ABSTRACT. In this paper I take a closer look at a recently published paradox by Göran Sundholm involving the notion of knowability. I point out that this paradox is not a novel, genuine paradox, but rather an important variant of the Knower Paradox. I briefly discuss further variations of the Knower Paradox, and in a final section I try to show that it is not unproblematic to assume that knowability is factive. There are several different notions and conceptions of knowability. The most straightforward ones are non-factive. Anti-realists as Sundholm and Tennant obviously use a different notion of knowability which they assume to be factive. But this notion has not yet been made sufficiently clear to rely on it as common ground in debates between realists and anti-realists.

1 Another Knowability Paradox

The modern debates between realists and anti-realists are often centred on the notion of knowability. Typical anti-realists defend an equivalence claim between truth and knowability, whereas realists deny such a thesis. In recent years one focus of this debate has been the so-called Knowability Paradox or Church-Fitch Paradox, and the literature on it has exploded. This paradox consists in a formal derivation that seems to show that the anti-realistic knowability principle ‘Every truth is knowable’ implies the seemingly much stronger claim that every truth is actually known, a thesis unacceptable even for all but the hardest anti-realists.

Now, in a short note Göran Sundholm (2008) has presented another paradox

---

1 Originally the argument was published in a paper by Frederic Fitch (1963), who attributed it to an anonymous referee. So, for a long time the paradox was named after Fitch, until Joe Salerno recently found out that the anonymous referee was Alonzo Church and proposed the name Church-Fitch Paradox. Cf. Salerno (2009b).

2 For an overview see the online entry in the Stanford Encyclopedia of Philosophy by Berit Brogaard and Joe Salerno, and for a recent collection of papers see Salerno (2009a). Even a monograph has been written on this subject: Kvanvig (2006). I have contributed to this debate also myself, Rückert (2004).
that essentially involves the notion of knowability. Here, Sundholm’s argument:  

“We consider:

(⋆) This sentence is unknowable.
1 Assume that (⋆) can be known. Assumption

Then

2 (⋆) is true.

whence

3 This sentence cannot be known. From (⋆) and 2; T-Schema
4 Contradiction. From 1 and 3

Therefore

5 (⋆) cannot be known. From 1 and 2, without assumptions

That is

6 This sentence is unknowable. (⋆) is demonstrated on no assumptions

Therefore

7 (⋆) is known. From 7, what is demonstrated is known
8 Contradiction. From 5 and 7 on no assumptions.”

Before moving on, I’d like to point out three decisive features of this argument:

Sundholm does not seem to be worried by the original knowability paradox at all, because he thinks that it does not apply to his own conception of knowability, but he is rather silent about why he thinks that it does not:

“Personally, even though I admit of constructivist tendencies, I have never felt in the slightest threatened by the Church/Fitch reasoning. To my mind, its rendering of the knowability of truth simply does no justice to what is involved in constructivist knowability.” (Sundholm (2008), pp.375-376).

a) *Self-reference.* The starting point of the paradox, the sentence ‘This sentence is unknowable,’ is self-referential, it speaks about itself, claiming that this very sentence is not knowable. As remarked by Sundholm in a footnote,⁵ instead of using an indexical formulation with ‘this,’ one could also start the paradox with ‘$S \overset{\text{def}}{=} \text{cannot be known}’. The diagonal lemma guarantees that there is such a sentence.

b) *Factivity of knowability.* The step from (1) to (2) in the reasoning uses factivity of knowability, i.e. that every knowable sentence is also true. So, Sundholm assumes without any further comment that knowability, like knowledge, is a factive notion. I’ll come back to this in section 5.

c) *Necessitation.* Here, by necessitation I mean the transition from a sentence having been demonstrated or proved to the sentence being known, and as every actually known sentence is *a fortiori* also knowable, to the sentence being knowable (cf. (7) and (8) in Sundholm’s derivation).

2 Naming the Paradox

Sundholm begins and ends his short note with comments about the naming of paradoxes. At the beginning of his text he says:

“The naming of paradoxes is usually a straight-forward matter: they are named after their inventors.”⁶

And he concludes his text with the following footnote:

“Only the delicate question now remains: by what name should one call this novel (?) paradox . . .?”⁷

Of course, taken together these two passages suggest to name the paradox that I have presented in section 1 after Sundholm himself, and as it is a paradox that involves the notion of knowability, I propose to call it Sundholm’s Paradox of Knowability from now on.

3 A Novel Paradox?

But, the naming of the paradox certainly wasn’t the most ‘delicate’ question anyway.⁸ With the title of his note, “A Novel Paradox?” and, indirectly, with his last sentence (note the question mark after “novel”!) Sundholm formulates a more important question: Is Sundholm’s Paradox of Knowability really a new, genuine paradox, or is it rather a mere variant of another paradox.

---

⁸Sundholm’s use of ‘delicate’ in the quote above is at least partly ironic, of course.
Maybe Sundholm himself considered his own paradox to be a variant of the Liar. Or better, a variant of the Strengthened Liar. This is suggested by his second footnote:

“From a constructivist point of view this is very close to the Liar since, in some suitable version or other, truth is nothing but knowability.”

Indeed, for philosophers with constructivist views who equate truth with knowability or replace truth by knowability, Sundholm’s Paradox is a variant of the Liar paradox. On the other hand, the derivation of Sundholm’s Paradox is slightly more complex than the one of the Liar.

Although, in his paper, Sundholm mentions more than a dozen of other paradoxes, it is surprising that the most pertinent other paradox is missing: the so-called Knower Paradox which can be attributed to Montague, and which itself is a generalisation of the Liar Paradox. Here, the most simple version of the Knower Paradox:

Let $T$ be a theory extending Robinson Arithmetic, and $K$ a unary predicate of $L(T)$ that is factive $(K\phi \rightarrow \phi)$ and satisfies necessitation (if $T \vdash \phi$ then $T \vdash K\phi$). Then $T$ is inconsistent.

Proof.

1. $T \vdash \phi \rightarrow \neg K\phi$ [There is such a $\phi$ because of the diagonal lemma]
2. $T \vdash K\phi \rightarrow \phi$ [Factivity of $K$]
3. $T \vdash K\phi \rightarrow \neg K\phi$ [By (1) and (2)]
4. $T \vdash \neg K\phi$ [By (3)]
5. $T \vdash \phi$ [By (1) and (4)]
6. $T \vdash K\phi$ [By (5) and Necessitation for $K$]
7. Contradiction [Between (4) and (7)]

This paradox is called the Knower Paradox because $K$ might well be the knowledge predicate, but it need not be. Any predicate that satisfies Factivity and Necessitation will do. Égré (2005), p.18, calls factive predicates ‘knowledge predicates.’ This might be a bit misleading as not only epistemic predicates like ‘knowledge’ satisfy Factivity. Among others, the truth predicate itself is trivially factive.

Obviously, Sundholm’s derivation exactly mirrors the one of the Knower Paradox, using ‘knowability’ as the interpretation of $K$. This becomes even more apparent if we paraphrase what is going on in the formal derivations by using natural language. For the Knower Paradox, there is the following formulation by Égré (2005), pp.18-19, adapted from Tymoczko (1984) (my emphases are bold):

\[9\text{Whereas the Liar results from the sentence 'This sentence is false,' the Strengthened Liar originates from 'This sentence is not true.'} \]
\[10\text{Sundholm (2008), p.375.} \]
\[11\text{Cf. Kaplan and Montague (1960) and Montague (1963).} \]
\[12\text{Cf. Égré (2005), p.18.} \]
“Suppose someone knows this statement to be true; then this statement is true, otherwise it couldn’t be known; therefore, ‘nobody knows this statement to be true’ is true, that is nobody knows this statement to be true. So nobody knows this statement to be true. This is what the statement says, hence it is true. But hold on! I have just proved this statement to be true. Hence someone (at least me) knows this statement to be true! Now this contradicts what has just been established.”

To capture the reasoning behind Sundholm’s Paradox we only have to slightly modify this passage:

“Suppose someone can know this statement to be true; then this statement is true, otherwise it couldn’t be knowable; therefore, ‘nobody can know this statement to be true’ is true, that is nobody can know this statement to be true. So nobody can know this statement to be true. This is what the statement says, hence it is true. But hold on! I have just proved this statement to be true. Hence someone (at least me) knows (and a fortiori can know) this statement to be true! Now this contradicts what has just been established.”

So, Sundholm’s Paradox is not really a novel, genuine paradox, it is rather a variant of the Knower Paradox, but it is an important variant as the notion of knowability plays such an important role in recent debates.

Next, we will have a look at some more variations of the Knower Paradox before we turn our attention again to the notion of knowability.

4 Variations of the Knower Paradox

The Knower Paradox is often discussed in connection with related results about provability. For example, assuming certain properties of the provability predicate $P$, Lőb’s theorem says: $^\text{13}$

$$T \vdash P\phi \rightarrow \phi \text{ only if } T \vdash \phi.$$  

Put slightly differently: $^\text{14}$ Given a theory $T$ formulated within a sufficiently rich language, and $P$ a provability predicate for $T$, then, if $P$ is factive ($P\phi \rightarrow \phi$ for every sentence $\phi$), $T$ is inconsistent.

Given the following correspondences

- $\phi$ has been proved $\approx \phi$ is known
- $\phi$ is provable $\approx \phi$ is knowable

it might make more sense, instead of discussing the results about provability in connection with the proper Knower Paradox involving the notion of knowledge, rather to compare them to the knowability version of the Knower Paradox.

Another question is whether there are versions of the Knower Paradox that don’t assume the relevant predicate to be factive. As one of the main differences between

$^\text{13}$Cf. Lőb (1955).
$^\text{14}$Cf. Gödel (1933).
knowledge and belief is that knowledge is factive whereas belief is not, this leads us to the so-called Believer Paradox and variations of it. Indeed, there is a result by Thomason (1980) involving the notion of belief and corresponding assumptions about this notion which is, in a certain sense, a generalisation of the Knower Paradox.\footnote{For a proof of the result, see Égré (2005), pp.23-24. Other variants of the Believer Paradox are due to Turner (1990) and Koons (1992).}

It is interesting to remark that we have two knowability paradoxes: the Church-Fitch Paradox and Sundholm’s Paradox. Sundholm’s Paradox is a version of the Knower Paradox which makes use of factivity of knowledge/knowability and factivity of knowledge is also used in the standard version of Church-Fitch. Now, as we have just seen there are variations of the Knower Paradox, Believer Paradoxes, not relying on factivity but on weaker principles. The question is whether there is also a variation of Church-Fitch involving the non-factive notion of belief instead of the factive notion of knowledge? And indeed, recently Michael Fara has published a variant of the Church-Fitch Paradox for belief.\footnote{Cf. Fara (2007).}

5 Knowability and Factivity

Let’s come back to the notion of knowability. So far, we have always presupposed that knowability is factive, that knowability entails truth. But, is this justified? Is it clear that knowability is factive? I doubt it. Whereas it is pretty clear that knowledge is factive,\footnote{That knowledge is factive is doubtful according to some variants of contextualism and subjectivism. But potential counterexamples only concern formulas with nested knowledge operators. Cf. Rebuschi and Lihoreau (2008).} it might be different with knowability.

The most straightforward understanding of knowability analyses it as the composition of a standard possibility operator and a standard knowledge operator. But, neither if the possibility operator is supposed to express metaphysical possibility, nor if it is supposed to express epistemic possibility, the resulting conception of knowability turns out to be factive.

If we analyse knowability as $\Diamond K$ with $\Diamond$ expressing metaphysical possibility, a proposition is knowable iff it could have been the case that it was known. Or, using possible worlds talk: if there is a possible world such that in that world somebody knows the proposition. Of course, given factivity of knowledge, then the proposition has to be true in the respective possible world, but as this required possible world need not be the actual world, it might well be that the proposition isn’t true in the actual world (despite being known in another possible world).

Things do not look much different when we turn our attention to epistemic modalities. In order to consider the possible worlds as epistemic possibilities instead of metaphysical possibilities, we can use the following heuristic: an epistemically possible world is a possible state of the world that is compatible with all that
Sundholm’s Paradox of Knowability 295

one knows (or believes). For a proposition to be knowable (from my perspective) then means that somebody knowing this proposition is compatible with all that I know. If I know a proposition to be false, it follows that this proposition is not knowable in the relevant sense, of course. But, what about cases in which I neither know the proposition to be true nor know it to be false? Then it is compatible with my knowledge that the proposition might be true and also known, thus it is knowable, even if it is in fact false.

So, we have examined two conceptions of knowability and we have seen that both are non-factive. The question remains whether there are other sensible conceptions of knowability that are factive. One might appeal to two main sources in order to motivate the belief in the existence of factive knowability:

a) ordinary language use;

b) anti-realists’ use of knowability.

a) Knowability in Ordinary Discourse

Knowability expert Joe Salerno is convinced that there are uses of ‘knowability’ (and related expressions) in natural language that clearly involve a factive notion of knowability. Let’s have a look at one of his examples:

‘Does a factive conception of knowability figure in ordinary use? There is some reason to think so. ‘Knowable’ and related terms such as ‘discoverable,’ ‘observable,’ and ‘verifiable’ all seem to operate factively in ordinary discourse. Consider the following example, a dialog between colleagues A and B:

A: We could be discovered.
B: Discovered doing what?
A: Someone might discover that we’re having an affair.
B: But we are not having an affair!
A: I didn’t say that we were.

A’s remarks sound contradictory. In this context the factivity of ‘someone might discover that’ explains this fact. So there is some reason to believe that knowability and related modalities are factive in ordinary use.”

I agree that A’s remarks are definitely odd. Whether they are contradictory in a strict sense is a different question, though. If it is possible to explain the oddness of A’s remarks without having to appeal to a factive notion of knowability (or discoverability), the example loses much of its force. Indeed, I think, that the oddness of A’s remarks can be explained by bringing in pragmatic considerations. If A and B in fact don’t have an affair, as B claims, then it is completely irrelevant to speculate about a possible discovery of this (non-existing) affair. So, at least A is harshly violating Grice’s maxim of relevance.

But, even worse, if we assume that $A$ and $B$ know that they don’t have an affair (usually people know whether they have an affair with someone else or not!), then it is simply not true that it is epistemically possible for them that someone might know that they have an affair. Because somebody knowing that they have an affair would be incompatible with their knowledge. So, what $A$ says is plainly and rather obviously false, and $A$ can also be accused to violate Grice’s maxim of quality. Epistemic knowability of $p$ entails that the relevant persons don’t know $p$ to be false. But this does not amount to knowability of $p$ entailing $p$ to be true, as $p$ might be false and its falsity simply unknown to the relevant persons.

So, the example does not establish that there is a factive notion of knowability in ordinary discourse. The oddness of $A$’s remarks can be explained differently. The situation can be compared to examples in connection with Moore’s Paradox. Somebody says: “I believe that it rains, but it is not raining.” Such an utterance is definitely odd. But nobody would conclude because of such examples that the notion of belief as used in ordinary discourse be factive. Again, the oddness can be explained differently.

On the other hand, it is not too difficult to construct ordinary language examples involving true knowability claims with a clearly non-factive notion of knowability:

Steve is a candidate in a TV quiz show. He is asked questions and each time he can choose between four possible answers $A$, $B$, $C$ and $D$. The last question was about dinosaurs and Steve has chosen answer $B$. In the audience there are Mary, Steve’s wife, and Angela, Steve’s sister. The following dialog between them makes perfect sense:

Mary: Is it possible that Steve knows that $B$ is the correct answer?
Angela: Yes, that’s possible. In his youth Steve has read lots of books about dinosaurs, so he might have chosen $B$ because he knows that $B$ is the correct answer. But, of course, it is also possible that he had to guess. I myself I don’t know the correct answer, it might be $B$ or one of the other three.
Mary: Me neither. Let’s hope that Steve knew that $B$ was the correct answer or that he at least made a correct guess ...

It is clear that the notion of knowability involved in this example is non-factive. I don’t think that the notion of knowability involved in the Brogaard/Salerno example is any different. The difference between the examples is, what Mary and Angela are saying is perfectly reasonable, whereas what $A$ is saying is strange. But the strangeness and oddness of $A$’s utterances is definitely compatible with his
notion of knowability being non-factive, as the oddness can be explained differently.

b) Knowability and Anti-Realism

When developing their semantic and epistemological theories and views, many anti-realists or constructivists make ample use of the notion of knowability. But their notion is definitely different from the ones discussed above, combinations of a standard possibility operator (be it metaphysical or epistemic) and a standard knowledge operator. The notion of knowability used by anti-realists is factive.

Sundholm himself is quite explicit in equating truth with knowability and related notions:

“The true statements are the evidenceable, knowable, warrantable, justifiable, . . . ones.”

At another place he distinguishes truth in the actual sense from truth in the potential sense:

“When a judgement has become known, or evident, in virtue of its being demonstrated, it is clearly true. Truth is here taken in an actual sense of being known, or evident. However, before a mathematical judgement became known it could become known. In the potential sense a judgement is true when it is demonstrable (evidenceable, knowable, assertible, justifiable, warrantable).”

In his framework, Sundholm uses the notion of knowability to develop his ideas, but he does not say much how to exactly analyse and explicate this notion itself. Another anti-realist, Neil Tennant, is more explicit concerning the notion of knowability he uses and its factivity. To demarcate it from the other notions of knowability we discussed above, he calls it feasible knowability:

“If φ is any contingent falsehood (such as, say, ‘Grass is purple’), then it is not feasibly knowable that φ, in the sense of feasibility with which we are here concerned. Feasibility is not at all like the alethic modality of possibility. Another way of putting this last moral is to say that not only the epistemic operator K, but also its modalization ◻K, is ‘factive’.”

The idea behind feasible knowability maybe becomes a little bit clearer by invoking an analogy with dispositional predicates like ‘water-soluble.’ That a certain object is water-soluble does not mean that there is a possible world, and that in this possible world the object dissolves in water. After all, the object might have a different physical microstructure in another possible world. What is rather meant is that the object, its physical micro-structure unchanged, would dissolve if put into water. So, when considering other possibilities, certain features of the object have to remain fixed and unchanged, as for example the physical microstructure, in our case.

---

19 Sundholm (2004), p.452; my emphasis.
Correspondingly, an important idea behind feasible knowability seems to be that when considering other possibilities, the non-epistemic facts have to remain fixed and unchanged (so it would be excluded that the knowable proposition is false in the actual world, but true in a feasible possibility we consider), and only the epistemic facts, what we know or what we don’t know, might change. Such a conception of knowability is certainly interesting, but it still remains to be worked out in detail. Recently there has been quite some promising progress in such a direction.

Until this is accomplished (not only for knowability, but also for other notions used by anti-realists), misunderstandings are almost unavoi dable. The anti-realists develop their semantic-epistemological framework(s) by using certain modal expressions, but they disapprove of standard (non-anti-realistic) semantic theories for modalities like possible worlds semantics. So, when someone who was philosophically raised with possible worlds etc. reads the expositions of the framework(s) by anti-realists, he is inclined to understand the modal notions used by the anti-realists in a way that is unacceptable for them.

Sundholm suspects that Fichte might be right, that

“the point is moot whether there is a neutral background position from which the issue between realists and idealists [or anti-realists; H.R.] can be adjudicated.”

Nevertheless, even if there is no neutral background, we all should at least try to clarify the concurring positions and views as far as possible.

6 Acknowledgments

I am very thankful to Paul Égré (Paris) who not only made me aware of the similarities between Sundholm’s Paradox and the Knower Paradox in the first place, but who also provided me very helpful remarks and comments in personal communications. Finally, I have to thank the editors of this volume, Giuseppe Primiero and Shahid Rahman, for their almost endless patience.

Dear Göran, I am pretty sure that you will disagree with some (most?) things I said in this little paper. Nevertheless, I hope you will like it, if only (small) parts of it.

BIBLIOGRAPHY


22The main problems I see with such a conception concern the feasible knowability of epistemic facts themselves, and that the epistemic facts seem to supervene on the non-epistemic facts, such that changing the epistemic facts automatically also means changing at least some non-epistemic facts.


Sundholm’s Paradox of Knowability


