The Knowability Paradox: does logic come before metaphysics?

Giulia Casasole∗ Martina Orlandi†

Introduction

The legendary Theorem 5

The Knowability Paradox is a logical argument which states that if all truths are knowable, then all truths are actually known. In 1963 Frederich Fitch published ‘A Logical Analysis of Some Value Concepts’[3]. This brief article appeared on the *Journal of Symbolic Logic* and it immediately became a classic of philosophical logic. It is in this paper that Fitch presented the Knowability Paradox, whose contrapositive asserts that if there is an unknown truth, then there is a truth that can’t possibly be known. Fitch’s proof has been used to criticize the anti-realist position according to all the truths are knowable. Because we are not omniscient, individually and collectively, we are committed to conclude that it’s false that all truths are knowable.

(Theorem 5)

\[
\exists p(p \land \neg Kp) \vdash \exists p(p \land \neg \Box Kp)
\]

The famous Theorem 5 threatens the collapse of modal and epistemic dif-
ferences because if we take ignorance to be the cause of the fact that we don’t know some truths, then Theorem 5 turns a contingent ignorance into a necessary ignorance and shows that the existence of truths actually unknown implies the existence of truths necessarily unknown.

The contrapositive of Theorem 5 is usually referred to as the paradox.

(Knowability Paradox)

\[ \forall p (p \rightarrow \Diamond Kp) \vdash \forall p (p \rightarrow Kp) \]

The Knowability Paradox is said to prove that the anti-realist conception is unjustifiable. We just said that because we are not omniscient we are forced to admit that it is false that every truth is knowable. But the latter thesis is precisely what the anti-realist argues for. The anti-realist claims that reality depends on our epistemic abilities so it is reasonable to think that assertions which describe it correctly are knowable. If the proof built by Fitch is valid, then anti-realism is a philosophically unsustainable position.

But is this really the problem? Is it of a metaphysical or a logical nature?

According to Johnathan Kvanvig [4], concluding that it is false that all truths are known is not the real problem posed by Fitch. If that were the main problem, Fitch’s proof would only represents a simple counterexample to the thesis that all truths are knowable, but would not constitute a paradox.

The real problem posed by Fitch - Kvanvig says - is the inference from ‘All the truths are knowable’ to ‘All truths are known’. Truth implies possibility so the converse conditional is trivial.

Thus, the equivalence would follow:

(K) All truths are knowable if and only if all truths are known.

This result is clearly paradoxical. Creating an equivalence between the two
notions of ‘universally knowable truth’ and ‘universally known truth’, it would
destroy a modal distinction between notions intuitively different.

So the real problem posed by Fitch is firstly logical and secondly metaphys-
cical. If so, one wonders whether the rules implied in the proof are correct and, if
they are not, wether the solutions presented in these years are conclusive. But
before talking about this, let us briefly present the paradox in a formal setting.

The Paradox

Fitch’s proof involves quantifying over sentences. Our propositionals variables
$p$ and $q$ will take declarative statements as values. Let $K$ be the epistemic operator
‘it is known by someone at some time that’ and $\Diamond$ the modal operator ‘it is
possible that’. Let’s assume the knowability principle (KP) that all truths are
knowable by somebody at some time.

\[(KP) \quad \forall p(p \rightarrow \Diamond Kp)\]

It seems reasonable to accept that we are not omniscient either individually
or collectively, and so there is a truth that we don’t know:

\[(NonO) \quad \exists p(p \land \neg Kp)\]

If this claim is true, it has a true instance:

\[p \land \neg Kp \quad (1)\]

Now let’s consider the instance of (KP) obtained by substituting $p \land \neg Kp$
for the variable $p$:

\[(p \land \neg Kp) \rightarrow \Diamond K(p \land \neg Kp) \quad (2)\]
So it’s possible to know by *modus ponens*:

\[ \diamond K(p \land \neg Kp) \] (3)

But this is plainly false and we can prove it with the help of two very modest epistemic principles.

1. Knowing a conjunction implies knowing each of the conjuncts:

   (A) \[ K(p \land q) \vdash Kp \land Kq \]

2. Knowledge entails truth.

   (B) \[ Kp \vdash p \]

We will also need two simple modal principles.

3. The necessitation rule:

   (C) If \[ \vdash p \] then \[ \vdash \Box p \]

4. The interdefinability of \( \Box \) and \( \diamond \):

   (D) \[ \Box \neg p \iff \neg \diamond p \]

*Proof.*

\[
\begin{align*}
   K(p \land \neg Kp) & \quad \text{Assumption} \quad (4) \\
   Kp \land K\neg Kp & \quad \text{from (4), (A)} \quad (5) \\
   Kp \land \neg Kp & \quad \text{from (5), (B) to the right conjunct} \quad (6) \\
   \neg K(p \land \neg Kp) & \quad \text{from (4)-(6) by RAA, discharging (4)} \quad (7) \\
   \Box \neg K(p \land \neg Kp) & \quad \text{from (7), (C)} \quad (8) \\
   \neg \diamond K(p \land \neg Kp) & \quad \text{from (8), (D)} \quad (9) \\
\end{align*}
\]

\[\square\]
(9) contradicts (3). So, from (KP) and (NonO) a contradiction follows. It follows that ‘All truths are knowable’ is false, so we are non-omniscient.

\[ \neg \exists p (p \land \neg Kp) \quad (10) \]

It follows that all truths are actually known:

\[ \forall p (p \rightarrow Kp) \quad (11) \]

So who states that all truths are knowable by somebody at sometime is forced to conclude that every truth is effectively known by somebody.

**Possible solutions**

**The Intuitionistic Proposal**

Williamson states that Fitch’s result is not a refutation of anti-realism, but rather a reason for the anti-realist to accept intuitionistic logic. Intuitionistic logic provides a sort of solution of the Knowability Paradox by restricting the effect of epistemic rules. While the elimination of double negation is valid in classical logic, it is not in intuitionistic logic:

\[ \neg \neg p \vdash p \]

\[ \neg \forall P[x] \vdash \exists x \neg P[x] \]

Without the rule of double negation we cannot derive the conclusion \( \forall p (p \rightarrow Kp) \) from \( \neg \exists p (p \land \neg Kp) \)

More precisely:
\[\neg\exists p(p \land \neg Kp)\]

\[\forall p\neg(p \land \neg Kp)\]

\[\forall p(\neg p \lor \neg\neg Kp) \quad \text{(De Morgan)}\]

\[\forall p(\neg p \lor Kp)\]

So if we consider \(\neg\exists p(p \land \neg Kp)\) we can derive \(\forall p\neg(p \land \neg Kp)\), but without the elimination of double negation \(\neg(p \land \neg Kp)\) does not entail \(p \rightarrow Kp\).

**Proof.** Let’s suppose \(\neg(p \land \neg Kp)\) and suppose that \(p\) is true while \(Kp\) is false. Then \(\neg Kp\) is true, so \(p \land \neg Kp\) is true. And we have a contradiction. So \(\neg Kp\) is false and \(\neg\neg Kp\) is true. Then \(p \rightarrow \neg\neg Kp\) is true, but not \(p \rightarrow Kp\).

Why the intuitionistic proposal is not a solution

Williamson in [10], [11] argues that the intuitionist anti-realist must accept the assertion that \(\neg\exists p(p \land \neg Kp)\). But the thesis for which no truth is unknown is hard to swallow as well because the anti-realist cannot provide a proof for the non-omniscience thesis. Williamson replies that the intuitionist anti-realist may naturally express our non-omniscience as “not all truths are known”: \(\neg\forall p(p \rightarrow Kp)\) which is classically, but not intuitionistically, equivalent to the non-omniscience thesis, \(\exists p(p \land \neg Kp)\). That is because in intuitionistic logic the quantifier exchange rule, \(\neg\forall x P[x] \vdash \exists x\neg P[x]\), is not unrestrictedly valid. The expression of non-omniscience, \(\neg\forall p(p \rightarrow Kp)\), is only classically, and not intuitionistically, inconsistent with \(\neg\exists p(p \land \neg Kp)\). So the intuitionist anti-realist can consistently give expression to the truth that we are non-omniscient while
accepting the intuitionistic consequence derived at $\neg \exists p(p \land \neg Kp)$ admitting both that no truths are unknown and that not all truths are known.

**The Paraconsistent Proposal**

We have seen the problems of the intuitionistic revision and wh it cannot be a good way out to the paradox. But what if the intuitionistic logic is not the right one? What if the correct logic of knowability is paraconsistent?

A paraconsistent logic is a formal system which can deal with contradictions, since it rejects the ex absurdo quodlibet law, also known as the *explosion principle*. This inference rule, which is valid both classically and intuitionistically, allows us to derive the conclusion of an arbitrary statement $r$, once a contradiction has been asserted. Indeed a paraconsistent logic is inconsistency-tolerant and treats contradictions as, not only possible, but also informative. The proposal put forward by Routley [6] and Beall [1] does not trivialize the theory, quite the opposite, it describes perfectly the status of human knowledge, which is likely to be inconsistent.

In fact transposing Fitch’s reasoning in a paraconsistent logic would make it not paradoxical at all, and anti-realism could be saved, just by controlling the inferential power of the deductive system. This can be easily achieved by rejecting some inference rules used in Fitch’s proof (Page 4), like the *reductio ad absurdum* (Line 7), or the necessitation rule (Line 8), depending on the choice of the paraconsistent system used.

Beall succeeds in finding an independent evidence for thinking about the existence of true epistemic contradictions, with a strategy taken from a dialetheist point of view. The strength of his argument is provided by the lack of solutions to the *Knower paradox*, a self-referential semantic paradox which goes under the more general field of the anti-expertise paradoxes.
The heart of the reasoning is a self-referential sentence $K$ such that

($K$) No one knows this very sentence.

Let us assume that $K$ is known, for the sake of argument, and presume that knowledge entails truth. So $K$ is true, but $K$ says that $K$, itself, is unknown. So $K$ is both known and unknown. Then we have a proof which forces us to admit the falsehood of the former assumption. So it is known that $K$ is unknown, which is equivalent to say that it is known that $K$. So the assumption is proven to be true, and $K$ is both known and unknown.

Beall suggests that in the Knower paradox lies an instance for thinking that $K(p \land \neg Kp)$. This could be only accepted in a paraconsistent logic without triviality, and - without a proper reply to the Knower paradox - Fitch’s proof is ineffective against the Knowability principle (KP). We may conclude saying that nevertheless a paraconsistent logic constitutes a system weaker than the classical modal or intuitionistic ones, hence free from the danger of the knowability paradox.

The Metaphysics behind the logic of the Knowability Paradox

Now we are going to present some reflections raised by Fitch’s reasoning, briefly and with no pretense of exhaustiveness. Particularly, the metaphysical debate which involves realism and anti-realism, or verificationism, which whose roots date back to the origin of philosophy.

If reality depends on our epistemic skills, it is conceivable that the propositions which describe it correctly should be knowable. This is the centerpiece of the anti-realist position and it cannot be accepted in the light of Fitch’s paradox. Since there are actual unknown truths, there are unknowable truths. This
argument doesn’t work only in two cases.  

First, we can embrace a theist point of view and believe that there exists an omniscient being. In doing so we can accept the knowability conditional (2) as vacuously true, as the antecedent is false, and the paradox will not occur.  

The other way out is to admit that while all truths are knowable, there are some actual unknown truths. The latter position seems not to be so far from common sense. The basic notion for which we know that there are things we don’t know is not considered a paradox in the usual interpretation of natural language. If we adopt a paraconsistent logic we can plainly accept inconsistent states of reality.  

Behind this choice lie motivations of different kind.  

A forceful reason is that Dialetheism can support paraconsistent logic. Since if there are sentences such that both they and their negation are true, dialetheias, like in the case of the liar and the knower paradoxes, the logic behind knowability must be inconsistency-tolerant and non explosive. Of course there are argument against the Dialetheism, but we are not going to examine them in this circumstance, since our intention is just to grasp the meaning and the philosophical consequences of the knowability paradox.  

At the end of the day, if there is a logic and a metaphysical view that can support it, who can trustfully assert that is a paradox?  

Conclusion  

What can logic say about reality?  

If a paradox is determined by the surprise that it generates, then the knowability paradox can be surely considered that way. We said that the most paradoxical element of KP is the identification of a modal and non-modal claim.  

\[ p \rightarrow \Diamond Kp \]  

(12)
It’s trivial to notice that (12) logically implies (13) but the paradoxicality is due to the fact that the proof leads us to the converse implication as well, hence to the unbelievable conclusion:

\[ \vdash (p \rightarrow \Diamond Kp) \leftrightarrow (p \rightarrow Kp) \]

But why is this conclusion so surprising?

The history has related the problem of Fitch’s proof to the fact that if it is sound, then a particular version of anti-realism is false. According to Kvanvig this is not the point, though there may be good reasons to think he’s wrong. In \[4\] Kvanvig says that “what is paradoxical here is not tied in any way at all to a commitment to anti-realism. [...] The logical result in question requires no anti-realist commitments to prove, since it is a theorem; and denying anti-realist commitments does not eliminate whatever paradoxicality is involved in the equivalence formulated.”\[4\]

Basically the real issue is that the paradox is not a logical local problem, but a global one which involves everything and not only the restricted context of anti-realism. It’s a problem which deals with the collapse of the logical distinction, between actuality and possibility. So, the paradoxical core of the proof is the fact that this distinction which is deeply rooted in our conception of things, simply disappears. That is to say that “[...] this logical equivalence threatens some of our deepest modal conventions.”\[4\]

In our opinion the main problem is that if we don’t want to abandon the anti-realist conception and keep on accepting epistemic logic as the correct modeling of human reasoning, then we cannot be anti-realist anymore.

But the question is: is this conclusion inescapable? What can logic say about
reality? If the KP entails that anti-realism is false, does this necessarily mean that it’s really false?

Apparently logic doesn’t say anything about reality. In fact, we are the creators of formal systems, as well as the axioms and the inference rules. The problem is that unfortunately (or not so much after all) we often use formalization as a schematization of certain phenomena in reality. In the case of the modeling of epistemic reasoning, our inconvenient is that we don’t deal with a concrete phenomenon but with an implicit and subjective one, and very difficult to describe.

The assertion that the KP doesn’t concern the local problem related to anti-realism is not fully justified. Anti-realism is a metaphysical position, but it is connected to a precise vision of the world which involves the whole reality.

Unfortunately KP seems to be a situation in which we must give up something: if we accept epistemic logic we cannot be anti-realist. But what if we don’t want to quit anti-realism? In this case, we must conclude that epistemic logic is not a good model of reasoning. The next step is that we are forced to choose another formalization where the paradox doesn’t occur. In this way we don’t have the paradox, even though we are not be able to claim that anti-realism can be supported. We just know that another formal system doesn’t support its confutation, and this result doesn’t improve our information about the problem.

Another point to emphasize is that the absence of the paradox in another formal system doesn’t mean that it is solvable. That is, is the epistemic logic inadequate or is the paradox unresolvable? As we see it, its dismissal in another formal system is not a real solution, but just a trick.
References


