A purely epistemological version of Fitch's Paradox

Samuel A. Alexander*
Department of Mathematics, the Ohio State University

Published in *The Reasoner* vol. 6, issue 4, pp. 59–60.

The *knowability thesis* is the idea that every truth is knowable at least in principle: $\phi \rightarrow \diamond K(\phi)$ (I will always implicitly quantify ϕ and ψ over the set of all formulas, except within deductions, where they stand for fixed formulas). Fitch's Paradox is the fact that, along with other basic epistemic assumptions, the knowability thesis implies the *omniscience principle*, $\phi \rightarrow K(\phi)$. Though the knowability thesis seems reasonable, the omniscience principle is absurd. This seems a devastating blow against anti-realism. As for the "other assumptions," there is infinite variation, to the point that almost all the major papers employ slightly different assumptions. Still, the Church-Fitch argument does not fundamentally change, and can always be glossed as follows ($*$):

1. Formally verify (using the "other assumptions") the absurdity of Moore's Paradox (usually because $K(\phi \wedge \neg K\phi)$ implies $K(\phi)$ and $\neg K(\phi)$).

2. Conclude $\neg \diamond K(\phi \wedge \neg K\phi)$.

3. If $\phi \wedge \neg K\phi$ is true, then by the knowability thesis, $\diamond K(\phi \wedge \neg K\phi)$.

4. Therefore, $\phi \wedge \neg K\phi$ can not be true. So $\phi \rightarrow K(\phi)$.

By the *weak omniscience principle* I mean the schema $\phi \rightarrow K(K(\phi))$, and by the *purely epistemic knowability thesis* I mean the schema $\phi \rightarrow \neg K(\neg K(\phi))$. The latter implies the former by the standard Fitch's Paradox, given the usual other assumptions. My aim is to show that the implication holds given a more barren set of other assumptions, by an argument which is qualitatively different than ($*$). The plausibility of purely epistemic knowability will be discussed below.

We make the following assumptions:

- $\wedge$: $K(\phi \wedge \psi) \rightarrow K(\phi) \wedge K(\psi)$.

- Purely Epistemic Knowability (PEK): $\phi \rightarrow \neg K(\neg K(\phi))$.

- Rule of Necessitation: From a deduction of ϕ, we may deduce $K(\phi)$.

*Email: alexander@math.ohio-state.edu

From these we deduce weak omniscience as follows.

1. Assume $K(\phi \wedge \neg K(K(\phi)))$.
2. By $\wedge$, $K(\phi)$ and $K(\neg K(K(\phi)))$.
3. By PEK applied to $K(\phi)$, we have $\neg K(\neg K(K(\phi)))$.
4. Contradiction. Discharge 1 and conclude $\neg K(\phi \wedge \neg K(K(\phi)))$.
5. By Rule of Necessitation, conclude $K(\neg K(\phi \wedge \neg K(K(\phi))))$.
6. Assuming $\phi \wedge \neg K(K(\phi))$, we would have $\neg K(\neg K(\phi \wedge \neg K(K(\phi))))$ by KEP. That would contradict 5, so $\neg(\phi \wedge \neg K(K(\phi))$, or equivalently, $\phi \rightarrow K(K(\phi))$.

This argument is qualitatively different for four reasons. First, it factors through a *weak* Moore's paradox: "It's raining, and I don't know that I know it's raining." Second, the Moore contradiction is not obtained by stripping away modal operators (which seems impossible without additional assumptions) but rather by piling new modal operators on! Third, it never directly uses any consistency assumption on K, neither $K(\phi) \rightarrow \phi$ nor even the weaker $\neg(K(\phi) \wedge K(\neg\phi))$. Finally, it makes no use of modalities of possibility or necessity.

If we assume that all necessities are known (contrapositively, all unknowns are unnecessary) then it follows that PEK is stronger than the usual knowability thesis:

$$\neg K(\neg K(\phi)) \rightarrow \neg\Box(\neg K(\phi)) \text{ and } \neg\Box(\neg K(\phi)) \rightarrow \diamond(K(\phi)).$$

In a sense, PEK is the polar opposite of the *negative introspection axiom* (sometimes called **5**), which says $\neg K(\phi) \rightarrow K(\neg K(\phi))$. Is an assumption like PEK really plausible? Not in the contexts where Fitch's Paradox is normally discussed (human knowledge), so this note is at best an interesting curiosity in the bigger picture of Fitch's Paradox. But Purely Epistemic Knowability is somewhat plausible in the area of machine knowledge. A machine can be programmed to mechanically "know" formulas in a crude epistemic language (too weak, say, for Kurt Gödel's incompleteness theorems), and its knowledge can be closed under modus ponens and can include various epistemic axioms and rules, and furthermore it can extend to include contingent facts such as "the fifth user-input is a 0". We speak, as always, of idealized knowledge: the machine cannot reason that "I do not know that the fifth user-input is 0 because I haven't received five inputs yet": it has no way of talking about how many inputs it has received (the machine may be what socket programmers call *blocking*: if its programming instructs it to query the nth input before that input is received, the machine freezes until receiving the input). Neither can the machine deduce later inputs based solely on earlier ones, assuming the user has free will. The only way for the machine to conclude "I do not know that the fifth user-input is 0" is for the machine to observe, say, that the fifth user-input is 1. But this is all just a very drawn out way of articulating the Purely Epistemic Knowability thesis.