



Realizing impossibilities

Jonathan Mai¹

Received: 6 September 2023 / Accepted: 3 November 2023 / Published online: 21 December 2023
© The Author(s) 2023

Abstract

It is common in epistemic modal logic to model the epistemic states of agents via box operators in the normal logic $S5$. However, this approach treats agents as logically omniscient by requiring their knowledge to be closed under classical logical consequence. A promising way of avoiding logical omniscience consists in extending epistemic models with impossible states, that is states, where complex formulas are not evaluated recursively. However, this approach faces the dual problem of logical ignorance by modeling agents as not even minimally logically competent. In this paper I will outline an epistemic logic that combines impossible states with dynamic realization modalities akin to the dynamic announcement operators from public announcement logic (PAL). I will show that this epistemic logic avoids both the problem of logical omniscience and the problem of logical ignorance. Furthermore, I prove that so-called successful updates in my logic can be characterized in the same way as in PAL and that a similar logic due to Johan van Benthem can be simulated in my logic. Finally, I will compare my epistemic logic with a similar one, which has recently been advanced by JC Bjerring and Mattias Skipper.

Keywords Epistemic logic · Problem of logical omniscience · Impossible worlds · Dynamic epistemic logic · Formal epistemology

1 Logical omniscience and open worlds

It is common in epistemic modal logic to model the epistemic states of agents via box operators in the normal logic $S5$. We quickly go through the syntax and semantics of the single agent version of $S5$ which will also fix notation.

T.C. : Hyperintensional Formal Epistemology.

✉ Jonathan Mai
jonathan.mai@gmx.net

¹ Ruhr University Bochum, Bochum, Germany

For a finite or countably infinite set of propositional variables (or atoms) Φ the language of *basic epistemic logic* for one agent is generated by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi_1 \wedge \varphi_2) \mid K\varphi \quad (p \in \Phi)$$

The intended reading of $K\phi$ is that the agent knows that ϕ . We let an *epistemic model* be a tuple $M = (S, R, V)$, where $S \neq \emptyset$ is the set of epistemic states of the agent, $R \subseteq S \times S$ models the relation of epistemic accessibility and $V : \Phi \rightarrow \wp(S)$. We let an S5 epistemic model be an epistemic model, whose accessibility relation is an equivalence relation. The satisfaction relation between states and formulas in an S5 model is defined in the usual way with the clause for the K -operator being

- $M, s \models K\varphi \iff M, t \models \varphi$, for all $t \in S$ with sRt .

Let φ be *valid* in an epistemic model $M = (S, R, V)$ ($M \models \varphi$), if $M, s \models \varphi$, for all $s \in S$. Finally, we call a formula φ *valid* in a class \mathcal{M} of epistemic models ($\mathcal{M} \models \varphi$), if $M \models \varphi$, for all $M \in \mathcal{M}$; in case \mathcal{M} is the class of all epistemic models we write $\models \varphi$, instead of $\mathcal{M} \models \varphi$.

It is well known that epistemic S5 is highly unsuited for formalizing the epistemic states of agents that are resource bounded, that is, who have limited time and memory for computing what they know.¹ The reason for this representational mismatch is that according to S5 the knowledge of agents is subject to closure conditions which require *inter alia* that agents know everything that logically follows from what they know. Even worse, these closure conditions already hold for the smallest normal epistemic logic K, which is characterized by the class of all epistemic models. The collection of these closure conditions can be addressed as the *problem of logical omniscience*. Here are some of the most salient closure conditions, which we represent as closure conditions on the set of K-validities.

- $\models K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ (Distribution)
- $\models \varphi \Rightarrow \models K\varphi$ (Knowledge of validities)
- $\models \varphi \rightarrow \psi \Rightarrow \models K\varphi \rightarrow K\psi$ (Closure under valid implications)
- $\models \varphi \leftrightarrow \psi \Rightarrow \models K\varphi \leftrightarrow K\psi$ (Closure under valid equivalences)
- $\models (K\varphi \wedge K\psi) \rightarrow K(\varphi \wedge \psi)$ (Closure under adjunction)
- $\models K\varphi \rightarrow K(\varphi \vee \psi)$ (Closure under weakening)

There are many different proposals of how best to solve the problem of logical omniscience (for a survey see Fagin (1995a), Sim (1997)). On the impossible worlds approach epistemic possibilities are not logically transparent to agents: Among the epistemic alternatives there may be some that, unbeknownst to the agent, are logically impossible in the sense of not respecting classical consequence. In particular, agents may consider states epistemically possible that verify contradictions. Technically, this approach is implemented by extending the set of states with non-standard states that do

¹ It is worth pointing out, however, that epistemic S5 is a natural environment for representing an external notion of knowledge, that is, a notion of knowledge according to which agents need not be able to compute formulas they know. Consequently, this logic forms the backbone of the successful interpreted systems approach in computer science which uses epistemic temporal logics to model the knowledge dynamics in distributed systems such as processor networks (Fagin, 1995a).

not represent models from classical propositional logic and by restricting the definition of logical notions to standard states. As non-standard states lie beyond the reach of (classical) logic it is common to call them (*logically*) *impossible worlds* (Hintikka, 1975).²

There are two main versions of the impossible worlds approach. According to the open world approach, impossible worlds are not closed under any non-trivial consequence relation.³ Consequently, impossible worlds can be taken to be arbitrary functions from formulas to (classical) truth-values (Cresswell, 1970, 1972; Rantala, 1982a, ?; Wansing, 1989; Jago, 2014; Priest, 2016). The relevantist approach, on the other hand, holds that impossible worlds are closed under the consequence relation of some relevant logic, FDE being the first choice among proponents of the approach. Thus, after relevantism, impossible worlds can be equated with models from the various model-theories for FDE in existence (Levesque, 1984; Priest, 1992; Fagin, 1995b; Levesque & Lakemeyer, 2000).

Relevantist approaches are troubled by the fact that they treat agents as perfect reasoners in a relevant logic. If one is interested in representing the epistemic states of a knowledge based agent, that is a system whose knowledge is the deductive closure of a knowledge base (a finite set of purely propositional formulas), closure under relevant consequence may be acceptable. However, it is evident that the epistemic states of human agents are not closed under classical consequence nor under relevant consequence. Thus, the relevantist approach simply reproduces the problem of logical omniscience, but now with a logic weaker than classical logic (Fagin & Halpern, 1988).

Open world approaches are more promising, since they allow epistemic states to resist any non-trivial deductive closure. However, a dual problem arises here: Open world approaches seem to deprive agents of even the most basic logical competence. Let's call impossible worlds that are not closed under any non-trivial consequence relation *open worlds*. Any minimally logically competent reasoner a who knows that $\varphi \wedge \psi$, should be in a position to know φ . But not so on the open world approach, or so it appears: Suppose that open world s is an epistemic possibility of a that does not verify φ and suppose further that every epistemic alternative of a makes $\varphi \wedge \psi$ true. On the open world approach this implies that a knows that $\varphi \wedge \psi$ and that a does not know φ . It seems to follow that open world approaches model agents that are logically ignorant, that is agents, which are not even minimally logically competent. We call this apparent consequence the *problem of logical ignorance* for open world approaches (Lewis, 2004), (Jago, 2014). In the next section we will propose a solution to this problem. This proposal augments the open world approach with techniques from dynamic epistemic logic.

2 DEL meets the impossible

Every adequate solution to the problem of logical ignorance must answer the question which properties of the open world approach lead to the problem and how these

² The term 'non-normal states' is also in use.

³ On the open world approach even the most anarchic states are closed under the identity relation on the set of formulas.

properties can be eliminated. My claim is that the problem arises on account of the fact that open world approaches lack the means to represent informational or epistemic actions such as observation, communication or inference. More precisely, resource bounded agents such as humans have the cognitive resources to update their epistemic states in such a way that, after the update, they can realize that worlds they took to be epistemically possible before the update are in fact logically impossible and so do not count as epistemic alternatives any more.

I call the kind of epistemic action that induces the elimination of impossible worlds *realizations*. Intuitively, realizations consist in recognizing that certain formulas should be true in every world, given that certain other formulas are true in all the possible worlds. Realizations may be connected to acts of inference, but I will not discuss this issue here. Realizations of formulas φ can be formalized as dynamic modalities $[\varphi]$ in a logic that takes a basic epistemic logic with open worlds as its static base and adds machinery from dynamic epistemic logic (DEL) on top. In particular, the dynamic component is a slight modification of the simplest dynamic epistemic logic, namely public announcement logic or PAL (Plaza, 2008; Gerbrandy & Groeneveld, 1997; van Ditmarsch, 2008). The resulting logic will deliver validities that tell us that agents are not logically ignorant. In this way we can solve the problem of logical ignorance while at the same time avoiding logical omniscience. Here are the formal details.

We extend the language of basic epistemic logic with a clause for realization modalities. So, for a set of atoms Φ the *realization language* for one agent, denoted by \mathcal{L} , is given by the BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi_1 \wedge \varphi_2) \mid K\varphi \mid [\varphi_1]\varphi_2 \quad (p \in \Phi)$$

The intended interpretation of $[\varphi]\psi$ is that after every realization of φ , ψ holds. Let's turn to the structures on which the realization language will be interpreted.

Definition 1 (i) An *open world frame* is a tuple $F = (W, W^*, R)$ such that W, W^* are sets with $W \neq \emptyset, W \cap W^* = \emptyset$ and $R \subseteq (W \cup W^*)^2$.

(ii) A *valuation* on an open world frame $F = (W, W^*, R)$ is a pair (V, V^*) such that $V : W \rightarrow 2^\Phi$ and $V^* : W^* \rightarrow 2^\Phi$.

(iii) An *open world model* is a tuple $M = (F, (V, V^*))$, where F is an open world frame and (V, V^*) is a valuation on F .

Points from the set $W \cup W^*$ are called *states* or *worlds*; points from W are addressed as *possible* states and the points from W^* are called *impossible* states.

Satisfaction in states from open world models can then be defined as follows.

Definition 2 Let $M = (W, W^*, R, V, V^*)$ be an open world model.

(i) For every $w \in W$ we set:

- $M, w \models p \iff V(w)(p) = 1$;
- $M, w \models \neg\varphi \iff M, w \not\models \varphi$;
- $M, w \models (\varphi \wedge \psi) \iff M, w \models \varphi$ and $M, w \models \psi$;
- $M, w \models K\varphi \iff M, v \models \varphi$, for all $v \in W \cup W^*$ with wRv ;
- $M, w \models [\varphi]\psi \iff M, w \not\models \varphi$ or $M \uparrow_\varphi, w \models \psi$.

(ii) For every $w^* \in W^*$ and every $\varphi \in \mathcal{L}$ we let

- $M, w^* \models \varphi \Leftrightarrow V^*(w^*)(\varphi) = 1.$

In the last clause of (i) $M \upharpoonright_\varphi$ is the submodel of M generated by the set of impossible states where φ is true. More precisely $M \upharpoonright_\varphi = (W, W_\varphi^*, R_\varphi, V, V_\varphi^*)$, where

- $W_\varphi^* = \{w^* \in W^* : M, w^* \models \varphi\};$
- $R_\varphi = R \cap (W \cup W_\varphi^*)^2;$
- $V_\varphi^* = V^* \upharpoonright W_\varphi^*.$

As usual, validity in models is restricted to possible states. We let φ be *valid* in open world model M ($M \models \varphi$), if $M, w \models \varphi$, for every $w \in W$. φ is *valid* in an open world frame F ($F \models \varphi$), if φ is valid in every model M based on F , that is every $M = (F, (V, V^*))$ for some (V, V^*) on F . φ is *valid* in a class F of frames, if $F \models \varphi$, for every $F \in F$. In case F is the class of all open world frames we write $\models \varphi$ instead of $F \models \varphi$. Let's call the resulting logic *realization logic*.

It is obvious that realization logic avoids the problem of logical omniscience by invalidating the epistemic closure conditions mentioned in section 2. But it also avoids the problem of logical ignorance for open world approaches. To show this, we need the following simple lemma. Let a *factual* formula be any formula containnig no occurrences of K or of $[\varphi]$, for any formula φ .

Lemma 3 *Let $M = (W, W^*, R, V, V^*)$ be an open world model, $v \in W \cup W^*$ and ψ be factual. Then it holds that*

$$M, v \models \psi \Leftrightarrow M \upharpoonright_\psi, v \models \psi.$$

Proof If $v \in W$ the result follows by a straightforward induction on the structure of formulas. If $v \in W^*$ the result is immediate by the definition of V^* . □

Now, consider the argument concerning conjunction from section 1. The argument aimed to show that on open world approaches agents may know conjunctions without being in an position to know one of the conjuncts. Realization logic, however, yields validities, which tell us that agents knowing some conjunction with factual conjuncts know either conjunct after realizing that the respective conjunct is true. This, I take it, amounts to an agent being in a position to know the conjunct: After performing a certain epistemic action, the agent knows the conjunct. So agents are not logically ignorant when it comes to conjunction reduction, at least when restricted to factual formulas. The validities showing this are the formulas

- $K(\varphi \wedge \psi) \rightarrow [\varphi]K\varphi$, where φ is factual;
- $K(\varphi \wedge \psi) \rightarrow [\psi]K\psi$, where ψ is factual.

That the first of these formuals is valid in every open world model can be shown as follows. Suppose $M, w \models K(\varphi \wedge \psi)$ and $v \in W \cup W_\varphi^*$ with $wR_\varphi v$. If $v \in W$ we have $M, v \models \varphi$ and so by Lemma 3, $M \upharpoonright_\varphi, v \models \varphi$. The same follows, if $v \in W_\varphi^*$. So, $M, w \models [\varphi]K\varphi$. The validity of the second formula is shown analogously.

More generally, realization logic validates the following relativized versions of the K-validities characterizing logical omniscience (again with partial restrictions to factual formulas).

Proposition 4 For every $\varphi, \psi \in \mathcal{L}$ it holds that:

- (a) $\models K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow [\psi]K\psi)$, if ψ is factual.
- (b) $\models \varphi \Rightarrow \models [\varphi]K\varphi$, if φ is factual.
- (c) $\models \varphi \rightarrow \psi \Rightarrow \models K\varphi \rightarrow [\psi]K\psi$, if ψ is factual.
- (d) $\models \varphi \leftrightarrow \psi \Rightarrow \models [\varphi]K\varphi \leftrightarrow [\psi]K\psi$, if φ and ψ are factual.
- (e) $\models (K\varphi \wedge K\psi) \rightarrow [\varphi \wedge \psi]K(\varphi \wedge \psi)$, if φ and ψ are factual.
- (f) $\models K\varphi \rightarrow [\varphi \vee \psi]K(\varphi \vee \psi)$, if φ and ψ are factual.

Proof The only non-trivial case is (d). We prove its left-to-right direction. Let M be any model based on a frame F and let $M, w \models [\varphi]K\varphi, \psi$. Furthermore, consider any $v \in W \cup W_\psi^*$ with $wR_\psi v$. We show that $M \upharpoonright_\psi, v \models \psi$. If $v \in W_\psi^*$ the result is immediate by Lemma 3. Suppose $v \in W$. Then we have that $wR_\varphi v$. Since $M, w \models \psi$, it holds that $M \upharpoonright_\varphi, w \models \varphi$. Two applications of Lemma 3 then yield $M \upharpoonright_\psi, v \models \psi$. \square

Each of these relativized validities of realization logic says that agents can update their epistemic state in such a way that after the update they know the respective K-validity. In other words, although agents' epistemic states are not subject to the closure conditions constituting logical omniscience, agents are in a position to know the validities characterizing these conditions. Thus, on our approach agents are not logically ignorant, nor are they logically omniscient.

3 Successful updates

In the present section and the next we look more closely at some of the properties of realization logic. Here we discuss successful updates, which is a standard theme in PAL (Gerbrandy, 1998; van Ditmarsch, 2008). It is tempting to assume that after the realization of any formula, the formula holds. However, in general this is not the case. First let's fix some terminology.

Definition 5 Let $\varphi \in \mathcal{L}$.

- (i) An open world model $M_1 = (W_1, W_1^*, R_1, V_1, V_1^*)$ is a *submodel* of open world model $M_2 = (W_2, W_2^*, R_2, V_2, V_2^*)$ if the following conditions hold:
 - $W_1 \subseteq W_2$ and $W_1^* \subseteq W_2^*$;
 - $R_1 = R_2 \cap (W_1 \cup W_1^*)^2$;
 - $V_1 = V_2 \upharpoonright W_1$ and $V_1^* = V_2^* \upharpoonright W_1^*$.
- (ii) φ is *closed under submodels*, if for all open world models M_1 and all $w \in W_1 \cup W_1^*$ it holds that: If $M_1, w \models \varphi$ and M_2 is a submodel of M_1 with $w \in W_2 \cup W_2^*$, then $M_2, w \models \varphi$.
- (iii) φ is a *successful update*, if $\models [\varphi]\varphi$.

Not all formulas are successful updates. For a start, note that the satisfaction of a formula that contains the K -operator can change due to the elimination of impossible worlds, where the formula in question does not hold. A simple example is the model $M = (\{w\}, \{w^*\}, \{(w, w^*)\}, V, V^*)$ with $V^*(w^*)(p) = V^*(w^*)(\neg Kp) = 0$ and V

arbitrary. It is evident that $M, w \models \neg Kp$, but $M \upharpoonright_{\neg Kp}, w \not\models \neg Kp$. So, $M, w \not\models [\neg Kp]\neg Kp$.

What are conditions for a formula to be a successful update? We can give a necessary syntactic condition, which coincides with that which holds for PAL. We start with another simple lemma.

Lemma 6 *Every formula closed under submodels is a successful update.*

Proof Let M be any model based on an open world frame F and let $w \in W$ with $M, w \models \varphi$. Since $M \upharpoonright_{\varphi}$ is a submodel of M and $w \in W \cup W_{\varphi}^*$ we have that $M \upharpoonright_{\varphi}, w \models \varphi$. □

Next we define the set of formulas which will turn out to contain only formulas whose satisfaction is preserved under submodels.

Definition 7 The set of *stable* formulas is the least subset of \mathcal{L} generated by the BNF

$$\varphi ::= p \mid \neg p \mid (\varphi_1 \wedge \varphi_2) \mid K\varphi \mid [\neg\varphi_1]\varphi_2 \quad (p \in \Phi)$$

Now, we can prove the main result concerning successful updates for realization logic.

Proposition 8 *Every stable formula is closed under submodels.*

Proof A straightforward induction on the structure of stable formulas. We show the case involving realization modalities. Assume towards a contradiction that $M, w \models [\neg\psi]\chi$ and $N, w \not\models [\neg\psi]\chi$, where N is a submodel of M that contains w . Since $N, w \models \neg\psi$, the induction hypothesis implies that $M, w \models \neg\psi$ and so $M \upharpoonright_{\neg\psi}, w \models \chi$. Because $N \upharpoonright_{\neg\psi}$ is a submodel of $M \upharpoonright_{\neg\psi}$ and w is contained in $N \upharpoonright_{\neg\psi}$, we have that $N \upharpoonright_{\neg\psi}, w \models \chi$. Contradiction, since $N \upharpoonright_{\neg\psi}, w \not\models \chi$. □

Corollary 9 *Every stable formula is a successful update.*

4 Van Benthem’s access logic

Johan van Benthem has formulated a dynamic epistemic logic that is meant to model realizations in our sense. Indeed, our logic follows some suggestions van Benthem makes in passing as to how to reformulate his own logic more along the lines of PAL by using impossible worlds (van Benthem, 2008, 2011). Van Benthem conjectures that the resulting logic should be equivalent with his own. In this section we show that van Benthem is wrong and that his logic can be simulated in ours. First let’s briefly outline the syntax and semantics of van Benthem’s logic, which I will call (for reasons that will become clear shortly) *access logic*.

The language \mathcal{L}^a of access logic extends the language of realization logic by adding a unary sentential operator I with the syntactic clause that $I\varphi$ is a formula, for any factual formula φ . Sets that contain only factual formulas are themselves called factual. Models for the access language are tuples $M = (W, W^a, \sim, \pi, \cdot)$, where

- $W \neq \emptyset$ is set of possible states;
- $W^a \subseteq W \times \wp(\mathcal{L}^a)$ is the graph of a total function from states to factual sets;
- $\sim \subseteq W \times W$ is an equivalence relation modelling epistemic indistinguishability;
- π is a valuation sending atoms to subsets of W .

If $(w, X) \in W^a$, (w, X) is termed an *access state* and X its *access set*. Models in addition obey the restriction that epistemically indistinguishable states have the same access set: If $w \sim v$ and $(w, X), (v, Y) \in W^a$, then $X = Y$.

The recursive definition of the satisfaction relation between models, access states and formulas has the following main clauses:

- $M, (w, X) \models p \Leftrightarrow w \in \pi(p)$;
- $M, (w, X) \models K\varphi \Leftrightarrow M, (v, X) \models \varphi$, for all $v \in W$ with $w \sim v$;
- $M, (w, X) \models I\varphi \Leftrightarrow \varphi \in X$ and $M, (w, X) \models \varphi$;
- $M, (w, X) \models [\varphi]\psi \Leftrightarrow M, (w, X) \not\models K\varphi$ or $M \upharpoonright_{\varphi}^a, (w, X \cup \{\varphi\}) \models \psi$.

Here, for an access model $M = (W, W^a, \sim, \pi)$, let

$$M \upharpoonright_{\varphi}^a := (W, \{(v, Y \cup \{\varphi\}) : (v, Y) \in W^a\}, \sim, \pi).$$

Additionally, the satisfaction relation is subject to the requirement that $M, (w, X) \models \psi$, for all $\psi \in X$, which we call the *access requirement*. Validity of formulas in an access model is defined as satisfaction in every access state of the model.

Some remarks may be in order: First $I\varphi$ has as its intended interpretation that the agent is explicitly informed that φ , while $K\varphi$ is to be understood as saying that the agent is implicitly informed that φ . Secondly, access states are ways in which the agent cognitively accesses epistemic states, the access consisting in factual formulas that represent aspects of the state for the agent. Finally, note that under the access requirement the satisfaction conditions of formulas of the form $I\varphi$ can be simplified to $M, (w, X) \models I\varphi$ iff $\varphi \in X$.

Obviously, implicit information is subject to logical omniscience. Explicit information, on the other hand, can easily avoid any of the omniscience closure conditions simply by syntactically filtering out unwanted formulas. In addition, access logic can solve the problem of logical ignorance by providing validities that tell us that agents can extend their access to states in such ways as to know the omniscience formulas in question. For instance, the following restriction of the distribution axiom is valid in all access models.

$$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow [\psi]I\psi),$$

where ψ is factual.

So, access logic appears to be similar to realization logic. However, these logics are not equivalent, contrary to what van Benthem conjectured. To see this note that the following recursion axiom is valid in every access model.

$$[\varphi]K\psi \leftrightarrow (K\varphi \rightarrow K[\varphi]\psi)$$

While the left-to-right direction is valid in every open world model, this is not the case for the right-to-left direction. Here is a counter model: Let $M = (\{w\}, \{w', w'', w'''\}, R, V, V^*)$ with wRw', wRw'', wRw''' and $V^*(x)([p]q) = V^*(x)(p) = 1$, for $x = w', w'', w'''$ and $V^*(w')(q) = 0$.

Even though realization logic and access logic are not equivalent in the sense of having exactly the same validities in the intersection of their languages, we may ask whether the logics can simulate each other. More precisely: Does every access model induce an open world model such that both have exactly the same validities from the access language? And conversely, does every open world model induce an access model such that both models validate the same formulas from the realization language? We can answer the first question in the affirmative; the second is still open.

The proof that follows will proceed by induction on a complexity measure that is standardly used in completeness proofs for PAL (van Ditmarsch, 2008).

Definition 10 Let $c : \mathcal{L}^a \rightarrow \mathbb{N}$ be defined by recursion as follows:

- $c(p) = 1$;
- $c(\neg\varphi) = 1 + c(\varphi)$;
- $c(\varphi \wedge \psi) = \max\{c(\varphi), c(\psi)\}$;
- $c(K\varphi) = 1 + c(\varphi)$;
- $c(I\varphi) = 1 + c(\varphi)$;
- $c([\varphi]\psi) = (4 + c(\varphi)) \cdot c(\psi)$.

It is easy to show that the ordering of \mathcal{L}^a induced by this complexity measure preserves the subformula ordering of this language.

Now, we can turn to the embedding result.

Proposition 11 Every access model M induces an open world model M' such that for all $\varphi \in \mathcal{L}^a$ we have:

$$M \models \varphi \Leftrightarrow M' \models \varphi.$$

Proof For any access model $M = (W, W^a, \sim, \pi)$ let $M' = (W, W^*, R, V, V^*)$ with

- $W^* = W^a$;
- $R = \sim \cup \{(w, (v, X)) : (v, X) \in W^a \text{ and } w \sim v\}$;
- $V(p)(w) = 1 :\Leftrightarrow w \in \pi(p)$, for every $w \in W$;
- $V^*(\varphi)(w, X) = 1 :\Leftrightarrow M, (w, X) \models \varphi$, for every $(w, X) \in W^*$.

We can then define satisfaction in states from M' as follows. Let $M', (w, X) \models \varphi :\Leftrightarrow V^*(\varphi)(w, X) = 1$. For every $w \in W$:

- $M', w \models p :\Leftrightarrow V(w)(p) = 1$;
- $M', w \models \neg\varphi :\Leftrightarrow M', w \not\models \varphi$
- $M', w \models (\varphi \wedge \psi) :\Leftrightarrow M', w \models \varphi, \psi$;
- $M', w \models K\varphi :\Leftrightarrow M', v \models \varphi$, for all $v \in W \cup W^*$ with wRv ;
- $M', w \models I\varphi :\Leftrightarrow M', w \models \varphi$ and $\varphi \in X$, for the unique X with $(w, X) \in W^*$.
- $M', w \models [\varphi]\psi :\Leftrightarrow M', w \not\models K\varphi$ or $M'^a \upharpoonright_\varphi, w \models \psi$.

Here, $M'^a \upharpoonright_{\varphi} = (W, \{(w, X \cup \{\varphi\}) : (w, X) \in W^*\}, R_{\varphi}, V, V_{\varphi}^*)$, where $R_{\varphi}, V_{\varphi}^*$ are the obvious restrictions.

By induction on the complexity of access formulas φ we show that for all $w \in W$:

$$M, (w, X) \models \varphi \Leftrightarrow M', w \models \varphi.$$

The base case $c(\varphi) = 1$ is trivial. Suppose that $c(\varphi) = n + 1$ and that the result holds for all formulas ψ with $c(\psi) \leq n$. If φ is a negation, a conjunction or has the form $I\psi$, the result follows straightforwardly from the induction hypothesis. Let φ have the form $K\psi$ and assume that $M, (w, X) \models K\psi$ and that $v \in W \cup W^*$ with wRv . If $v \in W$, we have $w \sim v$ and so by the induction hypothesis it follows that $M', v \models \psi$. If $v \in W^*$, we have $v = (v', Y) \in W^a$ and $w \sim v'$. By the induction hypothesis this implies that $M', v \models \psi$. The other direction is shown similarly. Finally assume that φ has the form $[\psi]\chi$. Let $M, (w, X) \models [\psi]\chi$ and $M', w \models K\psi$. Since $c(K\psi) < c([\psi]\chi)$ the induction hypothesis implies that $M, (w, X) \models K\psi$. So, $M^a \upharpoonright_{\psi}, (w, X \cup \{\psi\}) \models \chi$. Since $c(\chi) < c([\psi]\chi)$ the induction hypothesis implies that $M'^a \upharpoonright_{\psi}, w \models \chi$. So, $M', w \models [\psi]\chi$. The other direction is shown analogously. \square

It may be conjectured that realization logic can be simulated in access logic as well. The reason for this conjecture lies in two facts: Firstly, as van Benthem himself points out (van Benthem, 2011), access logic is a dynamization of the general awareness logic of Fagin and Halpern (1988). Realization logic itself is a dynamization of the open world approach. Secondly, it is an established result that the general awareness approach and the open world approach can simulate each other (Wansing, 1990; Thijsse, 1993). Future work has to show whether the conjecture concerning access logic and realization logic is correct.⁴

5 Bjerring and Skipper's logic

The idea that non-omniscient yet non-ignorant agents are best modeled via epistemic logics that combine impossible worlds with techniques from dynamic epistemic logic is not entirely new. Although a relative newcomer to the scene of epistemic logic, the idea has already been spelled out in a variety of different logical systems (Rasmussen,

⁴ One anonymous reviewer suggested that the open world counter model to the recursion axiom defined above in conjunction with the embedding result might yield a proof for the claim that every open world model induces an equivalent access model, thereby settling the second embedding question I raised above. The referee's idea seems to be that the open world counter model induces an equivalent access model, which thus also would be a counter model to the recursion axiom. It is not clear how such a proof should work. Firstly, the embedding result I proved, guarantees that every access model corresponds to an equivalent open world model. So the result cannot be used to infer that the *open world* counter model induces an equivalent access model. Secondly, even if we could construct an equivalent access model for the specific open world counter model, this would be far from having proved an embedding result, because the counter model is a particular model and not an arbitrarily chosen one.

2015; Bjerring & Skipper, 2018; Solaki, 2019, 2022).⁵ In this section I situate realization logic with respect to this research. However, since these logics involve a certain amount of mathematical complexity and space limitations should not be overextended, it is not possible to compare realization logic with all of these in a way that would do each of them justice. Instead, I will contrast realization logic with one of the most influential systems among those mentioned, namely the logic advanced by Bjerring and Skipper in Bjerring and Skipper (2018).⁶

Bjerring and Skipper’s logic (in what follows abbreviated by ‘BS logic’) combines an open world approach with machinery from dynamic epistemic logic in order to model non-omniscient agents that are logically competent. Logical competence is intuitively spelled out via a notion of trivial consequence that is relative to deductive resources. The deductive resource of an agent is determined by a pair consisting of a set of inference rules R and a natural number n , where n is the greatest natural number m , such that the agent can easily perform m inference steps by applying rules from R . A formula φ is then defined to be a trivial logical consequence from a formula set Σ relative to a resource (R, n) , if φ can be deduced from Σ by at most n applications of inference rules from R . Finally, an agent a with associated resource (R, n) is defined to be logically competent, if a can deduce ψ , where ψ is any formula that is a trivial consequence of a ’s belief set relative to (R, n) .

To formalize this intuitive notion of logical competence the language of basic epistemic logic is extended by countably infinitely many deduction operators to the language \mathcal{L}^{BS} of BS logic:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi_1 \wedge \varphi_2) \mid B\varphi \mid \langle n \rangle\varphi \mid [n]\varphi \quad (p \in \Phi, n \in \mathbb{N})$$

Here, B is a belief operator and the intended reading of $\langle n \rangle\varphi$ ($[n]\varphi$) is that after some (every) sequence of n inference steps, φ holds.

Turning to semantic matters, let a *doxastic model* for one agent be a tuple $M = (W^p, W^i, f, V)$, where

- W^p, W^i are non-empty sets of possible and impossible worlds, respectively, and $W := W^p \cup W^i$;
- $f : W \rightarrow \wp(W)$ is an accessibility function sending every world to the set of worlds that are doxastic alternatives for the agent;
- $V : W \rightarrow \wp(\mathcal{L}^{BS})$ is a valuation function.

To define the semantic clauses for the deduction modalities, we need some definitions. Let $\Sigma \cup \Gamma \cup \{\psi\} \subseteq \mathcal{L}^{BS}$. Say that $\Sigma \vdash_R^n \psi$, if the agent can deduce ψ from Σ in at most n steps via rules from R (where (R, n) is the resource of the agent). Let $\Sigma \vdash_R^n \Gamma$,

⁵ The interesting paper (Sedlár, 2021) develops a general logical framework that allows to embed many epistemic logics for non-omniscient agents. This framework arises from generalising neighbourhood semantics for modal logic in such a way that epistemic or doxastic contents are explicitly represented without building any specific assumptions about them into the generalised models. It would be interesting to investigate whether the framework is strong enough to simulate the various dynamic open world approaches mentioned above. This might give the task of comparing them to each other a more solid logical basis.

⁶ For instance, the strong influence of Bjerring and Skipper (2018) is recorded in Solaki (2019) by describing Bjerring and Skipper’s approach as ‘the closest antecedent’ of Solaki et al.’s proposal.

if $\Sigma \vdash_R^n \psi$, for all $\psi \in \Gamma$. For a doxastic model M , any $w \in W$ and $n \in \mathbb{N}$ define the *n-radius* of w by

$$w^n := \begin{cases} \{w\}, & \text{if } w \in W^p; \\ \{w' \in W^i : V(w) \subseteq V(w') \wedge V(w) \vdash_R^n V(w')\}, & \text{if } w \in W^i. \end{cases}$$

The elements of w^n are the *n-expansions* of w . Note that impossible worlds can have more than one *n-expansion*.

For a doxastic model M let $C : \wp(\wp(W)) \rightarrow \wp(\wp(W))$ be a function, called a *choice function*, such that, for all $X \subseteq \wp(W)$ we have: $C(X)$ is the set of all choice sets for X , if X is a set of non-empty and pairwise disjoint sets; otherwise, let $C(X) = \emptyset$.⁷ Deduction modalities $\langle n \rangle, [n]$ are interpreted via relations \sim^n between pointed doxastic models, that is, pairs (M, w) consisting of doxastic models M and worlds $w \in W$. The idea is that pointed models characterize doxastic states of an agent and that $(M, w) \sim^n (M', w')$, if (M', w') is a doxastic state that the agent can reach from the state (M, w) by performing at most n inferential steps. This idea is captured formally by letting $(M, w) \sim^n (M', w')$, if the set of doxastic alternatives of w in M is replaced by a choice set of the *n-expansions* of these alternatives. If $M = (W^p, W^i, f, V)$ and $w \in W$, let $g : W \rightarrow \wp(W)$ be an (w, n) -variation of f , if for every $v \in W$ we have:

$$g(v) := \begin{cases} c, & \text{if } v = w; \\ f(v), & \text{else,} \end{cases}$$

where $c \in C(\{u^n : u \in f(w)\})$. $\mathcal{F}^n(M, w)$ denotes the set of all (w, n) -variations of f . Finally, let $M = (W^p, W^i, f, V), M' = (W'^p, W'^i, f', V'), w \in W$ and $w' \in W'$. Then define $(M, w) \sim^n (M', w')$ iff $w = w', W = W', V = V'$ and $f' \in \mathcal{F}^n(M, w)$. In the definition of the satisfaction relation between worlds and formulas the belief operator and the boolean connectives are treated analogously to the case of realization logic. As regards deduction modalities, a pointed model (M, w) satisfies $\langle n \rangle \varphi$ ($[n] \varphi$) iff φ is satisfied in some (every) \sim^n -successor of (M, w) . Logical consequence is defined as preservation of satisfaction in every possible world from every doxastic model.⁸

For BS logic we can prove the following result:

- If $\{\varphi_1, \dots, \varphi_n\} \vdash_R^n \varphi$, then $\{B\varphi_1, \dots, B\varphi_n\} \models \langle n \rangle B\varphi$,

where \models denotes logical consequence as defined above. It is evident that BS logic can model non-omniscient agents. The result above is supposed to show that BS logic

⁷ As usual, a choice set for a set of sets Y is a set U with $U \subseteq \bigcup Y$ and $U \cap y$ a singleton, for every $y \in Y$. The axiom of choice guarantees that every set of non-empty and pairwise disjoint sets admits a choice set.

⁸ The satisfaction definition given by Bjerring and Skipper in addition contains clauses for two different falsity relations, depending on whether the worlds involved are possible or impossible. In possible worlds falsity is non-truth, while in impossible worlds falsity is truth of the negation. These clauses are immaterial for the following comparison between BS logic and realization logic and therefore are not considered any further here.

also captures logical competence. Assume that (R, n) is the resource of the agent, (M, w) satisfies $B\varphi$, for some $w \in W^p$, and $\varphi \vdash_R^n \psi$. By the result it then follows that (M, w) satisfies $\langle n \rangle B\psi$. Thus, according to BS logic, agents have the ability to believe formulas ‘immediately’ (that is, after performing some easy inference steps) which trivially follow from their beliefs.

Let’s turn to a comparison between BS logic and realization logic, which will also highlight some shortcomings of the former. Firstly, the semantics for the deduction modalities of BS logic involves what in the AI literature is sometimes called a syntactic approach (Levesque, 1984): The satisfaction conditions of deduction formulas essentially refer to syntactic objects like formulas and inference rules. That’s because the \sim^n -successors of pointed models (M, w) are determined by the set of n -radii of the doxastic alternatives to w and every such n -radius is in turn determined by the resource (R, n) of the agent. In contrast, realization logic appeals only to non-syntactic objects when it comes to the satisfaction conditions of realization formulas. After all, the submodels where the operand formulas of realization modalities are evaluated are just tuples of sets.

Sometimes it is objected to syntactic approaches that they make ad hoc assumptions about what syntactic objects are assigned to what agents as their deductive resources or as their objects of belief (Levesque, 1984). However, every serious logic that tries to avoid logical omniscience uses a semantics for epistemic or doxastic operators that involves syntactic objects to a more or less apparent degree: Awareness logics use awareness functions, which output sets of formulas, impossible world approaches use impossible states that may be formally equated with sets of formulas⁹ and syntactic approaches use syntactic entities as deductive resources or as objects of knowledge or belief.

Secondly, in BS logic the semantics of the deduction modalities is evidently much more complex than the semantic treatment of realization modalities in realization logic. This property of BS logic is potentially problematic as it may complicate the study of its metalogical features such as expressiveness, succinctness or computational complexity. The same aspect of BS logic could also limit its use as an applied logic to describe resource bounded reasoning. For the relative complexity of the semantics makes model building much more difficult than in the case of realization logic.

Thirdly and relatedly, the usual formalisms of dynamic epistemic logic use model transformations that are familiar from classical model theory and that therefore are well understood mathematically. Realization logic sticks to this way of formalizing dynamics by using a simple mechanism of submodel construction. BS logic, on the other hand, employs model transformations whose model theoretic underpinnings are not entirely clear (at least to me). This has to do with the liberal use of choice functions in BS logic. Bjerring and Skipper seem to presuppose that in their semantics choice functions never output the empty set. But this is the case only if, for every world w , the set containing exactly the n -radii of successors of w is a set of non-empty and pairwise disjoint sets. It is clear that every n -radius of any world is non-empty. However, it is not straightforward to see that a situation never occurs, where the set of n -radii of

⁹ In Routley and Routley (1972) impossible states are defined as sets of formulas called ‘set-ups’. Note that the Routleys are advocates of relevantism, so that set-ups are closed under some relevant logic.

successors has the form $\{\{w\}, \{w, v\}, \{v\}\}$, for some worlds w, v with $w \neq v$. Such a set does not admit a choice set.

The conceptual idea behind BS logic, namely that logically competent agents can update their doxastic states via deduction actions, is a very promising one. Nonetheless, I think that the formalism Bjerring and Skipper use to implement the idea is not optimal. It is worth pointing out that there are regular dynamic logics used in AI, such as descriptive dynamic logic (Sierra et al., 2002), that are specifically designed to model deduction actions via program modalities and that are easily adaptable to an open world approach. Indeed, I have recently worked out an epistemic logic that combines open worlds with machinery from descriptive dynamic logic that allows a simple proof of Bjerring and Skipper's central result mentioned above. This logic has a comparatively simple semantics and is based on well understood mathematical formalisms (Mai, 2022).

6 Coda on hyperintensionality

What is the hyperintensional picture that emerges from the current approach? I take propositions to be entities that play certain roles, like being the meanings of declarative sentences and being the primary truth bearers. Let's call a proposition *hyperintensional*, if its identity is not determined by its truth-conditions. Now, there has been a lot of philosophical debate on whether propositions are hyperintensional and what hyperintensionality precisely consists in. I think that the intuitive notion of proposition is a quite messy one with different conceptual desiderata pulling in different directions. It is questionable that we can single out one precisely defined kind of objects comprising all and only the propositions (Lewis, 1986).

My take is that we may choose to model propositions in different ways, depending on the context and our interests. In the case of propositions expressed by attitude reports there are many contexts where we are interested in the linguistic ways speakers express their attitudes. In these contexts we should regard the propositions expressed by the attitude reports in question as hyperintensional propositions. My favourite way of modeling hyperintensional propositions that are attitude contents is to treat them as sets of possible or impossible worlds and that may well be the best modeling we have at present.

However, there are also contexts, where the choice of words speakers use to give voice to their attitudes is immaterial to our interests. In these cases we may model the attitude contents expressed as coarse-grained objects such as sets of possible worlds. Different types of propositions are suited for different purposes and I do not see why we should favour one type over all the others. So, according to the present approach attitude reports at least sometimes report attitude relations to hyperintensional propositions, which are sets of possible or impossible worlds. Realization logic is meant to model the reasoning of resource bounded agents regarding some of these attitude relations.

Acknowledgements Different versions of this paper were presented on various occasions, where I received many valuable comments from the audience. I want to thank Macej Kleczek and Ede Zimmermann for pointing out a fatal flaw in a forerunner of the present logical system. Thanks also go to Aybüke Özgün for providing excellent hints on how a completeness proof for realization logic might be constructed.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests The author did not receive support from any organization for the submitted work. The author has no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bjerring, J. C., & Skipper, M. (2018). A dynamic solution to the problem of logical omniscience. *Journal of Philosophical Logic*, 48, 501–21.
- Cresswell, M. J. (1970). Classical intensional logics. *Theoria*, 36, 347–72.
- Cresswell, M. J. (1972). Intensional logics and logical truth. *Journal of Philosophical Logic*, 1, 2–15.
- Fagin, R., et al. (1995). A nonstandard approach to the logical omniscience problem. *Artificial Intelligence*, 79, 203–40.
- Fagin, R., et al. (1995). *Reasoning about knowledge*. MIT Press.
- Fagin, R., & Halpern, J. Y. (1988). Belief, awareness and limited reasoning. *Artificial Intelligence*, 34, 39–76.
- Gerbrandy, J.D. (1998). *Bisimulations on Planet Kripke*. Ph.D. thesis, University of Amsterdam.
- Gerbrandy, J. D., & Groeneveld, W. (1997). Reasoning about information change. *Journal of Logic, Language, and Computation*, 6, 147–69.
- Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4, 475–84.
- Jago, M. (2014). *The impossible*. OUP.
- Levesque, H. J. (1984). A logic of implicit and explicit belief. In *Proceedings of the Fourth International Conference on Artificial Intelligence* (pp. 198–202). AAAI Press.
- Levesque, H. J., & Lakemeyer, G. (2000). *The logic of knowledge bases*. MIT Press.
- Lewis, D. (1986). *On the plurality of worlds*. Blackwell.
- Lewis, D., et al. (2004). Letters to priest and beall. In B. Armour-Garb (Ed.), *The law of non-contradiction* (pp. 176–177). OUP.
- Mai, J. (2022). *Impossibilities in action*. Unpublished manuscript.
- Plaza, J. 2007[1989]. Logics of public communications. *Synthese*, 158, 165–79.
- Priest, G. (1992). What is a non-normal world? *Logique et Analyse*, 35, 291–302.
- Priest, G. (2016). *Towards non-being* (2nd ed.). OUP.
- Rantala, V. (1982). Impossible worlds semantics and logical omniscience. In I. Niiniluoto & E. Saarinen (Eds.), *Intensional logic* (pp. 106–115). Helsinki.
- Rantala, V. (1982). Quantified modal logic: Non-normal worlds and propositional attitudes. *Studia Logica*, 41, 41–65.
- Rasmussen, M. S. (2015). Dynamic epistemic logic and logical omniscience. *Logic and Logical Philosophy*, 24, 377–99.
- Routley, R., & Routley, V. (1972). The semantics of first degree entailment. *Noûs* 6, 335–359.
- Sedlár, I. (2021). Hyperintensional logics for everyone. *Synthese*, 198, 933–56.
- Sierra, C., et al. (2002). A descriptive dynamic logic and its application to reflective architectures. In J.-J.C. Meyer & J. Treur (Eds.), *Handbook of defeasible reasoning and uncertainty management systems* (Vol. 7, pp. 197–220). Springer.
- Sim, K. (1997). Epistemic logic and logical omniscience: A survey. *International Journal of Intelligent Systems*, 12, 57–81.

- Solaki, A., et al. (2019). The logic of fast and slow thinking. *Erkenntnis*, 86, 733–62.
- Solaki, A. (2022). The effort of reasoning: Modelling the inference steps of boundedly rational agents. *Journal of Logic, Language and Information*, 31, 529–53.
- Thijssse, E. (1993). On total awareness logics. In M. de Rijke (Ed.), *Diamonds and defaults* (pp. 309–347). Kluwer.
- van Benthem, J., et al. (2006). One is a Lonely Number: on the Logic of Communication. In Z. Chatzidakis (Ed.), *Logic Colloquium '02* (pp. 96–129). A.K. Peters. ASL and.
- van Benthem, J. (2008). Merging observational and access dynamics in logic. *Journal of Logic Studies*, 1, 1–16.
- van Benthem, J. (2011). *Logical dynamics of information and interaction*. Cambridge UP.
- van Ditmarsch, H., et al. (2008). *Dynamic epistemic logic*. Springer.
- Wansing, H. (1989). Bemerkungen zur Semantik nicht-normaler möglicher Welten. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 35, 551–7.
- Wansing, H. (1990). A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica*, 49, 523–39.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.