Memo: Arguments against $A \to \mathsf{K}\,A$

March 16, 2021

The trivial solution to Church-Fitch

If an intuitionistic logic of knowledge validates the co-reflection principle ($A \to \mathsf{K}\,A$) the knowledge paradox - the inference from $A \to \Diamond\mathsf{K}\,A$ to $A \to \mathsf{K}\,A$ - is seemingly trivially solved, since the arguments conclusion, which on a classical reading states omniscience , now can be read as *proofs can be checked*.

However there is the issue what the intuitionist actually means by *proof*. For example if proofs are seen as platonic objects, that is for a statement to be true, we only require that a proof exists, regardless of its actual construction. But when proofs must only exist as platonic objects, $A \to \mathsf{K}\,A$ is no longer justified, since for A to be true, the proof π must no longer be actually constructed but only exist, which makes proof-checking implausible (Murzi 2010).

If proofs are mental constructions, the question is wether they are types or tokens (Williamson 1988). The high-level argument made in Murzi (2010) is, that both lead to different problems. If proofs are types, $A \to \mathsf{K}\,A$ is not valid and while co-reflection is valid when proofs are tokens, this leads to truth having a temporal component, against which there are objections. We will first outline Williamson's proposal of proofs as types.

Proofs as types

Williamson introduces proof types as a *ontologically neutral* (Williamson 1988) concept: Two proof tokens are of same proof-type if they have the same structure and conclusion, but may occur at different times (we shall denote similiarity of proof tokens by $\sim$). Proof types are under this conception nothing more than proof tokens grouped together by sameness (like equivalence classes). For example, two similar proofs of the pythagorean theorem carried out at different times would be different proof tokens of the same type, where different proofs of the pythagorean theorem would count both as different proof types and tokens. There is no such thing as a proof type of the Pythagorean theorem (Murzi 2010) , unless all proofs of the Pythagorean theorem where similar in structure.

Talk of proof types can always be reduced to talk of similar proof tokens. This also applies to the BHK-semantics, consider the case of implication. The usual explanation is, that a proof of $A \to B$ is a function (in the sense of a procedure or construction) transforming proofs of A into proofs of B. So if proofs are types, a function mapping proof types to proof types, how can this be reduced to proof tokens? Williamson (1988) suggests to interpret the conditional as a **unitype** function between proof tokens, that is a function which preserves similiarity of tokens i.e. if p_1 and p_2 are similar, so are $f(p_1)$

and $f(p_2)$. Formally a function f between proof tokens is unitype iff

$$\forall \pi, \rho : \pi \sim \rho \implies f(\pi) \sim f(\rho).$$

His objection against $A \to \mathsf{K}\, A$ argues that there is no unitype function mapping proof tokens of A to proof-tokens of $\mathsf{K}\, A$. He interprets a proof of $\mathsf{K}\, A$ as a proof that there exists a time t at which A has been proven. Under this interpretation the function f can't be unitype: Consider a statement p with two proofs π_1, π_2 proven at different times i.e. $t(\pi_1) \neq t(\pi_2)$. Under Williamsons reading of the K operator, from both π_1 and π_2 a proof of $\mathsf{K}\, p$ can be obtained, however the resulting proofs are not similar, since one asserts that p has been proven at $t(p_1)$ while the other asserts that p has been proven at $t(p_2)$. But since π_1 and π_2 are similar, but $f(\pi_1)$ and $f(\pi_2)$ are not, f can not be unitype.

However this argument hinges on the K-operator introducing a time component, on the time being a part of the proof of $\mathsf{K}\, A$. As Usberti 2016 and Artemov and Protopopescu 2016 correctly observe, as soon as the temporal reference is dropped from the reading of the K-operator e.g. $\mathsf{K}\, p$ is read as p has been proven or as conclusive evidence that p has been proven, the objection no longer works.

Other counterarguments / objections

Forever unknwown truths

A second argument Murzi (2010) brings up, is that even defenders of strict anti-realism, might have to accept the existence of forever unknown truth e.g. a statement of the form $p \wedge \neg \mathsf{K}\, p$, which is impossible in IEL.

Consider a statement which is decidable (i.e. $p \vee \neg p$) but all evidence for deciding it has been lost - the example in the literature (Dummet, Reply to Wolfgang Künne) being the number of hairs on Dummets head at a specific time and date. The evidence being lost, in IEL terms, would mean that no concrete verifiable evidence of either p or $\neg p$ exists, i.e. $\neg \mathsf{K}\, p \wedge \neg \mathsf{K} \neg p$. Now it would be possible to prove $(p \wedge \neg \mathsf{K}\, p) \vee (\neg p \wedge \neg \mathsf{K} \neg p)$.

I guess the weakest point of this argument is the inference of $p \vee \neg p$ from the fact that there existed a decision procedure at the time. Dummet (the argument is presented in Murzi 2010) rejects the inference by analyzing it using counterfactuals. According to Dummet , when asserting that p is decidable, you really assert a counterfactual, that if Dummet's hairs had been counted, they would have found to be even or odd. This can be expressed as $\chi \;\Box\!\!\to\; (p \vee \neg p)$ where χ is the proposition that the hairs have been counted and $\Box\!\!\to$ is the counterfactual conditional. But from this it is impossible to obtain $(\chi \;\Box\!\!\to\; p) \vee (\chi \;\Box\!\!\to\; \neg p)$, which would be needed to assert the disjunction.

To see that this inference is wrong, consider the usual definition of $\Box\!\!\to$ as strict implication i.e. $\phi \;\Box\!\!\to\; \psi :\iff \Box(\phi \to \psi)$. Now

$\chi \mathrel{\Box\!\!\rightarrow} (p \vee \neg p)$ only asserts that at every world where χ is true, either p or $\neg p$ are true; however for asserting $\chi \mathrel{\Box\!\!\rightarrow} p \vee \psi \mathrel{\Box\!\!\rightarrow} \neg p$, either at every world where χ is true p is true or at every world where χ is true $\neg p$ is true (i.e. the first formula is valid in a model where at some worlds only χ, p and at other worlds only $\chi, \neg p$ are true - while the second formula isn't.

Conclusion

While the objection against viewing proofs as types seems to work well with Dummet's definition of the K-operator - that objection against co-reflection seems to fail with IEL as correctly observed by Artemov.

The argument that anti-realists are committed to the existence of forever unknown truths can be criticised too. In summary, the considered objections against an intuitionistic solution to the knowability paradox, which criticize co-reflection in an intuitionistic setting do not seem forceful.

Future goals / todo:

1. Are there other problems with identifying proofs with proof-types?
2. Argument from Murzi - not yet considered: Paradox of Idealization

References

Artemov, Sergei, and Tudor Protopopescu. 2016. "Intuitionistic epistemic logic". *Review of Symbolic Logic* 9 (2): 266–298. ISSN: 17550211. doi:`10.1017/S1755020315000374`. arXiv: `1406.1582`.

Murzi, Julien. 2010. "Knowability and bivalence: Intuitionistic solutions to the Paradox of Knowability". *Philosophical Studies.* ISSN: 00318116. doi:`10.1007/s11098-009-9349-y`.

Usberti, Gabriele. 2016. "The Paradox of Knowability from an Intuitionistic Standpoint". doi:`10.1007/978-3-319-22686-6_7`.

Williamson, Timothy. 1988. "Knowability and constructivism". *Philosophical Quarterly.* ISSN: 14679213. doi:`10.2307/2219707`.