Memo: Intuitonistic epistemic logic and Fitch's knowability paradox

August 28, 2020

We give an introduction into intuitonistic epistemic logic and apply it to Church Fitch's knowability paradox. In a follow-up memo, we will also apply it to fallibilism.

Motivation

Epistemology is the subfield of philosophy concerned with the study of knowledge. At the core are questions like "How to react to sceptic arguments?", "How is our perception related to the outside world?", "Is justified true belief knowledge?". These issues have long been debated, for example the concept of knowledge as justified true belief can already be traced to Plato's dialogue Meno .

A more recent development in epistemology has been the emergence of a subfield named *formal epistemology*, pioneered by Hintakka's influential paper "On knowledge and belief" (Hintikka (1962)). Formal epistemology tries to attack problems from epistemology using formal methods (e.g logic, mathematics, computer science).

In this memo we will give a short introduction into single-agent epistemic propositional logic, give a natural deduction calculus and Kripke-style model semantics for it and apply it to Fitch's knowability paradox following Artemov u. Protopopescu (2016) .

Epistemic Propositional logic

A common motivation for introducing propositional logic in undergraduate philosophy courses is to be able to make a distinction between valid and logically valid arguments, where the former are *just valid*, the latter are by virtue of their form. One possible motivation for developing epistemic propositional logic is then, to have a logic in which arguments of the form "I know that φ, therefore φ" are (logically) valid.

The approach we will present will model single-agent propositional knowledge by introducing a new modal operator K to the language, with the intended semantics that $\mathsf{K}\,\varphi$ is true if and only if the agent knows the proposition φ. We assume a set of propositional variables $P := \{p_i | i \in \mathbb{N}\}$, where every p_i is a fixed proposition. For example p_2 could denote the proposition "All bachelors are unmarried". We can now define the language of epistemic propositional logic with a knowledge operator. [1].

[1] There are developments where an additional operator B for belief is introduced.

Definition 1. *The set of formulas of epistemic propositional knowledge with a knowledge operator $\mathcal{L}_K$ is generated by the following grammar:*

$$\varphi, \psi := \varphi \circ \varphi \mid p_i \mid \neg\varphi \mid \mathsf{K}\,\varphi \quad (i \in \mathbb{N})$$

, where $\circ$ is any binary operator taken from the set $\{\rightarrow, \leftrightarrow, \wedge, \vee\}$.

An intuitionistic conception of knowledge

In an intuitionistic setting, asserting the truth of proposition A, means asserting that there is a verifiable proof for it (Moschovakis (2018)). Under this interpretation a proof of a proposition $A \vee B$ consists of either a proof of A or a proof of B. A proof of an implication $A \rightarrow B$ is a description of a scheme to translate a proof of A into a proof of B [2].

[2] Such a scheme might be given by a computable function, etc.

How does an epistemic operator like K fit into this? If the agent knows A, he must have conclusive evidence to believe that A is true. Under an intuitionistic conception of knowledge a proof of $\mathsf{K}\,A$ therefore is conclusive evidence that A is provable.

Under this reading of the knowledge operator, as proofs are "a special and most strict kind of verification" (Artemov u. Protopopescu (2016)), the co-reflection principle

$$A \rightarrow \mathsf{K}\,A$$

is immediately justified to be valid. [3] However, this conception renders the *knowledge principle* $\mathsf{K}\,A \rightarrow A$ unplausible, since it seems impossible to (constructively) provide a proof given a verification that there exists one. Does an intuitionistic account of knowledge therefore reject the fact that knowledge is factive?

[3] According to Artemov this also fits well with a type-theoretic reading, where asserting that $\mathsf{K}\,A$ is true can be seen as asserting that the type A is inhabited.Brogi (2020) claims to refute this.

There are two parts to the answer: First an example from Artemov u. Protopopescu (2016), why the knowledge principle is unplausible under an intuitonistic conception of knowledge. In computer science, especially in cryptography, there is the concept of zero-knowledge proofs, where one person (the prover) wants to convince another person (the verifier) of a proposition without revealing its proof. For example, Alice might want to convince Bob that she has solved a Sudoku game without revealing the solution. There are zero-knowledge-protocols which solve this problem. But this is some sense a refutation of the knowledge principle in an intuitonistic setting. To see this, consider the formula $\exists x : S(x)$ where $S(x)$ should be true if and only if x is a Sudoku solution. The knowledge principle would give $\mathcal{K}(\exists x : S(x)) \rightarrow \exists x : S(x)$, thus under an intuitonistic interpretation of existential quantification (a witness has to be provided explicitly) the solution would have to be obtainable from the zero-knowledge proof.

Second, rejecting the knowledge principle is not the same as rejecting activity of the knowledge operator, since the knowledge principle is just one of many possible ways to state this property [4]. Note that factivity of knowledge, often phrased as "Everything what is known is true", can also be interpreted as "You can't know false propositions". But knowing nothing false can also be expressed as

[4] Most of the different ways to state it are equivalent classically, but *not* intuitonistically.

$$\mathsf{K}\,A \rightarrow \neg\neg A.$$

If it is known, it is impossible for it to be false. While classically equivalent to the knowledge principle, we opt to accept this principle of *intuitonistic introspection*.

Deduction rules. We have talked about formulas being intuitonistically (or classically) equivalent without having given provability semantics. While one is somewhat used to reasoning within classical logic without explicitly recalling every definition, we have to fix which axioms and principles for provability semantics with regard to epistemic logic to accept.

It seems plausible to extend the axioms of intuitonistic propositional logic by axioms regarding the operator K.

Definition 2 (Natural deduction for IEL)**.**

$$\text{A}\ \frac{\varphi \in \Gamma}{\Gamma \vdash \varphi} \qquad \text{E}\ \frac{\Gamma \vdash \bot}{\Gamma \vdash \varphi} \qquad \text{II}\ \frac{\Gamma \cup \{\varphi\} \vdash \psi \quad \Gamma \vdash \varphi}{\Gamma \vdash (\varphi \to \psi)}$$

$$\text{IE}\ \frac{\Gamma \vdash (\varphi \to \psi) \quad \Gamma \vdash \varphi}{\Gamma \vdash \psi} \qquad \text{KImp}\ \frac{A \vdash \mathsf{K}\,(\varphi \to t)}{A \vdash \mathsf{K}\,\varphi \to \mathsf{K}\,t} \qquad \text{IntRefl}\ \frac{A \vdash s}{A \vdash \mathsf{K}\,s}$$

$$\text{DIL}\ \frac{A \vdash s}{A \vdash s \vee t} \qquad \text{DIR}\ \frac{A \vdash t}{A \vdash s \vee t} \qquad \text{DE}\ \frac{A, s \vdash \psi \quad A, t \vdash \psi \quad A \vdash s \vee t}{A \vdash q}$$

$$\text{CI}\ \frac{A \vdash s \quad A \vdash t}{A \vdash s \wedge t} \qquad \text{CEL}\ \frac{A \vdash s \wedge t}{A \vdash s} \qquad \text{CER}\ \frac{A \vdash s \wedge t}{A \vdash t}$$

Kripke models

While we are now able to use the knowledge operator in our formulas and have taken a look at the intuitonistic approach modeling knowledge as verifiable evidence of provability, we have yet to develop a proof-calculus and semantics for it.

Consider an agent, as an ideal reasoner. He knows, not just a single situation, but considers multiple situations to be possible, these ways the worlds could be are the so-called *possible worlds*. From a perspective of a single possible world, only a subset of all possible worlds might be considered as possible developments.

Standard Kripke models for modal logic with operators for necessity and possibilty consist of a set of possible worlds $\mathcal{W}$, an accessibility relation $\mathcal{R}$ and a valuation $\mathcal{V}_w : P \to \{0, 1\}$ for each world w. This is an adequate representation, if one is using set theory as the basis, as we will formalize this development inside a proof assistant based on type theory, we instead opt to represent the valuation function as a function $\mathcal{V}_w : P \to \mathbb{P}$, mapping a propositional variable to a proposition; with the intended reading that p_i is true at world w if $\mathcal{V}_w(p_i)$ is true , i.e. is provable, proposition.

Truth of formulas can then be defined relative to a world, if a model $\mathcal{M}$ validates a formula φ at a world w, this is denoted as $\mathcal{M}, w \vDash \varphi$. For conjunction and disjunction entailment is easy to define. A formula is possibly true, if there exists (at least) one $\mathcal{R}$-accessible world, where it is true and necessarily true if all $\mathcal{R}$-accessible worlds entail it.

But can knowledge be interpreted in regard to possible worlds? Hintikka (1962) suggests to understand these models as modelling an ideal researcher. This researcher knows a set of *epistemically possible* worlds (just like the possible worlds in the standard modal logic), worlds which he believes could be the actual world. These worlds represent the development she has made while researching, they represent her knowledge at different possible stages (e.g. stages in time). If a world w_1 is related to w_2, it means she would consider w_2 a possible world from the perspective of world w_1.[5] Each world has a valuation function, which maps propositions to truth values. We require, that the valuation is monotonic in respect to the order on possible worlds, since an ideal reasoner, once she has acquired knowledge of a proposition cannot disprove it. This relation on the set of worlds is additionally required to be a preorder.

[5] As there is controversy around wether the term *possible world* (and the possible ontological consequences, if we as Lewis does, adopt the position, that all of them actually exist) it is possible to think of these as possible states of affairs instead.

Satisfaction of formulas can be defined relative to a model and a world. We again use the notation $\mathcal{M}, w \vDash \varphi$ to denote, that φ is true at world w in the model. While the formal definition will be given later, to give an intuition, let's consider the cases for φ being a propositional variable, a conjunction and (last but not least) a formula involving K.

- A propositional variable p is validated by a model and a world w, if and only if $V_w(p) = 1$.
- A conjunction $\varphi \wedge \psi$ is validated if and only if both conjuncts are validated.
- If the formula is $\mathsf{K}\,\phi$, we need an additional idea: We view knowledge of a proposition in an **indistinguashability**-way (Rendsvig u. Symons (2019)). The idea is, that if I cannot conceive a possible world where φ is false, I know that φ is true.

  Thus the reasoner knows $\mathsf{K}\,\varphi$ if at all related worlds φ holds.

If the so-called cognition relation (where two worlds are related if one is conceivable from the standpoint of the other) is the relation we take for checking in the K-clause, we no longer have an intuitonistic conception of knowledge (e.g. the knowledge principle is always valid). Therefore, a second relation, the verification relation is introduced. It is a subset of the cognition relation (this ensures $A \to \mathsf{K}\,A$ is valid), with the additional condition, that if we are in a world w_1 and w_2 is a possible verification world from w_2's perspective, it must also be from w_1's perspective.

Definition 3. *A Kripke-Model for IEL^- is a quadruple $(W, \leq, \leq_v, \mathcal{V})$ consisting of*

- *a set of possible worlds W*
- *a binary cognition preorder $\leq$ on W*
- *a binary validation relation $\leq_v$ on W*

- *a valuation function* $\mathcal{V} : W \to P \to \mathbb{P}$. *We write* $\mathcal{V}_w(p)$ *instead of* $V(w, p)$.

Furthermore the following facts hold:

- $\leq_v \subseteq \leq$
- *Whenever* $u \leq v$ *and* $v \leq_v w$, $u \leq_v w$ *must hold.*
- *The valuation is monotonic in respect to* $\leq$ *i.e.* $w \leq w' \to \forall p, \mathcal{V}_w(p) \to \mathcal{V}_{w'}(p)$.

IEL models have the additional constraint, that there can't be any blind worlds for verification, i.e $\forall w \exists u : w \leq_v u$..

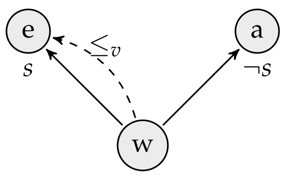
We can now define satisfaction of a formula relative to a model and a world. We use the common notation $\mathcal{M}, w \Vdash \varphi$ to denote that φ is satisfied in $\mathcal{M}$ at world w. The relation $\vDash$ is defined inductively (by induction on the formula):

- $\mathcal{M}, w \vDash p_i ::= \mathcal{V}_w(p_i)$
- $\mathcal{M}, w \vDash \varphi \wedge \psi := M, w \vDash \varphi \wedge \mathcal{M}, w \vDash \psi$
- $\mathcal{M}, w \vDash \varphi \vee \psi := M, w \vDash \varphi \vee \mathcal{M}, w \vDash \psi$
- $\mathcal{M}, w \vDash \mathsf{K}\,\varphi := \forall w \leq_v w' : \mathcal{M}, w' \vDash \varphi$
- $\mathcal{M}, w \vDash \varphi \to \psi := \forall w \leq w' : \mathcal{M}, w' \nvDash \varphi \vee M, w \vDash \psi$
- $\mathcal{M}, w \vDash \varphi \leftrightarrow \psi := \mathcal{M}, w \vDash \varphi \to \psi \wedge \mathcal{M}, w \vDash \varphi \to \psi$

Together with the derivation rules, this system is **sound** and **consistent** for both IEL and IEL^-. A proof for soundness and completeness will be given in a second memo.

We conclude this section on models and the part of this memo introducing IEL by introducing a small example (due to Artemov u. Protopopescu (2016)).

People in Europe believed, before discovering Australia, that all swans are white. We can model this using three worlds: In the above



model e represents Europe, a represents Australia and w the world, from the persepctive of a reasoner in Europe. In this model, $\mathcal{M} \vDash \mathsf{K}\,s$ since $\mathcal{M}, e \vDash s$ and all other worlds are blind. But $\mathcal{M} \nvDash s$, since $\mathcal{M}, a \nvDash s$. While being a valid IEL^--model, it is not a valid IEL-models since there is a blind world.

Fitch's knowability paradox

We know many things. We know how to write letters, we know that the earth is round, we know that bachelors are unmarried young men, we know that $5 + 7 = 12$ and so much more. Yet, we would not expect a single person to be able to know everything. And most of us even believe, that there are truths out there not yet discovered. So our intuition might be, that while every truth can possibly be known, no single persons or even humans collectively can not know all truths [6]. But Fitch's knowledge paradox states the contrary:

[6] The operator K is sometimes read as "there is a person who knows".

> If every truth is knowable, all truths can be known.

The knowability paradox is often framed as an argument against anti-realists and verificationists, who both believe that truth coincides with knowability (Marton (2006)). It was first published by Fitch Fitch (1963), but earlier accounts are attributed to Alonzo Church.

Intuitonistic epistemic logic is not the only solution that has been proposed to the paradox. There are solutions which change the verificationist principle $A \to \Diamond \mathsf{K}\, A$, some other restrict it to a specific set of propositions (e.g. to *cartesian* (Tennet) or *basic* propositions (Dummet)) and some which change the underlying logic.

The paradox can be expressed in bi-modal[7] logic as the formula

[7] A modal logic with operators for necessity ($\Box$), possibility ($\Diamond$) and knowledge (K).

$$(A \to \Diamond \mathsf{K}\, A) \to (A \to \mathsf{K}\, A)$$

which can be proven.

The maybe most interesting part about the proof is, that it does not assume many principles about the knowledge operator (e.g. it does not matter if K is read as "somebody knows" or "an agent knows" and even believing instead of knowing would work). The reason for this can be best seem by considering Fitch's original proof (cf. Fitch (1963)).

He introduces the concept of a *truth-class*, a name for any set of true propositions. If α is a truth-class, the proposition αp is true if and only if $p \in \alpha$.[8] By definition, the implication $\alpha p \to p$ holds.

[8] Note, that the knowledge operator is a truth-class, regardless which interpretation is picked.

A truth-class is called closed under conjunction elimination, if $\alpha(p \wedge q) \vdash \alpha p \wedge \alpha q$ (or alternatively $p \wedge q \in \alpha \implies p \in \alpha \wedge q \in \alpha$) . In a similar sense, call a truth-class *omniscient* if it contains every true proposition and non-omniscient otherwise. We can now state the result in terms of truth-classes:

Lemma 1. *If α is a non-omniscient truth-class closed under conjunction elimination, there is a proposition p, s.t. $p \wedge \neg \alpha p \notin \alpha$.*

Proof. Assume the contrary $p \wedge \neg \alpha p \in \alpha$. Therefore by conjunction elimination, we obtain $\alpha p \wedge \neg \alpha \alpha p$. By applying factivity of truth-classes to the second conjunct, we obtain $\alpha p \wedge \neg \alpha p$, a contradiction. $\Box$

Applying this to knowledge we get the following result, stated as Theorem 5 in Fitch (1963) .

> If there is some true proposition which nobody knows to be true, there is some true proposition which nobody can now to be true.

Fitch's paradox is the contrapositive of the result above: If there is no proposition, which can't be known (if every truth can be known), all true propositions are known. When faced with a result which misaligns with our intuition, we have some options: We can bite the bullet and have to justify why our intuitions are wrong or why the theory is nevertheless useful or we can develop a different theory which might account for the differences. It is a good idea to take a step back and analyze which principles regarding knowledge were used in producing the proof.

1. Classical axioms such as double-negation elimination.
2. Factivity of knowledge (more general of truth-classes).
3. Conjunction elimination.
4. Non-omniscience of knowledge.

Since an intuitonistic conception of knowledge rejects (2) (and (1)), it might be a suitable solution regarding the knowability paradox. The Church-Fitch-paradox rests on understanding

- $A \to \Diamond \mathsf{K}\, A$ as all truths can be known
- $A \to \mathsf{K}\, A$ as all truths are known

but when read under an intuitionistic conception, they are to be read as

- All constructive provable truths can be verified
- proof yields verification

and the paradox is resolved since it can be interpreted as, "If all proofs can be verified, every proof yields verification."

To further support the claim, that the paradox is not paradoxical when using an intuitionistic conception of knowledge, note that the conclusion of the argument is always provable in IEL.

References

[Artemov u. Protopopescu 2016] ARTEMOV, Sergei ; PROTOPOPESCU, Tudor: Intuitionistic epistemic logic. In: *Review of Symbolic Logic* 9 (2016), Nr. 2, S. 266–298. `http://dx.doi.org/10.1017/S1755020315000374`. – DOI 10.1017/S1755020315000374. – ISSN 17550211

[Brogi 2020] BROGI, Cosimo P.: Curry-Howard-Lambek Correspondence for Intuitionistic Belief. (2020), Nr. January, 1–15. `http://arxiv.org/abs/2006.02417`

[Fitch 1963] FITCH, Frederic B.: A Logical Analysis of Some Value Concepts. In: *The Journal of Symbolic Logic* 28 (1963), Nr. 2, S. 135–142

[Hintikka 1962] HINTIKKA, Jaakko: *Knowledge and Belief*. https://philpapers.org/rec/HINKAB. Version: 1962

[Marton 2006] MARTON, Peter: Verificationists versus realists: The battle over knowability. In: *Synthese* 151 (2006), Nr. 1, S. 81–98. http://dx.doi.org/10.1007/s11229-004-6269-4. – DOI 10.1007/s11229–004–6269–4. – ISSN 00397857

[Moschovakis 2018] MOSCHOVAKIS, Joan: Intuitionistic Logic. In: ZALTA, Edward N. (Hrsg.): *The Stanford Encyclopedia of Philosophy*. Winter 201. Metaphysics Research Lab, Stanford University, 2018

[Rendsvig u. Symons 2019] RENDSVIG, Rasmus ; SYMONS, John: Epistemic Logic. In: ZALTA, Edward N. (Hrsg.): *The Stanford Encyclopedia of Philosophy*. Summer 201. Metaphysics Research Lab, Stanford University, 2019