Acta Analytica (2021) 36:443–449
https://doi.org/10.1007/s12136-020-00456-8

# What Does It Take to Know that You Know?

**Noah Gordon**[1]

Received: 22 July 2020 / Accepted: 24 November 2020 / Published online: 12 January 2021


**Abstract**
In some recent work, (Williams *Analysis, 75*(2), 213–217, 2015, *Logos and Episteme, 7*(1), 83–94, 2016) John N.Williams defends a new objection to the defeasibility theory of knowledge. But the objection is of wider interest, since Williams also suggests that this style of objection may undermine other theories of knowledge. I distinguish two versions of Williams' objection. I then show that the first version relies on false conceptual principles, and the second relies on a specific and dubious conception of the goal of the analysis of knowledge.

**Keywords** Defeasibility theory · Analysis of knowledge · Conceptual analysis · Metaphysical analysis · Hyperintensionality

The defeasibility theory of knowledge is undergoing something of a revival of interest as of late.[1] In fact, defeasibility theorists have recently made some bold claims—the loss of interest in the defeasibility theorist is a purely sociological matter, and no serious objections to the theory remain unanswered (de Almeida and Fett 2016: 158, 167–168). This supposedly spotless record is threatened by a recent objection from Williams (2015, 2016). But this objection also has broader relevance since it is suggested that other theories of knowledge may be vulnerable to it as well (Williams 2015: 216). I argue that defeasibility theorists and others have nothing to fear. By attending to some lessons about the hyperintensionality of knowledge ascriptions, we can show that one version of the objection relies on false conceptual principles. I also consider a close variant of this objection and show that it depends on an idiosyncratic conception of the goal of the analysis of knowledge.

[1]For evidence of this revival, see Williams (2015), de Almeida and Fett (2016), Williams (2016), Borges (2016), de Almeida (2017), Klein (2017), and Fitelson et al. (2019).

✉ Noah Gordon
noahgord@usc.edu

1 University of Southern California, Los Angeles, CA, USA

The early defeasibility theory (EDT) began as an anti-Gettier theory which adds a characteristic fourth condition to the traditional JTB analysis of knowledge (Williams 2015: 213):

EDT[2]: *S* knows that *P* if and only if

(i) *P* is true.
(ii) *S* believes that *P*.
(iii) *S*'s belief that *P* is justified.
(iv) There is no defeater for *S*'s justification for believing that *P*.

The crucial notion of a *defeater* bears elaboration in a full development of the defeasibility theory. Following Williams (2015: 213), I give a simplistic gloss on this notion as follows[3]:

A defeater *D* of your justification for believing that *P* is a truth such that believing it would render your belief that *P* unjustified.

For example, although I may be justified in believing at noon that it is raining at my house (on the basis of checking a weather app), there may exist a defeater for my justification, namely, the truth that my lawn is not wet. Since I fail condition (iv), I cannot know that it is raining, according to the early defeasibility theory (regardless of whether it is in fact raining or not).

Although the defeasibility theory has evolved over time, Williams' objection is supposed to target all versions of the defeasibility theory (Williams 2015: 215–216). For simplicity, I confine my discussion to EDT, although the same points apply to other versions of the defeasibility theory mutatis mutandis.

Williams' objection to the defeasibility theory is that it cannot account for a certain kind of second-order knowledge, namely, knowing that you know that *P*, where your knowledge that *P* is a posteriori. Consider a mundane example (Williams 2015: 214–215). You perceive rain falling outside, on the basis of which you come to know that it is raining. Introspecting, you also come to know that you know this.

According to Williams, the defeasibility theory cannot handle even these mundane cases. The crux of the issue is that Williams thinks that, on the defeasibility theory, you would have to know that condition (iv) is satisfied to know that you know it is raining (Williams 2015: 215, Williams 2016: 85–86). In other words, you would have to know that there is no defeater for your belief that it is raining. But according to Williams, you cannot plausibly know this. You cannot know a priori that there is no defeater for your belief, but you also cannot know it by inspecting each truth individually and discerning that it is not a defeater for your belief, since you are not omniscient. So you just cannot know that condition (iv) is satisfied at all, and so cannot know that you know it is raining. The general form of Williams' objection then is as follows:

(1) If EDT is true, then *S* knows that *S* knows that *P* only if *S* knows that there is no defeater for *S*'s belief that *P*.

[2] As is unfortunately typical in presentations of "theories" or "analyses" of knowledge, I do not specify what modal status this biconditional has. I return to this important issue later.

[3] For a more sophisticated characterization of an epistemic defeater in the context of the defeasibility theory, see Klein (1986: 261-269).

(2) If *S*'s knowledge that *P* is a posteriori, then *S* cannot know that there is no defeater for *S*'s belief that *P*.
(3) Therefore, if EDT is true, then *S* cannot know that *S* knows that *P*, if *S*'s knowledge that *P* is a posteriori.
(4) But for some *S* and some *P*, *S* knows *P* a posteriori and *S* knows that *S* knows that *P*.
(5) So EDT is false.

I am not concerned here with Williams' claim that we cannot know that condition (iv) is satisfied in cases where our knowledge is a posteriori (i.e., premise 2).[4] Instead, I will argue that, even if the defeasibility theory is true, we need not know that condition (iv) is satisfied to know that we know (so premise 1 is false).[5]

Why does Williams think that we must know that condition (iv) is satisfied to know that we know? His main reason is that it is entailed by a certain conceptual principle that he finds plausible (Williams 2015: 215). However, Williams' discussion fluctuates between two different construals of this principle. I will argue that the first construal is false and that the second construal cannot support Williams' argument without assuming a specific and dubious version of the defeasibility theory. The first construal, which follows Williams' explicit formulation, is as follows:

> CP: If the satisfaction of a condition at least partly constitutes an instance of a concept, then knowing that such an instance obtains requires you to know that the condition is satisfied.

Williams gives the following instance to illustrate CP. Satisfying the condition *being three-sided* at least partly constitutes being an instance of the concept *triangle*. According to CP, then, knowing that something is a triangle requires knowing that it is three-sided. This sounds plausible. Now if EDT is correct, then being an instance of the concept *knowledge* is at least partly constituted by satisfying condition (iv). So it follows from CP that to know that something is an instance of knowledge, one must know that it satisfies (iv). In this way, premise (1) follows from CP.

In later work, Williams (2016): 85-86) suggests another reason to believe that we must know that condition (iv) is satisfied, namely, the *knowing that you know* principle:

> *Knowing that you know*: If you know that you know that *P*, then you know the content of each necessary condition of your knowing that *P*.

[4] Borges (2016) criticizes this premise, but I largely agree with the reply given by Williams (2016). However, I remain agnostic about this premise.
[5] Borges (2016) argues that we need not know a priori that condition (iv) is satisfied to know that we know. But the argument from CP does not require that stronger claim (Williams 2016: 90–93). Unfortunately, Williams' presentation of the argument encourages the stronger reading since he supplies only examples where the satisfaction of condition (iv) can be known a priori. In my view, this is more evidence that Williams is not sensitive to the distinction between CP and CP*, introduced below. An argument from CP* would require a priori knowledge of the satisfaction of (iv), if conceptual truths are knowable only a priori.
[0] This justification assumes that being a necessary condition of falling under a concept entails partly constituting being an instance of that concept, in Williams' sense.

Since by EDT, condition (iv) is a necessary condition of knowing that *P*, premise (1) immediately follows. Williams does not say why *knowing that you know* is plausible, but I suggest that he finds this principle plausible since it is basically an instance of CP where the concept in question is *knowledge.*[6]

It is easily seen by the *knowing that you know* principle why you might think that this style of objection could threaten other theories of knowledge. By either of these principles, it would follow from a given analysis of knowledge that you must know that your beliefs satisfy each necessary condition of that analysis in order to know that you know. So, as Williams (2015): 216) notes, by these principles it follows that any analysis incorporating a safety condition on knowledge[7] has the consequence that we must know that our beliefs are safe to know that they amount to knowledge. Williams does not argue against other analyses of knowledge on this basis, but thinks that this question should be pursued (2015: 216).

I do not think that this question should be pursued. Both of the principles that Williams relies on to pursue his challenge are false. While the example involving *being three-sided* and *triangularity* seems plausible, a very nearby example quickly shows CP to be false. Instances of *triangularity* are also at least partly constituted by satisfying the condition *having interior angles that sum to 180° when located in Euclidean spaces*. In fact, this is both a necessary and a sufficient condition of being a triangle. It is not plausible that one can know that a figure is a triangle only if one knows that its interior angles sum to 180° in Euclidean spaces. The mathematical neophyte who knows nothing about the properties of interior angles of geometric figures can surely know and identify some figures as triangles without knowing that these figures' interior angles sum to 180° in Euclidean spaces.

Consider another simple example involving the concept *water*. The satisfaction of the condition *being* $H_20$ at least partly constitutes being an instance of water. According to CP, then, knowing that some substance is water requires one to know that it is $H_20$. But even before humans discovered the chemical composition of water, we knew that the substance in our cup was water. We were not, unbeknownst to us, already in possession of the knowledge of the chemical composition of water. So there is another counterexample to CP.

The *knowing that you know* principle is simply an application of CP. It is no better than the following analogous principle about water:

If you know that a substance is water, then you know the content of each necessary condition of its being water.

The above counterexample shows the above to be false. So we should doubt *knowing that you know* as well. If Williams' conceptual principles were correct, then the folk would have to study chemistry to know that they are drinking water. Similarly, they would have to study epistemology to know that they know. (This might be a welcome result for some epistemologists but I doubt the folk would appreciate it.)

These simple counterexamples are illustrations of a very general phenomenon—knowledge and belief contexts are hyperintensional. My knowledge or belief that *P*

[6] This justification assumes that being a necessary condition of falling under a concept entails partly constituting being an instance of that concept, in Williams' sense.

[7] *S*'s belief that *P* is safe just in case if *S* were to believe *P*, then *P* would not be false.

does not guarantee that I also know whatever is intensionally equivalent with *P*. The following is a very familiar and basic illustration:

(A) *S* knows that what is in the cup is water.
(B) *S* knows that what is in the cup is $H_2O$.

Since (A) can be true while (B) is false, knowledge ascriptions are hyperintensional. And since I can know that something is water without knowing that it satisfies a condition (*being* $H_20$) that is intensionally equivalent with *being water*, it of course follows that I can know that it is water without it satisfying a given necessary condition of its being water. There is no reason to think that this general lesson about the hyperintensionality of knowledge contexts fails to apply when the conditions and concepts within the scope of the knowledge ascription involve knowledge itself. As illustration, consider the very similar pair below:

C *S* knows that *S* knows that *P*.
D *S* knows that *S* has a safe belief that *P*.

Even if *knowledge* and *safe belief* were intensionally equivalent, (C) could be true while (D) is false. Therefore *knowing that you know* and the conceptual principle upon which it is based are both false.

However, perhaps the reason that Williams finds CP attractive is that he sometimes misleadingly takes it to be equivalent to a slightly different principle that may actually be true.[8] By attending to this principle, we can bring out another lesson about the target of the analysis of knowledge. Below I state the principle; the places where it differs from CP are in italics:

> CP*: If the satisfaction of a condition at least partly constitutes *a concept*, then knowing that an instance of that concept obtains *and falls under that concept* requires you to know that the condition is satisfied.

It seems true, for instance, that if *being three-sided* is constitutively a part of the very concept of *triangularity*, then knowing that something falls under the concept *triangle* requires knowing that it is three-sided. By contrast, the same may not be true for the pairs *triangle*/*having interior angles summing to 180°* or *water*/$H_20$. While the concept $H_20$ is necessarily coextensive with the concept *water*, it is not constitutively a part of that concept. So knowing that something falls under *water* does not require knowing that it falls under $H_20$.

If we were to run Williams' objection to the defeasibility theory using CP*, we would need the defeasibility theorist to claim that condition (iv) is constitutively a part of the very concept of knowledge. This would take the analysis of knowledge represented by EDT to be an instance of the very strong, old-fashioned brand of *conceptual analysis*. EDT would then be a conceptual truth and have the modal status of conceptual necessity.

[8] See the sentence immediately after he introduces CP for evidence of this (Williams 2015: 215).

By contrast, suppose that EDT were taken to be a *metaphysical analysis* of the nature of knowledge. EDT would then be a metaphysical truth about the necessary and sufficient conditions for knowledge and have the weaker status of metaphysical necessity. No claim about the constitution of the concept of knowledge would be made, and the defeasibility theory would be immune to any conceptual objection based on CP*.

It is unfortunate that so-called analyses of knowledge are often put forth without any clarification on the aim of the analysis—is it a semantic claim about the schema "*S* knows that *P*," or a conceptual claim about the concept *knowledge*, or a metaphysical claim about knowledge itself?[9] Nonetheless, I think it is most plausible to take defeasibility theorists (and many others in the business of analyzing knowledge) to be aiming for metaphysical, not conceptual analyses.[10] It should be noted that the mere fact that the analysis of knowledge is largely carried out as an a priori project does not show that it consists in conceptual analysis. The project of mathematics is also done a priori, but mathematical truths are not conceptual analyses. The geometrical analysis of the properties of triangles, for example, does not yield conceptual truths. To put the point in slightly less fashionable Kantian terms, like geometry, the analysis of knowledge may be a source of synthetic a priori truths rather than analytic a priori truths.

On the other hand, if all a priori analysis is conceptual analysis, then it should be a conceptual truth that triangles have interior angles that sum to 180° in Euclidean spaces. By CP*, it would then follow that we cannot know that something is a triangle without knowing that it has interior angles that sum to 180° in Euclidean spaces. This, as we have already seen, is an absurdity. In that case, the argument based on CP* is hopeless anyway.

I do not have any knockdown argument to show that EDT or other analyses of knowledge are generally not put forth as conceptual analysis; however, I will give a quick consideration in favor of this claim. If the proposed analyses are conceptual, then it seems that to possess the concept of knowledge, one must possess each of the component concepts that constitutes this complex concept. But even young children seem to possess the concept of knowledge, whereas it is not all that plausible that such children possess concepts such as *justification* and *defeater* in the defeasibilist's sense.[11]

---

[9] This complaint is echoed by Ichikawa and Steup (2018):

> In practice, many epistemologists engaging in the project of analyzing knowledge leave these metaphilosophical interpretive questions unresolved; attempted analyses, and counterexamples thereto, are often proposed without its being made explicit whether the claims are intended as metaphysical or conceptual ones.

[10] See Sosa (2017) for an explicit endorsement of the metaphysical project in contrast to the conceptual one.

[11] In this vein, consider the following somewhat monstrous more recent version of the distinctive defeasibilist fourth condition on knowledge:

> If [justifier] *e* is true, then there is no genuine defeater of the propositional justification of any of the propositions in the evidential path up to and including *e* and there is no genuine defeater of the propositional justification of any proposition between *e* and *h*; if *e* is false, then there is no genuine defeater of the propositional justification of any of the propositions in the evidential path up to and including t and there is no genuine defeater of the propositional justification of any proposition between *t* and *h*, where t is defined by Conditions 1–7. (Klein 2008: 49–50)

Is it at all plausible that young children possess these concepts and that this condition is conceptually prior to *knowledge*?

But it will suffice for my purposes here to note that the variant of Williams' objection based on CP* depends on taking the theories of knowledge as analyzing the concept of knowledge. If we are instead interested in analyzing knowledge itself, Williams' objection has no purchase.

**Acknowledgments** My thanks to Janet Levin, James Van Cleve, and an anonymous referee for helpful comments.

## References

Borges, R. (2016). A failed twist to an old problem: A reply to John N. Williams. *Logos and Episteme, 7*(1), 75–81.

de Almeida, C. (2017). Knowledge, benign falsehoods, and the Gettier problem. In R. Borges, C. de Almeida, & P. Klein (Eds.), *Explaining knowledge: New essays on the Gettier problem* (pp. 292–311). Oxford University Press.

de Almeida, C., & Fett, J. R. (2016). Defeasibility and Gettierization: A reminder. *Australasian Journal of Philosophy, 94*(1), 152–169.

Fitelson, B., Borges, R., & Braden, C. (Eds.). (2019). *Themes from Klein*. Springer International Publishing.

Ichikawa, J. J., & Steup, M. (2018). The analysis of knowledge. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Klein, P. (1986). Immune belief systems. *Philosophical Topics, 14*(1), 259–280.

Klein, P. (2008). Useful false beliefs. In Q. Smith (Ed.), *Epistemology: New essays* (pp. 25–63). Oxford University Press.

Klein, P. (2017). The nature of knowledge. In R. Borges, C. de Almeida, & P. Klein (Eds.), *Explaining knowledge: New essays on the Gettier problem* (pp. 35–56). Oxford University Press.

Sosa, E. (2017). The metaphysical Gettier problem and the X-Phi critique. In R. Borges, C. de Almeida, & P. Klein (Eds.), *Explaining knowledge: New essays on the Gettier problem* (pp. 232–242). Oxford University Press.

Williams, J. N. (2015). Not knowing you know: A new objection to the defeasibility theory of knowledge. *Analysis, 75*(2), 213–217.

Williams, J. N. (2016). Still a new problem for defeasibility: A rejoinder to Borges. *Logos and Episteme, 7*(1), 83–94.