefore, also $\Gamma[B_\star, \Box] = \Delta[B_\star, \Box]$.*

Proof $\Gamma[\Box] = \Delta[\Box]$ follows from the axioms and rules of $\mathsf{S5}_\Box$. For $\Gamma[B_\star] \subseteq \Delta[B_\star]$, let $\varphi \in \Gamma[B_\star]$. This means that there is $\psi \in \mathcal{L}$ such that $B_\star\psi \in \Box(\psi \to \varphi) \in \Gamma$. Then, by $\mathsf{Ax3}_{B_\star}$ and $\mathsf{S5}_\Box$, we have $\Box B_\star\psi \wedge \Box\Box(\psi \to \varphi) \in \Gamma$. As $\Gamma \sim_\Box \Delta$, we obtain that $B_\star\psi \wedge \Box(\psi \to \varphi) \in \Delta$, thus, $\varphi \in \Delta[B_\star]$. The other direction follows similarly since $\sim_\Box$ is symmetric. We then also have $\Gamma[B_\star, \Box] = \Gamma[\Box] \cup \Gamma[B_\star] = \Delta[\Box] \cup \Delta[B_\star] = \Delta[B_\star, \Box]$. □

Given a mcs Γ_0 of L, the canonical model for Γ_0 is a tuple $\mathcal{M}^c = \langle [\Gamma_0]_\Box, \mathcal{O}^c, T^c, \oplus^c, t^c, \nu^c\rangle$ where

- $\mathcal{O}^c = \{\{\Delta \in \mathcal{X}^c : \Gamma_0[B_A, \Box] \subseteq \Delta\}, \{\Delta \in \mathcal{X}^c : \Gamma_0[B_P, \Box] \subseteq \Delta\}\}$. To simplify the notation, we denote

$$\begin{aligned}
O &:= \{\Delta \in \mathcal{X}^c : \Gamma_0[B_A, \Box] \subseteq \Delta\}\\
O^\cap &:= \{\Delta \in \mathcal{X}^c : \Gamma_0[B_P, \Box] \subseteq \Delta\}.
\end{aligned}$$

Since $\Gamma_0[\Box] \subseteq \Gamma_0[B_\star, \Box]$, we guarantee that $O, O^\cap \subseteq [\Gamma_0]_\Box$. Moreover, since $\Gamma[B_A, \Box] \subseteq \Gamma[B_P, \Box]$, we have $O^\cap \subseteq O$. Therefore, $\bigcap \mathcal{O}^c = O^\cap$.

- $T^c = \{\mathfrak{a}, \mathfrak{b}_P, \mathfrak{b}_A\}$, where $\mathfrak{a} = \{q \in \mathsf{Prop} : \neg B_P \overline{q} \in \Gamma_0\}$ and $\mathfrak{b}_P = \{q \in \mathsf{Prop} : B_P \overline{q} \wedge \neg B_A \overline{q} \in \Gamma_0\}$, and $\mathfrak{b}_A = \{q \in \mathsf{Prop} : B_A \overline{q} \in \Gamma_0\}$.
- $\oplus^c : T^c \times T^c \to T^c$ such that the corresponding strict partial order $\sqsubset^c$ is $\mathfrak{b}_A \sqsubset^c \mathfrak{b}_P \sqsubset^c \mathfrak{a}$ (see Fig. 2).
- $t^c : \mathcal{L} \cup \mathcal{O}^c \to T^c \cup \mathcal{P}(T^c)$ such that, for every $a \in T^c$ and $q \in \mathsf{Prop}$, $t^c(q) = a$ iff $q \in a$, and t^c extends to $\mathcal{L}$ by $t^c(\varphi) = \oplus^c Var(\varphi)$. Moreover, $t^c(O^\cap) = \{\mathfrak{b}_P\}$ and $t^c(O) = \{\mathfrak{b}_A\}$.
- $\nu^c : \mathsf{Prop} \to \mathcal{P}([\Gamma_0]_\Box)$ such that $\nu^c(p) = \{\Gamma \in [\Gamma_0]_\Box : p \in \Gamma\}$.

In order to show that the canonical model $\mathcal{M}^c$ for Γ_0 is a topic-sensitive subset space model, we need the following auxiliary lemmas.

Lemma 7 *Given a mcs Γ, for all finite $\Phi \subseteq \Gamma[B_\star]$ and $\Phi' \subseteq \Gamma[\Box]$, we have $\bigwedge \Phi \in \Gamma[B_\star]$ and $\bigwedge \Phi' \in \Gamma[\Box]$.*

Proof Given a finite $\Phi' \subseteq \Gamma[\Box]$, $\bigwedge \Phi' \in \Gamma[\Box]$ follows via a standard argument since $\Box$ is a normal modal operator. We only show the case for $\Gamma[B_\star]$. Let $\Phi = \{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma[B_\star]$. This means that, for each φ_j with $1 \leq j \leq n$, there is a $\psi_j \in \mathcal{L}$ such that $B_\star \psi_j \wedge \Box(\psi_j \to \varphi_j) \in \Gamma$. Thus, $\bigwedge_{1 \leq j \leq n} B_\star \psi_j \wedge \bigwedge_{1 \leq j \leq n} \Box(\psi_j \to \varphi_j) \in \Gamma$. Then, by $\mathsf{C}_{B_\star}$, we obtain that $B_\star(\bigwedge_{j \leq n} \psi_j) \in \Gamma$. By $\mathsf{S5}_\Box$, we also have $\Box(\bigwedge_{j \leq n} \psi_j \to \bigwedge_{j \leq n} \varphi_j)$. Therefore, $\bigwedge \Phi \in \Gamma[B_\star]$. □

Lemma 8 *Given a mcs Γ, both $\Gamma[\Box]$ and $\Gamma[B_A]$ are consistent. Moreover, $\Gamma[B_A, \Box]$ is consistent.*

Proof Consistency of $\Gamma[\Box]$ follows via a standard argument since $\Box$ is an $\mathsf{S5}$ operator, in particular, since $\neg\Box\bot$ is a theorem of L.

To show that $\Gamma[B_A]$ is consistent, assume, toward contradiction, that $\Gamma[B_A]$ is not consistent, i.e., $\Gamma[B_A] \vdash \bot$. This means that there is a finite subset $\Phi = \{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma[B_A]$ such that $\vdash \bigwedge \Phi \to \neg\varphi_j$ for some $j \leq n$. By Lemma 7, we have that $\bigwedge \Phi \in \Gamma[B_A]$, thus, there is a $\psi \in \mathcal{L}$ such that $B_A\psi \in \Gamma$ and $\Box(\psi \to \bigwedge \Phi) \in \Gamma$. Since $\vdash \bigwedge \Phi \to \neg\varphi_j$, by $\mathsf{S5}_\Box$, we also have $\Box(\psi \to \neg\varphi_j) \in \Gamma$. Hence, $\neg\varphi_j \in \Gamma[B_A]$ too. As $\varphi_j \in \Gamma[B_A]$, we also have a $\psi' \in \mathcal{L}$ with $B_A\psi' \in \Gamma$ and $\Box(\psi' \to \varphi_j) \in \Gamma$. From $\Box(\psi \to \neg\varphi_j) \in \Gamma$ and $\Box(\psi' \to \varphi_j) \in \Gamma$, by $\mathsf{S5}_\Box$, we obtain that $\Box(\psi \to \neg\psi') \in \Gamma$. As $B_A\psi' \in \Gamma$, by $\mathsf{Ax1}_{B_A}$ and Lemma 3.2, $B_A\overline{\neg\psi'} \in \Gamma$. Therefore, $B_A\overline{\neg\psi'} \in \Gamma$, $\Box(\psi \to \neg\psi') \in \Gamma$, $B_A\psi \in \Gamma$, by $\mathsf{Ax2}_{B_A}$, imply that $B_A\neg\psi' \in \Gamma$, contradicting the consistency of Γ: $B_A\psi' \in \Gamma$ implies $\neg B_A\neg\psi' \in \Gamma$, by D_B. Therefore, $\Gamma[B_A]$ is consistent. Suppose now, toward contradiction, that $\Gamma[B_A, \Box]$ is inconsistent. Recall that $\Gamma[B_A, \Box] = \Gamma[B_A] \cup \Gamma[\Box]$ and both $\Gamma[\Box]$ and $\Gamma[B_A]$ are consistent. Therefore, $\Gamma[B_A, \Box]$ being inconsistent implies (by Lemma 7) that there is $\psi \in \Gamma[B_A]$ and $\varphi \in \Gamma[\Box]$ such that $\vdash (\varphi \wedge \psi) \to \bot$, i.e., $\vdash \varphi \to \neg\psi$, while both φ and ψ are consistent. Then, by $\mathsf{S5}_\Box$ and since $\Box\varphi \in \Gamma$, we obtain that $\Box\neg\psi \in \Gamma$. Moreover, $\psi \in \Gamma[B_A]$ implies that there is a $\chi \in \mathcal{L}$ such that $B_A\chi \wedge \Box(\chi \to \psi) \in \Gamma$. This implies that $\Box(\neg\psi \to \neg\chi) \in \Gamma$. Then, by $\mathsf{K}_\Box$, we have that $\Box\neg\psi \to \Box\neg\chi \in \Gamma$. Thus, as $\Box\neg\psi \in \Gamma$, we obtain that $\Box\neg\chi \in \Gamma$.

Moreover, by $B_A\chi \in \Gamma$, $\mathsf{Ax1}_{B_A}$, and Lemma 3.2, we have $B_A\overline{\neg\chi} \in \Gamma$. Then, as $\Box\neg\chi \wedge B_A\overline{\neg\chi} \in \Gamma$, by Lemma 3.1, we have $B_A\neg\chi \in \Gamma$, contradicting the consistency of Γ: $B_A\chi \in \Gamma$ implies $\neg B_A\neg\chi \in \Gamma$, by D_{B_A}. Therefore, $\Gamma[B_A, \Box]$ is consistent. □

Lemma 9 *Given a mcs Γ_0, the canonical model $\mathcal{M}^c = \langle [\Gamma_0]_\Box, \mathcal{O}^c, T^c, \oplus^c, t^c, \nu^c\rangle$ for Γ_0 is a topic-sensitive subset space model.*

Proof Observe that, since Γ_0 is consistent, by axiom Inc, topics $\mathfrak{a}$, $\mathfrak{b}_P$, and $\mathfrak{b}_A$ are mutually disjoint. This implies that the canonical topic assignment function t^c is well-defined. Moreover, both $\mathcal{O}^c$ and, for each $O \in \mathcal{O}^c$, $t^c(O)$ are finite. Finally, we show that $\mathcal{O}^c \neq \{\emptyset\}$: by Lemma 8, we know that $\Gamma_0[B_A, \Box]$ is consistent. Therefore, by Lemma 5, there is a mcs Δ such that $\Gamma_0[B_A, \Box] \subseteq \Delta$. Therefore, $\Delta \in O \neq \emptyset$. □

Lemma 10 *For any mcs Γ and $\varphi \in \mathcal{L}$, $B_\star\overline{\varphi} \in \Gamma$ iff $B_\star\overline{p} \in \Gamma$ for all $p \in Var(\varphi)$.*

Proof The direction from left-to-right follows from Lemma 3.2. For the opposite direction, let $Var(\varphi) = \{p_1, \ldots, p_n\}$ and observe that $\overline{\varphi} := \overline{p_1} \wedge \cdots \wedge \overline{p_n}$. If $B_\star\overline{p_i} \in \Gamma$ for all $p_i \in \{p_1, \ldots, p_n\}$, then $\bigwedge_{i\leq n} B_\star\overline{p_i} \in \Gamma$ (by Lemma 4.4). Then, by $\mathsf{C}_{B_\star}$, we obtain that $B_\star(\bigwedge_{i\leq n}\overline{p_i}) \in \Gamma$, i.e., $B_\star\overline{\varphi} \in \Gamma$. □

Corollary 11 *Given the canonical model $\mathcal{M}^c = \langle [\Gamma_0]_\Box, \mathcal{O}^c, T^c, \oplus^c, t^c, \nu^c\rangle$ for Γ_0, for any mcs $\Gamma \in [\Gamma_0]_\Box$ and $\varphi \in \mathcal{L}$,*

1. *$B_A\overline{\varphi} \in \Gamma$ iff $t^c(\varphi) \sqsubseteq \mathfrak{b}_A$, and*
2. *$B_P\overline{\varphi} \in \Gamma$ iff $t^c(\varphi) \sqsubseteq \mathfrak{b}_P$.*

Proof We prove item (2) only, item (1) follows similarly.

$B_P\overline{\varphi} \in \Gamma$ iff $B_P\overline{q} \in \Gamma$ for all $q \in Var(\varphi)$ (Lemma 10)
iff $B_P\overline{q} \in \Gamma_0$ for all $q \in Var(\varphi)$ ($\mathsf{Ax3}_{B_P}$ and $\Gamma \in [\Gamma_0]_\Box$)
iff $t^c(q) = \mathfrak{b}_A$ or $t^c(q) = \mathfrak{b}_P$ for all $q \in Var(\varphi)$
(by the definitions of $\mathfrak{b}_P$, $\mathfrak{b}_A$, and t^c)
iff $t^c(\varphi) \sqsubseteq \mathfrak{b}_P$ (by the definition of $(T^c, \oplus^c)$ and $t^c(\varphi)$, and $\mathfrak{b}_A \sqsubseteq^c \mathfrak{b}_P$)
□

Lemma 12 *For every mcs Γ and $\varphi \in \mathcal{L}$, if $\Gamma[B_\star, \Box] \vdash \varphi$ and $B_\star\overline{\varphi} \in \Gamma$, then $B_\star\varphi \in \Gamma$.*

Proof Suppose that $\Gamma[B_\star, \Box] \vdash \varphi$ and $B_\star\overline{\varphi} \in \Gamma$. Recall that $\Gamma[B_\star, \Box] = \Gamma[B_\star] \cup \Gamma[\Box]$. Then, the first assumption means that there are finite sets $\Phi \subseteq \Gamma[\Box]$ and $\Phi' \subseteq \Gamma[B_\star]$ such that $\vdash (\bigwedge\Phi \wedge \bigwedge\Phi') \rightarrow \varphi$. By Lemma 7, we know that $\bigwedge\Phi' \in \Gamma[B_\star]$. This means that there is a ψ such that $B_\star\psi \wedge \Box(\psi \rightarrow \bigwedge\Phi') \in \Gamma$. Again by Lemma 7, we have $\bigwedge\Phi \in \Gamma[\Box]$, i.e., $\Box(\bigwedge\Phi) \in \Gamma$. Then, by $\mathsf{S5}_\Box$, we obtain that $\Box(\psi \rightarrow \bigwedge\Phi) \in \Gamma$. Therefore, as $\Box(\psi \rightarrow \bigwedge\Phi') \in \Gamma$ as well, we have $\Box(\psi \rightarrow (\bigwedge\Phi \wedge$

$\bigwedge \Phi')) \in \Gamma$. Then, since $\vdash (\bigwedge \Phi \wedge \bigwedge \Phi') \rightarrow \varphi$, we have $\Box(\psi \rightarrow \varphi) \in \Gamma$. Hence, by $\mathsf{Ax2}_{B_\star}$ together with $B_\star \psi \in \Gamma$ and $B_\star \overline{\varphi} \in \Gamma$, we obtain that $B_\star \varphi \in \Gamma$. □

Lemma 13 (Truth Lemma) *Let* Γ_0 *be a mcs of* L *and* $\mathcal{M}^c = \langle [\Gamma_0]_\Box, \mathcal{O}^c, T^c, \oplus^c, t^c, v^c \rangle$ *be the canonical model for* Γ_0. *Then, for all* $\varphi \in \mathcal{L}$ *and* $\Gamma \in [\Gamma_0]_\Box$, *we have* $\mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash \varphi$ *iff* $\varphi \in \Gamma$.

Proof First observer that $(\Gamma, O_{\mathfrak{b}_A})$ is a well-defined world-memory pair in $\mathcal{M}^c$: $\Gamma \in [\Gamma_0]_\Box$, $O \neq \emptyset$ (see the proof of Lemma 9), and $\mathfrak{b}_A \in t^c(O) = \{\mathfrak{b}_A\}$ (by the definition of t^c). Due to latter two, we in fact have that $(\Delta, O_{\mathfrak{b}_A})$ is a well-defined world-memory pair of $\mathcal{M}^c$, for all $\Delta \in [\Gamma_0]_\Box$.

The proof follows by induction on the structure of φ.

Base case: $\varphi := p$

$$\begin{aligned} \mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash p \quad & \text{iff } \Gamma \in v^c(p) \qquad \text{(by the semantics)} \\ & \text{iff } p \in \Gamma \qquad \text{(by the definition of } v^c\text{)} \end{aligned}$$

The cases for the Booleans are standard. We here prove the cases $\varphi := \Box\psi$ and $\varphi := B_\star \psi$.

Case $\varphi := \Box\psi$

($\Leftarrow$) Suppose that $\Box\psi \in \Gamma$ and let $\Delta \in [\Gamma_0]_\Box$. The latter implies that $\Delta \sim_\Box \Gamma$. Therefore, $\psi \in \Delta$. Then, by the induction hypothesis (IH), we have $\mathcal{M}^c, (\Delta, O_{\mathfrak{b}_A}) \Vdash \psi$. As Δ has been chosen arbitrarily from $[\Gamma_0]_\Box$, we conclude that $\mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash \Box\psi$.

($\Rightarrow$) Suppose that $\mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash \Box\psi$ and, toward contradiction, that $\Box\psi \notin \Gamma$. The latter implies that $\Gamma[\Box] \cup \{\neg\psi\}$ is consistent. Therefore, by Lemma 5, there is a mcs Δ such that $\Gamma[\Box] \cup \{\neg\psi\} \subseteq \Delta$. Observe, by Lemma 6, that $\Delta \in [\Gamma_0]_\Box$ (since $\Gamma_0[\Box] = \Gamma[\Box]$). And, as $\neg\psi \in \Delta$, we also obtain that $\psi \notin \Delta$. Then, by IH, we have that $\mathcal{M}^c, (\Delta, O_{\mathfrak{b}_A}) \nVdash \psi$, contradicting the initial assumption $\mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash \Box\psi$. Therefore, $\Box\psi \in \Gamma$.

Case $\varphi := B_P \psi$

($\Leftarrow$) Suppose that $B_P \psi \in \Gamma$. Then, by $\mathsf{Ax1}_{B_P}$, we also have that $B_P \overline{\psi} \in \Gamma$. This implies, by Corollary 11.2, that $t^c(\psi) \sqsubseteq^c \mathfrak{b}_P$. Observe that $\oplus^c(\bigcup_{O \in \mathcal{O}^c} t^c(O)) = \oplus^c\{\mathfrak{b}_P, \mathfrak{b}_A\} = \mathfrak{b}_P \oplus^c \mathfrak{b}_A = \mathfrak{b}_P$, so we satisfy the topicality component of the semantic clause for $B_P\psi$. In order to complete the proof, we need to show that $\bigcap \mathcal{O}^c \subseteq [\![\psi]\!]^O_{\mathfrak{b}_A}$. As stated before, $\bigcap \mathcal{O}^c = O^\cap$. So, we need to show that $\{\Delta \in \mathcal{X}^c : \Gamma_0[B_P, \Box] \subseteq \Delta\} \subseteq [\![\psi]\!]^{O_{\mathfrak{b}_A}}$. Let $\Sigma \in O^\cap$. This means that $\Gamma_0[B_P, \Box] \subseteq \Sigma$. As $B_P\psi \in \Gamma$, by $\mathsf{Ax3}_{B_P}$ and the fact that $\Gamma \sim_\Box \Gamma_0$, we also have $B_P\psi \in \Gamma_0$. Moreover, $\Box(\psi \rightarrow \psi) \in \Gamma_0$ (by $\mathsf{S5}_\Box$). We then have that $\psi \in \Gamma_0[B_P]$, thus, $\psi \in \Gamma_0[B_P, \Box]$. Hence, $\psi \in \Sigma$. Then, by IH, we have $\mathcal{M}^c, (\Sigma, O_{\mathfrak{b}_A}) \Vdash \psi$, i.e., that $\Sigma \in [\![\psi]\!]^{O_{\mathfrak{b}_A}}$. As Σ has been chosen arbitrarily from $O^\cap$, we conclude that $O^\cap \subseteq [\![\psi]\!]^{O_{\mathfrak{b}_A}}$. Since $t^c(\psi) \sqsubseteq^c \mathfrak{b}_P$ as well, we obtain that $\mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash B_P\psi$.

($\Rightarrow$) Suppose that $\mathcal{M}^c, (\Gamma, O_{\mathfrak{b}_A}) \Vdash B_P\psi$, i.e., that $\bigcap \mathcal{O}^c = O^\cap = \{\Delta \in \mathcal{X}^c : \Gamma_0[B_P, \Box] \subseteq \Delta\} \subseteq [\![\psi]\!]^{O_{\mathfrak{b}_A}}$ and $t^c(\psi) \sqsubseteq^c \oplus^c(\bigcup_{O \in \mathcal{O}^c} t^c(O)) = \mathfrak{b}_P$. By Corollary 11.2, the latter means that $B_P\overline{\psi} \in \Gamma$. Then, by the former and the IH, we have that whenever $\Gamma_0[B_P, \Box] \subseteq \Delta$, then $\psi \in \Delta$. This implies that $\Gamma_0[B_P, \Box] \vdash \psi$.

Otherwise, $\Gamma_0[B_P, \Box] \cup \{\neg\psi\}$ would be consistent, thus, by Lemma 5, there would exist a mcs Δ' such that $\Gamma_0[B_P, \Box] \cup \{\neg\psi\} \subseteq \Delta'$, contradicting the fact that if $\Gamma_0[B_P, \Box] \subseteq \Delta$ then $\psi \in \Delta$. By Lemma 6, $\Gamma_0[B_P, \Box] \vdash \psi$ means $\Gamma[B_P, \Box] \vdash \psi$. Since $B_P\overline{\psi} \in \Gamma$, by Lemma 12, we obtain that $B_P\psi \in \Gamma$.

Case $\varphi := B_A\psi$: Follows similarly to the proof of case $\varphi := B_P\psi$, using Corollary 11.1. □

Corollary 14 ⊥ *is complete with respect to the class of topic-sensitive subset space models.*

Proof Let $\varphi \in \mathcal{L}$ such that $\nvdash \varphi$. This mean that $\{\neg\varphi\}$ is consistent. Then, by Lindenbaum's Lemma (Lemma 5), there exists a mcs Γ_0 such that $\varphi \notin \Gamma_0$. Therefore, by Lemma 13, we conclude that $\mathcal{M}^c, (\Gamma_0, O_{\mathfrak{b}_A}) \nVdash \varphi$, where $\mathcal{M}^c$ is the canonical model for Γ_0. □

Acknowledgements This research is published within the project 'The Logic of Conceivability', funded by the European Research Council (ERC CoG), grant number 681404. We thank the anonymous reviewers of the Journal of Philosophical Logic for their valuable comments. Versions of this paper were presented at the Cognitive Lunch, Indiana University, April 7, 2021; the workshop 'Beyond the Impossible', University of Padua, October 22, 2021; the London Group for Formal Philosophy (UCL/KCL/BBK), November 18, 2021; the Department of Philosophy, University of Frankfurt, November 30, 2021; the Moral Science Club, University of Cambridge, January 25, 2022; the Department of Philosophy, Kansas State University, March 25, 2022. Thanks to the audiences for helpful discussion.

References

1. Anderson, J. (1983). *Architecture of cognition*. Cambridge: Harvard University Press.
2. Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.
3. Baddeley, A. (2002). Is working memory still working? *European Psychologist*, *7*, 85–97.
4. Balbiani, P., Fernández-Duque, D., & Lorini, E. (2019). The dynamics of epistemic attitudes in resource-bounded agents. *Studia Logica*, *107*, 457–488.
5. Baltag, A., & Renne, B. (2016). Dynamic epistemic logic. In E. N. Zalta (Ed.) *The stanford encyclopedia of philosophy. Metaphysics Research Lab, Stanford University, winter 2016 edn*.
6. Barsalou, L. (1992). *Cognitive psychology*. Hillsdale: Erlbaum.
7. Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Stanford: CSLI Publications.
8. van Benthem, J., & Velázquez-Quesada, F. R. (2010). The dynamics of awareness. *Synthese*, *177*(1), 5–27.
9. Berto, F. (2018). Aboutness in imagination. *Philosophical Studies*, *175*, 1871–1886.
10. Berto, F. (2019). Simple hyperintensional belief revision. *Erkenntnis*, *84*, 559–575.
11. Berto, F., & Hawke, P. (2018). Knowability relative to information. *Mind*, *130*, 1–33.
12. Bjorndahl, A., & Özgün, A. (2020). Logic and topology for knowledge, knowability, and belief. *The Review of Symbolic Logic*, *13*(4), 748–775.

13. Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic volume 53 of cambridge tracts in theoretical computer scie*. Cambridge: Cambridge University Press.
14. Busby, E., Flynn, D., & Druckman, J. (2018). Studying framing effects on political preferences. In P. D' Angelo (Ed.) *Doing news framing analysis, volume II, pp. 67–90. Routledge*.
15. Crowder, R. (1993). Short-term memory: where do we stand? *Memory and Cognition*, *21*, 142–45.
16. Dabrowski, A., Moss, L. S., & Parikh, R. (1996). Topological reasoning and the logic of knowledge. *Annals of Pure and Applied Logic*, *78*(1), 73–110.
17. Van Ditmarsch, H., Van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic*. Dordrecht: Springer.
18. Dretske, F. (1970). Epistemic operators. *The Journal of Philosophy*, *67*, 1007–1023.
19. Druckman, J. (2001a). Evaluating framing effects. *Journal of Economic Psychology*, *22*, 96–101.
20. Druckman, J. (2001b). Using credible advice to overcome framing effects. *Journal of Law Economics, and Organization*, *17*, 62–68.
21. Elgot-Drapkin, J., Miller, M., & Perils, D. (1991). Memory, reason, and time: the step-logic approach. In R. Cummins (Ed.) *Philosophy and AI: essays in the interface, pp. 79–103. MIT Press, Cambridge*.
22. Eysenck, M., & Keane, M. (2015). *Cognitive psychology*. New York: Psychology Press.
23. Fagin, R., & Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence*, *34*(1), 39–76.
24. Fine, K. (2016). Angellic content. *Journal of Philosophical Logic*, *45*(2), 199–226.
25. Gächter, S., Orzen, H., Renner, E., & Stamer, C. (2009). Are experimental economists prone to framing effects? A natural field experiment. *Journal of Economic Behavior and Organization*, *70*, 443–46.
26. Giordani, A. (2019). Axiomatizing the logic of imagination. *Studia Logica*, *107*, 639–657.
27. Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior*, *37*(2), 321–339.
28. Hansson, S. (1999). *A textbook of belief dynamics: theory change and database updating*. Dordrecht: Kluwer.
29. Hawke, P. (2016). Questions, topics and restricted closure. *Philosophical Studies*, *73*(10), 2759–2784.
30. Hawke, P., Özgün, A., & Berto, F. (2020). The fundamental problem of logical omniscience. *Journal of Philosophical Logic*, *49*, 727–766.
31. Hawthorne, J. (2005). The case for closure. In E. Sosa (Ed.) *New directions in semantics, pp. 26–43. Blackwell, Oxford*.
32. Hintikka, J. (1962). *Knowledge and belief. An introduction to the logic of the two notions*. Ithaca: Cornell University Press.
33. Holliday, W. (2015). Fallibilism and multiple paths to knowledge. In T. Gendler, & J. Hawthorne (Eds.) *Oxford studies in epistemology, volume 5, pp. 97–144. Oxford University Press*.
34. Humberstone, L. (2000). Parts and partitions. *Theoria*, *66*, 41–82.
35. Kahneman, D. (2011). *Thinking: fast and slow*. London: Penguin.
36. Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–50.
37. Konolige, K. (1986). What awareness isn't: a sentential view of implicit and explicit belief. In J. Y. Halpern (Ed.) *Theoretical aspects of reasoning about knowledge, pp. 241–250. Morgan Kaufmann*.
38. Kripke, S. (2011). Nozick on knowledge. In *Philosophical troubles: collected papers, volume 1. Oxford University Press*.
39. Lawlor, K. (2005). Living without closure. *Grazer Philosophische Studien*, *697*, 25–49.
40. Levin, I., Gaeth, G., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes*, *88*, 411–429.
41. Lewis, D. (1988). Relevant implication. *Theoria*, *54*(3), 161–174.
42. Lorini, E. (2020). Rethinking epistemic logic with belief bases. *Artificial Intelligence*, *282*, 103233.
43. Lorini, E., & Song, P. (2022). A computationally grounded logic of awareness. Journal of Logic and Computation, Online first:https://doi.org/10.1093/logcom/exac035.
44. Miyake, A., & Shah, P. (1999). *Models of working memory*. Cambridge: Cambridge University Press.
45. Moss, L. S., & Parikh, R. (1992). Topological reasoning and the logic of knowledge. In *Proc. of the 4th TARK, pp. 95–105. Morgan Kaufmann*.
46. Nozick, R. (1981). Philosophical explanations. Harvard University Press.

47. Özgün, A., & Berto, F. (2021). Dynamic hyperintensional belief revision. *The Review of Symbolic Logic*, *14*(3), 766–811.
48. Pacuit, E. (2017). *Neighbourhood semantics for modal logic*. Dordrecht: Springer.
49. Perry, J. (1989). Possible worlds and subject matter. In *The problem of the essential indexical and other essays, pp. 145–160. CSLI Publications*.
50. Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.
51. Rott, H. (1998). "Just because": taking belief bases seriously. *Lecture Nores in Logic*, *13*, 387–408.
52. Schachter, D., & Tulving, E. (1994). *Memory systems*. Cambridge: MITy Press.
53. Schipper, B. (2015). Awareness. In H. Van Ditmarsch, J. Halpern, W. Van der Hoek, & B. Kooi (Eds.) *Handbook of epistemic logic, pp. 79–146. College Publications, London*.
54. Scott, D. (1970). Advice on modal logic. In K. Lambert (Ed.) *Philosophical problems in logic, pp. 143–73. Reidel, Dordrecht*.
55. Sharon, A., & Spectre, L. (2017). Evidence and the openness of knowledge. *Philosophical Studies*, *174*, 1001–1037.
56. Squire, L. (1987). *Memory and brain*. New York: Oxford University Press.
57. Stalnaker, R. (1984). *Inquiry*. Cambridge: MIT Press.
58. Thaler, R. (2008). Nudge: improving decisions about health, wealth, and happiness. Yale University Press.
59. Velázquez-Quesada, F. R. (2014). Dynamic epistemic logic for implicit and explicit beliefs. *Journal of Logic Language and Information*, *23*(2), 107–140.
60. Weiss, M. A., & Parikh, R. (2002). Completeness of certain bimodal logics for subset spaces. *Studia Logica*, *71*(1), 1–30.
61. Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
62. Yablo, S. (2014). *Aboutness*. Princeton: Princeton University Press.
63. Yalcin, S. (2018). Belief as question-sensitive. *Philosophy and Phenomenological Research*, *97*(1), 23–47.