Knowledge Without Complete Certainty

Hans van Ditmarsch[1] and Louwe B. Kuijer[2]

[1] CNRS, LORIA, France (hans.van-ditmarsch@loria.fr)
[2] University of Liverpool, United Kingdom (Louwe.Kuijer@liverpool.ac.uk)

Abstract. We present an epistemic logic ELF (Epistemic Logic with Filters) where knowledge does not require complete certainty. In this logic, instead of saying that an agent knows a particular fact if it is true in every accessible world, we say that it knows the fact if it is true in a sufficiently large set accessible worlds. On a technical level, we do this by enriching the standard Kripke models of epistemic logic with a set of filters: a sufficiently large set of worlds is one that is in the filter. We introduce semantics for ELF, and give a sound and complete proof system.

1 Introduction

In the standard Kripke semantics for epistemic logic, we say that an agent a *knows* a proposition φ if and only if φ is true in every world that is *epistemically accessible* for a [15,12]. In other words, according to such semantics a knows that φ if and only if φ is true in every world that is consistent with a's observations.

Unfortunately, it is generally not very hard to invent worlds that are consistent with a's observations where φ is false. Along a general line of skepticism going back to (at least) Descartes' evil demon (*le mauvais genie*) and that continues to flourish in logical and epistemological circles [25,10], consider the following example, adapted from Harman [14].

> Alice is sitting at her desk, writing a logic paper. Strictly speaking, the skeptical scenario where she is merely a "brain in a vat" that wrongly believes itself to be sitting at a desk is consistent with Alice's observations. In theory, this means that the possible world where Alice is a brain in a vat is epistemically accessible for her, so according to the standard Kripke semantics she does not *know* that she is sitting at her desk. Still, we would typically like to say that Alice does know this fact.

There are several solutions to this problem. Firstly, we could bite the bullet and conclude that Alice does not, and cannot, know that she is sitting at her desk. This skeptic's choice is internally consistent, but results in a rather trivial notion of (unobtainable) knowledge. So while we acknowledge that the skeptics may be correct, that is not the kind of knowledge that we are interested in here.

Secondly, we could say that the scenario where Alice is a brain in a vat should not be considered a proper possible world, and therefore should not be among

Alice's epistemic alternatives. Doing so can be justified from a *contextualist* point of view [28,11,21], which states that the conditions for knowledge depend on context. Whenever skeptical scenarios are irrelevant, they are excluded by the context. As long as we are modeling a context where skeptical scenarios are excluded, we may (and must) omit the worlds where Alice is a brain in a vat, allowing us to conclude that Alice knows that she is sitting at her desk.

This second solution is the most practical one, and commonly used in epistemic logic. Unfortunately, this solution is not always available. If we want to reason about whether Alice knows that she is not a brain in a vat, then clearly the world where she is in fact a brain in a vat is relevant to our context. So it cannot be omitted. What are the consequences? In the epistemic logical setting, and in particular in the modal logical propositional modeling of it, the typical notion to fall short of knowledge within a given context is called *belief*, and the minimal difference between belief and knowledge is that belief unlike knowledge may be *false* (incorrect). Even within that restriction there is a wide gap between *defeasible belief* [19] and so-called *conviction* [26]. Defeasible beliefs may be defeated, i.e., the agents may be willing to change their beliefs after further evidence or consideration. But false convictions remain false forever. When modeling *certain* knowledge this rather Platonic focus on *true* knowledge is peculiar. Why should one particular exception of the rule matter more than any other exception? Many works have been dedicated to the difference between (modal) knowledge and belief [18,16], and in particular on notions of knowledge closer to belief [15,27], *fallible knowledge* [2] and (the more dynamically motivated) *safe belief* [3]. They all fall short of modeling certain knowledge, because the set of accessible worlds where φ is false is always *too big*, even when there is only one.

In this paper we therefore choose a third solution: we *do* include the world where Alice is a brain in a vat, as well as other skeptical worlds. But we say that we *know* φ even if there are accessible worlds where φ is false, as long as the set of accessible φ worlds is *sufficiently large*.

Much like the second one, the third solution is justified by contextualism. The key observation is that our context as modelers may be different from the context of the agents being modelled. We are interested in whether Alice knows she is not a brain in a vat, so our context does not allow us to omit the worlds where she is a brain in a vat. But as long as Alice's context allows her to ignore such skeptical worlds, she can know she is not a brain in a vat even though the worlds are accessible.

The remaining question, then, is to decide on what we mean by the set of counterexamples being "small". A simple numerical ("up to n counterexamples") or finite fraction ("up to $\frac{n}{m}$ of the possible worlds may be counterexamples") rule would not solve Alice's problem: we can create infinitely many skeptical scenarios, so for every $n \in \mathbb{N}$ there are more than n counterexamples, and the ratio of worlds where she is a brain in a vat divided by those where she is not is $\frac{\infty}{\infty}$ and therefore not a finite fraction. Note also that a numerical or finite fraction threshold is vulnerable to the lottery paradox [20], while Alice's example above and further examples below are immune to finite lottery paradoxes.

A more promising approach would be to say that the number of counterexamples is small if the set of counterexamples has measure zero. Such a notion of knowledge, employing Keisler's infinitesimals for measure zero sets [17], has been proposed in [1] for modeling knowledge revision. Measure theory is unnecessarily heavy machinery for our current purpose, however: we do not need an exact measure of all sets of worlds, we only need to know which sets are small. We therefore prefer a very similar but somewhat more lightweight approach: we use *filters*. The notion goes back to [7], and it is frequently used within modal logic [6], also for default reasoning [4]. We say that the set of counterexamples is small if its complement is a member of the filter.

The structure of the rest of this paper is as follows. In Section 2 we formally define the syntax and semantics of our logic Epistemic Logic with Filters (ELF). Then, in Section 3 we present detailed examples. In Section 4 we provide a sound and complete axiomatization for ELF. Section 5 compares our framework to the class of non-normal modal logics known as regular modal logics.

2 Syntax and Semantics

Before defining the language, models and semantics, let us first define filters.

Definition 1. *Let S be a set. Then $F \subseteq 2^S$ is a* filter *if*

- $F \neq \emptyset$,
- *for every* $X_1, X_2 \in F$, *we have* $X_1 \cap X_2 \in F$,
- *for every* $X_1 \in F$ *and every* $X_2 \subseteq S$, *if* $X_1 \subseteq X_2$ *then* $X_2 \in F$.

A filter F is proper *if* $\emptyset \notin F$.

We have no use for improper filters, so for the remainder of this paper we assume all filters to be proper.

A filter serves to identify which subsets of S are small or large, with $X \subseteq S$ being *large* if $X \in F$ and $X \subseteq S$ being *small* if $S \setminus X \in F$. Note that, by the fact that for $X_1, X_2 \in F$ we have $X_1 \cap X_2 \in F$, the intersection of two large sets is itself large. Typical examples of filters include (i) the co-finite subsets of S (if S is infinite),[3] (ii) the sets of full measure in a measure space and (iii) for a fixed $C \subseteq S$, all sets that contain C.

The latter kind of filter, where $F = \{X \subseteq S \mid C \subseteq X\}$, is called a *principal filter*. In that case, the set C can be considered to be the set of *important*, or *relevant* worlds.[4] So in that case it may not be quite accurate to say that a set $X \in F$ is necessarily large. It is, however, *sufficiently* large in the sense that it contains all important worlds. Another way to think of this is that while C may have a small cardinality, the fact that they are important gives C a larger weight, so any set containing C has large weight.

The language of ELF is the same as that of standard single-agent modal logic.

[3] An epistemic modal use of that is the *Majority Logic* of [24].
[4] This is similar to the approach advocated in [21], see also Remark 3.

Definition 2. *Let At be a countable set of propositional atoms. The language* $\mathcal{L}$ *is given by the following normal form, where* $p \in At$*:*

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid \Box\varphi,$$

As usual, we omit parentheses where this should not cause confusion, and use $\wedge, \rightarrow, \leftrightarrow, \Diamond$ and $\bigwedge$ as abbreviations.

The models of ELF are based on the usual Kripke models, but they are enriched with filter structures.

Definition 3. *A* model *is a tuple* $\mathcal{M} = (W, R, \mathcal{F}, V)$*, where* W *is a set of worlds,* $R \subseteq W \times W$ *is an accessibility relation,* $\mathcal{F} : W \rightarrow \{F \mid F \text{ is a filter on } W\}$ *assigns to each world a filter and* $V : At \rightarrow 2^W$ *is a valuation.*

We write $R(w)$ *for* $\{w' \mid (w, w') \in R\}$*.*

Now we can define the semantics.

Definition 4. *The* satisfaction relation $\models$ *is defined recursively by*

$$\begin{array}{ll} \mathcal{M}, w \models p & \Leftrightarrow w \in V(p) \\ \mathcal{M}, w \models \neg\varphi & \Leftrightarrow \mathcal{M}, w \not\models \varphi \\ \mathcal{M}, w \models \varphi \vee \psi & \Leftrightarrow \mathcal{M}, w \models \varphi \text{ or } \mathcal{M}, w \models \psi \\ \mathcal{M}, w \models \Box\varphi & \Leftrightarrow [\![\varphi]\!]_{\mathcal{M}} \cap R(w) \in \mathcal{F}(w) \end{array}$$

We use $\models \varphi$ *and* $\Gamma \models \varphi$ *in the usual way to denote* φ *being valid and* φ *being entailed by* Γ*, respectively.*

Note that we have $\mathcal{M}, w \models \Box\varphi$ if there is a large set of accessible φ worlds. This is not exactly the same as there being a small set of accessible $\neg\varphi$ worlds: the accessible φ worlds being large always implies that the accessible $\neg\varphi$ worlds are small, but if $R(w) \notin \mathcal{F}(w)$ it is possible for $[\![\neg\varphi]\!]_{\mathcal{M}} \cap R(w)$ to be small without $[\![\varphi]\!]_{\mathcal{M}} \cap R(w)$ being large. The reason for this "largeness requirement" is that we consider knowledge to require some amount of intellectual effort and honesty.

It is generally held (e.g., [9,13,22]) that a necessary[5] condition for knowing φ is that φ is a justified true belief. So, in particular, for an agent to know φ it must be the case that there is a justification for the agent to believe φ. In the case of standard epistemic logic, this justification derives from the fact that all accessible worlds satisfy φ. Here, in ELF, the justification derives from the slightly weaker condition that the accessible $\neg\varphi$ worlds are negligible compared to the accessible φ worlds.

In order to obtain this justification it does not suffice that the set of accessible $\neg\varphi$ worlds is small in an absolute sense; if the agent considers three possible worlds and all three of them satisfy $\neg\varphi$, then it would be strange to say that they are justified in believing φ simply because there are few counterexample. Instead, the set of accessible counterexamples must be small compared to the set of accessible φ worlds. But even that is not quite enough; if there are no accessible $\neg\varphi$ worlds and at least one accessible φ world, then one could argue

[5] But, unless one uses a very strong notion of justification, not sufficient [13].

that the set of $\neg\varphi$ worlds is small compared to the set of φ worlds. In some cases, we would endorse the claim that this single accessible φ world, in the absence of accessible $\neg\varphi$ worlds, provides a justification for believing φ. But in other cases, the fact that there is only one world that the agent considers possible can betray a lack of effort and imagination by the agent.

We want the model $\mathcal{M}$ to describe what the agent knows, not what the agent *thinks* they know. This means that the model is drawn from the perspective of an outside observer who knows the agent's mental state, not from the perspective of the agent themselves. So the relation R describes objectively which worlds the agent considers possible. But the mental state is itself of course a subjective opinion of the agent: we are objectively describing a subjective mental state. If the agent has never thought of a world w_2, it would therefore be inaccurate to say that the agent considers w_2 possible, even if w_2 happens to be consistent with the agent's observations.

As a result, if the agent does not consider a world w_2 to be accessible, this could be either because the agent has thought of w_2 and determined it to be incompatible with their information, or because the agent never though of w_2. So if the agent considers only one world to be accessible, this could be because the agent is lazy, and didn't think of any other worlds. In that case, even if the only accessible world satisfies φ, this would not be a justification for believing φ. In order for the agent to be justified in believing φ, they should first consider sufficiently many worlds.

Note that the agent must consider *sufficiently* many worlds. This is not a cardinality requirement: in some situations, a finite number of worlds might be sufficient, while in another case even a continuum of worlds might not be enough. Instead, sufficiency is determined from the perspective of the objective, outside observer who designs the model. Specifically, the model designer determines sufficiency using the filter function $\mathcal{F}$.

If the agent considered sufficiently many worlds, so $R(w) \in \mathcal{F}(w)$, and all but a negligible amount of these worlds satisfy φ, so $[\![\varphi]\!]_{\mathcal{M}} \in \mathcal{F}(w)$, this yields the justification for the agent's belief that φ. These two conditions together are equivalent to $[\![\varphi]\!]_{\mathcal{M}} \cap R(w) \in \mathcal{F}(w)$, our condition for knowledge.

Remark 1. Note that we allow the filter $\mathcal{F}(w)$ to depend on the world w. This is because, otherwise, it would be impossible to have $\mathcal{M}, w_1 \models \Box\varphi$ and $\mathcal{M}, w_2 \models \Box\neg\varphi$. After all, $\mathcal{M}, w_1 \models \Box\varphi$ requires $[\![\varphi]\!]_{\mathcal{M}} \in \mathcal{F}(w_1)$ and $\mathcal{M}, w_2 \models \Box\neg\varphi$ requires $[\![\neg\varphi]\!]_{\mathcal{M}} \in \mathcal{F}(w_2)$. So if $\mathcal{F}(w_1) = \mathcal{F}(w_2)$, then we would have $[\![\varphi]\!]_{\mathcal{M}} \cap [\![\neg\varphi]\!]_{\mathcal{M}} \in \mathcal{F}(w_1)$, which is a contradiction since $\emptyset \notin \mathcal{F}(w_1)$.

Remark 2. The semantics presented above do not guarantee that knowledge in ELF satisfies certain properties that knowledge is often considered to have, such as truthfulness and introspection. ELF is, in this sense, similar to the basic modal logic K. And, like K, ELF can be extended with axioms and frame properties to guarantee truthfulness and introspection.

Remark 3. We recall example (*iii*) of a filter F given by $F = \{X \subseteq S \mid C \subseteq S\}$. This can be seen as an implementation of the contextualist view from [21] of knowledge as truth in all relevant accessible worlds. Among yet other precisions, Lewis writes:

> Then S knows that P iff S's evidence eliminates every possibility in which not-P — Psst! — except for those possibilities that conflict with our proper presuppositions. [21, page 554]

In other words, a proposition (called P) is known (by an agent S) if it is true in the intersection of the accessible worlds and the relevant worlds. On the assumption that $R(w) \in F(w)$, this corresponds exactly to the semantics of ELF when S is the set of relevant worlds.

ELF is more general however, because we do not require that $R(w) \in F(w)$ and not every filter is of the form $\{X \subseteq S \mid C \subseteq S\}$.

Now that we have defined the semantics of our logic, we can consider a few examples in some detail.

3 Examples

Example 1. Bob is a mathematics student. On an exam, he writes a proof by case distinction for a proposition p in some mathematical theory $\mathfrak{T}$. Because Bob is not very experienced in writing proofs, however, he is not certain that his case distinction is exhaustive. But even though Bob does not know this, his case distinction is in fact exhaustive and his proof is correct.

We will represent Bob's situation by a pointed model $\mathcal{M}, w$, where $M = (W, R, \mathcal{F}, V)$. The possible worlds of $\mathcal{M}$ are closely related to the models of $\mathfrak{T}$. Specifically, for every model $\mathcal{T}$ of $\mathfrak{T}$, there is a world where $\mathcal{T}$ is the "true" model. Having one such world per model is not quite enough, however, because there are other facts that may differ per world. In particular, if two worlds w_1 and w_2 have the same model $\mathcal{T}$ but Bob's beliefs differ between w_1 and w_2, then they must be different worlds. This can be represented by considering these worlds to be pairs $w = (\mathcal{T}, i)$, where i is simply some index used to differentiate between worlds with the same model of $\mathfrak{T}$.[6]

In Bob's proof, he considered the cases $q_1, \cdots, q_n$. Because the case distinction is in fact exhaustive, every model of $\mathfrak{T}$ satisfies at least one of these cases. So for every $(\mathcal{T}, i)$, there is at least one j such that $\mathcal{M}, (\mathcal{T}, i) \models q_j$. Furthermore, since the proof is correct, any world that satisfies one of these cases also satisfies p. So we have $\mathcal{M}, (\mathcal{T}, i) \models p$.

In order to represent Bob's uncertainty about whether his case distinction is exhaustive, we need some further worlds where none of the cases apply. While there are no models of $\mathfrak{T}$ that fall outside the case distinction, Bob thinks that there might be. So we need to add a number of worlds of the form $(\mathcal{N}, i)$, where $\mathcal{M}, (\mathcal{N}, i) \not\models q_j$ for every j. Here $\mathcal{N}$ is objectively not a model of $\mathfrak{T}$, but Bob is not certain that it is not a model. Because none of the cases apply, it is uncertain

[6] Because we require W to be a set, as opposed to a class, we may have to restrict ourselves to the models of $\mathfrak{T}$ in some set-theoretic universe $\mathcal{U}$, where $W \notin \mathcal{U}$.

whether these worlds satisfy p. It is possible for p to be true there, but it is also possible for p to be false in these worlds.

The worlds that Bob considers possible are those that fall inside his case distinction. This is true for every world, so $R(w') = V(q_1) \cup \cdots \cup V(q_n)$ for every $w' \in W$. The filters represent the "relevant" worlds, in the sense that one is justified in believing a proposition after verifying that it holds in every world of the filter. Bob's belief in p is justified if p is true in every model of $\mathfrak{T}$, so $\mathcal{F}(w) = \{F \mid C \subseteq F\}$, where C is the set of worlds of the form $(\mathcal{T}, i)$. We have $R(w) \cap [\![p]\!]_{\mathcal{M}} \supseteq C$, and therefore $\mathcal{M}, w \models \Box p$. So Bob knows that p is true.

However, even though Bob's case distinction was exhaustive, he is uncertain about this. So in some of the accessible worlds $w' = (\mathcal{T}, i)$ his case distinction is not exhaustive. In such a world we have $\mathcal{F}(w') = \{F \mid C' \subseteq F\}$, where C' contains not only the worlds of the form $(\mathcal{T}, j)$, but also some worlds of the form $(\mathcal{N}, j)$. In these worlds, we have $R(w') \cap [\![p]\!]_{\mathcal{M}} \notin \mathcal{F}(w')$ and therefore $\mathcal{M}, w' \not\models \Box p$. Note that it does not matter whether p holds in the worlds $C' \setminus C$. Even if p happens to be true in all of C', the fact that his case distinction was non-exhaustive means that his belief in the truth of p would be unjustified.

Example 2. Suppose that we are about to draw a random real number uniformly from the interval $[0, 1]$. This situation can be modeled in the following way:

- For every $x \in [0, 1]$ there is a world w_x where x is the number that is drawn.
- Every world is accessible from every other world, i.e., $R = \{(w_x, w_y) \mid x, y \in [0, 1]\}$.
- The large sets are those that have full measure, i.e., for every $x \in [0, 1]$, we have $\mathcal{F}(w_x) = \{w_Y \mid \mu(Y) = 1\}$, where μ is the Lebesgue measure and $w_Y = \{w_y \mid y \in Y\}$.

Under these circumstances, we can say that we know that the drawn number x will be irrational, since the rationals have measure 0. Note that this knowledge is fallible: even in those worlds where we will draw a rational number, we know that the number will be irrational. Such failure is infinitely unlikely, however.

Example 3. Claire is a software engineer, who is demonstrating a program to a client. The program has been given its input, and is now running. Claire tells the client that she knows that the program will terminate and return the output "TRUE". In saying so, she ignores a number of possible worlds. In particular, if there is a power failure then the program will not terminate at all. Claire has thought of such possibilities, but she considers the conversation with the client to have a number of underlying unspoken assumptions, including the assumption that there will be no power failure. So while there are possible worlds where the program is interrupted by power failure or some other outside factor, the unspoken assumptions render such worlds irrelevant.

The set of worlds W of our model is given by $W = W_1 \cup W_2$, where W_1 is the set of worlds where the program will be allowed to run normally and W_2 is the set of worlds where the program will be interrupted by some outside event, such as a power cut or a meteorite strike. The accessibility relation is given by

$R = W \times W$. Finally, for every world w the filter $\mathcal{F}(w)$ is the set of all sets containing the relevant worlds. In this case, as discussed above, we consider the relevant worlds to be those where the program is allowed to run uninterrupted, so $\mathcal{F}(w) = \{F \mid W_1 \subseteq F\}$. We let p stand for "the program terminates and returns TRUE", so $V(p) = W_1$.

For any world w of this model $\mathcal{M}$ we have $\mathcal{M}, w \models \Box p$. Note that, as in the previous example, this knowledge is fallible: Claire knows p in every world, including those where a power failure occurs. Unlike the previous example, however, such failure is not necessarily infinitely unlikely. The probability of power failures et cetera is low, but not infinitely so, after all. But this possibility of failure does not stop Claire from knowing p, under the conversational assumptions.

Example 4. As above, except now the possibility of a power failure or other outside event has not crossed Claire's mind. The accessibility relation is now given by $R = W \times W_1$. But because W_1 contains all relevant worlds, Claire still knows that the program will return TRUE.

Example 5. As above, except that Claire is now less careful in considering all possible executions of her program. Instead of considering all possible executions W_1, she makes some implicit assumptions and only thinks of $W_1' \subset W_1$. We have $R = W \times W_1'$. The set of relevant worlds remains the same, however: $\mathcal{F}(w) = \{F \mid W_1 \subseteq F\}$.

In this situation, Claire does not know that the program will return TRUE, because $R(w) \notin \mathcal{F}(w)$. Note that this is independent of whether the program returns TRUE in the relevant worlds that she failed to consider: Claire's belief that the program will return TRUE is not justified, so even if she happens to be right she doesn't *know* that the program will return TRUE.

4 Axiomatization

We introduce the proof system **WKL**. The **W** in **WKL** stands for "weak", since **WKL** is strictly weaker than **KL**, which is obtained by adding the axiom **L** to the standard proof system **K** for modal logic.[7]

Definition 5. *The proof system* **WKL** *is given by the following rules and axiom schemata.*

P *all substitution instances of propositional tautologies*
K $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
L $\neg\Box\bot$
RM *if* $\varphi \to \psi$ *is a theorem, infer* $\Box\varphi \to \Box\psi$
MP *from* $\varphi \to \psi$ *and* φ*, infer* ψ*.*

[7] The axiom **L** is, using the other axioms and rules, interderivable with the axiom **D**, given by $\Box\varphi \to \Diamond\varphi$. One could, therefore, think of **WKL** as "weak **KD**" instead of "weak **KL**". Our reason for preferring **L** over **D** in this context is that **L** more closely follows the semantical constraint that $\emptyset \notin \mathcal{F}(w)$.

Definition 6. *A formula φ is a* theorem *of* **WKL***, denoted* $\vdash \varphi$ *if it can be derived in a finite number of steps using the rules and axioms of* **WKL***. A formula φ is* entailed by *a set Γ of formulas, denoted $\Gamma \vdash \varphi$ if φ can be derived in a finite number of steps using the rules and axioms of* **WKL** *and using Γ as premises.*

Note that **WKL** does not have a necessitation rule, i.e., we cannot infer from $\vdash \varphi$ that $\vdash \Box\varphi$. Instead, we use a strictly weaker *monotonicity rule* **RM**. In particular, $\Box\top$ is not provable in **WKL**.[8] ELF is therefore not a normal modal logic, although it is a *regular modal logic* [23]. In Section 5 we discuss ELF's position in the landscape of non-normal modal logics.

Soundness of **WKL** follows immediately from the semantics.

Lemma 1 (Soundness). *For all $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, if $\Gamma \vdash \varphi$ then $\Gamma \models \varphi$.*

Completeness of **WKL** is shown in the usual way, i.e., by constructing a canonical model and proving that every consistent formula is satisfied in that model (see for example [5]). Some of the following lemmas can be proven in the exact same way as the corresponding lemmas in other completeness proofs. We therefore omit the proofs of those lemmas.

We start with a lemma that allows us to switch between three different characterizations of entailment.

Lemma 2. *The following are equivalent.*

1. $\Gamma \vdash \varphi$
2. *there is a finite subset Γ' of Γ such that $\Gamma' \vdash \varphi$*
3. *there is a finite subset Γ' of Γ such that $\vdash \bigwedge \Gamma' \to \varphi$*

As usual, maximal consistent sets will serve as worlds for the canonical model.

Definition 7. *A set Γ of formulas is* consistent *if $\Gamma \nvdash \bot$,* maximal *if for every formula φ either $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$ and* maximal consistent *if it is both maximal and consistent.*

Lemma 3 (Lindenbaum lemma). *Let Γ be a consistent set. Then there is a maximal consistent set Δ such that $\Gamma \subseteq \Delta$.*

Definition 8. *If Γ is a set of formulas, then $\Box^{-1}\Gamma = \{\varphi \mid \Box\varphi \in \Gamma\}$.*

The proof of the following lemma is slightly more complicated than usual, since we only have access to the monotonicity rule **RM** as opposed to the more powerful necessitation rule. We therefore provide a detailed proof.

Lemma 4. *If Γ is consistent, then so is $\Box^{-1}\Gamma$.*

[8] Note that $\not\models \Box\top$ in ELF, since $\mathcal{M}, w \not\models \Box\top$ when $R(w) \notin \mathcal{F}(w)$.

Proof. Suppose towards a contradiction that $\Box^{-1}\Gamma \vdash \bot$. Then there is a finite subset of $\Phi \subseteq \Box^{-1}\Gamma$ such that

$$\vdash \bigwedge \Phi \to \bot.$$

By **RM**, this yields

$$\vdash \Box \bigwedge \Phi \to \Box\bot. \tag{1}$$

Now, note that by repeatedly applying **K** and **MP**, we also have

$$\vdash \bigwedge \Box\Phi \to \Box \bigwedge \Phi. \tag{2}$$

Together, (1) and (2) imply that

$$\Box\Phi \vdash \Box\bot$$

and therefore by **L** and the fact that $\Box\Phi \subseteq \Gamma$

$$\Gamma \vdash \bot,$$

contradicting the consistency of Γ. □

Lemma 5. *If Γ is maximal consistent and $\Gamma \vdash \varphi$, then $\varphi \in \Gamma$.*

Proof. By maximality, either $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$. Since $\Gamma \cup \{\neg\varphi\} \vdash \bot$ it follows from consistency that $\varphi \in \Gamma$. □

Now, let us define the canonical model.

Definition 9. *The* canonical model $\mathcal{M}^c = (W^c, R^c, V^c, \mathcal{F}^c)$ *is given by:*

- W^c *is the set of maximal consistent sets of formulas,*
- $\mathcal{F}(w) = \{F \subseteq W^c \mid S(w) \subseteq F\}$, *where* $S(w) = \{w' \mid \Box^{-1}w \subseteq w'\}$,
- *if* $\Box\top \in w$, *then* $R^c(w) = S(w)$ *otherwise* $R^c(w) = \emptyset$,
- $V^c(p) = \{w \in W^c \mid p \in w\}$.

Lemma 6 (Truth Lemma). *For every $w \in W^c$ and every formula φ, $\mathcal{M}^c, w \models \varphi$ if and only if $\varphi \in w$.*

Proof. By induction on the complexity of φ. If φ is atomic, then the lemma follows immediately from the definition of V^c. So assume as induction hypothesis that φ is not atomic and that the lemma holds for all strict subformulas of φ.

We continue by a case distinction on the main connective of φ. If it is a Boolean connective, then the lemma is once again trivial. So let us consider the interesting case, $\varphi = \Box\psi$.

Suppose $\Box\psi \in w$. By the definition of $S(w)$, we have $[\![\psi]\!]_{\mathcal{M}^c} \supseteq S(w)$. Furthermore, by **P** we have $\vdash \psi \to \top$, so by **RM** we have $\vdash \Box\psi \to \Box\top$. Since w is maximal and consistent, this implies that $\Box\top \in w$. So $R^c(w) = S(w)$. We therefore have $[\![\psi]\!]_{\mathcal{M}^c} \cap R^c(w) = S(w) \in \mathcal{F}(w)$, so $\mathcal{M}^c, w \models \Box\psi$.

Suppose, on the other hand, that $\Box\psi \notin w$. We distinguish two sub-cases. First, suppose that $\Box\top \notin w$. Then $R^c(w) = \emptyset \notin \mathcal{F}(w)$, so $\mathcal{M}^c, w \not\models \Box\psi$.

The other case is if $\Box\top \in w$ but $\Box\psi \notin w$. In this case, suppose towards a contradiction that $\Box^{-1}w \cup \{\neg\psi\}$ is inconsistent. Then there is a finite subset Φ of $\Box^{-1}w$ such that $\Phi \cup \{\neg\psi\}$ is inconsistent. It follows that $\vdash \bigwedge \Phi \to \psi$ and therefore $\vdash \Box \bigwedge \Phi \to \Box\psi$. Since $\Box\varphi \in w$ for every $\varphi \in \Phi$, we have $\Box \bigwedge \Phi \in w$ and therefore $\Box\psi \in w$, contradicting our assumption.

So $\Box^{-1}w \cup \{\neg\psi\}$ is consistent, and can therefore be extended to a maximally consistent set w'. By the definition of $\mathcal{F}$, we have that $w' \in F$ for every $F \in \mathcal{F}(w)$. Since $w' \notin [\![\psi]\!]_{\mathcal{M}^c}$, it follows that $[\![\psi]\!]_{\mathcal{M}^c} \cap R(w) \notin \mathcal{F}(w)$, and therefore $\mathcal{M}^c, w \not\models \Box\psi$. □

Completeness now follows immediately.

Lemma 7 (Completeness). *For all $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, if $\Gamma \models \varphi$ then $\Gamma \vdash \varphi$.*

Proof. If $\Gamma \not\vdash \varphi$ then $\Gamma \cup \{\neg\varphi\}$ is consistent, so by Lemma 3 there is a maximal consistent set $w \supseteq \Gamma \cup \{\neg\varphi\}$. By Lemma 6 this implies that $\mathcal{M}^c, w \models \psi$ for every $\psi \in \Gamma$ and $\mathcal{M}^c, w \not\models \varphi$. Therefore, $\Gamma \not\models \varphi$. □

We have now proven both soundness and completeness.

Theorem 1. *For all $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{K}$, we have $\Gamma \models \varphi$ if and only if $\Gamma \vdash \varphi$.*

Remark 4. In $\mathcal{M}^c$, every filter is of the form $\mathcal{F}(w) = \{F \subseteq W^c \mid S(w) \subseteq F\}$. So all filters in the canonical model are principal filters. It follows that the proof system is also sound and complete for the class of models

$$\mathfrak{M} := \{\mathcal{M} = (W, R, \mathcal{F}, V) \mid \forall w \in W : \mathcal{F}(w) \text{ is principal}\}.$$

5 Comparison to other non-normal modal logics

ELF is a so-called non-normal modal logic. In this section, we therefore compare the semantics of ELF to the commonly used neighborhood semantics, and the proof system **WKL** to other proof systems for non-normal modal logics.

In neighborhood semantics [8,23], a model is a tuple $\mathcal{M} = (W, \mathcal{N}, V)$, where W is a set of worlds, $\mathcal{N} : W \to 2^{2^W}$ is a neighborhood function that assigns to each world a set of sets of worlds and $V : At \to 2^W$ is a valuation. We then say that $\mathcal{M}, w \models \Box\varphi$ if and only if $[\![\varphi]\!]_{\mathcal{M}} \in \mathcal{N}(w)$.

The semantics of ELF can be reduced to neighborhood semantics. This is not very surprising, both because the formalism of neighborhood semantics is sufficiently versatile to encompass almost everything done in modal logic, and because the filter $\mathcal{F}(w)$ already looks a lot like a neighborhood function $\mathcal{N}$. Still, translating from ELF to neighborhood semantics is not entirely trivial; after all, whether $\mathcal{M}, w \models \Box\varphi$ depends not only of $\mathcal{F}(w)$ but also on $R(w)$, so $\mathcal{F}$ is not exactly the neighborhood function that we are looking for. Instead, given an

ELF model $\mathcal{M} = (W, R, \mathcal{F}, V)$ we find a neighborhood model $\mathcal{M}' = (W, \mathcal{N}, V)$ by taking

$$\mathcal{N}(w) = \begin{cases} \mathcal{F}(w) & \text{if } R(w) \in \mathcal{F}(w) \\ \emptyset & \text{otherwise} \end{cases}$$

Proposition 1. *$\mathcal{M}, w \models \varphi$ if and only if $\mathcal{M}', w \models \varphi$.*

Proof. By induction. As base case, suppose that φ is an atom p. Since $\mathcal{M}$ and $\mathcal{M}'$ have the same valuation, we have $\mathcal{M}, w \models p \Leftrightarrow \mathcal{M}', w \models p$. Suppose then as induction hypothesis that φ is not atomic, and that for every strict subformula ψ of φ we have $\mathcal{M}, w \models \psi \Leftrightarrow \mathcal{M}', w \models \psi$. We continue by case distinction on the main connective of φ.

If the main connective of φ is Boolean, then it follows immediately from the induction hypothesis that $\mathcal{M}, w \models \varphi \Leftrightarrow \mathcal{M}', w \models \varphi$. Suppose then that $\varphi = \Box\psi$. Then

$$\begin{aligned} \mathcal{M}, w \models \Box\psi \Leftrightarrow [\![\psi]\!]_{\mathcal{M}} \cap R(w) \in \mathcal{F}(w) \Leftrightarrow [\![\psi]\!]_{\mathcal{M}} \in \mathcal{F}(w) \text{ and } R(w) \in \mathcal{F}(w) \\ \Leftrightarrow [\![\psi]\!]_{\mathcal{M}} \in \mathcal{N}(w) \Leftrightarrow [\![\psi]\!]_{\mathcal{M}'} \in \mathcal{N}(w) \Leftrightarrow \mathcal{M}', w \models \Box\psi \end{aligned}$$

This completes the case distinction and thereby the induction step. □

Note that while $\mathcal{F}(w)$ is always a filter, $\mathcal{N}(w)$ need not be one. After all, a filter is by definition non-empty, whereas $\mathcal{N}(w)$ is empty whenever $R(w) \notin \mathcal{F}(w)$.

The neighborhood function $\mathcal{N}$ is, in the terminology of [23], consistent, closed under (binary) intersection and closed under supersets. From the fact that $\mathcal{N}$ is consistent, it immediately follows that $\models \neg\Box\bot$, from the fact that $\mathcal{N}$ is closed under intersection it follows that $\models (\Box\varphi \wedge \Box\psi) \to \Box(\varphi \wedge \psi)$ and from the fact that $\mathcal{N}$ is closed under supersets it follows that $\models \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.

We continue by comparing the proof system **WKL**, where we recall Definition 5 on page 8, to other proof systems for non-normal modal logics. We write **WK** for the proof system containing **P**, **K**, **RM** and **MP**. So **WK** is **WKL** minus the axiom **L**.

A *regular modal logic* [8] contains the following axioms and rules.

P	all propositional tautologies
Dual	$\Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$
M	$\Box(\varphi \wedge \psi) \to (\Box\varphi \wedge \Box\psi)$
C	$(\Box\varphi \wedge \Box\psi) \to \Box(\varphi \wedge \psi)$
RM	from $\varphi \to \psi$, infer $\Box\varphi \to \Box\psi$
MP	from $\varphi \to \psi$ and φ, infer ψ

We refer to the proof system containing exactly these six axioms and rules as the minimal regular modal logic **MRML**.

The proof system **WK** is an alternative presentation of a regular modal logic [8, Exercise 8.13a, page 241], i.e., a formula is provable in **MRML** if and only if it is provable in **WK**. It follows that **WKL** is a regular modal logic.

The axiom **L** is not provable in **MRML**. This can, for example, be seen by noting that **MRML** is sound and complete for relational models with impossible worlds (see, e.g., [23]), and that **L** is not valid on those models. So **WKL** is a strictly stronger proof system than **MRML**.

Regular modal logics have been studied quite extensively, see the aforementioned [23] for an overview. But the extension of a regular modal logic with the axiom **L** specifically has not, to the best of our knowledge, been studied before.

6 Conclusion

We have introduced ELF, an epistemic logic that uses filters in order to represent situations where an agent knows (or has a justified belief) that a proposition φ is true even though there are some epistemically accessible worlds where φ is false. We have shown that the proof system **WKL** is sound and complete for ELF. This proof system is similar to **KD**, except that the necessitation rule of that proof system is replaced by a strictly weaker monotonicity rule.

In the basic version of ELF that we discussed in this paper, the properties of truthfulness, positive introspection and negative introspection are not valid. As with normal modal logics, we can enforce these properties by restricting to a smaller class of models. However, unlike normal modal logic, there is not something as elegant and general as correspondence, and there are also additional properties to consider, such as $\Box\top$ (**N**), the dual of our $\neg\Box\bot$ (**L**) axiom. We can then create a sound and complete proof system for ELF on such smaller classes of models by adding a number of axioms to **WKL**. Such technical explorations are relevant, as intuitive scenarios involving certainty and knowledge often satisfy, or fail to satisfy, such constraints. Due to space constraints we must leave the reporting of such frame conditions and axioms for future work.

Acknowledgements

We are grateful to Ramanujam, who suggested the idea of defining knowledge through filters to us. We also thank two anonymous reviewers for their helpful comments. Hans van Ditmarsch is also affiliated to IMSc, Chennai, as associate researcher.

References

1. Aucher, G.: How our beliefs contribute to interpret actions. In: Pechoucek, M., Petta, P., Varga, L. (eds.) Proc. of CEEMAS 2005: Multi-Agent Systems and Applications IV. pp. 276–285. LNCS 3690, Springer (2005)
2. Baltag, A., Bezhanishvili, N., Özgün, A., Smets, S.: Justified belief and the topology of evidence. In: Proc. of WoLLIC. pp. 83–103. Springer (2016), LNCS 9803
3. Baltag, A., Smets, S.: A qualitative theory of dynamic interactive belief revision. In: Proc. of 7th LOFT. pp. 13–60. Texts in Logic and Games 3, Amsterdam University Press (2008)

4. Ben-David, S., Ben-Eliyahu, R.: A modal logic for subjective default reasoning. Artificial Intelligence **116**, 217–236 (2000). https://doi.org/10.1016/S0004-3702(99)00081-8
5. Blackburn, P., van Benthem, J., Wolter, F. (eds.): Handbook of Modal Logic. Elsevier (2006)
6. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, Cambridge (2001), cambridge Tracts in Theoretical Computer Science 53
7. Cartan, H.: Théorie des filtres. Comptes Rendus de l'Académie des Sciences de Paris **205**, 595–598 (1937)
8. Chellas, B.: Modal Logic: An Introduction. Cambridge University Press (1980)
9. Chisholm, R.: Perceiving: A Philosophical Study. Cornell University Press, Ithaca, USA (1957)
10. Conitzer, V.: A puzzle about further facts. Erkenntnis (2018). https://doi.org/10.1007/s10670-018-9979-6, online first
11. DeRose, K.: Contextualism and knowledge attributions. Philosophy and Phenomenological Research **52**, 913–929 (1992). https://doi.org/10.2307/2107917
12. van Ditmarsch, H., Halpern, J., van der Hoek, W., Kooi, B. (eds.): Handbook of epistemic logic. College Publications (2015)
13. Gettier, E.: Is justified true belief knowledge? Analysis **23**, 121–123 (1963). https://doi.org/10.2307/3326922
14. Harman, G.: Thought. Princeton University Press, Princeton, USA (1973)
15. Hintikka, J.: Knowledge and Belief. Cornell University Press (1962)
16. van der Hoek, W.: Systems for knowledge and belief. Journal of Logic and Computation **3**(2), 173–195 (1993). https://doi.org/10.1093/logcom/3.2.173
17. Keisler, H.: Elementary Calculus: an Approach Using Infinitesimals. Prindle Weber & Schmidt (1986)
18. Kraus, S., Lehmann, D.: Knowledge, belief and time. Theoretical Computer Science **58**, 155–174 (1988). https://doi.org/10.1016/0304-3975(88)90024-2
19. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**, 167–207 (1990). https://doi.org/10.1016/0004-3702(90)90101-5
20. Kyburg, H.: Probability and the Logic of Rational Belief. Wesleyan University Press, Middletown, CT (1961)
21. Lewis, D.: Elusive knowledge. Australasian Journal of Philosophy **74**, 549–567 (1996). https://doi.org/10.1080/00048409612347521
22. Nozick, R.: Philosophical Explanations. Harvard University Press, Cambridge, USA (1981)
23. Pacuit, E.: Neighborhood Semantics for Modal Logic. Springer (2017)
24. Pacuit, E., Salame, S.: Majority logic. In: Proc. of Ninth KR. pp. 598–605 (2004)
25. Putnam, H.: Reason, Truth and History. Cambridge University Press, Cambridge (1982)
26. Segerberg, K.: Irrevocable belief revision in dynamic doxastic logic. Notre Dame Journal of Formal Logic **39(3)**, 287–306 (1998). https://doi.org/10.1305/ndjfl/1039182247
27. Stalnaker, R.: On logics of knowledge and belief. Philosophical Studies **128**(1), 169–199 (2005). https://doi.org/10.1007/s11098-005-4062-y
28. Unger, P.: Philosophical relativity. Oxford University Press, New York, USA (1984)