Self-Referential Gettier Sentences

Volker Halbach and Leon Horsten*

7th December 2021

We construct new examples of justified true belief that do not constitute knowledge. In contrast to normal Gettier examples, if our counterexamples are used, the demonstration that justified true belief is not sufficient for knowledge requires only that belief and knowledge are both conceived as predicates and that basic syntactic reasoning is available.

THE EXTENSIONAL ARGUMENT

If knowledge is defined as justified true belief, then the following equivalence will hold for all declarative sentences A:

(JTB) $$\mathsf{K}\,[A] \leftrightarrow \mathsf{JB}\,[A] \wedge A$$

In this equivalence K and JB are predicates. We defend the conception of knowledge and justication as predicates below from pp. 6. For a sentence A, the expression $[A]$ may be conceived as a designator for the proposition expressed by A in our fixed language or for the sentence A, depending on what kind of objects are assumed to be knowable or justifiable.

The condition $\mathsf{JB}\,[A]$ can be split up further into a belief and a justification condition; we use JB only for brevity. It is crucial for what follows that JB is understood as justified belief, not as another condition such as warrant or some causal or other external condition. Later we invoke a further assumption on JB that we can defend only for justified true belief, but not for other conditions.

* We thank Tim Williamson for discussions about the topics in this paper and New College for hosting Leon Horsten as visiting fellow in Michaelmas term 2017. We are also grateful to various anonymous referees who have helped to improve this paper.

In (JTB) the truth condition is not expressed using a predicate, but only by adding the sentence A. Consequently (JTB) becomes a schema. If a predicate for truth were used, (JTB) could be stated as a single universally quantified principle, using truth in the sense of truth-theoretic deflationism as generalising device:

$$\forall x\,(\mathsf{K}\,x \leftrightarrow \mathsf{JB}\,x \wedge \mathsf{T}\,x)$$

As we argue below, this is the only way to obtained (JTB) as a standard first-order definition. However, here we dispense with the truth predicate in order to dispel any suspicions that our reasoning below involves truth-theoretic paradoxes. Our account can be recast with a truth predicate and T-sentences for all truth-free instances. [1]

The left-to-right direction of (JTB)

(Nec) $$\mathsf{K}\,[A] \rightarrow \mathsf{JB}\,[A] \wedge A$$

is usually seen as less problematic. In particular the factivity of knowledge (knowledge implies truth), which is entailed by it, is widely taken to be analytic of the concept of knowledge.

Gettier (1963) provides counterexamples against their sufficiency, that is against the right-to-left direction:

(Suf) $$\mathsf{JB}\,[A] \wedge A \rightarrow \mathsf{K}\,[A]$$

[1] There is a certain irony in the fact that we treat truth as a trivial sentential operator and justification and belief as predicates. This is exactly the opposite of Gettier's (1963, p. 121) original formulation:

S knows that P IFF (i) P is true
(ii) S believes that P, and
(iii) S is justified in believing that P.

The letter P in Gettier's formulation seems to be a propositional variable, because it is combined with 'that', except for the truth condition (i) where it functions as a normal objectual variable. If a propositional variable is used, the truth predicate in (i) becomes superfluous: If P is a propositional variable, P by itself is perfectly sufficient as truth condition. We suspect that Gettier added the truth predicate, because it would be disappointing if a definition of knowledge of true justified belief did not contain the word 'true'. Strangely, most epistemologists, for instance Ichikawa and Steup (2014), have stuck to Gettier's unhappy mix of predicates and sentential operators. For a proper explicit definition all notions – truth, justification, belief, and knowledge – ought to be treated as predicates (Halbach 2016). Our version (JTB), in contrast, is to be understood as a schema.

For our arguments only K has to be a predicate, because Knower cannot be obtained with an operator. JB in contrast can be replaced with a corresponding sentential operator.

We produce a new counterexample to (Suf) by reasoning that relies purely on syntactic assumptions. In particular, we assume that there is a closed term g such that $g = [\neg \mathsf{K} g]$ is provable. It does not matter whether the term is obtained using Gödel's diagonal lemma or some other way. Abbreviating $\neg \mathsf{K} g$ as G, this identity implies the following equivalence:

(Knower) $$G \leftrightarrow \neg \mathsf{K}[G]$$

Such a sentence G can be derived in the usual settings using the Gödel diagonal lemma. The sentence is known to lead to contradiction with certain assumptions on the knowledge predicate.[2] We do not rely on any specific assumption on the knowledge predicate implicit in (JTB) other than the Factivity of knowledge.

The unproblematic direction (Nec) of (JTB) implies the factivity of knowledge, that is, $\mathsf{K}[A] \rightarrow A$.

	$\mathsf{K}[G] \rightarrow G$	Factivity
	$\mathsf{K}[G] \rightarrow \neg G$	Knower
(1)	$\neg \mathsf{K}[G]$	two preceding lines
(2)	G	Knower
(3)	$\mathsf{JB}[G]$	crucial assumption

The last step labelled (3) is the crucial assumption in our argument. We have proved G and we have come to believe G on the basis of this proof. Moreover, all the premisses in the proof of G – diagonalization and (JTB) – are justified. Therefore we have $\mathsf{JB}[G]$. This is an empirical fact: If for whatever reason we failed to believe G, $\mathsf{JB}[G]$ would not be true. We certainly do not assume a general rule that allows us to infer $\mathsf{JB}[A]$ from the availability of a proof of A from justified premisses. For the moment being we assume the step from (2) to (3) is sound. We return to this point to later.

The sentence G is a Gettier sentence: It is a belief that is true (2) and justified (3), but not known (1). From JTB we have so far used only the factivity of knowledge, which follows from its left-to-right direction (Nec). It can be shown that (Nec), (3), and (Knower) are jointly consistent. Assuming also the sufficiency direction (Suf) of (JTB) leads to a contradiction, because as a Gettier sentence G is a counterexample to the sufficiency of true justified belief for knowledge:

	$\mathsf{JB}[G] \wedge G$	from (2) and (3)
(4)	$\mathsf{K}[G]$	(Suf)

The last line is a contradiction with (1).

[2] The sentence is called *Knower* in analogy to the *Liar* sentence. It is not the diagonal sentence usually used in the derivation of Montague's and Kaplan's (1960) Knower Paradox.

Technically our paradox is a variant of the simple knower paradox.[3] The latter shows that factivity $\mathsf{K}[A] \to A$ and the rule of 'necessitation' for K are inconsistent with diagonalization. The derivation proceeds in the same way as our paradox, except that from the proof of G, that is (2), the last line (4) follows directly by necessitation. In our version we conclude $\mathsf{JB}[G]$ from (2) and then use the sufficiency direction of (JTB) to arrive at $\mathsf{K}[G]$. What is gained by the detour through $\mathsf{JB}[G]$ and (JTB)?

Our paradox dispenses with the notoriously controversial rule of necessitation for knowledge. Defences of the necessitation rule for K rely on detours like ours that are usually not formally spelled out: On the basis of the proof for G, one might believe G. Moreover, the proof supplies a justification for the belief. From these observations one might then conclude that G is known, that is $\mathsf{K}[G]$. We are not aware of justifications of necessitation for K that are fundamentally different. Thus, if this is how necessitation for K is defended, our proof merely made the steps in the defence explicit in the formal derivation. What has gone more or less unnoticed is that this defence makes use of the sufficiency direction (Suf) of (JTB). If we reject the sufficiency of justified true belief for knowledge, the defence of the rule of necessitation for K collapses. Therefore the rejection of the sufficiency of justified true belief for knowledge is a possible solution to the simple knower paradox unless there is another defence of necessitation.

In our proof necessitation is broken down in at least two steps: From the proof of G we infer $\mathsf{JB}[G]$ and then use the sufficiency direction (Suf) of (JTB) to conclude $\mathsf{K}[G]$. One might try to retain the sufficiency direction (Suf) by rejecting the transition from the proof of G to $\mathsf{JB}[G]$. This was our crucial empirical assumption. In the next section we show that only a much weaker assumption is needed, which puts further pressure on (Suf).

THE INTENSIONAL ARGUMENT

There is a further objection to our crucial assumption: Given the strange derivation, nobody may ever believe G. In response we weaken our assumption. From G we merely conclude that is possible to be justified in the belief G. This weakened version is sufficient for our argument.

[3] This is Montague's (1963) paradox for K, which is not be confused with the Kaplan–Montague (1960) paradox. Obviously Montague's paradox is a slight strengthening of the liar paradox. Since Tarski (1936) was the first to present the liar paradox in a setting as above, we call it the Montague–Tarski paradox. It can be applied to necessity, knowledge, apriority, and various other notions.

This can be seen as follows. We assume that the metaphysical necessity operator $\Box$ is governed by the S5 rules of propositional modal logic and we formalise necessity as an *operator* in order to dispel the worry that our argument might turn on the well-known difficulties of treating necessity as a predicate.

Under this reading the following modalized version of (JTB) ought to be correct, if (JTB) is an adequate definition of knowledge:

$$(\mathrm{JTB}_\Diamond) \qquad \Box(\mathsf{K}[A] \leftrightarrow \mathsf{JB}[A] \wedge A)$$

This implies the modalized form $\Box(\mathsf{K}[A] \rightarrow A)$ of factivity, which in turn implies the following:

$$(5) \qquad \Diamond\mathsf{K}[A] \rightarrow \Diamond A$$

We assume a modalized version of diagonalization, that is, we assume that there is a diagonal term $g_\Diamond$ such that the equations $g_\Diamond = [\neg\Diamond\mathsf{K}[g_\Diamond]]$ and $\Box(g_\Diamond = [\neg\Diamond\mathsf{K}[g_\Diamond]])$ are provable. We abbreviate $\neg\Diamond\mathsf{K}[g_\Diamond]$ as $G_\Diamond$

$$(\mathrm{Knower}_\Diamond) \qquad G_\Diamond \leftrightarrow \neg\Diamond\mathsf{K}[G_\Diamond]$$

Then we reason as follows:

	$\neg G_\Diamond \rightarrow \Diamond\mathsf{K}[G_\Diamond]$	Knower$_\Diamond$
	$\neg G_\Diamond \rightarrow \Diamond G_\Diamond$	(5)
(6)	$\neg G_\Diamond \rightarrow G_\Diamond$	S5
(7)	$G_\Diamond$	

The penultimate line (6) follows because $G_\Diamond$ is of the form $\neg\Diamond A$; and in S5 $\Diamond\neg\Diamond A$ is equivalent to $\neg\Diamond A$. In S5 $\neg\Diamond A$ is also equivalent to $\Box\neg\Diamond A$, which yields the following:

$$(8) \qquad \Box G_\Diamond$$

We now use the modalized version of our crucial assumption: $G_\Diamond$ is demonstrated from justified premisses, so it can be justifiedly believed:

$$(9) \qquad \Diamond\mathsf{JB}[G_\Diamond]$$

Finally, since $G_\Diamond \leftrightarrow \Box\neg\mathsf{K}[G_\Diamond]$ holds by (Knower$_\Diamond$), line (7) implies the following:

$$(10) \qquad \Box\neg\mathsf{K}[G_\Diamond]$$

Together (8), (9), and (10) imply the following in S5:

$$(11) \qquad \Diamond(G_\Diamond \wedge \mathsf{JB}[G_\Diamond] \wedge \neg\mathsf{K}[G_\Diamond])$$

This means there can be a counterexample to (JTB). This is inconsistent with the modalized version (JTB$_\Diamond$) of (JTB).

We have added this section in response to several requests from referees. Clearly, traditionally in philosophical logic belief and knowledge are conceived as sentential operators, not as predicates, at least since Hintikka (1962). The consequences of applying diagonalization to epistemic predicates have often been seen as strong arguments in favour of the operator view. The availability of possible-worlds semantics for sentential operators and the discovery of the paradoxes by Kaplan, Montague, and others helped the operator approach to an almost complete victory over the predicate approach advanced earlier by Carnap, Quine, and others. Our observations above may be suspected to be only further evidence for the superiority of the operator conception of belief and knowledge. In this section we can sketch only some reasons why the epistemologist should treat knowledge as predicate.[4]

Before providing reasons for preferring the predicate approach to knowledge, we emphasize that we think that belief, justification, possibility, necessity, and truth should also be conceived as predicates. The reason for the unhappy mix of sentential operators and the knowledge predicate above is an attempt to block objections to the effect that the use of a truth or necessity predicate is at the root off our observations. If preferred, JB can also be seen as an operator in the derivations above. In that case, $\mathsf{JB}\,[A]$ is to be replaced with $\mathsf{JB}\,A$ in the formulae above.

Conversely, it is not hard to see that justified belief, necessity, and truth can all easily be treated as predicates in our arguments above. Basically this requires only to replace A with $\mathsf{T}\,[A]$ and $\Diamond A$ with $\mathsf{P}\,[A]$, where P is a predicate for possibility and A some sentence. Roughly, whatever can be done with sentential operators can also be done with corresponding predicates.[5]

[4] Here we cannot provide a comprehensive bibliography on the treatment of epistemic notions as predicates, but mention only some recent contributions: Stern (2016) provides a book-length discussion of the predicate approaches to modalities. Halbach (2016) defends the use of predicates for knowledge, justification, and belief in the debate about the Gettier problem. (Tucker 2017) is a recent contribution to the discussion about whether knowledge and related notions should be ascribed to sentences or more coarse-grained objects. Pailos and Rosenblatt (2015) focus on solutions of multi-modal paradoxes where the interaction of different predicates leads to problems. *De re*-modalities are discussed by Halbach (2021). Halbach and Leigh (2022) provide a very general analysis of the paradoxes of modal predicates and their background assumptions. This monograph contains also an account of possible-worlds semantics for modal predicates, which was introduced by Asher and Kamp (1989) and Halbach et al. (2003).

[5] The qualification 'roughly' is required. Especially in modal contexts predicates and operators may be thought to behave very differently. For instance, if A is a tautology, A is necessary, while $\mathsf{T}\,[A]$ may be thought to be contingent because the sentence $[A]$ has its meaning

The main reason for treating notions as predicates rather than sentential operators is that this allow us to express quantified claims. For instance, factivity of knowledge can be pressed as the single sentence $\forall x\,(\mathsf{K}x \to \mathsf{T}x)$, expressing that whatever is known is true. With operators the best we can do is to formulate a schema $\mathsf{K}A \to A$. For truth we can omit the operator, because it would be the trivial operator governed by the axiom $\mathsf{T}A \leftrightarrow A$, as we did already in (JTB)). Using a schema instead of a single quantified sentence may not be thought to be too problematic until we try to negate factivity. With predicates is obvious, but with operators we would have to use $\mathsf{K}A \wedge \neg A$, but not as a schema, but rather a form that is existentially quantified sentence in the metalanguage.

Only with predicates the traditional definition of knowledge as justified true belief becomes a normal first-order definition, namely $\forall x\,(\mathsf{K}x \leftrightarrow \mathsf{JB}x \wedge \mathsf{T}x)$ or, even better, with separate predicates for belief and justification. Most epistemologists do not seem to care that the usual formulations of the definition fail to be definitions already for purely formal reasons. In footnote 1 we complained already about the mix of predicates and operators in Gettier's original version hat has been copied by many others.

The usual rejoinder is that propositional, 'substitutional', or second-order quantifiers can be used to express quantification. We believe that there are problems about these additional devices. At any rate, the definition of knowledge assumes then a format that is completely different from those of other definitions. Moreover, arguments are required for radical disparity of the logical form of 'Whatever is known is true' and 'Whatever is green is coloured'. With operators there are then no longer objects of belief and knowledge in the usual sense, and knowledge becomes incomparable with normal first-order notions. For instance, when we state that whatever is known is provable, then we probably have to treat 'provable' as a sentential operator as well, contrary to what has been done by logicians for nearly a century.

We stop the discussion here, although there is a lot more to say. The main claims of this paper are unaffected by this discussion. Just how interesting they are depends on the discussion whether knowledge is a predicate and can thus be diagonalized.

CONCLUSION

Our Gettier sentences G and $G_\Diamond$ refute the adequacy of the traditional tripartite definition of knowledge as true justified belief. We do not suggest that they can be used to refute the adequacy of arbitrary other definitions of knowledge. In

only contingently. Here we cannot go into this discussion.

particular, if justification is replaced with another condition, the premisses of analogous arguments may no longer be plausible: If JB is substituted with some other notion W, it is at least not clear why we should be able to conclude $\mathsf{W}[G]$ in what we called the crucial assumption; it is also not clear why we should be able to infer the modal variant $\mathsf{W}[G_\diamond]$ from (7). In fact, if our arguments applied to arbitrary conditions, they would overgenerate: They would imply that we have counterexamples against (JTB) and ($\mathrm{JTB}_\diamond$) with JB replaced with K (see Huemer 2005 and Halbach 2016). Presumably, a counterexample to $\mathsf{K}[A] \leftrightarrow \mathsf{K}[A] \wedge A$ would be a counterexample to the factivity of knowledge.

Whether our arguments can be applied to a given definition of knowledge depends on whether the analogues of the transition from (2) to (3) can be vindicated, that is, from G to $\mathsf{JB}[G]$ (or their modal counterparts). For instance our reasoning does apply to *no false lemmata* solutions (Clark 1963), because the derivations of (3) and (9) do not contain false lemmata.

The general upshot of our paper is that any adequate solution of the Gettier problem that implies $\Box(\mathsf{K}[A] \leftrightarrow \mathsf{W}[A] \wedge A)$ for some condition W will have to block the move analogous to the step from (7) to (9) on top of handling all the other known Gettier examples.

The reader might feel that our arguments against the definition of knowledge as true justified belief are frivolous: From the technical perspective our arguments are mere variations of the Montague or liar paradox. However, in contrast to the usual versions of the paradoxes, we do not require a step that takes us directly from ϕ to $\mathsf{K}[\phi]$, which strikes us as implausible for K as knowledge. In our setting, JTB and $\mathrm{JTB}_\diamond$ are required for this step and we show that the derivation of the contradiction can be blocked by rejecting JTB and $\mathrm{JTB}_\diamond$. Instead of rejecting them, we might also drop other assumptions or weaken our underlying logic. For instance, the diagonal lemma might blocked in some way, classical logic might be abandoned, or the formalization of knowledge or justified belief as predicates might be rejected.[6]

In general, we are sceptical that there is one single grand solution to all the paradoxes that arise from treating epistemic and modal notions as predicates.

[6] In particular, the formalization of JB as a predicate may be objected, because the following *restricted* factivity for JB in the form

$$\mathsf{JB}[\neg\mathsf{JB}[A]] \rightarrow \neg\mathsf{JB}[A] \tag{12}$$

already leads to a contradiction. Jan Heylen pointed out to us that if the laws of the minimal modal logic K hold for a predicate JB, and (12) holds, then a contradiction follows. Some philosophers, including Williamson (2000, p. 274), take (12) to be a plausible principle. But one might regard the paradox for rational belief that it entails to be good evidence that the principle is false (see Horsten and Schuster for a further discussion).

A more piecemeal approach may be more successful. At any rate, we should try to see whether we can avoid inconsistencies by giving up JTB and perhaps other principles, before we set out to rewrite epistemology in a nonclassical logic.

BIBLIOGRAPHY

Nicholas Asher and Hans Kamp. Self-Reference, Attitudes, and Paradox. In Gennaro Chierchia, Barbara H. Partee, and Raymond Turner, editors, *Properties, Types and Meaning*, volume 1, pages 85–158. Kluwer, Dordrecht, 1989.

Michael Clark. Knowledge and grounds: A comment on Mr. Gettier's paper. *Analysis*, 24:46–48, 1963.

Edmund Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963.

Volker Halbach. Prolegomena zu einer jeden künftigen Definition von Wissen, die als Lösung des Gettierproblems wird auftreten können. In Wolfgang Freitag, Holger Sturm Hans Rott, and Alexandra Zinke, editors, *Von Rang und Namen: Essays in Honour of Wolfgang Spohn*, pages 147–172. Mentis, Münster, 2016.

Volker Halbach. The fourth grade of modal involvement. In Carlo Nicolai and Johannes Stern, editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*, pages 331–356. Routledge, New York, 2021.

Volker Halbach and Graham Leigh. *The Road to Paradox: A Guide to Syntax, Truth, and Modality*. Cambridge University Press, 2022. to appear.

Volker Halbach, Hannes Leitgeb, and Philip Welch. Possible Worlds Semantics For Modal Notions Conceived As Predicates. *Journal of Philosophical Logic*, 32:179–223, 2003.

Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press, 1962.

Leon Horsten and Daniela Schuster. On the pure logic of justified belief. *Synthese*. to appear.

Michael Huemer. Logical Properties of Warrant. *Philosophical Studies*, 122(2): 171–182, 2005.

Jonathan Jenkins Ichikawa and Matthias Steup. The Analysis of Knowledge. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Center for the Study of Language and Information, Stanford, spring 2014 edition, 2014.

David Kaplan and Richard Montague. A Paradox Regained. *Notre Dame Journal of Formal Logic*, 1:79–90, 1960. Reprinted in (Montague 1974, 271–85).

Richard Montague. Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability. *Acta Philosophica Fennica*, 16:153–67, 1963. Reprinted in (Montague 1974, 286–302).

Richard Montague. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven and London, 1974. Edited and with an introduction by Richmond H. Thomason.

Federico Pailos and Lucas Rosenblatt. Solving multimodal paradoxes. *Theoria*, 81:192–210, 2015.

Johannes Stern. *Toward Predicate Approaches to Modality*, volume 44 of *Trend in Logic*. Springer, Cham, 2016.

Alfred Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica Commentarii Societatis Philosophicae Polonorum*, 1, 1936. translated as 'The Concept of Truth in Formalized Languages' in Tarski (1956).

Alfred Tarski. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Clarendon Press, Oxford, 1956.

Dustin Tucker. Montagovian paradoxes and hyperintensional content. *Studia Logica*, 105:153–171, 2017.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.