

# KK is Wrong Because We Say So

Penultimate Draft, forthcoming in *Mind*

Simon Goldstein and John Hawthorne

## 1. Against KK

The KK principle says that if you know something, then you know that you know it. KK is wrong because we say so.

Doubting Dudley says that he knows he has hands but he doesn't know that he knows ... that he knows that he has hands. In logician speak, he says that he doesn't '100-know' that he has hands. (You 1-know something if you know it. For  $n$  greater than 1, you  $n$ -know something if you know that you  $n-1$  know it.)

Dudley only says things that he believes. And he never believes things that he knows are false. Since he says he doesn't 100-know that he has hands, he believes he doesn't 100-know he has hands. So he doesn't know that he 100-knows that he has hands. But if KK is true, and if he knows he has hands, then it follows that he does know that he 100-knows that he has hands. So KK is wrong.<sup>1</sup>

We don't even need to say that KK is wrong. KK is wrong if we don't say that it's right. Agnostic Agnes is asked whether she 100-knows she has hands, and she refrains from answering, because she is unsure. Agnes doesn't refrain from answering questions when she knows the answer. Since she refrains from answering the question of whether she 100-knows she has hands, she does not know that she 100-knows she has hands. So KK is wrong.

Here is an analogy. Agnes doesn't know exactly what love is. Is loving someone a matter of wanting their life to go well? Or do you also have to enjoy spending time with them? Agnes has a friend, Boris, who she doesn't like spending time with, but who she wishes the best for. Since Agnes doesn't know exactly what love is, Agnes doesn't know whether she loves Boris. In this respect, knowledge is like love. Agnes doesn't know exactly what knowledge is. Since she doesn't know exactly what knowledge is, she doesn't know whether she 100-knows that she has hands. Knowledge is one relation among others. If you don't know whether a condition is required for something to be related to you in a certain way, and a particular thing obviously doesn't satisfy that condition, then you won't know whether the thing is related to you in that way. KK denies this. So KK is wrong.

---

<sup>1</sup>Dudley may say that KK fails without saying exactly where it fails. For example, Dudley may say the conjunction *he doesn't 100-know he has hands and he doesn't 100-know he has feet, and either he knows he has hands or he knows he has feet*. He says the conjunction, so he believes the conjunction, so he doesn't know the conjunction is false. But if KK is true, and if he does know he has hands, then he does know that the conjunction is false.

The defender of KK has one way out. Deny that Dudley and Agnes know they have hands! If they don't know they have hands, then they don't know that they 100-know that they have hands. The KK defender can say that no one who disagrees with them knows anything. This response is consistent; but it is radically combative.<sup>2</sup>

## 2. The Position to Know

Weaker principles than KK are also wrong, more or less because we say so. One principle says that if you are in a position to know something, then you're in a position to know that you are in a position to know it. What is it to be in a position to know? Roughly: if you were to believe, without any changes to your evidence or reliability, then you would know (Willard-Kyle 2020).

This 'Weak KK' principle avoids our first argument. The weaker principle does not imply that Dudley 100-knows he has hands. So he can correctly believe that he doesn't 100-know he has hands. This is compatible with him 'being in a 100-position to know' that he has hands. (You are in a 1-position to know  $p$  when you are in a position to know  $p$ . For  $n$  greater than 1, you are in an  $n$ -position to know  $p$  when you are in a position to know that you are in an  $(n-1)$ -position to know  $p$ .)

Alas, Weak KK is false, now because we justifiably say so. Justified Justine says that she is not in a 100-position to know she has hands. She says so because she has good evidence that she is not in a 100-position to know she has hands (she's read a series of brilliant papers denying KK). In fact, she's justified in believing that she is not in a 100-position to know she has hands. But if she's justified in believing something, then she is not in a position to know that it is false. After all, imagine that she suddenly formed the belief that she is in a 100-position to know she has hands, while retaining the same evidence she has now. She wouldn't know that she is in a 100-position to know she has hands, because her evidence says she is not in this position.<sup>3</sup>

Again, consider love. Agnes is not in a position to know what love is, because she doesn't have good evidence one way or the other about what exactly it is. For this reason, she is not in a position to know whether she loves Boris. The position to know is like love. Agnes is not sure exactly what it takes to be in a 100-position to know that she has hands. So she is not in a position to know that she is in a 100-position to know that she has hands.<sup>4</sup>

---

<sup>2</sup> Some defenders of KK may deny that the principle is literally true; instead, it is a useful generalization that permits exceptions. Although false, it allows for simple and powerful models of human behavior. For those who embrace KK in this modeling spirit, the upshot of our argument is that KK cannot be used to model the behavior of anyone who is disagreeing with KK.

<sup>3</sup> Moreover, if she suddenly formed the belief that she is in a 100-position to know, this belief would not obviously be the result of a reliable method. Rather, Justine's method would be something like: making a guess about whether she is in a 100-position to know. She could easily have guessed either way, and she could easily have guessed falsely.

<sup>4</sup> Weak KK is different from the principle that if you know, then you are in a position to know that you know. We think this principle also fails, for familiar reasons.

Imagine now that Dudley is wrong about how confident you have to be to count as believing. He thinks you need to have a credence of .9, but really you only need to have a .7 credence. Now Dudley has a .7 credence that he has hands, and

As used by philosophers, ‘position to know’ is jargon. Some may stipulate that Justine is in a position to know that she is in a 100-position to know she has hands, despite her evidence to the contrary. Moreover, there are some views according to which knowledge has little to do with evidence, merely requiring reliability or even just true belief (Beddor and Pavese 2018, Goldman 2004). One problem with such responses is that Weak KK doesn’t offer the same explanatory benefits as KK. If you’re going to accept something like KK, you should just accept KK; Weak KK doesn’t explain the same things that KK can explain.

Consider assertion. Cohen and Comesaña 2013, Greco 2014, and Das and Salow 2018 argue that KK explains why it is strange to say the ‘dubious’ assertion *I have hands but I don’t know that I know I have hands*. First, assertion is governed by a knowledge norm: I should only say what I know (Williamson 2000). Second, KK predicts that I can’t know a dubious assertion: if I know it, then (assuming, plausibly, that knowledge distributes over conjunctions), I know I have hands; so by KK I know that I know I have hands; but then it isn’t true that I don’t know that I know I have hands.

But Weak KK doesn’t predict that the dubious assertion is unknowable. Since I know I have hands, I am in a position to know I have hands. So, by Weak KK I am in a position to know that I am in a position to know that I have hands. But this is consistent with my failing to actually know that I know I have hands.

As a second example, consider complex iterations: “without bringing in heavy-duty philosophical theory, there is no natural way to interpret ... *I grant that Jane knows that she knows that she knows what time the movie starts, but does she know that she knows that she knows that she knows what time the movie starts?*” (Greco 2014). KK explains why the question is uninterpretable. But Weak KK does not explain this, since Weak KK allows Jane to 3-know something without 4-knowing it.

As a last example, consider rational coordination. Greco 2015a argues that some cases of rational coordination require that a group of agents has common knowledge. (A group has common knowledge of  $p$  iff each member of the group knows that each member of the group knows that ... each member of the group knows that  $p$ , for

---

knows it. But he doesn’t know that he knows, because he thinks he doesn’t believe. In what sense is Dudley in a position to know that he knows he has hands?

Here’s another analogy. Imagine Dudley is also wrong about how tall you have to be to count as being tall. He thinks you have to be 6 foot 1, but really you only have to be 6 feet. Now imagine Dudley is 6 feet tall. Dudley is tall. But he is not in a position to know that he is tall, because he is wrong about what it takes to be tall. We think belief and knowledge are just like being tall. Dudley can believe and know he has hands. But he will not be in a position to know that he believes or knows he has hands, if he is wrong about what it takes to believe and hence know.

Similar challenges also threaten Weak KK. Imagine Dudley is wrong about how much inductive support you need in order to know. He has seen 30 black ravens, and this is enough to know that all ravens are black. But Dudley falsely believes that he needs to see 40 black ravens in order to gain inductive knowledge in this way. Dudley is in a position to know that all ravens are black. But he is not in a position to know that he is in a position to know that all ravens are black, because he is in the grip of a false theory about what it takes to gain inductive knowledge.

arbitrarily many iterations.) But common knowledge of  $p$  requires each member of the group to  $n$ -know that  $p$ , for every  $n$ . If generals on opposite hilltops are trying to coordinate their attack, and will only succeed if both of them attack simultaneously, then they will only attack if they have common knowledge that they will attack. This requires that each general  $n$ -knows they will attack, for every  $n$ . KK then explains how each general can know so much: it follows immediately from knowing. By contrast, if KK is false and Weak KK is true, then this amount of knowledge would be difficult to acquire. Each level of knowledge would be a further achievement, not guaranteed by knowing alone, and so common knowledge would be difficult to achieve.<sup>5</sup>

One reaction from the defender of Weak KK may be to seek alternative explanations of the various data points above. For example, return to the infelicity of dubious assertions of the form *p, but I don't know whether I know p*. One natural reaction is to try to explain this infelicity in terms of pragmatic principles about being in a position to know.<sup>6</sup> For example, suppose that there were a defeasible rule that you shouldn't say *I don't know whether p* if you are in a position to know  $p$  and are interested in the question. In that case, asserting the dubious assertion would implicate that you are not in a position to know whether you know  $p$ . Given Weak KK, this would contradict knowledge of  $p$ .

We are resistant to this response, because we do not agree that there is a pragmatic rule that you shouldn't say *I don't know whether p* if you are in a position to know  $p$  and are interested in the question. We'll give three counterexamples to this rule. First, imagine that you are sitting in a math exam, and asked whether some complex mathematical claim  $p$  is true.  $p$  is interesting and important in the context. From your knowledge, you could in fact deduce whether  $p$ , and so you are in a position to know it (at least according to some theories). But it is perfectly fine to say that you don't know whether  $p$ . There is no implicature that you are therefore not in a position to know it.

Second, imagine that you are looking at a red wall in normal lighting conditions, but that you are in the grip of skeptical doubts, and so on this global basis are unsure whether you know. At least one kind of defender of Weak KK will say that you can still know that you are looking at a red wall in this case. You are also in a position to know that you know you are looking at a red wall; your doubts merely block you from forming a belief. In this case, you can reasonably say you don't know whether you know the wall is red. But there is no implicature that you aren't in a position to know whether you know the wall is red (and if there was, this would threaten Weak KK anyways).

---

<sup>5</sup> Here, one interesting project for the defender of Weak KK might be to lay down systematic rules for the automatic accumulation of beliefs, to ensure that once a group of agents achieves some amount of iterated belief and knowledge, they will automatically acquire additional iterated beliefs so that their common position to know becomes knowledge. Nothing automatically suggests itself. After all, the defender of Weak KK will not want to accept the BB principle (that if you believe  $p$ , then you believe that you believe  $p$ ), since in the presence of Weak KK that will imply KK.

<sup>6</sup> Thanks to an anonymous referee for suggesting the strategy here.

Third, imagine you are grappling with threats to the closure of knowledge under competent deduction (you've just read your Nozick and Dretske). You are looking at a zebra, know that it is a zebra, and know that you know it is a zebra. But since you are unsure whether closure is valid, you are unsure whether you know that it is not a cleverly disguised mule. You say that you don't know whether you know it is not a cleverly disguised mule. If the pragmatic principle above was correct, this would require for felicity that you are not in a position to know that you know it is not a cleverly disguised mule. If knowledge really is closed under deduction, then it follows by Weak KK that you do not know it is zebra after all. For all these reasons, we reject the pragmatic principle that uttering *I don't know whether I know that p* in settings where p is relevant requires for felicity that the speaker is not in a position to know whether they know that p. Of course, there may be other pragmatic explanations we have not unearthed. We leave this as future work for the defender of Weak KK.

### 3. Fragmentation

Our argument had two premises: Dudley only says what he believes, and Dudley never believes what he knows is false.<sup>7</sup> Defenders of KK may try to reject either premise. We'll start with the second premise. Perhaps people can believe something in one sense or 'guise' while knowing it in another sense. On this proposal, Dudley believes under one guise that he does not 100-know that he has hands. He knows under another guise that he does 100-know that he has hands.

The details of this response depend on what guises are. One theory is that guises are fragments. People believe and know things relative to different fragments. Each fragment is activated in different situations. Each fragment is logically closed and internally consistent, and each fragment satisfies KK (Greco 2015b).

For example, perhaps Dudley and Agnes have a classroom fragment and an outside fragment. In the classroom, they have beliefs that guide their philosophical musings; in the outside world, they have other beliefs that guide their ongoing relations with the external world. The beliefs of one fragment can contradict the beliefs of another; but each fragment is internally consistent.

The fragmenter wants to claim that people only deny KK relative to their classroom fragment. But when they are in the outside world, they do not deny KK. Sadly, the fragmenter cannot claim this. The problem is that the defender of KK says that each fragment satisfies KK.

Tim thinks he doesn't 100-know he has hands, because he thinks he doesn't 100-know any empirical truth about the world. The fragmenter interprets Tim as believing this only relative to the classroom fragment. But here's the problem. Presumably Tim knows some empirical truth about the world relative to the classroom fragment. As Tim shouts at you about KK, he points to his hands as an example of something he doesn't

---

<sup>7</sup> More carefully: we assume that there is a class of agents who deny KK in the way that Dudley does. Our premises are that people in this class only say what they believe, and never believe what they know to be false.

100-know, and he tells you that he does have hands. But if KK holds in the classroom fragment, and if Tim knows he has hands in the classroom fragment, then it immediately follows that he knows in the classroom fragment that he 100-knows he has hands. The KK fragmenter is forced to say that when Tim denies that he 100-knows any empirical truth about the world, he can only do so relative to a fragment that knows zero empirical truths about the world.

Imagine a tougher opponent to KK, who denies that they 100-know anything at all, including even the claim that they believe they fail to 100-know anything. Suppose that their classroom fragment knows some claim *p*. It follows that their classroom fragment knows they 100-know *p*. So what explains why this tougher person is in the classroom denying that they 100-know *p*? The KK fragmenter is forced to say that the tougher opponent knows nothing at all relative to the fragment that denies KK. Again, radically combative, but now with fragments.

A close cousin of the fragmenter is the contextualist (for a representative sample, see Greco 2014). The KK contextualist says that Dudley and his ilk are failing to notice shifts in the conversational context. When Dudley says that he knows he has hands, he is in an ordinary context (similar to the outside fragment). When Dudley denies that he 100-knows he has hands, he is in a more demanding context (similar to the classroom fragment). In both contexts, the KK principle is true. But different knowledge claims are true and false in each context. In the ordinary context, Dudley knows and 100-knows he has hands. In the more demanding context, Dudley neither knows nor 100-knows he has hands.

This contextualist faces similar challenges to the fragmenter. Again return to Tim, who denies 100-knowing any empirical claim about the world. This contextualist is forced to say that when Tim denies that he 100-knows any empirical truth about the world, he thereby enters a context in which it is false to say that he knows any empirical truths about the world. Similarly, the tough opponent who says he doesn't 100-know anything has uttered a 'knowledge bomb', creating a context in which he doesn't know anything.

As philosophers, we inquire into the nature of reality under conditions of profound uncertainty. It is possible that we will never truly know the answer to the questions we investigate. The proper response to this uncertainty is intellectual humility: recognition of the fact that we often do not know certain key fundamental truths about the nature of value, reality, and knowledge itself. Such humility sits in tension with KK. Even in its contextualist variant, KK tends to produce the result that in contexts in which humility is taken seriously, nothing at all is known.

The contextualist faces further challenges. First, they owe us an account of how the context shifts, and this is no easy task once KK is assumed. For example, contextualists of yore have flirted with the pragmatic principle that when you assert something, you tend to create a context in which you know what you have said (Lewis 1996). In this way, when Dudley starts to say he has hands, he begins to create a context in which he knows and thus 100-knows he has hands; but when he goes on to deny that he 100-knows he has hands, the context is destroyed.

Contextualists can use this kind of principle to explain the badness of asserting that one doesn't know anything: this assertion attempts to create a context in which you know you don't know anything; but there is no such context.

Once the contextualist accepts KK, this pragmatic principle overgenerates. We've suggested that there are contexts in which someone can reasonably assert 'I don't 100-know anything'. The contextualist pragmatics would say that this creates a context in which you know that you don't 100-know anything. This is in principle possible if KK is false. But if KK is true, there is no context in which you know (and thereby 100-know) that you don't 100-know anything.

Another interesting question for the contextualist concerns belief. We could have run our argument by beginning with the BB principle, that you believe you believe whatever you believe; in that case, we could imagine that Dudley goes around claiming that he doesn't 100-believe he has hands. For this reason and others, the KK contextualist should feel pressure to generalize their story from knowledge to belief (or to deny BB and also follow Lewis 1996 in denying that knowledge requires belief). Greco 2015b offers one solution: you believe  $p$  when you are in a state that is subjectively indistinguishable from (or feels the same from the inside as) knowing  $p$ . When the context raises the standards for knowing, it will automatically raise the standards for believing. Given KK, this leads to serious problems. Again imagine the tough opponent, who claims that they don't 100-know anything. Since they sincerely assert this, they believe it. By the proposed belief contextualist, it follows that they are in a state subjectively indistinguishable from knowing (by the standards of this context) that they don't 100-know (by the standards of this context) anything. By if KK holds in this context, it follows that they are in a state subjectively indistinguishable from 100-knowing that they don't 100-know anything. But there is no such state, because you can't 100-know that you don't 100-know anything. So there is no such belief. So there is no explanation of why the tougher opponent would say what they said.

#### 4. Sentential Guises

Another theory of guises is sentential. Lois Lane believes under one guise that Clark Kent does not fly. Under another guise, she knows that he does fly, because she knows that Superman flies. When Lois believes that Clark Kent does not fly, she does so by rejecting the sentence *Clark Kent flies*. When Lois knows that he does fly, she does so by accepting the sentence *Superman flies*. According to this proposal, a single proposition can be expressed using two different sentences. In such cases, you can believe the proposition is false while knowing it is true, because you reject one sentence while accepting another. To believe or know a proposition is to believe or know it under some guise or other.

As one application, the sentential theory can explain the limits of logical omniscience. On one view, there is only one logically necessary proposition, but it is expressed by infinitely many sentences. If you accept the sentence *either I have hands or I do not have hands*, then you believe every logical truth, because there is only one. But you don't believe the logical truth under every guise, because you do not accept some of the complex sentences that

express it. You know the logical truth under the guise *either I have hands or I do not have hands*. But you do not know it under other guises.

The defender of KK can't say exactly this. Dudley knows he has hands, because he accepts some sentence  $s$  which expresses the proposition that he has hands. In order to know that he knows he has hands, Dudley has to accept some sentence  $s^*$  which expresses the proposition that he knows he has hands. But what sentence could  $s^*$  be?  $s^*$  can't be  $s$ , because  $s$  expresses the distinct proposition that Dudley has hands. KK implies that accepting the sentence *I have hands* will require also accepting some distinct sentence, such as *I know I have hands*. The KK defender has their work cut out for them here.

Here is another challenge for this type of KK defender. There is a tolerant tradition about believing under guises (for discussion, see Percus and Sauerland 2003, Dorr 2014, and Lederman 2021 among others). In the tolerant tradition, it can be true to say that Lois doesn't know Clark Kent flies, even though she knows under a Superman guise that Clark Kent flies. Our reports about ignorance can be true, because in context there is quantifier domain restriction over guises. For example, in some natural contexts it is true to say that Lois doesn't know Clark Kent flies, because she does not know this under the guise *Clark Kent flies*. This tolerant tradition explains how we came to know that Hesperus is Phosphorus; at one time, we did not know that Hesperus was Phosphorus, and then later we learned that this was so.<sup>8</sup>

Now return to Dudley. This tolerant account will grant that (in some natural contexts) it is true to say that Dudley does not 100-know he has hands. After all, he does not know that he 99-knows he has hands under the guise *he 99-knows he has hands*. For this reason, the tolerant theorist will grant that it is true to say that: Dudley knows he has hands even though he doesn't 100-know he has hands. But at this point, the tolerant theorist has granted that it is true to say that KK is false.<sup>9</sup> If the KK defender seeks to avoid this result, they will be forced to deny the tolerant view, saying instead that it is strictly speaking false to say that Dudley doesn't 100-know he has hands. This in turn will push them to say that it is strictly speaking false to say that Lois Lane doesn't know Clark Kent flies (see Salmon 1986, Soames 1987).

Return to the analogy with love. Agnes is not sure what love is, so she's not sure whether she loves Boris. It's not like there are two guises, and she believes she loves Boris under one guise and believes under another guise that she doesn't love Boris. She just doesn't have an opinion.

---

<sup>8</sup> It's not just that it is true that Lois doesn't know that Clark Kent flies. She can also for the same reason know she doesn't know that Clark Kent flies, because she will know under a relevant guise that she doesn't know under a relevant guise that Clark Kent flies, because the relevant guises all involve the name *Clark Kent*.

<sup>9</sup> These tolerant concessions may undermine the explanatory power of KK. Return to the two generals engaged in rational coordination. If I want to explain their ability to coordinate, I must truly say that the first general knows that the second general knows that the first general knows ... that the second general knows they will attack. To do so, I need to say truly that the first general 100-knows that he will attack. But imagine that the first general is Dudley. How can I say this while also saying truly that Dudley doesn't 100-know he will attack? The tolerant theorist will not be able to explain in one breath Dudley's apparent ignorance and his ability to coordinate.



Knowledge is like love. If Agnes is agnostic about exactly what knowledge is, then she can fail to have an opinion about whether she 100-knows she has hands. It's not that under one guise she believes she 100-knows, and under another guise she believes she doesn't. Agnes just doesn't have an opinion under any guise about whether she 100-knows she has hands.

Or so we think. But we also think it is fruitful to explore how the defender of KK might develop a theory of sentential guises. In the next section, we do so.

## 5. Bridge

Say that a sentence is a knowledge attribution when it is of the form *you know q*, for some *q*. We think the defender of KK should accept the following principle:

**Bridge.** For any sentence *s*, you know under guise *s* that *p* only if there is some knowledge attribution *s'* where *s'* means that you know *p*, and you know under guise *s'* that you know *p*.

There is an ordinary guise by which someone knows that they know *p*. It is the guise involved in accepting the sentence *I know p*. The idea of Bridge is that first-order knowledge is defeated by uncertainty under the ordinary guise about whether you know *p*. For example, Dudley knows under the guise *he has hands* that he has hands. Bridge then says that Dudley knows under the guise *he knows he has hands* that he knows he has hands. On this proposal, Dudley can go around saying that he fails to 100-know he has hands, without destroying his first order knowledge of having hands. But when a more disreputable character goes around saying he doesn't know that he knows he has hands, this destroys his first order knowledge he has hands.

Bridge allows Dudley to fail to know under the guise *he knows he 100-knows he has hands* that he 100-knows he has hands. To see how, consider the following strengthening of KK:

**KK Identity.** The proposition that you know *p* is identical to the proposition that you know you know *p*.

KK Identity implies that the proposition that Dudley knows he has hands is identical to the proposition that Dudley knows he knows he has hands. This means that once Dudley knows under the guise *he knows he has hands* that he knows he has hands, it follows that he knows under the guise *he knows he has hands* that he 2-knows he has hands. Repeated applications of KK Identity imply that Dudley knows under the guise *he knows he has hands* that he 100-knows he has hands. But this is compatible with Dudley believing under the guise *he*

*doesn't 100-know he has hands* that he doesn't 100-know he has hands. In this way, Bridge and KK-Identity reconcile KK with the theory of sentential guises.<sup>10</sup>

Unfortunately, we think that this guise-relative version of KK does not do the same explanatory work as the naive KK principle. Return to dubious assertions. Imagine that Dudley says *I have hands, but I don't know that I know that I have hands*. Does the guise-sensitive KK theory explain why this sentence is strange?

In order to explain why the sentence is strange, we need to predict that Dudley knows that he does not know that the sentence is true. Otherwise, Dudley will not realize that the knowledge norm of assertion forbids saying a dubious assertion, and so Dudley would say it.

Here, the dialectic is complicated. First, Bridge and KK-Identity do imply that Dudley fails to know under any guise that: he has hands and doesn't 100-know he has hands.<sup>11</sup> But this isn't enough to explain the badness of dubious assertions. Here is an analogy. I can prove to you that Lois can't know that Clark Kent doesn't fly and Superman flies. My proof is that Clark Kent is Superman. But this proof is unavailable to Lois. And so asserting the sentence *Clark Kent doesn't fly and Superman flies* is not dubious. Lois will assert this because she believes under the guise *she knows Clark Kent doesn't fly* that she knows Clark Kent doesn't fly, and she believes under the guise *she knows Superman does fly* that she knows Superman does fly.

In an analogous way, this theory does not explain why it would be strange for Dudley to assert the sentence *I have hands without 100-knowing I have hands*. Dudley will utter a sentence if he believes it is permissible to do so. Let's grant that Dudley accepts the knowledge norm on assertion. So he believes it is permissible to assert the

---

<sup>10</sup> Here is one model of knowledge (generalizing Greco 2014 to guises) that accepts Bridge and KK Identity. You know  $p$  under guise  $g$  iff you believe  $p$  under guise  $g$ , conditions are normal for your believing  $p$  under guise  $g$ , and normally if you believe  $p$  under guise  $g$ ,  $p$  is true. In addition, normality freely iterates: if normally  $p$ , then normally normally  $p$ . Propositions are sets of worlds. Normality obeys one further condition, which corresponds to Bridge: conditions are normal for your believing  $p$  under guise  $g$  only if you there is some sentence  $s'$  which means that you know  $p$ , where you believe under  $s'$  that you know  $p$ . The resulting model validates KK. Suppose you know  $p$ . Then you know  $p$  under some guise  $g$ . So you believe  $p$  under guise  $g$ , conditions are normal for your believing  $p$  under guise  $g$ , and normally if you believe  $p$  under guise  $g$ ,  $p$  is true. Now we can show that you know that you know  $p$ . By the corresponding principle to Bridge, we can infer that you believe under the guise *you know  $p$*  that you know  $p$ . Since you do know  $p$ , the free iteration of normality implies that normally you know  $p$ . So normally, when you believe under the guise *you know  $p$*  that you know  $p$ , it is true that you know  $p$ . It follows that you know under the guise *you know  $p$*  that you know  $p$ . It follows that you know you know  $p$ . KK is valid. Finally, if propositions are sets of possible worlds, it follows that KK Identity is valid.

<sup>11</sup> Again, say that Dudley knows  $p$  iff he knows it under some guise. First, say that a sentence is assertable only if you know the proposition expressed under the guise of that sentence. Then the relevant question is whether Dudley can know that has hands without 2-knowing he has hands, under the guise *he has hands without 2-knowing he has hands*. Suppose so. Second, we assume that if you know a proposition under the guise of a conjunctive sentence, then you know each propositional conjunct under the guise of each conjunct. Now it follows that Dudley knows under the guise *he has hands* that he has hands. By Bridge, it follows that Dudley knows under a guise like *he knows he has hands* that he knows he has hands. But we can also infer that Dudley knows under the guise *he doesn't 2-know that he has hands* that he doesn't 2-know he has hands. By the factivity of knowledge, it follows that he doesn't 2-know that he has hands. This contradicts his knowing under some guise that he knows he has hands.

sentence *I have hands without 100-knowing I have hands* if he believes he knows under the guise *I have hands without 100-knowing I have hands* that he has hands without 100-knowing he has hands. But nothing in the above blocks Dudley from believing this. Imagine I ask Dudley the question *do you have hands?* He will answer yes, because he believes under the guise *I know I have hands* that he knows he has hands. Now imagine I ask Dudley the question *do you know you know you have hands?* He will answer no, because he believes under the guise *I know I don't know I know I have hands* that he knows he doesn't know he knows he has hands. So he will assert the sentence *I have hands and don't know I know I have hands*.<sup>12</sup>

## 6. Diagonalization

We have explored the prospects of denying the principle that when you believe *p*, you do not know that *p* is false. Now consider our other starting premise: that if someone utters the sentence *I don't 100-know I have hands*, then they believe that they don't 100-know that they have hands. One way to deny this premise appeals to the theory of diagonalization (Stalnaker 1978). According to this theory, our beliefs about the non-linguistic world are systematically conflated with beliefs about language. Dudley believes that he has hands. Dudley also believes the sentence *I have hands* is true. Philosophical confusion results from mistaking one kind of belief for the other.

For example, the diagonalizing KK-er will claim that Dudley knows he has hands, and knows that he 100-knows he has hands. Crucially, however, Dudley does not know that the sentence *I 100-know I have hands* is true. Dudley is in the grip of a false theory about the word *know*. He thinks that *know* expresses a relation that does not freely iterate. So Dudley does not know that *I 100-know I have hands* expresses a truth.

Dudley utters the sentence *I don't not 100-know I have hands*. He utters this sentence because he thinks the sentence is true, and Dudley tries his best to say what is true. But he does not actually believe he does not 100-know he has hands.<sup>13</sup>

We are unimpressed. Return to the case of love. Agnes is unsure of what love is, and so she is unsure of whether she loves Boris. Here, the diagonalizer could intervene, arguing that Agnes does know that she loves Boris; Agnes is merely ignorant of the fact that the sentence *Agnes loves Boris* is true. This proposal strikes us as veering towards madness. Does all ignorance about the nature of worldly relations turn out to be ignorance of language?

---

<sup>12</sup> Here's a further problem for Bridge. Some defenders of Bridge may want to embrace the tolerant theory of guises. Recall from footnote 6 that the tolerant theory grants that Lois Lane can know that she does not know that Clark Kent flies. In the setting of debates about KK, the tolerant theory will then grant that Dudley can know that: he knows he has hands without 100-knowing he has hands. Any such theory is straying far from a naive embrace of KK.

<sup>13</sup> Note that the defender of diagonalization must in general deny that people know the disquotation principle that *p* iff '*p*' is true. Otherwise, Dudley could infer from the fact that he knows he has hands that the sentence 'he knows he has hands' is true. In place of this principle, the defender of diagonalization should say that Dudley merely knows that the sentence '*p* iff '*p*' is true' is true.

At any rate, we feel strong pressure to analogize Agnes's attitudes towards love and knowledge. If she is ignorant about who she loves, then she can also be ignorant about what she knows.

Returning to a familiar theme, the diagonalization defense of KK risks undercutting the explanatory power of KK. Recall that the defender of KK seeks to explain why the dubious assertion *I have hands but I don't know that I know I have hands* is strange to say. But Dudley will say a sentence from his own language if he believes he knows the sentence is true, regardless of whether he believes the proposition it expresses. The theory under discussion successfully predicts that Dudley knows that he does not know that: he has hands and doesn't know that he knows he has hands. But the problem is that the theory allows Dudley to believe that he knows that the sentence *I have hands but I don't know that I know I have hands* is true. In that case, he may well assert it.<sup>14 15</sup>

## 7. Closure

We think KK is false because we say so. Some may worry that our objection overgenerates. Does knowledge fail to satisfy every interesting structural condition, simply because we say so? No. But the question is illuminating, and one good case study is the closure principle, which says (roughly) that knowledge is preserved by valid arguments. Consider a naive version of closure, which says that if you know p, and if p necessitates q, then you know q.

Fred is looking at Zach the zebra. Fred knows that Zach is a zebra, and says so. But, enamored with anti-closure literature, he also says that he doesn't know that Zach is not a cleverly disguised mule. As long as KK fails, there is no challenge yet for naive closure. Fred knows Zach is a zebra, and therefore knows that Zach is not a cleverly disguised mule. But he may not know that he knows that Zach is not a cleverly disguised mule, and so he may falsely believe that he does not know Zach is not a cleverly disguised mule, and so he may say that he does not know that Zach is not a cleverly disguised mule.

But now suppose that Fred knows that he knows Zach is a zebra. If naive closure is correct, then knowing that Zach is a zebra necessitates knowing that Zach is not a cleverly disguised mule. By a second order application of naive closure, if Fred knows that he knows Zach is a zebra, then he knows that he knows Zach is not a cleverly disguised mule. But if Fred knows that he knows Zach is not a cleverly disguised mule, then he does not believe that he doesn't know Zach is not a cleverly disguised mule, and so he wouldn't say that he didn't know Zach was

---

<sup>14</sup> Likewise, the defender of diagonalization will predict that people will happily go around saying things like *I grant that Jane knows that she knows that she knows what time the movie starts, but does she know that she knows that she knows that she knows what time the movie starts?*

<sup>15</sup> The diagonalization strategy faces similar problems to the tolerant account of guises. Imagine that Chip has tried many times to multiply  $7 \times 43$ , and he can't figure it out. He realizes this about himself. The diagonalization account is designed to explain why it sounds good to say that Chip knows he doesn't know that  $7 \times 43 = 301$ . By similar reasoning, the diagonalization account will predict that it sounds good to say that Dudley knows that he knows that he has hands without 100-knowing he has hands. But if the diagonalization account grants that the claim 'People can know KK is false' typically communicates a truth, then one starts to lose a grip on the sense in which it vindicates KK.

not a cleverly disguised mule (assuming Fred is like Dudley, and does not go around believing things he knows to be false).

To explain Fred's saying *I don't know Zach is not a cleverly disguised mule*, the defender of naive closure must deny that Fred knows that he knows Zach is a zebra (or else set off on the adventures through fragments, sentential guises, and diagonals familiar from earlier in the paper). Unlike the defender of KK, however, the closure defender can allow that Fred knows that Zach is a zebra. The idea is that Fred's mistaken views about closure destroy his higher order knowledge, but do not destroy his first order knowledge.<sup>16</sup>

Return again to love. Imagine that in addition to accepting a false theory of knowledge, Fred also accepts a false theory of love. Fred believes that you love someone only if you write about them in your diary. Fred doesn't own a diary, but he does actually love his friends and family. In this case, Fred loves people without knowing that he loves people. In this respect, knowledge may be like love. Fred knows Zach is not a cleverly disguised mule, but Fred believes a false theory of knowledge. Because of this, Fred doesn't know that he has knowledge. But unbeknownst to him, he does have knowledge.

We've now shown that the defender of naive closure must deny any second-order knowledge to anyone who denies naive closure. For those who find this response radically combative, one option is to accept a weaker form of closure.

For example, one strategy is to relativize closure to guises. To see how this would work, return to the problem of logical omniscience. Naive closure implies that if you know anything, then you know every logical truth, since every proposition necessitates every logical truth. To account for putative failures to know every logical truth, we might again appeal to the idea that you can know a proposition under one guise and not another, for example by accepting one sentence but not another.

One guise-relative closure principle goes as follows, where  $s$  and  $s'$  are sentences that express the propositions that  $p$  and that  $q$ : if you know under the guise  $s$  that  $p$ , and you know under the guise  $s$  necessitates  $s'$  that [ $p$  necessitates  $q$ ], then you know under the guise  $s'$  that  $q$ .

Imagine that you don't accept the sentence *it is not not not false that either I have hands or I do not have hands*, but you do accept the sentence *I have hands or I do not have hands*. You know under the guise *either I have hands or I do not have hands* but not under the guise *it is not not not false that either I have hands or I do not have hands* that it is not not not false that either I have hands or I do not have hands. Guise-relative closure allows

---

<sup>16</sup> The closure defender must reject a bit more than Fred's own second-order knowledge. They must also reject Fred having interpersonal knowledge as well. Imagine that Fred sees George looking at Zach the zebra. Fred asserts that George doesn't know that Zach is not a cleverly disguised mule. By parity of reasoning to our previous argument, it follows from naive closure that Fred doesn't know that George knows that Zach is a zebra. So Fred doesn't know what George knows. Generalizing, naive closure implies that closure deniers lack knowledge about what anyone knows.

this. You don't know under the guise *either I have hands or I do not have hands* necessitates *it is not not not false that either I have hands or I do not have hands* that [either I have hands or I do not have hands necessitates it is not not not false that either I have hands or I do not have hands]. For this reason, you don't know under the guise *it is not not not false that either I have hands or I do not have hands* that it is not not not false that either I have hands or I do not have hands, even though you know under the guise *either I have hands or I do not have hands* that either I have hands or I do not have hands (and that it is not not not false that either I have hands or I do not have hands).

Guise-relative closure makes two interesting predictions about Fred. First, guise-relative closure implies that Fred knows under the guise *Zach is not a cleverly disguised mule* that Zach is not a cleverly disguised mule. After all, Fred knows under the guise *Zach is a zebra* necessitates *Zach is not a cleverly disguised mule* that Zach is a zebra necessitates Zach is not a cleverly disguised mule.

Second, guise-relative closure is compatible with Fred possessing second-order knowledge about zebras, and lacking second-order knowledge about mules. Since Fred rejects guise-relative closure, Fred does not know under the guise [*Fred knows that Zach is a zebra*] necessitates [*Fred knows that Zach is not a cleverly disguised mule*] that ([Fred knows that Zach is a zebra] necessitates [Fred knows that Zach is not a cleverly disguised mule]). Since Fred does not know this, Fred can know under the guise [*Fred knows that Zach is a zebra*] that Fred knows that Zach is a zebra, without knowing under the guise [*Fred knows that Zach is not a cleverly disguised mule*] that Fred knows that Zach is not a cleverly disguised mule. In this way, guise-relative closure escapes our challenge.

Here, the dialectic is familiar. While guise-relative closure can escape our challenge, it does not have the explanatory strength of naive closure. Recall that defenders of KK use that principle to explain the infelicity of dubious assertions like *I have hands but don't know that I know I have hands*. Similarly, defenders of closure use this principle to explain the infelicity of abominable conjunctions like *Zach is a zebra, but I don't know whether Zach is not a cleverly disguised mule* (see DeRose 1995). Just as with Bridge, however, guise-relative closure does not explain the infelicity of the abominable conjunction. Unlike naive closure, guise-relative closure allows that Fred can know that he knows Zach is a zebra without knowing that he knows Zach is not a cleverly disguised mule. Guise-relative closure also allows Fred to believe that he knows he is in this predicament, and so allows him to reasonably assert abominable conjunctions.<sup>17</sup>

Returning to naive closure, it is also worth considering more robust challenges to the principle. Before, we merely imagined that Fred asserted that he did not know that the animal he encountered was not a cleverly disguised mule. But it is worth considering whether closure is threatened by Fred taking even more forceful attitudes. For example, imagine a case where Fred actively doubts that the animal is not a cleverly disguised mule.

---

<sup>17</sup> Another reason to accept closure concerns the relationship between knowledge and evidence. If evidence simply is knowledge, and if rational agents conditionalize their credences on all and only the evidence, then knowledge must satisfy naive closure. After all, when you conditionalize on a proposition, you also conditionalize on anything necessitated by it. Guise-relative closure threatens to sever this close connection between knowledge, evidence, and rational credence.

In that case, naive closure implies that Fred simultaneously knows and yet also doubts that the animal is not a cleverly disguised mule.<sup>18</sup>

To make sense of cases like this, it's worth first thinking about what doubt is. The case is hard to interpret if doubt were merely a matter of low to middling credence: if Fred is probabilistically coherent and assigns high probability to the animal being a zebra, then he must also assign high probability to the animal not being a cleverly disguised mule. Instead, we suggest thinking of doubt in terms of the absence of outright belief, as distinct from credence. The situation then is that Fred is unwilling to believe that the animal is not a cleverly disguised mule.

So far, we've focused on naive closure, and the more forceful version of doubting Fred is indeed a threat to the principle. (On the other hand, we are unsure how dialectically effective this threat is, since defenders of naive closure for knowledge will also accept naive closure for outright belief, and so may deny that Fred so doubts. If belief is closed, then Fred's belief that the animal is a zebra implies that he believes the animal is not a cleverly disguised mule. By contrast, our own challenges have focused on the linguistic behavior of Dudley et al and what follows from it, without assuming from the beginning that the agents have a particular cognitive attitude.) But, analogously to our discussion of Weak KK, most proponents of closure actually accept more sophisticated principles. These principles are not immediately challenged by Fred's unwillingness to believe.

First, we can consider a Weak Closure principle, in analogy to Weak KK, which says that if you know  $p$  and  $p$  necessitates  $q$ , then you are in a position to know  $q$ .<sup>19</sup> Is Weak Closure threatened by Fred's doubts? Consider the Blockage principle that if you are unwilling to believe that  $p$ , then you are not in a position to know  $p$ . Blockage would be enough to make trouble: since Fred is unwilling to believe that the animal is not a cleverly disguised mule, Blockage implies that he is not in a position to know it is not a cleverly disguised mule; so Weak Closure fails.

The argument depends on Blockage. But we think Blockage is probably false.<sup>20</sup> After all, *being in a position* does not in general pattern in this way: someone can be in a position to apply for a promotion, because they meet all the relevant criteria, and yet be unwilling to actually initiate the application. Being in a position to know  $p$  has something to do with being in the state where if you did competently deduce  $p$  from your evidence, you would

---

<sup>18</sup> Thanks to an anonymous referee for suggesting this argument.

<sup>19</sup> One threat to Weak Closure comes for those who accept the principle that if you know that you will never know  $p$ , then you are not in a position to know that  $p$ . Imagine that  $p$  implies  $q$  by a complex route, and you know  $p$  but have not yet deduced  $q$  from  $p$ , and are wondering about  $q$ . Now you learn that you are about to die. You know that you will never know  $q$ . Nothing about learning you are about to die should threaten your knowledge of  $p$ . But Weak Closure would then imply that you are in a position to know that  $q$ , contradicting the principle that if you know that you will never know  $q$ , then you are not in a position to know that  $q$ .

<sup>20</sup> Similarly, we reject the principle that if you doubt something, then you are not in a position to know it. This is far too strong. If we think of doubt as the absence of belief, then this principle would be tantamount to saying that if you are in a position to know  $p$ , then you believe  $p$ . This is very close to equating knowledge with being in a position to know.

know it. Someone suggests p: I am unwilling to believe p without checking whether it follows from my axioms. But I might know the axioms, and thereby be in a position to know p nonetheless.

For a specific challenge to Blockage, imagine (perhaps contra KK) someone who knows p without being sure whether they know p. Imagine that the person has no doubts about closure, and imagine that closure is in fact valid. As the person is unsure whether they know p, they may be unwilling to proceed and infer p or q. For this reason, they may be unwilling to believe p or q. But as a matter of fact they are in a position to know p or q.

Can the argument against Weak Closure be reconstructed without reliance on Blockage? In our discussion of Weak KK, we relied on an alternative to Blockage: if you are justified in believing p, then you are not in a position to know p is false (call this Justification Blockage). Unlike Blockage, however, Fred's doubts do not conspire with Justification Blockage to threaten Weak Closure. Fred may be unwilling to believe that the animal is not a cleverly disguised mule. But he is not *justified* in believing it is in fact a cleverly disguised mule. After all, he knows and is justified in believing that it is a zebra. So we cannot infer that he is not in a position to know that it is not a cleverly disguised mule. Tellingly, we do think there are no plausible versions of Fred's case in which he is justified in believing that it is a cleverly disguised mule while also knowing it is a zebra.

Still, one interesting feature of Blockage is it allows us to systematically characterize the class of structural principles in epistemology that are vulnerable to these kinds of problems. Let R be any relation between a person S and a proposition p that is logically compatible with being unwilling to believe p. For each such R, we can then consider the structural principle saying that if S R s p, then S is in a position to know p. For example, in the case of naive closure, R is the relation of there being some q which is known by S and which necessitates p. In the case of Weak KK, p is itself of the form 'S knows q', and R is the relation that S stands to p iff p. If Blockage is true, then any structural principle of this form is false. Since S R s p is logically compatible with being unwilling to believe p, there will be a possible case where S R s p and where S is unwilling to believe p. By Blockage, in that case we have S R s p even though S is not in a position to know p. So the relevant structural principle fails. Similar remarks apply to Justification Blockage. Now let R be any relation between a person S and a proposition p that is logically compatible with being justified in believing not p, and consider any structural principle in epistemology saying that if S R s p, then S is in a position to know p. Again, we can infer the structural principle must fail, because there will be a possible case in which S R s p and yet S is justified in believing not p; Justification Blockage then implies that S is not in a position to know p. Crucially, however, naive closure is not an instance of this newest schema (although Weak KK is), because S satisfying the relation to p is incompatible with justifiably believing p is false.

Finally, it's worth flagging that many structural principles in epistemology are not threatened at all by Blockage. As one example, consider the thesis that safely believing p is sufficient for being in a position to know p (or to know) p. Safely believing that p is logically incompatible with being unwilling to believe p. So Blockage has no force at all against the principle that safe belief is sufficient for knowledge.



It is also worth considering another challenging case for closure principles, which relies on a different blockage principle.<sup>21</sup> Imagine now a version of Agnostic Agnes in the case of Closure, who is looking at Zach the zebra, and refrains from asserting that Zach is not a cleverly disguised mule. What explains Agnes's refusal to assert? One tempting thought is that Agnes does not assert because she does not know that Zach is not a cleverly disguised mule. This is an immediate challenge to naive closure, at least if Agnes knows that Zach is a zebra.

On further reflection, however, naive closure can be defended from the immediate challenge. When discussing KK, we did appeal to the principle that "Agnes doesn't refrain from answering questions when she knows the answer." But once we have given up KK, we can consider a more complex principle: Agnes doesn't refrain from answering questions when she knows that she knows the answer. This more complex principle allows that Agnes can refrain from answering a question when she knows the answer, if she suspects that she may not know the answer.<sup>22</sup> Crucially, however, defenders of KK cannot appeal to this strategy, because given KK the more complex principle is equivalent to the simple one. To summarize, then, Agnostic Agnes does not pose the same challenges for naive closure as it did for KK, because we can explain Agnes's speech behavior in terms of higher order ignorance.

With further resources, we can also create a challenge for Weak Closure.<sup>23</sup> There may be 'level-bridging' principles that connect higher order ignorance to first-order states. For example, consider the level-bridging principle that if you are justified in believing that you aren't in a position to know p, then you are justified in suspending belief about p. And now consider an analogue of our Justified Justine, who in this case has strong evidence that she does not know that Zach is not a cleverly disguised mule. If the level-bridging principle is correct, it follows that Justine is justified in suspending belief about whether Zach is not a cleverly disguised mule. Finally, consider the Justified Suspension Blockage principle, which says that if you are justified in suspending belief about p, then you are not in a position to know p. Given both the level-bridging principle and Justified Suspension Blockage, Justified Justine would pose a challenge to Weak Closure.

Still, the challenge to Weak Closure is importantly different from our earlier challenge to Weak KK, precisely because it relies on the level-bridging principle that if you're justified in believing that you aren't in a position to know p, then you are justified in suspending belief about p. Importantly, it isn't enough that Justine is *justified* in believing that the level-bridging principle is true. This would merely mean that Justine is *believing* that she is justified in suspending belief about whether Zach is a cleverly disguised mule. And this alone would not threaten Weak Closure. (Here, it may also be worth noting that it is difficult in this setting to reconcile the level bridging principle with a Bayesian conception of justification. On that conception, once Justine is justified in believing that Zach is a zebra, she is thereby justified in believing Zach is not a cleverly disguised mule, since the rational credence in the latter claim must be at least as high.)

---

<sup>21</sup> Thanks to an anonymous referee for suggesting this challenge.

<sup>22</sup> For further discussion of more complex norms governing assertion, see among others DeRose 2002, Benton 2013, and Williamson 2013,

<sup>23</sup> Thanks again to an anonymous referee.

There is yet another route to challenging Weak Closure on the basis of blockage principles. Instead of considering zebra cases, consider cases involving logical confusion. Imagine that Sam the logic student is checking whether an extremely long sentence  $s$  is a logical truth. Sam asks his professor for help, and the professor incorrectly tells Sam that  $s$  is a contradiction. This gives Sam good evidence that  $s$  is false. But imagine that  $s$  is actually a logical truth, and that if Sam were to sit down for ten minutes and attempt to prove it, he would succeed. In this case, Justification Blockage implies that Sam is not in a position to know  $s$ , since he is justified in believing  $s$  is false. But Weak Closure implies that Sam is in a position to know  $s$ , since everything he knows necessitates it. Cases like this may motivate the defender of Weak Closure to set off again on our earlier adventures through fragments, diagonals, and sentential guises.

Another reaction to this case would be to deny Justification Blockage. This could itself threaten part of our argument against Weak KK. Here, the crucial question is what it takes to be in a position to know. There are two possible conceptions. On the conservative conception, when we consider what an agent is in a position to know, we hold fixed the agent's evidence and justificatory condition, and consider how changes in their beliefs would lead to knowledge. On the liberal conception, when we consider what an agent is in a position to know, we consider how further episodes of reasoning could change their evidence to produce knowledge. These two conceptions make different predictions about what Sam is in a position to know. The liberal conception allows that Sam is after all in a position to know  $s$ , since he is in a position to deduce  $s$ , and this act of deduction would defeat his misleading evidence that  $s$  is false, and thereby destroy his previous justification in the negation of  $s$ . The conservative conception blocks Sam from being in a position to know, because the acts of deduction available to him would change his evidence. For these reasons, the liberal conception of being in a position to know is incompatible with blockage principles, because an agent's current justified beliefs are not relevant to the knowledge they might gain from further reasoning. (This would threaten part of our argument against Weak KK. But we already noted that even if some conceptions of being in a position to know would avoid our direct challenge, on these theories Weak KK would fail to explain the things that KK explains.) By contrast, these cases of logical confusion do not threaten the combination of Justification Blockage with the conservative conception of being in a position to know.

Another way to handle these logical confusion cases and the previous challenges is to rely on other closure principles. As one example, Williamson and others have accepted a Deduction Closure principle, which says that if you know  $p$  and competently deduce  $q$  from  $p$ , then you know  $q$ . The cases above produce no threat to Deduction Closure. Fred is unwilling to believe that the animal is not a cleverly disguised mule, which means that he has not competently deduced that the animal is not a cleverly disguised mule from his belief that the animal is a zebra. In this way, Blockage produces no threat to Deduction Closure, since agents who are unwilling to believe a conclusion have thereby failed to competently deduce it. Likewise, Justification Blockage can also be safely combined with Deduction Closure, without threat from zebra cases. Finally, Sam's logical confusion can also be safely combined with Deduction Closure. If Sam were to competently deduce  $s$ , his evidence would change, and he would no longer be justified in believing that  $s$  is a contradiction. In this new evidential state, he

would come to know *s*. It is also worth noting how Deduction Closure avoids terminological disputes about being in a position to know. Whether this notion is liberal or conservative has no bearing on Deduction Closure, because it replaces being in a position to know with the more direct question of what the agent knows if they do competently deduce a conclusion.

Deduction Closure also has good potential to retain and even enhance the explanatory power of naive closure. Like naive closure, it also has the potential to explain the infelicity of abominable conjunctions like *Zach is a zebra but I don't know whether Zach is not a cleverly disguised mule*. The picture is that when we consider the utterance of such abominable conjunctions, we naturally focus on agents who have competently deduced that Zach is not a cleverly disguised mule from the premise that Zach is a zebra. After all, the utterance of the conclusion makes it clear that the agent is consciously reasoning about these questions, and the inference is obvious. In other ways, Deduction Closure may even enhance the explanatory power of naive closure. Kripke famously defended closure by pointing out that competent deduction is one of the very best ways of extending

our knowledge. But according to naive closure, explicit deduction is merely a reenactment of what one already knows.<sup>24 25</sup>

---

<sup>24</sup> An anonymous referee proposes one more KK principle, analogous to Williamson's competent deduction version of closure. Reflection KK says that if you know p, then if you come to believe that you know p by competently reflecting on whether you know p, then you know that you know p. The notion of competent deduction is well studied and familiar from logic classes. By contrast it is rather less clear what the notion of competent reflection amounts to. But we won't try to survey the options for precisification here. Rather we confine ourselves to some general remarks about the strategy of developing a Reflection KK principle that is analogous to the competent deduction version of closure.

Overall, our arguments address this principle similarly to how they address Weak KK, which says that if you know p, then you are in a position to know that you know p. In fact, one way to understand 'being in a position to know that you know p' is precisely that if you were to come to believe that you know p by competently reflecting on whether you know p, then you know p.

In response to Weak KK, our first response is to consider Justified Justine, who is justified in believing she doesn't know p, on the basis of powerful evidence. Justified Justine is likewise potentially a threat to Reflection KK. If Justified Justine did form a belief that she knows p, she would *not* know that she knows p, at least if Justification Blockage is true. The problem is that she has too much evidence against knowing, so holding the facts of justification fixed she can't know..

To save Reflection KK from Justified Justine, one strategy is to make it harder to competently reflect. On this picture, Justified Justine's evidence against knowing p would make it impossible for her to form a belief that she knows p via *competent* reflection. On this more 'expensive' interpretation of competence, Reflection KK is compatible with Justification Blockage, and with the case of Justified Justine. But in this case, Reflection KK is dramatically weaker than KK, and weaker than Weak KK. Here, particularly relevant will be our second line of argument against Weak KK, arguing that Weak KK does not play the same explanatory role as KK. This line of argument will apply even more so to Reflection KK. Notice, however, that while Reflection KK cannot play the same explanatory role as KK, Deduction Closure can retain, and perhaps even surpass, the explanatory benefits of naive closure.

There are also further respects in which Reflection KK looks worse than Deduction Closure. Imagine that Fred is in the grips of Dretskean epistemology, and so believes he isn't in a position to know that Zach is not a cleverly disguised mule. Nonetheless, imagine that by quirk or whim Fred nonetheless does deduce that Zach is not a cleverly disguised mule. In such a case, it does seem to us that Fred could come to know that Zach is not a cleverly disguised mule, despite Fred's false views regarding the nature of knowledge. But analogous structures in the case of iterated knowledge generate different judgments. Imagine that you know you have hands, but that your degree of commitment to the proposition that you have hands is only just strong enough to count as believing. Imagine that you falsely believe that your commitment level is slightly below the required threshold for belief, and so falsely believe that you don't believe that you have hands. But again imagine that by quirk or whim you nonetheless 'reflect' that you know you have hands. After all, you do in fact believe and know you have hands, and so this belief might be able to activate whatever reflection processes you possess. In this case, we judge that you do not know that you know you have hands. In this way, Reflection KK faces challenges that Deduction Closure avoids.

<sup>25</sup> Finally, it is worth considering a final Weaker Closure principle, which says that if you know p, and you know that p necessitates q, then you are in a position to know that q. Weaker Closure is weaker than Weak Closure, because it requires

We've gone into some detail here to give a sense of the range of principles that are or are not threatened by the kinds of considerations in this paper. As the detail of our discussion here suggests, there is no simple recipe for constructing a general threat to all systematic principles in epistemology from either ignorance about the nature of knowledge or even direct unwillingness to believe. Rather, different principles in epistemology must be looked at with careful philosophical attention to see whether they are compatible or not with ignorance about the nature of knowledge. We think a valuable future research program would try to identify further principles that are vulnerable to the kinds of challenges raised in this paper.

## 8. Externalism

We are hardly the first to object to KK. Of the many objections to the principle, the one closest to our own concerns 'externalism'. According to externalists about knowledge, knowledge of  $p$  requires a condition that is 'external' the agent, in the sense that the agent can know  $p$  and believe they know  $p$  without knowing that the condition obtains.<sup>26</sup> For example, a common externalist condition is that knowledge requires believing via a reliable process. People can have a reliable belief without reliably believing they have one, and so without knowing they have one. On these grounds, many externalists have rejected KK.<sup>27</sup>

This objection is interesting, but it is different from our objection. Our question is not whether you know that a particular externalist condition obtains. Our question is whether you know that the condition is required for knowledge! The challenge isn't that knowledge requires reliably formed belief. The challenge is that if I don't know whether knowledge requires reliably formed belief, then I may not know what I know.

Our challenge strengthens the existing externalist critique. Bird and Pettigrew 2021 argue that in the cases of interest, you can know that the condition is necessary for knowledge without knowing that the condition obtains. Given a suitable closure condition, they show that this leads to KK (see their S4.1). This raises the question: what happens to KK when the relevant agents are ignorant about the nature of knowledge, for

---

the further condition that you know that  $p$  necessitates  $q$ . Weaker Closure escapes the threat of Sam the logic student, which undermined Weak Closure. Sam was justified in falsely believing that  $s$  is not a logical truth, and so Justification Blockage implied that Sam is not in a position to know  $s$ . But notice that Sam does not know that  $s$  is necessitated by other propositions he knows, and so this final closure principle does not imply that he knows  $s$ . More generally, Justification Blockage does not obviously threaten Weaker Closure. But Justified Suspension Blockage and the level bridging principle would lead to the failure of Weaker Closure. Finally, notice that Weaker Closure has no analogue in the case of KK, since there is no analogue to knowing that  $p$  necessitates  $q$ .

<sup>26</sup> See Bird and Pettigrew 2021 for a definition, and Okasha 2013 for further discussion.

<sup>27</sup> For endorsement of the challenge, see Williams 1995, Dretske 2004, Bird 1998, Pagin 2012, Schantz 2004, and Hemp 2006.

example by failing to know that a condition is necessary for knowledge? As we've argued at length, KK fails in these conditions as well.<sup>28 29</sup>

## Bibliography

- Beddor, Bob & Pavese, Carlotta (2018). Modal Virtue Epistemology. *Philosophy and Phenomenological Research* 101 (1):61-79.
- Benton, Matthew (2013). Dubious objections from iterated conjunctions. *Philosophical Studies*, 162:355–358.
- Bird, Alexander and Pettigrew, Richard (2021) "Internalism, Externalism, and the KK Principle" *Erkenntnis* 86: 1713-1732
- Cohen, Stewart & Comesaña, Juan (2013). Williamson on Gettier Cases and Epistemic Logic. *Inquiry: An Interdisciplinary Journal of Philosophy* 56 (1):15-29.
- Conee, Earl (2016) "Review of: Sanford Goldberg (ed.) Externalism, Self-Knowledge, and Skepticism", Notre Dame Philosophical Reviews
- Das, Nilanjan & Salow, Bernhard (2018). Transparency and the KK Principle. *Noûs* 52 (1):3-23.
- DeRose, Keith (1999). Solving the Skeptical Problem. In Keith DeRose & Ted A. Warfield (eds.), *Skepticism: a contemporary reader*. New York: Oxford University Press.
- DeRose, Keith. Assertion, knowledge, and context (2002). *Philosophical Review*, 111(2): 167–203. doi: 10.1215/00318108-111-2-167
- Dorr, Cian (2014). Transparency and the Context-Sensitivity of Attitude Reports. In Manuel García-Carpintero & Genoveva Martí (eds.), *Empty Representations: Reference and Non-Existence*. Oxford University Press. pp. 25-66.
- Goldman, Alvin (2004). *Pathways to Knowledge: Private and Public*. Oxford University Press USA.
- Greco, Daniel (2014). Could KK Be OK? *Journal of Philosophy* 111 (4):169-197.
- Greco, Daniel (2015a). Iteration Principles in Epistemology I: Arguments For. *Philosophy Compass* 10 (11):754-764.
- Greco, Daniel (2015b). Iteration and Fragmentation. *Philosophy and Phenomenological Research* 91 (3):656-673.
- Lederman, Harvey (2021). Fine-grained semantics for attitude reports. *Semantics and Pragmatics* 14 (1).
- Lewis, David K. (1996). Elusive knowledge. *Australasian Journal of Philosophy* 74 (4):549 – 567.
- Okasha, Samir (2013) "On a flawed argument against the KK principle" *Analysis* 73: 80–86

---

<sup>28</sup> There are a few spots in the existing literature that make an initial version of our argument. First, Bird and Pettigrew 2021 S4.3 note that internalism fails to entail KK because of potential ignorance about the analysis of knowledge. Second, Conee 2016 briefly observes that KK fails for agents who falsely believe that indefeasible evidence is required for knowing: they will know things without being justified in believing that they have indefeasible evidence, and so without being justified in believing that they know. Finally, Okasha 2013 also discusses threats to KK from subjects who don't know the correct theory of knowledge.

<sup>29</sup> This research was supported by the John Templeton Foundation, through the grant Humble Knowers, #62607.

- Percus, Orin and Uli Sauerland (2003). On the LFs of attitude reports. In Proceedings of the Conference “SuB7 – Sinn und Bedeutung”, ed. M. Weisgerber, 228–242. Arbeitspapier Nr. 114, FB.
- Salmon, Nathan. (1986). *Frege’s puzzle*. Ridgeview.
- Soames, Scott (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics* 15 (1):47-87.
- Stalnaker, Robert (1978). Assertion. *Syntax and Semantics (New York Academic Press)* 9:315-332.
- Willard-Kyle, Christopher (2020). Being in a Position to Know is the Norm of Assertion. *Pacific Philosophical Quarterly* 101 (2):328-352.
- Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.
- Timothy Williamson (2013). Response to cohen, comesañá, goodman, nagel, and weatherson on gettier cases in epistemic logic. *Inquiry*, 56(1):77–96.