Fragile Knowledge

Simon Goldstein
Australian Catholic University, Australia
simon.d.goldstein@gmail.com

Abstract

This paper explores the principle that knowledge is fragile, in that whenever S knows that S doesn't know that S knows that p, S thereby fails to know p. Fragility is motivated by the infelicity of dubious assertions, utterances which assert p while acknowledging higher order ignorance of p. Fragility is interestingly weaker than KK, the principle that if S knows p, then S knows that S knows p. Existing theories of knowledge which deny KK by accepting a Margin for Error principle can be conservatively extended with Fragility.

1 Introduction

Sosa 2009 introduces the phenomenon of 'dubious assertion', infelicitous utterances concerning higher order ignorance. In dubious assertions, an agent asserts a claim while raising doubts about her higher order epistemic standing with respect to p.

(1) #p, but I don't know whether I know that p.

This paper explains the infelicity of dubious assertions by defending a new principle about knowledge, Fragility. I say that knowledge is fragile, so that it cannot withstand the knowledge of higher order ignorance:

(2) **Fragility.** If S knows that S doesn't know that S knows p, then S doesn't know p.

Fragility implies that (1) is unknowable, and hence infelicitous given a knowledge norm on assertion.

The paper proceeds in several parts. §2 reviews extant work on dubious assertions. §3 explores Fragility in greater detail, considering why one might accept the principle, comparing a few alternative formulations of Fragility, and explaining how Fragility is related to the unknowability of dubious assertions. One main point is that Fragility is interestingly weaker than the KK principle:

(3) **KK.** If S knows that p, then S knows that S knows that p.

Defenders of KK have recently used dubious assertions to motivate the validity of KK. This paper suggests that such an argument is inconclusive. Dubious assertion can be explained without resorting to KK, as long as we accept Fragility.

To explore Fragility in more detail, §4 characterizes Fragility within epistemic logic, to show that Fragility is interestingly weaker than KK. §5 strengthens the case for Fragility by showing that Fragility can be added to extant theories of knowledge which reject KK on the basis of Margin for Error principles, where knowing p requires that p couldn't easily have been false. In particular, §5 develops a theory of knowledge which validates Fragility while invalidating KK and respecting a version of the Margin for Error requirement. §6 extends the theory to other types of dubious assertion.

2 Dubious assertions

The central data point for this paper, from Sosa 2009, is the infelicity of sentences which assert p while reporting higher order ignorance about p:

(1) #p, but I don't know whether I know that p.

To illustrate the infelicity of dubious assertions, Greco 2014 imagines an extended discourse in which an agent asserts p while later implying that they don't know that they know p.

(4) A: When did Queen Elizabeth die?
B: She died in 1603.
A: How do you know that?
B: I didn't say I know it.
A: So you're saying you don't know when Queen Elizabeth died?
B: I'm not saying that either. I'm saying she died in 1603. Maybe I know that she died in 1603, maybe I don't. Honestly, I've got no idea. But you didn't ask about what I know, did you? You just asked when she died. (Greco 2014 p. 667)

Such discourses sound incoherent, and for the same reason conjunctions like (1) are infelicitous.

The literature contains a few different reactions to dubious assertions. Sosa 2009 uses the data to challenge the knowledge norm of assertion (defended in Williamson 2000 for example).

(5) **KA.** S ought: assert p only if S knows p.

KA can explain the infelicity of Moore paradoxical sentences like:

(6) #p, but I don't know that p.

Such sentences are unknowable, and hence unassertable by KA. But Sosa 2009 suggests that KA undergenerates with respect to (1). The problem is that many defenders of KA reject the thesis that knowledge freely iterates.

(7) **KK.** If S knows that p, then S knows that S knows that p.

If KK fails, then there are agents who know p without knowing that they know p. But there seems to be no barrier to such agents knowing that they are in just this predicament. In that case, KA allows them to assert (1).

By contrast, other recent work (Stalnaker 2009 p. 404, Cohen and Comesaña 2013, Greco 2014, Greco 2015, and Das and Salow 2018) embraces the knowledge norm of assertion and uses the infelicity of (1) to motivate KK.[1] If KK is valid, then (1) is unknowable. For if S knows (1), then S knows p, and so by KK knows that she knows p. But if S knows (1) then she also knows that she doesn't know whether she knows p. But this contradicts the Factivity of knowledge.

(8) **Factivity.** If S knows p, then p.

Finally, Benton 2013 and Williamson 2013a offer explanations of the infelicity of (1) which rely on KA without KK. For example, Benton 2013 suggests that while asserting (1) satisfies KA, it violates secondary rational requirements that follow from KA. When agents are subject to a norm, they incur a secondary requirement to perform actions they believe satisfy the norm. Conversely, if they believe that they do not satisfy the norm in acting, then they are criticizable. To explain the infelicity of (1), this proposal could be enriched with the requirement that whenever someone fails to know whether they satisfy the primary norms for performing an action, they violate the secondary norms for performing that action. Similarly, Williamson 2013a analogizes assertions like (1) to paradoxical utterances like:

(9) Stand to attention!—and I don't know whether I have authority to order you to stand to attention.

These secondary explanations of (1) may ultimately succeed (although see Greco 2014 and Greco 2015 for skepticism).[2] But the rest of this paper pursues a more direct approach.

This paper holds fixed the knowledge norm of assertion and the infelicity of (1) and its ilk. The paper defends the thesis that sentences like (1) are infelicitous because they are unknowable. To explain the unknowability of (1), the paper develops and defends the following principle:

(2) **Fragility.** If S knows that S doesn't know that S knows p, then S doesn't know p.

Before proceeding, it's worth flagging that the phenomenon of dubious assertion extends beyond the data point in (1), in two respects. First, we get similar infelicities when we replace ignorance with other epistemic states, including

[1] See Smithies 2012 for analogous arguments in the case of justification.

[2] In particular, Greco 2015 suggests that dubious assertions like (1) are irrational to believe, not just bad to assert. Fragility can explain this further fact if we assume a weak form of a knowledge norm on belief: that it is irrational to believe anything which is *a priori* guaranteed to be unknowable. Similarly, suppose we accept the Reduction principle, discussed below, that S is justified in believing p if and only if for all S knows, S knows p. In that case if Fragility is known by S then S does not justifiedly believe (1), since S knows that (1) is unknowable.

doubt, belief, and justification. Second, similar assertions are dubious which involve even higher iterations of knowledge. For simplicity, the next few sections focus on (1). Once I have developed the theory in detail, I then explore more complex examples in §6.

3 Fragility

The thesis of this paper is that knowledge is fragile:

(2) **Fragility.** If S knows that S doesn't know that S knows p, then S doesn't know p.

This section explicates Fragility by exploring a few equivalent formulations. (2) says that knowledge is fragile, because (2) articulates a connection between knowledge and defeat. If you learn that you don't know that you know p, you learn that you are in some way epistemically defective with respect to p. If you learn that you are epistemically defective with respect to p, this knowledge defeats your knowledge of p. Knowledge of p is fragile in the face of evidence that one is not epistemically ideal with respect to p.[3]

Thinking about Fragility in terms of defeat helps clarify the relationship between Fragility and KK. Fragility allows that an agent can know p without knowing that she knows p. But things are different if the agent becomes aware that they are in such a predicament. If an agent learns they knew p while failing to know that they knew p, something changes in their epistemic position. New information about their non-ideal status leads to a failure of their knowledge of p.

To better illustrate Fragility, consider an example of higher order ignorance: the unwitting historian (Radford 1966, Feldman 2005).

> Jean insists that she knows nothing about English history. As a matter of fact, she studied English history in secondary school, but doesn't recall taking the course. She hasn't forgotten the content of what she learned, however. If you force her to guess as to matters such as when William the Conqueror landed in England, the dates of Queen Elizabeth's reign, and so on, she'll reliably respond correctly. But if told that her answers are correct, she'll be quite surprised, as she takes herself to have no way of knowing these facts. (Greco 2014 p. 658.)

Jean the unwitting historian is plausibly an example of higher order ignorance. Although she knows when Queen Elizabeth ruled, she doesn't know that she knows this. Fragility implies that there is something unstable about Jean's predicament. If Jean is apprised of her higher order ignorance, she either loses

[3]Here, I assume that anything an agent knows is part of her evidence. This is weaker than the principle that an agent's knowledge is exactly their evidence, a principle embraced in Williamson 2000.

her first order knowledge or gains second order knowledge. On the other hand, Fragility allows that Jean can believe that she doesn't know that she knows p in the original case. It only insists that this belief is not knowledge.

It's worth considering one more consequence of Fragility. Greco 2014 observes that pragmatic accounts of dubious assertion such as Benton 2013 and Williamson 2013a predict that higher order ignorance gives rise to assertoric dilemmas: cases in which a speaker has no rationally permissible response to their interlocutor.[4]

> Given the views [Benton 2013] suggests, speakers will find themselves in a sort of awkward dilemma whenever they know that P without knowing that they know. In such cases, while they will be able to permissibly assert that P, if their permission to assert that P is challenged, they will not be able to permissibly defend themselves. It strikes me as implausible that our conversational norms allow for such situations. (Greco 2014, p. 667.)

Fragility has a similar consequence. For suppose S knows p and doesn't know that they know p. Now suppose that S asserts p, and their interlocutor asks them whether they know p. How can they respond? They cannot answer 'yes', for they do not know that they know p. They cannot answer 'no', for they don't know that they don't know p. Strangely, they also can't answer 'I don't know', because Fragility implies that they don't know that they don't know that they know p. If they believe they know, they may say so, as Fragility allows them to know that they think they know. But suppose that they do not in fact believe they know. As we'll see in §6, Fragility forbids them from knowing that they don't believe they know, since this would imply that they know they don't know they know. In sum, Fragility along with the knowledge norm of assertion implies that in such cases there is simply no permissible response to their interlocutor. The best they can do might be the following:

(10) A: When did Queen Elizabeth die?
B: Queen Elizabeth died in 1603.
A: How do you know that?
B: I didn't say I know it.
A: So you're saying you don't know when Queen Elizabeth died?
B: I'm not saying that either. I'm saying she died in 1603.
A: So you're saying you don't know whether you know when Queen Elizabeth died?
B: No, I'm not saying that. All I'm saying is Queen Elizabeth died in 1603.

Such cases are assertoric dilemmas. Greco 2014 suggests that such situations should be ruled out by our conversational norms and best epistemology. I see no reason for such a sanguine view. Sometimes there is no way to make the best of a bad situation. Once an agent has fallen into higher order ignorance, perhaps they simply have no good way of responding to forceful inquiry on the matter.

[4]Perhaps the forced march Sorites is another example of an assertoric dilemma.

On the other hand, one might think that the very act of inquiry in the above may be a way of escaping higher order ignorance. Perhaps once S is asked whether they knows p, their epistemic position changes. Either they come to know that they know p, or they come to know that they don't know that they know p. In the latter case, Fragility implies that they also lose their knowledge of p. In such a case, S can assert that they don't know that they know p, and then retract their previous assertion of p. In this way, perhaps Fragility has a small advantage over pragmatic accounts of dubious assertion, since Fragility can predict how higher order defeat might resolve assertoric dilemmas.

To better understand Fragility, consider another of its equivalent forms:

(11) **Iterated Ignorance.** If S knows p and S doesn't know that S knows p, then S doesn't know that S doesn't know that S knows p.

In this form, Fragility encodes an iterative conception of higher order ignorance. Suppose that you know p but fail to know that you know p. In this case, you have higher order ignorance—ignorance about your knowledge. This ignorance iterates. Agents who know p without knowing that they know p are also agents who are ignorant of this fact.

For another formulation of Fragility, let's introduce the dual of knowledge, epistemic possibility. p is epistemically possible for S just in case it holds for all S knows, just in case p is consistent with what they know, just in case the agent does not know that p is false. Then Fragility embodies a kind of optimism about the epistemic possibility of iterated knowledge.

(12) **Optimism.** If S knows p, then for all S knows, S knows that S knows p.

When an agent knows p, they may fail to know that they know p. But Optimism says that even in such a case, it is epistemically possible for them that they know that they know. Optimism is optimistic, because it says that when we do know p, we can never rule out the possibility that we are in the better epistemic position of knowing that we know p. Optimism allows us to compare Fragility with KK straightforwardly. Fragility is strictly weaker than KK, since it replaces knowing that one knows with the epistemic possibility of knowing that one knows.

We can also understand Optimism in another way. Building on Lenzen 1978, Williamson 2000 (p. 46), Stalnaker 2006, Williamson 2013a, Rosenkranz 2018, and Carter and Goldstein 2021 we might define justified belief as a state that is epistemically indistinguishable from knowing.[5]

(13) **Reduction.** S is justified in believing p if and only if for all S knows, S knows p.

Given Reduction, Optimism and hence Fragility is equivalent to the JK principle:

(14) **JK.** If S knows p, then S is justified in believing that S knows p.

As Berker 2008 observes, Williamson 2000's arguments against KK do not

[5] For similar views, see Bird 2007 and Ichikawa 2014.

immediately extend to JK, since justified belief does not require safety from error. On the other hand, Reduction is a controversial thesis about justification, and so the connection between Fragility and BK is by no means conclusive.

I began the paper with a discussion of dubious assertion. The main data point is that dubious assertions are unknowable, so that:

(15) **Ignorance of the Dubious.** S doesn't know that: p and S doesn't know that S knows that p.

Suppose that anyone who knows a conjunction knows each conjunct, and vice versa. Then Ignorance of the Dubious is equivalent to Fragility. For suppose S knows the conjunction: p and S doesn't know that S knows that p. Then S knows p and S knows that S doesn't know that S knows that p, contradicting Fragility. Conversely, suppose that Fragility fails. Then there is some agent who knows p while knowing that they don't know that they know p. But then they can conjoin this knowledge, to learn that the conjunction p and S doesn't know that S knows that p, contradicting Ignorance of the Dubious.

As we saw above, defenders of KK have used Ignorance of the Dubious to motivate KK, because KK implies Ignorance of the Dubious. This paper undercuts that argument. In particular, we've now seen that Ignorance of the Dubious is equivalent to Fragility given modest assumptions. In the rest of the paper, I'll argue that Fragility is weaker than KK and interesting in its own right. §4 shows that a theory of knowledge can consistently embrace Fragility without accepting KK. §5 strengthens this argument by showing that Fragility is also compatible with a version of the Margin for Error principles that have motivated recent attacks on KK. Opponents of KK can strengthen their theories with Fragility in order to explain Ignorance of the Dubious. For this reason, Ignorance of the Dubious does not provide a compelling argument for KK.

4 Fragility in epistemic logic

My thesis is that knowledge is fragile. But what exactly does Fragility require of a theory of knowledge? Do we have any guarantee that Fragility is even a consistent principle, or that it really is weaker than KK? This section exploits familiar tools from epistemic logic to show that Fragility is consistent and weaker than KK.

I interpret the knowledge of a single agent as a modal necessity operator K, with epistemic possibility as its dual M. To do so, introduce an epistemic accessibility relation R, and say that Kp is true at world w just in case p is true at every world v R-accessible from w. Then p is true for all S knows (Mp) just in case there is some R-accessible world at which p is true.

Definition 1.

1. R is an epistemic accessibility relation which relates any world w to any world v that is epistemically possible for the agent in w.

2. Rw abbreviates $\{v \mid wRv\}$.

3. wR^2u if and only if $\exists v : wRv \ \& \ vRu$.

4. $\mathbf{A}$ abbreviates $\{w \mid w \models A\}$.

Definition 2.

1. $w \models p$ if and only if $w(p) = 1$

2. $w \models \neg A$ if and only if $w \not\models A$

3. $w \models A \wedge B$ if and only if $w \models A$ and $w \models B$

4. $w \models KA$ if and only if for every v in $Rw : v \models A$

5. $w \models MA$ if and only if for some v in $Rw : v \models A$

In epistemic logic, different properties of knowledge correspond to different properties of epistemic accessibility. Consider the requirement that knowledge is factive, so that anything known is true. In the framework above, Factivity corresponds to the reflexivity of epistemic accessibility: every world w is accessible from itself.

In this setting, the KK principle corresponds to the transitivity of accessibility. KKA is true at w just in case A is true at every world accessible from a world accessible from w. In other words, KKA is true at w just in case A is true throughout R^2w, where R^2 relates w and u just in case u can be reached by a world reached by w. In this way, iterated knowledge universally quantifies over an accessibility relation R^2 derived from epistemic accessibility. Now KK says that KA implies KKA. This means that whenever A is true throughout Rw, we are guaranteed that A is also true throughout R^2w. This itself is equivalent to the requirement that $R^2w \subseteq Rw$, so that u is accessible from w whenever u is accessible from v and v is accessible from w.

I now represent Fragility in epistemic logic. For simplicity, I focus on a particular form of Fragility, Optimism. In this formulation, Fragility says that if S knows p, then it is possible that S knows that S knows p.

(16) $KA \models MKKA$

Fragility corresponds to an interesting variant of transitivity. Fragility says that every world can see some world where every world reachable by two applications of accessibility can also be reached from the base world.

Definition 3. R is jump transitive if and only if $\forall w \exists v \in Rw : R^2v \subseteq Rw$.

Observation 1. Fragility is valid if and only if R is jump transitive.[6]

[6] By contraposition of Optimism: suppose R is jump transitive, and suppose that $w \models KMMA$. Then $\forall v \in Rw : R^2v \cap \mathbf{A} \neq \emptyset$. By jump transitivity, $\exists v^* \in Rw : R^2v^* \subseteq Rw$. So $Rw \cap \mathbf{A} \neq \emptyset$. So $w \models MA$. Conversely, suppose that R is not jump transitive. Then $\forall v \in Rw : \exists z \in R^2v : z \notin Rw$. Now let $\mathbf{A} = \{w \mid \exists v \in Rw : z \in R^2v \ \& \ z \notin Rw\}$. $w \models KMMA$,

Fragility corresponds to a coherent constraint on epistemic accessibility. This constraint is compatible with the reflexivity of accessibility, so that Fragility is compatible with Factivity. Fragility also has some small consequences for any modal operator that satisfies it: for example, it implies that accessibility is serial, so that every world sees some other world. This in turn corresponds to the requirement that KA implies MA, itself a consequence of Factivity.

Our characterization of Fragility allows us to consider the relationship between KK and Fragility. First, we can see that KK and Factivity imply Fragility. For suppose R is transitive and reflexive. Then every world w trivially sees a world v (in particular, itself) where $R^2v \subseteq Rw$. By contrast, Fragility does not imply KK. Epistemic accessibility can be jump transitive without being transitive.[7]

Of course, Fragility could be logically weaker than KK without being philosophically weaker. Perhaps KK is the only plausible theory of knowledge that validates Fragility, thereby explaining Ignorance of the Dubious. This response is unconvincing. We saw in the last section that Fragility can be understood in terms of at least three philosophical intuitions about knowledge: (i) that learning one's epistemic status with respect to p is non-ideal can defeat one's knowledge of p; (ii) that higher order ignorance of a certain kind iterates; and (iii) that knowing p always leaves open the possibility that one's epistemic status with respect to p is even better. The next section takes this defense a step further. I develop a theory of knowledge which combines Fragility with a version of the Margin for Error principles that motivate recent attacks on KK.[8]

5 Fragility and margins for error

Stemming from Williamson 2000, much recent criticism of KK relies on some kind of Safety or Margin for Error principle. This section develops a theory where knowledge satisfies both Fragility and a version of the Margin for Error principle. Fragility requires that the margin for error when appearance perfectly matches reality is sufficiently smaller than the margin for error at all other worlds. The result is that opponents of KK can explain dubious assertion by enriching their theory with further constraints on knowledge.

since $\forall v \in Rw : v \models MMA$, since for arbitrary such v we have $\exists z, u : vRu \ \& \ uRz \ \& \ z \in \mathbf{A}$. But $w \not\models MA$, since $\neg\exists v \in Rw : v \in \mathbf{A}$. This proof implicitly relies on the definition of an epistemic frame, a class of all interpretations of atomic sentences in which accessibility is jump transitive. For simplicity I suppress this complication by assuming that every set of worlds is the meaning of some sentence. In addition, this proof focuses on an equivalent version of Fragility, that $KMMA$ implies MA

[7] Jump transitivity concerns the relationship between Rw and R^2v for some v or other. Transitivity concerns the relationship between Rw and R^2w. Jump transitivity is also distinct from another weakening of transitivity we might call 'possible transitivity': that every world sees another world where accessibility is transitive ($\forall w \exists v \in Rw : R^2v \subseteq Rv$). These properties differ: jump transitivity compares the doubly accessible points at v with the singly accessible points at w, not the singly accessible points at v.

[8] For other work on weakenings of positive and negative introspection in epistemic logic, see San 2019.

5.1 Introduction

Safety says that knowledge is incompatible with the chance of being wrong.[9]

(17) **Safety.** If S knows that p, then S's belief that p could not easily have been false.

Williamson 2000 exploits Safety principles to undermine the KK principle. In particular, the relevant notion of 'easily being wrong' fails to iterate. Whether one easily could have been wrong concerns what happens at nearby possible worlds. But in order to know that one knows, Safety requires one to be epistemically successful not just at nearby worlds, but also at any worlds that are nearby a nearby world.[10]

In many cases of interest to opponents of KK, an agent believes p and couldn't easily have failed to so believe. In such cases, Safety is equivalent to the simpler Margin for Error principle:

(18) **Margin for Error.** If S knows that p, then p could not easily have been false.

Much debate about KK has concerned the exact relationship between Safety and Margin for Error. Defenders of KK have suggested that Margin for Error is inappropriately stronger than Safety. Opponents of KK have disagreed.[11] In the rest of this paper, we suppress this complexity, and consider the prospects for combining Margin for Error and Fragility.

To investigate this question precisely, we turn to Williamson 2013a's framework for exploring Margin for Error within epistemic logic.[12] Williamson 2013a introduces a special class of epistemic models, which connect margins for error to the difference between appearance and reality. We imagine the agent gaining information about the value of a parameter like temperature, tree height, or whatever. We then model the distinction between appearance and reality by a pair of values, r and a. r is the real value of the parameter, while a is the way the parameter appears to the agent. We only consider the agent's knowledge of the values of these parameters, and so let a possible world be a pair (r, a) of these two values, where $R(r, a)$ is the set of epistemic possibilities at world (r, a).

If the temperature is some real value r and apparent value a, then the agent's knowledge is constrained by a margin for error around a. In order to know that the value is in a certain range, this range must include the entire margin for error. This margin for error is large enough to include the real value r, but may include more as well.

To reach a precise theory of margins for error, Williamson 2013a proposes three constraints. First, we assume that appearances are luminous.

[9] For influential defenses of Safety, see Sosa 1999, Williamson 2000, and Pritchard 2005 among others. For recent criticism of Safety, see Goldstein and Hawthorne 2021.

[10] Although see Das and Salow 2018 for a way of reconciling Safety principles with KK.

[11] For representative samples of this debate, see Berker 2008, Srinivasan 2013, and Goldstein and Waxman 2020.

[12] See Goodman 2013 and Carter 2018 for interesting expansions of this framework.

(19) **Appearance Luminosity.** If $(r, a)R(r', a')$, then $a = a'$

Since accessible points never differ in the value of appearance, we usually confine our discussion below to the set of real values that are epistemically possible at a given world. We let $\mathsf{Real}(r, a)$ denote this set $(\{r' \mid \exists a : (r, a)R(r', a)\})$.

Second, the agent retains some ignorance even when appearance perfectly matches reality. As Williamson 2013a observes, the agent's 'perceptual apparatus is not perfectly discriminating' (p. 5).

(20) **Modesty.** $\{(a, a)\} \subset R(a, a)$

Even when appearance matches reality, there is some range of values around a which are epistemically possible.

The final constraint is that epistemic possibility is a function of the distance between appearance and reality. For simplicity, suppose that r and a are numbers. Then as the distance between r and a shrinks, the epistemically accessible worlds must also shrink.

(21) **Distance.** $R(r, a) \subseteq R(r', a)$ iff $|r - a| \leq |r' - a|$.

Distance implies that an agent's epistemic position improves as the distance between reality and appearance decreases. Holding fixed a and varying the real value r, we produce a nested series of spheres of epistemic accessibility, with $R(a, a)$ the innermost sphere, and with epistemic possibility increasing as r increases in distance from a.[13]

With these constraints in place, Williamson 2013a then offers a particular theory of accessibility. The actual margin for error around a at any point is the sum of the distance between r and a, and a fixed minimum margin for error c. At (a, a), the margin for error around a is simply c. As we move away from (a, a) to points (r, a) where appearance does not match reality, the margin grows from c to the sum $|r - a| + c$.

Definition 4.

1. Any world w is a pair (r, a) of a real value r and apparent value a.
2. The minimum margin for error, c, is a fixed positive constant.
3. $(r, a)R(r', a')$ if and only if $a = a'$ and $|r' - a| \leq |r - a| + c$.
4. $\mathsf{Real}(r, a) = \{r' : (r', a) \in R(r, a)\}$.
5. $\mathsf{Real}^2(r, a) = \{r'' : \exists r' \in \mathsf{Real}(r, a) : r'' \in \mathsf{Real}(r', a)\}$.

To illustrate this theory, consider Figure 1.

Here we represents an agent's knowledge of the temperature, using degrees of Fahreinheit. The temperature appears to be 75 degrees, and the margin for error is 5. This theory gives rise to characteristic epistemic differences between

[13] For further discussion of the psychological plausibility of Modesty and Distance, see Nagel 2013.

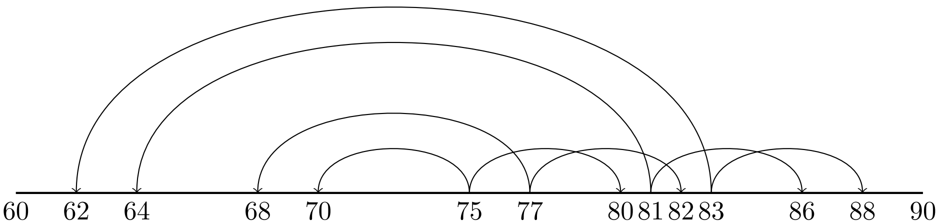

Figure 1: A model of Distance with $a = 75$, $c = 5$

the good case where reality matches appearance, and bad cases where it does not. So consider the good case $(75, 75)$. Arrows denote the upper and lower bounds of epistemic accessibility from the origin. Here $\mathsf{Real}(75, 75) = [70, 80]$ is the range of possible real values of the temperature. $\mathsf{Real}(80, 75) = [65, 85]$. So when the temperature is 75 degrees, the agent considers 80 degrees possible; and when it is 80 degrees, the agent treats 85 degrees as epistemically possible. But when the temperature is 75 degrees, the agent does not treat 85 degrees as epistemically possible.

Now consider Fragility. There are worlds (r, a) where Fragility holds locally, so that (r, a) accesses worlds (r', a) where $R^2(r', a) \subseteq R(r, a)$. For example, $(75, 75)$ is possible at $(80, 75)$, and $\mathsf{Real}^2(75, 75) = [65, 85] = [65, 85] = \mathsf{Real}(80, 75)$. But there are also worlds where Fragility does not hold. For example, $\mathsf{Real}^2(75, 75)$ is not included within $\mathsf{Real}(77, 75)$. But $(75, 75)$ is the strongest epistemic state accessible from $(77, 75)$. So at $(77, 75)$, the agent knows that the real value of the temperature is between 68 and 82, but she has also ruled out that she knows that she knows this.

We can generalize from this case. Fragility is incompatible with Appearance Luminosity, Modesty, and Distance. Appearance Luminosity and Modesty imply that (a, a) accesses some worlds distinct from (a, a). Let (r, a) be the furthest accessible world from (a, a). Distance implies that $\mathsf{Real}(a, a) \subset \mathsf{Real}(r, a)$. But Distance implies that $\mathsf{Real}(r, a) = \mathsf{Real}^2(a, a)$. So $\mathsf{Real}(a, a) \subset \mathsf{Real}^2(a, a)$. But Distance also implies that $\mathsf{Real}^2(a, a)$ is a proper subset of $\mathsf{Real}^2(r', a)$, for any (r', a) accessible from (a, a). So (a, a) cannot access a point (r', a) where $\mathsf{Real}^2(r', a) \subseteq \mathsf{Real}(a, a)$.

5.2 Fragility

If we want to validate Fragility, we must reject one of Appearance Luminosity, Modesty, and Distance. In the rest of this section, I hold fixed Appearance Luminosity and Modesty, and explore the prospects for rejecting Distance. Here, I follow both Cohen and Comesaña 2013 and Goodman 2013, although the former validates KK and the latter invalidates Fragility. In the rest of this section, I make room for Fragility by weakening Distance and allowing the possibility of varying margins for error. In this setting, Fragility corresponds to the requirement that the margin for error when appearance matches reality is sufficiently smaller than the margin at any other world. More precisely, Fragility corresponds to a simple epistemological principle: whenever an agent knows p in the good case where reality matches appearance, she knows that she knows p.

To begin with, we minimally weaken Distance so that as the distance between reality and appearance increases, the epistemic possibilities may stay the same without increasing.

(22) **Weak Distance.** $R(r, a) \subseteq R(r', a)$ only if $|r - a| \leq |r' - a|$.

Consider the worlds $(75, 75)$ and $(76, 75)$. In the former, reality matches appearances exactly; in the latter, reality is slightly different from appearance. Distance implies that $R(75, 75) \subset R(76, 75)$. By contrast, Weak Distance instead allows that $R(76, 75)$ may equal $R(75, 75)$.

When we replace Distance with Weak Distance, we can validate Fragility. Weak Distance implies that Fragility holds at every possible world if and only if KK holds in the good case where reality matches appearance (so that $R^2(a, a) = R(a, a)$).

Observation 2. Weak Distance implies that Fragility is valid if and only if KK holds at (a, a).[14]

This is a significant consequence of Fragility. Along with Factivity, it implies that any world (r, a) accessible from (a, a) has the same epistemic possibilities as (a, a). What is epistemically possible need not increase as reality diverges from appearance.

We can also understand Fragility in terms of the connection between knowledge and justification. Williamson 2013a introduces a notion of justified belief at (r, a) in terms of what is known at (a, a). Where S is doxastic accessibility, $S(r, a) = R(a, a)$. This is an internalist notion of justified belief which ignores the real value and depends only on the apparent value. Agents in the good case where appearance matches reality are fortunate enough to believe exactly what they know. Like Distance, Weak Distance ensures that knowledge implies justification, so that $R(a, a)$ is included in $R(r, a)$ for all (r, a).

Fragility has interesting connections to justification. An agent is justified in believing p just in case p is known in the good case. Fragility says that knowledge in the good case iterates. So given this theory of justification, Fragility is equivalent to the principle, endorsed in Stalnaker 2006, that an agent is justified in believing p if and only if she is justified in believing that she knows p.

Fragility has another consequence. Williamson 2013b defines a special class of Gettier cases, which structurally resemble fake barn cases. In this class of 'purely

[14] Fragility requires (a, a) to see some (r, a) where $R^2(r, a) \subseteq R(a, a)$. Since $|a - a| \leq |r - a|$, Weak Distance implies that $R(a, a) \subseteq R(r, a)$ and hence $R^2(a, a) \subseteq R^2(r, a)$. So Weak Distance implies that $R^2(a, a) \subseteq R^2(r, a) \subseteq R(a, a)$, and so whenever S knows A at (a, a), she also knows that she knows A. Conversely, suppose that the KK holds at (a, a), so that $R^2(a, a) \subseteq R(a, a)$. Weak Distance implies that $R(a, a) \subseteq R(r, a)$ for all (r, a). So every (r, a) sees (a, a), where $R^2(a, a) \subseteq R(r, a)$. So Fragility holds at every world.

veridical' Gettier case, an agent fails to know p despite having no false justified beliefs. Let $S(r, a)$ be the strongest claim that the agent believes justifiedly ($R(a, a)$). Then (r, a) has purely veridical Gettier cases just in case $S(r, a)$ is true at (r, a) even though $S(r, a)$ is strictly smaller than $R(r, a)$.

(23) Purely veridical Gettier cases exist at (r, a) if and only if $(r, a) \in S(r, a)$ and $S(r, a) \subset R(r, a)$.

Now a consequence of the previous observation is that Weak Distance implies that Fragility is valid if and only if there are no purely veridical Gettier cases. For suppose Fragility is valid. Then KK holds at (a, a), and so for every (r, a) in $S(r, a)$, we have that $S(r, a) = R(a, a) = R(r, a)$. Conversely, the absence of purely veridical Gettier cases implies that KK holds at (a, a), and so implies the validity of Fragility. In ruling out purely veridical Gettier cases, I agree with the theory in Cohen and Comesaña 2013, and depart from the theories in Williamson 2013a, Goodman 2013, Weatherson 2013, and Carter 2018.

Fragility also has consequences for margins for error. On some conceptions of margins for error, the validity of KK even locally at (a, a) is untenable. Williamson 2013b considers and rejects the possibility of cliff-edge knowledge at a world (r, a), where either S knows that the real value is at least r or S knows that the real value is at most r.[15]

(24) **Cliff-edge knowledge.** S has cliff-edge knowledge at (r, a) if and only if $\exists n : \mathsf{Real}(r, a) = [r, n]$ or $\mathsf{Real}(r, a) = [n, r]$.

Drawing on Goodman 2013, Williamson 2013b argues that cliff-edge knowledge violates a version of Safety, which says that reality could always have been slightly different while appearances remained the same.[16]

(25) $\forall(r, a)\exists c > 0 : [r - c, r + c] \subseteq \mathsf{Real}(r, a)$

We can distinguish this safety requirement from another, which simply holds that appearances give us unspecific evidence about real values.[17]

(26) $\forall(r, a)\exists c > 0 : [a - c, a + c] \subseteq \mathsf{Real}(r, a)$

Given Weak Distance, cliff-edge knowledge exists if KK is locally valid.[18] For suppose KK holds at (r, a), so that $R^2(r, a) \subseteq R(r, a)$. Factivity implies that

[15] Compare Stalnaker 2009, p. 406 offers a defense of cliff edge knowledge. For a response to this defense, see Hawthorne and Magidor 2010, p. 1092.

[16] Compare Weatherson 2013, p. 67.

[17] This is implied by a principle Goodman 2013 calls *Appearance Constraint*.

[18] Strictly speaking, this claim holds only if epistemic accessibility produces closed intervals of possible real values. Suppose instead that at $(70, 70)$ the accessible real values are in the open interval from 68 to 72, approaching but never hitting 68 and 72. Suppose that KK is locally valid at $(70, 70)$, with any accessible real value approaching 72 also having as epistemic possibilities the open interval from 68 to 72. In that case, the agent does not have cliff-edge knowledge, since there is always an ever-diminishing margin for error separating her from the value 72. But surely such an infinitesimal margin for error is small consolation for the opponent of cliff-edge knowledge.

$R^2(r,a) = R(r,a)$. Now let (r^*, a) be the highest (or lowest) world in $R(r,a)$. Since $R^2(r,a) = R(r,a)$, we know that $R(r^*, a) = R(r,a)$, and S has cliff-edge knowledge at (r^*, a): $\mathsf{Real}(r^*, a) = [n, r^*]$ for some choice of n. Since Fragility implies that KK holds at (a,a), Fragility thus implies that there is cliff-edge knowledge at the maximum and minimum of $\mathsf{Real}(a,a)$. At any such world, (25) fails (although (26) can still hold). In this way, one might think that cliff-edge knowledge is consistent with perceptual unspecificity but not with safety from error.[19]

5.3 Theories of knowledge

We've now explored in detail the various consequences of Fragility in a general framework for thinking about knowledge and margins for errors. In the rest of this section, we consider a few candidates for what knowledge could be, consistent with Fragility. Building on Stalnaker 2009, Cohen and Comesaña 2013 develop a theory consistent with Modesty and Weak Distance where KK and hence Fragility are unrestrictedly valid. Epistemic accessibility is defined relative to a fixed minimum margin for error c; but this minimum margin for error has different effects in three cases. When the real value r falls within the range $[a-c, a+c]$, epistemic accessibility simply produces the range of real values $[a-c, a+c]$. When the real value r falls below $a-c$, the possible real values are $[r, a+c]$. When the real value r rises above $a+c$, the possible real values are $[a-c, r]$.

(27) $$(r,a)R(r',a') \text{ if and only if } a = a' \text{ and } \begin{cases} r \leq r' \leq a+c & \text{if } r < a-c \\ a-c \leq r' \leq r & \text{if } r > a+c \\ a-c \leq r' \leq a+c & \text{otherwise} \end{cases}$$

On this theory, cliff-edge knowledge is pervasive. As in any theory of Fragility consistent with Weak Distance, cliff-edge knowledge occurs at $a-c$ and $a+c$. But cliff-edge knowledge also occurs at any value r below $a-c$ or above $a+c$.

On this theory, KK is valid. Within $[a-c, a+c]$, every world treats the same real values as possible: the range $[a-c, a+c]$. Above $a+c$, any real value can only see itself and any value lower, until reaching the minimum $a-c$. At any such world, epistemic accessibility is strictly included in the range $[a-c, r]$.

My task is to validate Fragility without KK. I will basically agree with Cohen and Comesaña 2013 about the behavior of epistemic accessibility within $\mathsf{Real}(a,a)$. But I offer a different theory of epistemic accessibility outside of this region. For Cohen and Comesaña 2013, accessibility outside of this region is pervaded by cliff-edge knowledge. For me, it will not be.

In Williamson 2013a, the actual margin for error at (r,a) is the sum of the distance $|r-a|$ and a fixed minimum margin for error c. I now depart from this theory and simply let the margin for error at (r,a) be some value $m(r,a)$

[19] The theory in Goodman 2013 satisfies the version of Margin for Error in (25), and also Weak Distance. For this reason, it contains no cliff-edge knowledge, and therefore invalidates Fragility.

determined as a function of r and a, subject to a variety of constraints. Then r' is a possible real value at (r, a) just in case the distance between r' and a is within the margin $m(r, a)$.

We can reach a substantive theory of knowledge by imposing a variety of constraints on margins for error. To validate Factivity, I assume that the margin at (r, a) is always at least as large as the distance between r and a. To validate Modesty, I assume that the margin at r and a is always positive. Within this framework, Distance corresponds to the requirement that $m(r', a)$ exceeds $m(r, a)$ if and only if the distance between r' and a is greater than that between r and a. I replace this requirement with Weak Distance, which now says that $m(r', a) \geq m(r, a)$ if the distance between r' and a is at least as large as that between r and a. This gives us the following class of models:

Definition 5.

1. The margin for error, $m(r, a)$, is a function of r and a.
2. $(r, a)R(r', a')$ if and only if $a = a'$ and $|r' - a| \leq m(r, a)$, where:
   (a) m is factive: $m(r, a) \geq |r - a|$.
   (b) m is modest: $m(r, a) > 0$.
   (c) m is weakly monotone: if $|r - a| \leq |r' - a|$, then $m(r, a) \leq m(r', a)$.

This theory is consistent with Fragility. Fragility then corresponds to the constraint that there is some region around a where the margin for error is constant.

Observation 3. Fragility is valid if and only if $\exists i \geq 0$: if $|r - a| \leq i$, then $m(r, a) = i$.[20]

My theory predicts that when appearance matches reality, the agent inhabits a kind of inner sanctum. For some distance around a, the margin for error is simply $i = m(a, a)$, the minimum margin for error. In the range of real values $[a - m(a, a), a + m(a, a)]$, the agent experiences automatic iterated knowledge. At any world in this area, the range of possible real values is just $\mathsf{Real}(a, a) = [a - m(a, a), a + m(a, a)]$. Here, we have a violation of Distance that respects Weak Distance. The agent is not omniscient at (a, a), but their epistemic position does not get worse for a small period of time as reality departs from appearance. In the most extreme case, the agent at $a + m(a, a)$ stands on the cliff of epistemic accessibility and knows that the real value is at most exactly what it is.[21]

[20] Follows immediately from our previous observation that Fragility is valid if and only if KK holds locally at (a, a).

[21] Another advantage of this theory is that it predicts that 'what you justifiably believe is known in all normal worlds with the same appearances' (Goodman 2013, building on Lasonen-Aarnio 2010). Williamson 2013b formulates a weak version of this principle:

(i) **Weak Disposition to Know.** For any r^*, there is some $0 < c^* \leq c$ where if $|a - r| \leq c^*$, then $R(r, a) \subseteq S(r^*, a)$.

KK fails at (r, a) when (r, a) can access a real value r' which can access a value unavailable to (r, a). This requires that r' is within $m(r, a)$ of a. In addition, it requires that $m(r', a)$ exceeds $m(r, a)$. So KK fails in the theory just in case there are some real values r and r' where $|r' - a| \leq m(r, a) < m(r', a)$. This condition is consistent with Fragility. For this reason, the framework allows me to validate Fragility without validating KK.

Since I validate Fragility, the theory makes a precise prediction about how epistemic accessibility behaves inside $R(a, a)$. But this leaves unsettled how epistemic accessibility behaves outside of $R(a, a)$. Perhaps the simplest option is to make a hybrid theory which agrees with Cohen and Comesaña 2013 within distance c from a, and agrees with Williamson 2013b after that.When r is beyond c from a, the margin $m(r, a)$ is the sum of $|r - a|$ and c. But within a distance of c from a, the margin $m(r, a)$ is fixed at c. In this way, c becomes a lower bound for the agent's epistemic power of discrimination, so that anywhere inside of c distance from a, the range of possible real values is just $[a - c, a + c]$.

(28) $$m(r, a) = \begin{cases} |r - a| + c & \text{if } |r - a| > c \\ c & \text{otherwise} \end{cases}$$

On this interpretation, Fragility can be thought of as imposing further barriers on knowledge. When an agent is at $(a + m(a, a), a)$, their epistemic position is already as strong as possible. Further improvements in the match between reality and appearance have no affect on their epistemic position, because they have already reached the limit of their epistemic power.

The hybrid theory validates Fragility. For any (r, a) can access the world (a, a), and $\mathsf{Real}^2(a, a) = \mathsf{Real}(a, a)$ is guaranteed by Weak Distance to be within $\mathsf{Real}(r, a)$. But the theory still respects Modesty by requiring (a, a) to access other worlds. In this way, the case where appearance matches reality is epistemically privileged without being epistemically ideal. Finally, this theory validates KK locally in the range $[a - c, a + c]$: when r is in this range and the agent knows A, the agent is guaranteed to know that they know A. But when r is outside this range, KK fails while Fragility remains valid. In this way, the hybrid theory is a minimal revision of Williamson 2013a which validates KK locally at (a, a) so that Fragility is valid.[22]

For an illustration of this theory, consider Figure 2.
Within $[70, 80]$, epistemic accessibility is transitive. So the very same range of real possibilities is known at $(75, 75)$, $(77, 75)$, and $(80, 75)$: namely, that the real value is between 70 and 80. But once the real value departs from what is

Since we define justified belief so that $S(r^*, a) = R(a, a)$, Weak Disposition to Know is valid on our theory. In particular, let c^* be the distance between a and the highest value in $R(a, a)$. Then $R(a, a)$ is believed and known throughout the inner sanctum within c^* distance of a, where reality and appearance are sufficiently similar. On the other hand, for criticism of Weak Disposition to Know, see Williamson 2013b p. 87.

[22] The hybrid view differs from that in Williamson 2013a and Cohen and Comesaña 2013 in that it is discontinuous: small changes in the divergence between appearance and reality can lead to a large change in what is known (when the real value moves just outside the range $[a - c, a + c]$).

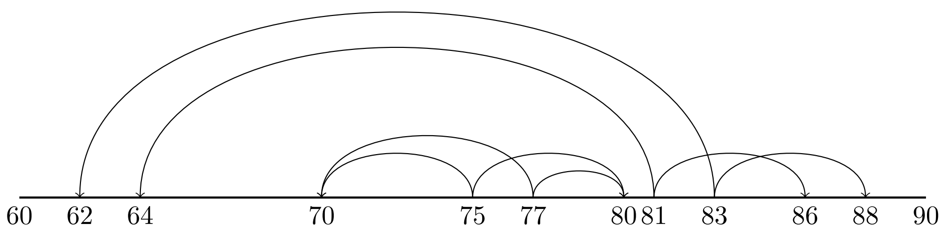

Figure 2: A model of Weak Distance with $a = 75$, $c = 5$

epistemically possible in the good case, accessibility is no longer transitive. As before, 83 is an epistemically possible real value at $(81, 75)$, and 88 is a possible real value at $(83, 75)$, yet 88 is not possible at $(81, 75)$. So the agent at $(81, 75)$ knows that the temperature is between 64 and 86 degrees, but doesn't know that she knows this.[23]

5.4 Improbable knowledge

The hybrid theory is not the only option. There is a reason to explore more dramatic departures from extant theories: we can thereby prevent improbable knowing. Williamson 2013b observes that Modesty and Distance generate cases of improbable knowing. At (a, a), $R(a, a)$ is the strongest known proposition. Modesty implies that $R(a, a)$ includes worlds besides (a, a), and Distance implies that $R(a, a)$ is not known at any such world, because at any such (r, a) we have $R(a, a) \subset R(r, a)$. In this way, Modesty and Distance generate improbable knowing: although $R(a, a)$ is known at (a, a), the agent at (a, a) considers it unlikely that $R(a, a)$ is known. In particular, at every epistemic possibility for the agent other than (a, a), $R(a, a)$ is not known.

To make this more precise, I follow Williamson 2011 and Williamson 2014 and introduce an evidential probability function Pr. I let the evidential probability $Pr_{(r,a)}$ at world (r, a) come from conditionalizing a prior Pr on $R(r, a)$, the agent's knowledge at (r, a). Improbable knowing occurs at (r, a) when there is a proposition p that is known at (r, a) while the probability that it is known falls below a threshold t. For any proposition p, let $\mathbf{K}p = \{(r', a') : R(r', a') \subseteq p\}$ be the set of worlds at which p is known. Then:

[23] An anonymous referee wonders about the status of further introspection principles. Consider the Geach rule, that $MKA \models KMA$. The referee observes that Geach and Fragility imply the 'Shift Symmetry' rule, that $KA \models KKMA$ (Symmetry says that $A \models KMA$; Shift Symmetry says that this rule applies when we add a K operator to the premise and conclusion). Here, I note that all of the models considered in this paper validate Shift Symmetry. Interestingly, this includes Williamson's Appearance/Reality models, which invalidate Fragility and yet validate Geach. An open question for future research is whether it is possible to modify Williamson's Appearance/Reality models to retain Geach while invalidating Fragility and Shift Symmetry.

(29) S has improbable$_t$ knowledge at (r, a) if and only if $\exists p : (r, a) \in \mathbf{K}p$ and $Pr_{(r,a)}(\mathbf{K}p) \leq t$.

Modesty and Distance imply that improbable knowledge is pervasive. At any world (r, a), $R(r, a)$ is known at (r, a), but is not known at any world (r', a) where the distance between r' and a exceeds that between r and a. This means that $R(r, a)$ is a case of improbable$_{\frac{1}{2}}$ knowing whenever the margin $m(r, a)$ is twice the distance between reality and appearance $|r - a|$.

When we replace Distance with Weak Distance, we can prevent improbable knowledge. As the distance between reality and appearance grows, the epistemic possibilities cannot diminish. But they may sometimes stay the same. To avoid improbable knowing, we can create bands of constancy. As we move from (r, a) to worlds (r', a) further from r but still inside $R(r, a)$, we can for a while retain the same epistemic possibilities, so that $R(r', a) = R(r, a)$.

(30) **Bands of constancy.** R has a band of constancy at (r, a) of length n if and only if $R(r + n, a) = R(r, a)$.

To avoid KK, however, we allow that there are some worlds $(r^*, a) \in R(r, a)$ where the epistemic possibilities expand, so that $R(r, a) \subset R(r^*, a)$.

We can use bands of constancy to prevent improbable knowing. Assume Pr is indifferent. Then we can guarantee that whenever S knows p at (r, a), the evidential probability that S knows p is at least t. This is simply a matter of ensuring that the band of constancy at (r, a) is sufficiently large.

Observation 4. If for every (r, a), R has a band of constancy at (r, a) of length $x > t \times m(r, a) - |a - r|$, then S lacks improbable$_t$ knowledge.[24]

For example, with $m(80, 75) = 10$, $\mathsf{Real}(80, 75) = [65, 85]$. Throughout $r = [70, 80]$, $R(80, 75)$ is known. But given Distance, $R(80, 75)$ is not known at any $r > 80$. So at $(80, 75)$, $R(80, 75)$ has an evidential probability of $\frac{1}{2}$. Since we reject Distance, we can create a band of constancy of length 3 beyond $(80, 75)$. This means that $R(83, 75) = R(80, 75) \subset R(84, 75)$. On the resulting theory, the evidential probability at $(80, 75)$ of knowing $R(80, 75)$ is at least $\frac{4}{5}$.

All that is left is to find an interpretation of knowledge on which it plausibly has bands of constancy. One option here, drawing on Goodman 2013, looks to normality.

[24] Take arbitrary (r, a). We must show that $R(r, a)$ is known throughout at least t proportion of worlds in $R(r, a)$. After all, if $R(r, a)$ is known there, so is any other proposition known at (r, a). Now suppose $r > a$. Given symmetry, we can then confine our attention to the status of $R(r, a)$ at real values above a. Weak Distance implies that $R(r, a)$ is known at any real value between a and r. To prevent improbable knowing, we must guarantee that $R(r + x, a) = R(r, a)$ for some distance x above r. In particular, we must show that $\frac{|a-r|+x}{m(r,a)} > t$, so that the region extending from a upwards beyond r to length x is greater than $m(r, a)$, the size of the region above a which is epistemically possible. In that case, the region in which S knows $R(r, a)$ will make up greater than t proportion of the epistemic possibilities at (r, a). This equation simplifies to $x > t \times m(r, a) - |a - r|$.

> If things are normal, then what you know is that they aren't extraordinary; if things aren't normal, you know less. (Goodman 2013, p. 46)

At any world, some worlds count as normal, some as extraordinary, and some as neither. Then we can say that at any world, what an agent knows is simply that things aren't extraordinary by the lights of that world. The resulting picture motivates bands of constancy. At any world (r, a), it would be extraordinary for r to be significantly further from a than it is. But if r were slightly further from a, the same values would be extraordinary. This gives us bands of constancy. As r moves further away from a towards the extraordinary, but before r becomes extraordinary, the standards for normality weaken, so that more worlds become normal and the extraordinary moves further away. In this way, KK fails (for more on the contingency of normality and its consequences for KK, see Carter 2018). Finally, to validate Fragility we distinguish the good case where reality meets appearance. In the good case, any world that isn't extraordinary has the same standards for normality. In this way, we experience no jump in possibility until we have moved into an extraordinary case. This gives us a realistic interpretation for our theory, validating Fragility and allowing bands of constancy once reality and appearance diverge sufficiently.

In this section, I've shown that it is possible to endorse Fragility while also accepting that knowledge is subject to a form of margin for error. To do so, we must allow that appearance can diverge from reality without creating further barriers to knowing. We must also allow that the margin for error when appearance meets reality is sufficiently smaller than other margins for error. In this way, opponents of KK may explain the infelicity of dubious assertions by validating Fragility. One cost of the theory is the existence of a case of cliff-edge knowledge, with automatic iterated knowledge in the inner sanctum where appearance approximates reality. One advantage of the theory is that it allows bands of constancy, preventing improbable knowing.

In the last part of the paper, I explore more complex dubious assertions, and show how to generalize Fragility to explain them.

6 Generalizations

6.1 Other attitudes

Some dubious assertions are more complex than (1), involving mixed attitudes of belief and knowledge. Sosa 2009 observes that each of the following is infelicitous:

(31)
- a. #p but I doubt that I know that p.
- b. #p but I believe that I don't know that p.
- c. #p but I have no justification for believing that I know that p.
- d. #p but I have (sufficient) justification for believing that I don't know that p.

Fragility implies that each of the conjunctions above is unknowable. In each case, the argument is roughly the same: we can show that the iterated state in the second conjunct of the dubious assertion is logically as strong as the state of not knowing that one knows. For this reason, assuming that knowledge is closed under simple deduction, anyone who knows any of these conjunctions knows the dubious assertion (1) with which we began.

Start with (31-a). Knowledge is incompatible with doubt. So if S doubts that S knows that p, then S doesn't know that S knows p. So if S knows that p and that S doubts that S knows p, then S knows that S doubts that S knows p. But since this last bit of knowledge implies that S doesn't know that S knows p, we now have that S knows that S doesn't know that S knows p. This contradicts Fragility, since we also have that S knows p. In short, this complex assertion is logically stronger than (1), our original dubious assertion. Since the weaker dubious assertion is unknowable, so is the stronger.

The same argument applies to each of the other dubious assertions above. For (31-b), we assume that if S believes that S doesn't know that p, then S doesn't know that S knows p. For (31-c), we assume that if S is not justified in believing that she knows p, then S doesn't know that she knows p. After all, knowledge requires justification. For (31-d), we assume that if S is justified in believing she doesn't know p, then she doesn't know that she knows p.

Fragility is a powerful principle. It has consequences for various patterns of iterations of belief, justification, and ignorance. In this way, Fragility provides a systematic theory of dubious assertion.

6.2 Higher orders

Another way to generalize the phenomenon of dubious assertion involves further iterations of knowledge. For example, perhaps the following assertions are infelicitous in the same sense as (1):

(32) a. p but I don't know that I know that I know that p.
b. p but I don't know that I know that I know that I know that p.
c. ...

Fragility alone does not predict that (32-a) and its ilk are unknowable. To do so, Fragility would have to imply:

(33) If S knows that S doesn't know that S know that S knows p, then S doesn't know that p.

But Fragility does not have this consequence. Note that the antecedent of (33) does not imply the antecedent of Fragility. This follows from the more general fact that an agent can be ignorant of having second order knowledge without being ignorant of having first order knowledge.

If we wish to predict that (32-a) and its ilk are unknowable, we can introduce strengthened versions of Fragility, such as (33). To better understand such stronger principles, let K^n abbreviates n consecutive occurrences of K. Then we can formalize equivalents of higher order principles like (33) with the following schema:

(34) **Fragility^n.** $KA \to MK^nA$

(33) is equivalent to the instance Fragility^3.

The results from above extend to further iterations. First, we can introduce the concept of the n-ancestral of R, which relates w and v just in case v can be reached from w through n applications of R. Then Fragility^n corresponds to a generalization of jump transitivity, where every world w can see some world v where any world accessible from v by the n-ancestral of R is accessible from w by R.

Definition 6.

1. (a) wR^1u if and only if wRu
   (b) wR^nu if and only if $\exists v : wR^{n-1}v$ & vRu
2. R is jump transitive^n if and only if $\exists v \in Rw : R^nv \subseteq Rw$

Observation 5. Fragility^n is valid if and only if R is jump transitive^n.[25]

There is a structural difference between KK and Fragility. Once KK is valid, so is any further iteration of KK. KK implies for example that:

(35) If S knows that p, then S knows that S knows that S knows that p.

For this reason, the validity of KK immediately implies that $R^nw \subseteq Rw$ for every n. So KK implies that Fragility^n is valid for every choice of n. So the validity of KK implies that every dubious assertion in (32) is unknowable and hence unassertable. By contrast, if we reject KK and accept Fragility, then in order to predict the unassertability of (32) we must accept each instance of Fragility^n as a separate constraint on knowledge.

This flexibility may be a bug or a feature, depending on the data. As Benton 2013 warns us, it is important to distinguish 'clashes' from 'clunks'. Perhaps at high levels of iteration, instances of (32) are not infelicitous in the same way as (1). They may instead simply be unparsable. Perhaps these conjunctions are knowable at some level of processing, but are so difficult to entertain consciously that they are strange to say.

We saw above that Fragility encodes the idea that an agent's knowledge of p is defeated by the information that her epistemic position with respect to p is not ideal. But here we might distinguish different degrees of epistemic ideality. Failing to know that one knows p is not ideal. Failing to know that one knows that one knows p is not ideal in another way. Perhaps the first failure defeats

[25] Suppose R is jump transitive^n and $w \models KM^nA$. Then $\forall v \in Rw : R^nv \cap \mathbf{A} \neq \emptyset$. By jump transitivity^n, $\exists v^* \in Rw : R^nv^* \subseteq Rw$. So $Rw \cap \mathbf{A} \neq \emptyset$. So $w \models MA$. Conversely, suppose that R is not jump transitive^n. Then $\forall v \in Rw : \exists z \in R^nv : z \notin Rw$. Let $\mathbf{A} = \{w \mid \exists v \in Rw : z \in R^nv \;\&\; z \notin Rw\}$. $w \models KM^nA$, since $\forall v \in Rw : v \models M^nA$. But $w \not\models MA$, since $\neg\exists v \in Rw : v \in \mathbf{A}$.

knowledge in a way that the second does not. We can express this distinction by developing a theory of knowledge in which jump transitivity is valid but jump transitivity^n is not valid for all n.

On the other hand, we also considered the prospects for reconciling Fragility with Margin for Error. Interestingly, the theory I developed predicts that Fragility is valid if and only if Fragility is valid at every order. On that theory, I replaced Distance by Weak Distance and generated an inner sanctum of worlds where reality is similar enough to appearance that epistemic possibility is the same as when appearance agrees exactly with reality. On that view, KK holds locally at the point where appearance matches reality, and so we have jump transitivity^n at every order.[26]

[26]Thanks to the audience of the 2019 Goethe Epistemology Meeting.

References

Matthew Benton. Dubious objections from iterated conjunctions. *Philosophical Studies*, 162:355–358, 2013.

Selim Berker. Luminosity regained. *Philosophers' Imprint*, 8(2):1–22, 2008.

Alexander Bird. Justified judging. *Philosophy and Phenomenological Research*, 74:81–110, 2007.

Sam Carter. Higher order ignorance inside the margins. *Philosophical Studies*, pages 1–18, 2018.

Sam Carter and Simon Goldstein. The normality of error. *Philosophical Studies*, January 2021.

Stewart Cohen and Juan Comesaña. Williamson on gettier cases and epistemic logic. *Inquiry*, 56:15–29, 2013.

Nilanjan Das and Bernhard Salow. Transparency and the kk principle. *Noûs*, 52 (1):3–23, 2018.

Richard Feldman. Respecting the evidence. *Philosophical Perspectives*, 19, 2005.

Simon Goldstein and John Hawthorne. Counterfactual contamination. *Australasian Journal of Philosophy*, March 2021.

Simon Goldstein and Daniel Waxman. Losing confidence in luminosity. *Noûs*, August 2020.

Jeremy Goodman. Inexact knowledge without improbable knowing. *Inquiry*, 56 (1):30–53, 2013.

Daniel Greco. Iteration and fragmentation. *Philosophy and Phenomenological Research*, 88(1):656–673, 2014.

Daniel Greco. Iteration principles in epistemology 1: Arguments for. *Philosophy Compass*, 10(11):754–771, 2015.

John Hawthorne and Ofra Magidor. Assertion and epistemic opacity. *Mind*, 119 (476):1087–1105, 2010.

Jonathan Jenkins Ichikawa. Justification is potential knowledge. *Canadian Journal of Philosophy*, 44:184–206, 2014.

Maria Lasonen-Aarnio. Unreasonable knowledge. *Philosophical Perspectives*, pages 1–21, 2010.

Wolfgang Lenzen. Recent work in epistemic logic. *Acta Philosophical Fennica*, 30, 1978.

Jennifer Nagel. Motivating williamson's model gettier cases. *Inquiry*, 56(1): 54–62, 2013.

Duncan Pritchard. *Epistemic Luck*. Oxford University Press, 2005.

Colin Radford. Knowledge–by examples. *Analysis*, 27:1–11, 1966.

Sven Rosenkranz. The structure of justification. *Mind*, 127(506):309–338, April 2018.

Weng Kin San. Kk, knowledge, knowability. May 2019.

Declan Smithies. Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research*, 85:273–300, 2012.

David Sosa. Dubious assertions. *Philosophical Studies*, 146(2):269–272, 2009.

Ernie Sosa. How to defeat opposition to moore. *Philosophical Perspectives*, 13: 141–154, 1999.

Amia Srinivasan. Are we luminous? *Philosophy and Phenomenological Research*, 90(2), 2013.

Robert Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128: 169–199, 2006.

Robert Stalnaker. On hawthorne and magidor on assertion, context, and epistemic accessibility. *Mind*, 118(470):399–409, 2009.

Brian Weatherson. Margins and errors. *Inquiry*, 56(1):63–76, 2013.

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

Timothy Williamson. Improbable knowing. In Trent Dougherty, editor, *Evidentialism and its Discontents*. Oxford University Press, 2011.

Timothy Williamson. Gettier cases in epistemic logic. *Inquiry*, 56(1):1–14, 2013a.

Timothy Williamson. Response to cohen, comesaña, goodman, nagel, and weatherson on gettier cases in epistemic logic. *Inquiry*, 56(1):77–96, 2013b.

Timothy Williamson. Very improbable knowing. *Erkenntnis*, 79:971–999, 2014.