

# Losing confidence in luminosity

Simon Goldstein<sup>1</sup> | Daniel Waxman<sup>2</sup>

<sup>1</sup> Dianoia Institute of Philosophy,  
Australian Catholic University

<sup>2</sup> Department of Philosophy, Lingnan  
University

## Correspondence

Daniel Waxman, Lingnan University.  
[danielwaxman@gmail.com](mailto:danielwaxman@gmail.com)

## Abstract

A mental state is luminous if, whenever an agent is in that state, they are in a position to know that they are. Following Timothy Williamson's *Knowledge and Its Limits*, a wave of recent work has explored whether there are any non-trivial luminous mental states. A version of Williamson's anti-luminosity appeals to a safety-theoretic principle connecting knowledge and confidence: if an agent knows  $p$ , then  $p$  is true in any nearby scenario where she has a similar level of confidence in  $p$ . However, the relevant notion of confidence is relatively underexplored. This paper develops a precise theory of confidence: an agent's degree of confidence in  $p$  is the objective chance they will rely on  $p$  in practical reasoning. This theory of confidence is then used to critically evaluate the anti-luminosity argument, leading to the surprising conclusion that although there are strong reasons for thinking that luminosity does not obtain, they are quite different from those the existing literature has considered. In particular, we show that once the notion of confidence is properly understood, the failure of luminosity follows from the assumption that knowledge requires high confidence, and does not require any kind of safety principle as a premise.

## 1 | ANTI-LUMINOSITY

A mental state is luminous if, whenever an agent is in that state, they are in a position to know that they are. Much recent debate, following Williamson (2000), concerns the question of whether there are any non-trivial luminous states.<sup>1</sup>

Williamson's discussion is rich, and has been read as offering not one anti-luminosity argument but two. The arguments share a common form: a sequence of cases is presented in which the agent undergoes gradual change in a mental state. The assumption that the mental state is luminous is then shown to conflict with safety principles connecting knowledge, truth, and some doxastic state. The precise formulation of the safety principles—and the doxastic states involved—varies between the two arguments. One involves the coarse state of 'all-or-nothing' belief. The other, which will be the primary focus of this paper, connects knowledge with the 'graded' notion of *degrees of confidence*, which Williamson understands as "degrees of outright belief" (p. 99).

Let us represent the epistemic state of a single agent using a knowledge operator  $K$ .<sup>2</sup> Consider a candidate luminous mental state, such as feeling cold, and let  $p$  be the proposition that the agent feels cold. Then luminosity about feeling cold says:

(1) **Luminosity.** If  $p$ , then  $Kp$ .

For simplicity, assume that subjective coldness comes in degrees, and can be measured on a scale from 0 to 1, where 1 is extremely cold and 0 is extremely hot. Now consider an example where an agent starts off feeling cold and ends up feeling hot. This can be modeled with a set of worlds  $\{w_1, w_{0.99}, \dots, w_0\}$ , where  $w_i$  is a world where the agent's degree of subjective coldness is  $i$ . (Note that higher subscripts indicate worlds where the agent is *colder*). Furthermore, assume that there is a cutoff  $c$  where the agent feels cold just in case her degree of subjective coldness is greater than  $c$ .<sup>3</sup>

(2) **Cutoff.**  $p$  is true at  $w_i$  iff  $i > c$  where  $0 < c < 1$

The anti-luminosity argument requires two premises about confidence. The first is that the agent's degrees of confidence are continuous, so that if the agent felt slightly colder, her confidence that she felt cold would be slightly higher. Let  $C_i(p)$  represent the agent's degree of confidence in  $p$  at world  $i$ . Let  $C_i(p) \approx C_j(p)$  signify that  $C_i(p)$  and  $C_j(p)$  are similar degrees of confidence.<sup>4</sup> The assumption is:

(3) **Continuity.**  $C_i(p) \approx C_j(p)$  iff  $|i - j| < 0.02$

Finally, we turn to the safety principle connecting knowledge and confidence. Here is Williamson on the required premise:

Suppose that at  $[w_i]$  one knows that one feels cold. Thus one is at least reasonably confident that one feels cold ... Moreover, this confidence must be reliably based, for otherwise one would still not know that one feels cold. Now at  $[w_{i-0.01}]$  one is almost equally confident that one feels cold... So if one does not feel cold at  $[w_{i-0.01}]$ , then one's confidence at  $[w_i]$  that one feels cold is not reliably based, for one's almost equal confidence on a similar basis a millisecond later that one felt cold is mistaken. (Williamson, 2000, p. 96)

The relevant safety principle is this: if an agent knows  $p$  at  $w_i$  and her degrees of confidence in  $p$  are similar between  $w_i$  and  $w_j$ , then  $p$  must also be true at  $w_j$ . Call this principle:

(4) **Confidence Safety.** If  $Kp$  is true at  $w_i$  and  $C_i(p) \approx C_j(p)$ , then  $p$  is true at  $w_j$ .

Now, Continuity implies that an agent's confidence at  $w_i$  is similar to her confidence at  $w_{i-0.01}$ . So if she knows that she feels cold in  $w_i$ , then she feels cold in  $w_{i-0.01}$ . This quickly leads to contradiction given Luminosity.

*Observation 1.* Luminosity, Cutoff, Continuity, and Confidence Safety are incompatible.

By Cutoff, there is some  $i$  such that  $p$  is true at  $w_i$  and false at  $w_{i-0.01}$ . By Continuity,  $C_i(p) \approx C_{i-0.01}(p)$ . By Confidence Safety,  $Kp$  is therefore false at  $w_i$ . So Luminosity fails at  $w_i$ .

The primary aims of our paper are to systematically investigate this argument, and the relation between confidence and anti-luminosity more generally. The question is far from a mere matter of detail. The anti-luminosity debate bears on fundamental issues both within epistemology—the internalism/externalism dispute, the nature of epistemic normativity, constraints on doxastic planning, etc—and elsewhere in philosophy—e.g. the existence of response-dependent concepts and the feasibility of assertibility-conditional semantic theories.<sup>5</sup> And in turn, we believe that confidence-theoretic considerations are crucial to the overall case for anti-luminosity; we share the view (argued in Berker (2008) and conceded by Srinivasan, 2013) that Williamsonian arguments involving all-or-nothing belief are dialectically ineffective against an important class of proponents of luminosity, namely those who hold that, for certain mental states, there is a constitutive connection between the state obtaining and being believed to obtain.<sup>6</sup> Much therefore turns on confidence.

The plan is as follows. We start from first principles and develop (§2) a precise theory of confidence to help us find our way through the issues. Following Williamson (2000), we understand an agent's degree of confidence in  $p$  as the degree to which an agent relies on  $p$  in practical reasoning. Departing from Williamson (2000), however, we offer an account of these degrees in terms of the objective chance that the agent relies on  $p$ . In §3, we advocate a principle linking knowledge and ideally reliable confidence: that if an agent's degrees of confidence are ideally reliable and her confidence in  $p$  exceeds a threshold, then she knows  $p$ . A principle of this sort is implicitly accepted by much of the contemporary literature; the debate thus turns, in large part, on what an ideally reliable confidence profile looks like. We go on to offer a new account, according to which an agent's degrees of confidence are ideally reliable just in case they correspond to the objective chances. We then (§4) return to the anti-luminosity argument and show that our theory of confidence leads to two surprising conclusions. The first is that, by considering the connection between ideally reliable confidence and knowledge, Confidence Safety can be shown to fail in a wide range of cases. Second, however, this should not be cause for celebration by proponents of Luminosity, since Luminosity can also be shown to fail in a wide range of cases by the principle that knowledge implies confidence.

By the end we hope to have accomplished two goals. One is to have provided a new and philosophically fruitful theory of confidence. The other is to have offered a novel anti-luminosity argument: one which does not rely on controversial and dubiously motivated safety principles like Confidence Safety, but rests instead on an independently plausible account of confidence and its connection to knowledge.

## 2 | CONFIDENCE

Confidence, as we will understand it, is a graded doxastic state, representing a certain kind of ‘strength’ or ‘degree’ of commitment to a proposition.<sup>7</sup> We follow Williamson (2000) in viewing confidence as closely tied to practical reasoning:

One believes  $p$  outright when one is willing to use  $p$  as a premise in practical reasoning ... Since using  $p$  as a premise in practical reasoning is relying on  $p$ , we can think of one’s degree of outright belief in  $p$  as the degree to which one relies on  $p$ [...]

The degrees of confidence mentioned in the argument for [anti-luminosity] should therefore be understood as degrees of outright belief. (Williamson, 2000, p. 98-9)

It is crucial to note that, despite some similarities between confidence and the more familiar graded doxastic state of ‘subjective probability’ or ‘credence’, the two are importantly distinct. The clearest way to see this is via a case adapted from Williamson (2000):

- (5) **Lottery Bag.** In a number of draws, a ball is drawn from a bag. In Draw <sub>$i$</sub> , the bag contains  $i$  red balls and  $100 - i$  black balls. In each draw, Billy knows the number of red and black balls in the bag. Billy reasonably assigns a credence of  $\frac{i}{100}$  to the proposition that the drawn ball is red. In each case, Billy is offered the opportunity to purchase a lottery ticket which costs 1 and which pays out \$1000 if the drawn ball is black. In Draw<sub>100</sub> he declines; in all other draws he purchases the ticket.

Let  $r$  be the proposition that the drawn ball is red. In Draw<sub>100</sub>, Billy knows  $r$ , his credence in  $r$  is 1, and his confidence in  $r$  is also 1. However, in Draw<sub>99</sub>, Billy does not know  $r$  and his credence in  $r$  is 0.99. What about his degree of confidence? Williamson suggests that, since Billy takes the bet, it is very low, perhaps even 0: Billy hardly relies at all on  $r$ , but rather relies on the proposition that *the probability of  $r$  is 0.99*, and this explains why he purchases the ticket. More generally:

One may assign  $p$  a high subjective probability without believing  $p$  outright, if the corresponding premise in one’s practical reasoning is just that  $p$  is highly probable on one’s evidence, not  $p$  itself. (Williamson 2000, p. 98)

Williamson’s description of Billy’s doxastic and epistemic position in Lottery Bag implies that that confidence is distinct from credence. In Draw<sub>99</sub>, he diagnoses Billy’s degree of confidence in  $r$  as very low, perhaps even 0; at the same time, Billy’s credence in  $r$  is 0.99. Thus credence and confidence come apart. As Williamson (2000) puts it, “one’s subjective probability can vary while one’s degree of outright belief remains zero” (p. 99).

Williamson’s gloss on confidence is suggestive, and his diagnosis of Lottery Bag is plausible. Nevertheless, for the purposes of evaluating the anti-luminosity argument, it would be desirable to have a more systematic and developed theory of confidence at our disposal. Developing such a theory is our aim in the rest of this section.

## 2.1 | A dispositional theory of confidence

We've said that an agent's degree of confidence in  $p$  is the degree to which they rely on  $p$  in practical reasoning. But how can reliance come in degrees? Our answer looks to dispositions: an agent's confidence in  $p$  is the degree to which they are disposed to rely on  $p$ .

The idea that dispositions come in degrees has independent motivation. One reason comes from comparative dispositional claims:

(6) Glass A is more fragile than Glass B.

A natural way to make sense of (6) is that it is true when Glass A is fragile *to a higher degree* than Glass B.

We propose to adopt a version of the view found in Manley and Wasserman (2008), where the degree of a disposition is explained in terms of the proportion of nearby worlds in which the disposition manifests.

In particular, we work with a simplification of the Manley-Wasserman view. The degree of a disposition is evaluated relative to some set of nearby worlds  $N$ . In general  $N$  will likely be context-sensitive; we will say more soon about how  $N$  might naturally be chosen in the epistemic cases with which we are concerned. Next, to simplify talk of 'proportions' of possible worlds, we assume an objective chance function  $Ch$  over the set of possible worlds. Intuitively, the proportion of  $\phi$  worlds in which  $\psi$  is true can be represented as the chance of  $\psi$  conditional on  $\phi$ , or  $Ch(\psi | \phi)$ . Then we propose:

(7) The degree to which  $o$  is disposed to be  $F$  is the objective chance that  $o$  is  $F$ , conditional on  $N$ .

This proposal explains (6): Glass A is more fragile than Glass B when, in nearby possible worlds, A has a higher objective chance of shattering than does B.<sup>8,9</sup>

This theory of degrees of disposition gives us a precise account of confidence. Let  $Rp$  be the proposition that the agent relies on  $p$ . Then we claim an agent's degree of confidence in  $p$  is simply the chance that they rely on  $p$  in nearby worlds:

(8) **Dispositionalism.**  $C(p) = Ch(Rp | N)$

In addition to degrees of confidence, we also introduce a notion of all-or-nothing confidence. For simplicity we assume a Lockean analysis, so that an agent is all-or-nothing confident in  $p$  (written  $Cp$ ) iff her degree of confidence in  $p$  is greater than a threshold  $l$ .<sup>10</sup>

(9) **Lockeanism.**  $Cp$  if  $C(p) > l$

We now apply Dispositionalism to Lottery Bag. In  $\text{Draw}_{100}$ , Billy's degree of confidence that the drawn ball is red is the chance that he relies on it being red in nearby worlds. In order for this value to be 1, Billy must rely on the ball being red at every nearby world. In §3, we offer a substantive gloss on nearness in the kinds of epistemic cases we are concerned with, in terms of subjective indiscernibility. For now, however, we help ourselves to the simplifying assumption that two worlds are near in Lottery Bag just in case they have the same lottery structure. Since Billy is certain of the setup of the case, this assumption seems both harmless and independently plausible.

All of the nearby worlds to  $\text{Draw}_{100}$  are thus worlds in which the bag contains 100 red balls. Now, turn to the question of whether Billy relies on the ball being red at nearby worlds. To generate a prediction here, we make two mild assumptions: (i) all nearby worlds preserve Billy's knowledge of the case and his rational betting dispositions; and (ii) Billy relies on the drawn ball being red iff he refrains from buying the lottery ticket that wins if the drawn ball is black. By (i), Billy does not buy the lottery ticket in any of the nearby worlds; so by (ii), he relies on the drawn ball being red in all nearby worlds. Thus his degree of confidence is 1.

Now consider  $\text{Draw}_{99}$ . Again, any nearby world also contains a draw from a bag with 99 red balls and 1 black ball. By (i), Billy makes the bet in all nearby worlds. So by (ii), no nearby world is one in which Billy relies on the drawn ball being red. Consequently, his degree of confidence that the drawn ball is red is 0.

Lottery Bag is an example of a case where degrees of confidence come apart from the more familiar notion of credence. For another such case, imagine the agent is holding a transparently fair coin in her hand, and is about to flip it twice. Her credence that the coin will land heads twice is 0.25. But she may have a very low degree of disposition to rely in practical reasoning on the coin landing heads twice, and hence a very low degree of confidence.

Thus, in cases like Lottery Bag, one's degree of confidence in our sense can systematically depart from the credence it would be rational to hold given knowledge of chance-facts. This may at first be somewhat surprising, but it is in keeping with the existing literature on the role of confidence in anti-luminosity arguments.<sup>11</sup> We hope the initial surprise will be reduced once it is noticed that confidence and credence are simply two different states, and that there is nothing incoherent about supposing that people have both. Moreover, nothing in our account of confidence prohibits us from adopting an entirely orthodox account of credence.

Dispositionalism is a precise and systematic theory of confidence which delivers the desired verdicts about Billy's degrees of confidence. What's more, it provides a plausible reduction of confidence to naturalistically unproblematic notions like dispositions and reliance. It is true that we have said little about what it is to rely on a proposition in practical reasoning. For the most part, we conduct our inquiry below by considering general principles about reliance which should hold on any reasonable theory. That said, in §3.2 we show that the main assumption about reliance necessary for our argument also follows from a plausible pragmatic theory of reliance.

### 3 | CONFIDENCE SAFETY AND IDEAL RELIABILITY

The heart of existing anti-luminosity arguments is the principle we have called Confidence Safety, which holds that knowledge requires 'safe' confidence. However, the principle is not simple to assess. On the one hand, safety-theoretic constraints on knowledge have gained currency in recent years, and perhaps Confidence Safety inherits some plausibility from the theoretical success of this general approach. But on the other hand, the differences between Confidence Safety and more commonly discussed safety principles are worth bearing in mind. The usual safety constraint on knowledge concerns all-or-nothing belief, and holds that a belief must be safe in order to constitute knowledge. Confidence Safety, by contrast, says that one's degree of confidence must be safe not just in the nearby scenarios where one's confidence is the same degree but also in those where it is slightly lower (or higher). It is no surprise, then, that opponents of Williamson's anti-luminosity argument have discerned something problematically soritical in the principle.<sup>12</sup>

To overcome this impasse, we offer a suggestion as to how a disciplined investigation into the relationship between confidence and knowledge can be conducted. Start with the platitude that an agent's degrees of confidence can be epistemically evaluated. Just as we might judge an agent's beliefs or degrees of credence to be (more or less, or outright) justified, or rational, or reliable, so too can we do the same for degrees of confidence. Now, return to the motivation behind Confidence Safety. Suppose one's confidence that  $p$  is high, and that there is a nearby situation in which one's confidence is almost the same but where  $p$  is false. As Williamson (2000, p. 96) puts it, "one's confidence that one feels cold... is not reliably based, for one's almost equal confidence on a similar basis a millisecond later that one felt cold is mistaken".

The thought, in short, is that certain degrees of confidence are too unreliable or unreliably based to constitute knowledge. A promising direction of investigation, therefore, is to theorize about *fully* or *ideally* reliable degrees of confidence. Let us dramatize the situation by considering an agent whose degrees of confidence are ideally reliable, in the sense of being not at all unreliable or unreliably based. Of course, we do not claim that any actual epistemic agents are remotely like this; nevertheless, considering an ideal agent of this kind will clarify the issues considerably, for such an agent's epistemic state will be invulnerable to the criticism that they are unreliable. We write  $C^1(\cdot)$ ,  $C^1$  and  $K^1$  to represent the degrees of confidence, outright confidence, and knowledge of an ideally reliable agent.

The following principle connects ideal reliability and knowledge. If an ideally reliable agent is (outright) confident in a proposition, and the proposition is true, then it is known:

(10) **Ideal Reliability.**  $(C^1 p \wedge p) \rightarrow K^1 p$

A few comments are in order.

First, those who hold safety-theoretic or more generally externalist accounts of knowledge will find Ideal Reliability appealing as a general epistemic principle—in particular, those who believe that the primary epistemic constraint on a doxastic state in order for it to constitute knowledge is that it be safely or reliably true.<sup>15</sup> If the primary threat to knowledge is unreliability, and we consider an agent whose degrees of confidence are ideally reliable by stipulation, then such an agent surely overcomes the threat of unreliability—and hence knows—whenever they are confident in a true proposition.

Second, even those who reject such an account of knowledge, such as those who believe that knowledge additionally requires some kind of internalist justification, should consent to the following principle:

(11) **Ideal Justified Reliability.**  $(C^1 p \wedge p \wedge J^1 p) \rightarrow K^1 p$

where  $J^1 p$  says that the ideally reliable agent is justified in believing  $p$ .

The difference between Ideal Reliability and Ideal Justified Reliability will prove irrelevant for the purposes of our discussion. Our main discussion (§§3&4) will consider a series of cases where an ideally reliable agent's subjective degree of coldness gradually changes. We see no reason why the setup cannot be tweaked, so that the agent possesses one's favourite kind of internalist justification whenever they are confident that they feel cold. If you like, imagine that whenever the threshold for confidence is reached, the agent undergoes perceptual or introspective seemings of the right sort, or otherwise gains internalist evidence that they feel cold, or comes to possess additional grounds of the kind that make for rational or blameless or epistemically permissible belief. Because the setup can be unproblematically tweaked in this way, we will henceforth ignore



any justification condition in the formulation of Ideal Reliability, assuming if necessary that confidence is correlated with justification of the relevant kind.

Third, Ideal Reliability is not an analysis of knowledge. It claims nothing about the knowledge of normal agents or agents in general, only ideally reliable ones. The spirit in which we offer the principle is nothing like a full analysis, but as an attempt to gain some theoretical purchase on the connection between confidence and knowledge.

Fourth, we only rely essentially on Ideal Reliability in one of our two main arguments. Later we will argue against both Luminosity and Confidence Safety. Ideal Reliability plays an essential role only in the second case. By contrast, our argument against Luminosity does not assume Ideal Reliability: we argue against Luminosity using the different assumption that knowledge implies confidence. So those skeptical of Ideal Reliability can still find much of interest in what follows.

Notice finally that Ideal Reliability has virtually no substantive implications on its own. In order for interesting consequences to be drawn, it must be supplemented with a substantive account of the confidence profile of an ideally reliable agent; this is the task that will occupy us in §3. Previous writers on luminosity have offered conflicting conceptions of ideal reliability and have put these conceptions to work in arguing for or against Confidence Safety. Defenders of Luminosity have suggested that the ideally reliable confidence in  $p$  is simply the degree to which  $p$  is true. By contrast, opponents of Luminosity have suggested that the ideally reliable agent's degree of confidence in  $p$  is 1 when  $p$  is true and 0 when  $p$  is false. In the next section, we argue that both of these proposals are misguided, and go on to offer our own view.

### 3.1 | Two theories of ideal reliability

#### 3.1.1 | Calibration

Leitgeb (2002), Berker (2008), Ramachandran (2009), and Cohen (2010) defend Calibration: the view that an ideally reliable agent's degree of confidence that she feels cold is precisely her degree of subjective coldness.<sup>14</sup> Here is a representative example:

Why should we withhold the honorific 'reliable' [in cases which Confidence Safety diagnoses as not being knowledge]? What if one's degree of confidence in its being the case that  $p$  perfectly tracks the underlying basis for its being the case that  $p$ , so that one's degree of confidence that  $p$  falls just short of belief at the precise point at which things fall just short of making it the case that  $p$ ? Why would *that* be a situation in which one's initial belief that  $p$  is not reliable enough to constitute knowledge? (Berker, 2008, p. 12)

Two things are happening here. One is that Berker is offering a substantive conception of ideal reliability. The other is that Berker is implicitly endorsing a version of Ideal Reliability (when he suggests that an ideally reliable agent's confidence ought to count as knowledge).

A commitment to both Calibration and Ideal Reliability can also be found in Ramachandran:

[Confidence Safety] rule[s] out luminosity in the hypothesized *perfect-calibration* situation, which is daft, because one couldn't be any more reliable. (Ramachandran, 2009 p. 668)



Let us assume that degrees of subjective coldness can be normalized to the same scale as degrees of confidence. For the sake of simplicity, let us also assume that the Lockean threshold for outright confidence,  $l$ , is the same as the threshold for feeling cold,  $c$ .<sup>15</sup>

(12) **Cutoff Identification.**  $c = l$

Calibration—the view that the ideal degree of confidence that one feels cold is simply one’s degree of subjective coldness—can then be stated. Where  $C_i^!$  is the ideally reliable agent’s confidence profile at a world where they are cold to degree  $i$ :

(13) **Calibration.**  $C_i^!(p) = i$

Many defenders of Luminosity have found Calibration to be an appealing conception of ideally reliable confidence, since it provides a way to resist the anti-luminosity argument. Here is one way of reconstructing their line of thought. (Note how it appeals to Ideal Reliability). From the background setup,  $p$  is true at  $w_i$  whenever  $i > c$ . But according to Calibration, whenever  $i > c$ , we have:

(14)  $C_i^!(p) = i > c = l$

Thus the agent is outright confident in  $p$  and hence, by Ideal Reliability, knows that  $p$ . It follows that Luminosity holds for the ideal agent. In addition, it is relevant that Confidence Safety fails: the ideally reliable agent knows  $p$  when they are just above the threshold for coldness. But there is a nearby case, just below the threshold, in which the agent’s confidence is almost the same and in which  $p$  is false. Thus Calibration can be used to argue that Confidence Safety fails.

### 3.1.2 | Vindication

In defense of anti-luminosity, Srinivasan (2013) argues that, contra Calibration, calibrated agents are systematically unreliable. Her main complaint is that calibrated agents have a fairly high degree of confidence that they feel cold even in cases where they do not feel cold. To see why, suppose the cutoff for feeling cold is 0.7; then a calibrated agent whose subjective degree of coldness is 0.69 would not feel cold, but would nevertheless have fairly high confidence—0.69—that they do feel cold.

For this reason Srinivasan offers a rival conception of ideally reliable confidence according to which it is fully invested in the truth. If the agent does in fact feel cold, then their ideally reliable degree of confidence in  $p$  is 1; and if not, it is 0.<sup>16</sup> We will call this view Vindication:

(15) **Vindication.**  $C_i^!(p) = \begin{cases} 1 & \text{if } p \text{ is true at } w_i \\ 0 & \text{otherwise.} \end{cases}$

Vindication may well be psychologically unrealistic for creatures like us, since it requires discontinuous assignments of confidence. This is no objection: the claim concerns ideal reliability, not the capacities of creatures like us.

How does Vindication bear on the anti-luminosity argument? The dialectic is subtle. Vindication cannot be used directly to rule out Luminosity. If an agent possesses vindicated degrees of confidence, then Ideal Reliability and Vindication imply that they know they feel cold iff they feel cold. So Luminosity holds for them. So Ideal Reliability and Vindication allow that Luminosity holds for certain agents, provided that they closely enough approximate ideal reliability. However, this is cold comfort to defenders of the claim that there are non-trivial luminous mental states for creatures similar to us. As Berker and Srinivasan note, actual physical systems rarely exhibit the requisite kind of discontinuity.

The real reason why Vindication appeals to opponents of Luminosity (applied to actual, as opposed to ideal, agents) is that it allows Confidence Safety to be defended against Calibrationist criticism. For although vindicated agents satisfy Luminosity, they *also* satisfy Confidence Safety. Consider a case where the agent is just above the threshold for coldness. There  $p$  is true, so the vindicated agent has confidence to degree 1 in  $p$ , and therefore knows  $p$ . But the only nearby worlds where the agent has a similar degree of confidence are those in which  $p$  is also true; in any nearby world on the other side of the threshold for feeling cold, the agent's degree of confidence shifts abruptly to 0. So, according to Vindication, ideally reliable agents provide no basis for rejecting Confidence Safety, and thus the anti-luminosity argument (for non-ideal agents) threatens.

The upshot of all of this is an impasse. The proponent of Luminosity accepts Calibration, and on that basis rejects Confidence Safety. The opponent of Luminosity accepts Vindication, and thus sees no argument against Confidence Safety. Both Calibration and Vindication possess a certain intuitive plausibility. If we think of one's degree of subjective coldness as the degree to which it is true that one feels cold, then Calibration seems highly attractive: who could be more reliable than an agent whose confidence tracks the degree to which the proposition in question is true? But if we think about things in an all-or-nothing way, then Vindication seems highly attractive: who could be more reliable than an agent whose confidence is always maximally invested in the truth?

We propose a simple resolution: both Calibration and Vindication are false. Confidence does not ideally track either degrees of truth or monadic truth, but rather tracks objective chance.

### 3.2 | A new theory of ideal reliability

Consider the following passage from Srinivasan:

A confidence just short of outright belief in  $p$  will yield some cases in which one uses  $p$  in one's practical reasoning despite not believing  $p$  outright...[such an agent] has some tendency to use that false proposition as a premise in her practical decision-making. If she were making several...decisions at the same time, we could expect to see [that agent] acting on [ $p$ ] when it is false. (Srinivasan, 2013, 21)

The compelling thought here is that there is something unreliable about an agent who satisfies Calibration. Such an agent tends to rely on  $p$  in practical reasoning even in nearby cases where  $p$  is false. (Indeed, Srinivasan makes the stronger claim that such an agent will be *too unreliable to know* that  $p$ . For our purposes, all that is needed is the weaker claim that such an agent is not ideally reliable. If true, this is enough to refute Calibration.)

But while we find the thought compelling, we wish to draw a different lesson than Srinivasan. We agree that an ideally reliable agent relies on  $p$  only if  $p$  is true: but we think that this condition

applies also to nearby scenarios. To express this idea, let  $R^1 p$  be the claim that the ideal agent relies on  $p$ . Then we propose:

(16) **Ideality.**  $Ch(R^1 p \leftrightarrow p \mid N) = 1$

Ideality captures a crucial modal aspect of reliability. Suppose  $p$  is true in the actual world but false in all nearby worlds. According to Vindication, an ideally reliable agent must rely on  $p$  in all the nearby  $\neg p$  worlds. But this is doubtful; an agent who is fully reliable with respect to  $p$  surely relies on  $p$  in a nearby world  $w$  if and only if  $p$  is true at  $w$ .

Although Ideality could reasonably be taken as a premise without further argument, it is worth mentioning that it is independently supported by a plausible conception of reliance. Recall that confidence is a measure of an agent's disposition to rely in practical reasoning. This suggests that the 'ideally reliable' degrees of confidence are those that produce the best practical reasoning—reasoning that best satisfies the agent's desires.

(17) **Pragmatism.**  $C^1(\cdot)$  is the degree of confidence function that maximizes utility in nearby cases.

But a very natural account of reliance can be modeled on the pragmatic accounts of belief found in Stalnaker (1984) and Weatherson (2005): S relies on  $p$  iff for all relevant choices A and B, S prefers A to B iff one prefers A and  $p$  to B and  $p$ . And given that rational agents maximize expected utility, this yields:

(18) **Reliance By Action.** A rational agent relies on  $p$  iff they perform whichever action maximizes expected utility given  $p$ .

Dispositionalism, Pragmatism, and Reliance By Action imply Ideality. By Dispositionalism and Pragmatism, the ideal degree of confidence is equal to the chance of relying on  $p$  that would maximize utility at nearby worlds. Now consider some nearby world  $v$  to  $w$ . Which propositions should an agent rely on at  $v$ , in order to maximize their utility? Reliance By Action says that to rely on  $p$  is to perform whichever action maximizes expected utility given  $p$ . Now imagine that the agent relies on all and only the claims that are true. In that case, they will perform whichever action maximizes their actual utility at  $v$ . This is exactly what Pragmatism requires. In other words, the modal profile of action that best maximizes utility is for the agent to rely on  $p$  at a world iff  $p$  holds there; that is to say, Ideality holds.<sup>17</sup>

### 3.3 | Objectivism

We now show that Ideality leads to a rival conception of ideal confidence, according to which it tracks objective chance:

(19) **Objectivism.**  $C_i^1(p) = Ch_i(p \mid N_i)$

Objectivism is a straightforward consequence of Dispositionalism and Ideality. An agent's confidence in  $p$  is the chance she relies on  $p$  in nearby worlds. An ideal agent relies on  $p$  in a nearby world iff  $p$ . So an ideal agent's confidence in  $p$  is simply the chance of  $p$  in nearby worlds.

We can now see the problem in the argument from Srinivasan (2013) quoted above. The argument assumes that when an agent has a fairly high degree of confidence in a false proposition, she will be somewhat disposed to rely on  $p$  in situations where  $p$  is false. But this is not necessarily so. Even though  $p$  is actually false,  $p$  may be true in the nearby worlds in which she relies on  $p$ . If an agent is ideally reliable, she will not be disposed to rely on whatever is actually true; rather, she will be disposed to rely on whatever is true *at the world in which she relies on it*.

### 3.4 | Consequences of Objectivism

Objectivism allows us to make a variety of interesting predictions about ideal degrees of confidence. To see these predictions, we introduce a pair of simplifying assumptions which together allow us to derive some results about ideal confidence. Not much turns on the particular assumptions we make, but they make the exposition considerably easier. In the appendix, we present more complex analogues of our results that hold independently of the simplifying assumptions.

The first assumption, Indifference, is that the chance function at any world  $i$  ( $Ch_i$ ) is indifferent about the worlds in our model: every degree of subjective coldness has the same prior degree of chance. The second assumption, Centering, concerns the nearness relation. In the gradual-change scenario being considered, it is compelling to think of worlds as being nearby to one another if their degrees of subjective coldness are within a distance parameter  $d$ .<sup>18,19</sup> We also assume that  $d < \frac{1}{2}$ , to avoid the trivializing consequence that the agent's confidence in  $p$  is  $\frac{1}{2}$  at every world.

Together, Indifference and Centering imply that the chances are symmetric. When an agent feels cold to some degree  $i$ , they are just as likely to feel warmer than to feel colder. This symmetry implies that when an agent feels cold, the chance that they feel cold is at least 0.5. For this reason, we'll refer to the combination of our two assumptions as Symmetry:

#### (20) Symmetry.

- a. **Indifference.**  $Ch_i(w_j) = Ch_i(w_k)$
- b. **Centering.**  $N_i = \{w_j \mid |i - j| < d\}$  where  $d < \frac{1}{2}$

In the presence of Symmetry, Objectivism precisely specifies the ideal agent's confidence. Given Objectivism, the ideally reliable agent's confidence in  $p$  is the chance of  $p$  in nearby worlds. By Centering, the nearby worlds to  $w_i$  are those in the interval  $(i - d, i + d)$ ; their chance measure is thus  $2d$ . The chance of  $p$  within this region is simply the proportion of such worlds where the degree of coldness is greater than  $c$ , the cutoff point for coldness.

*Observation 2.* Suppose Cutoff and Symmetry. Then Objectivism implies that:

$$C_i^!(p) = \begin{cases} 1 & \text{if } c < i - d \\ \frac{i+d-c}{2d} & \text{if } i - d < c < i + d \\ 0 & \text{if } c > i + d \end{cases}$$

(Proofs of observations can be found an appendix.)

We can also understand Observation 2 diagrammatically, as in Figure 1. An ideal agent's confidence in  $p$  corresponds to the region of nearby space where  $p$  holds:

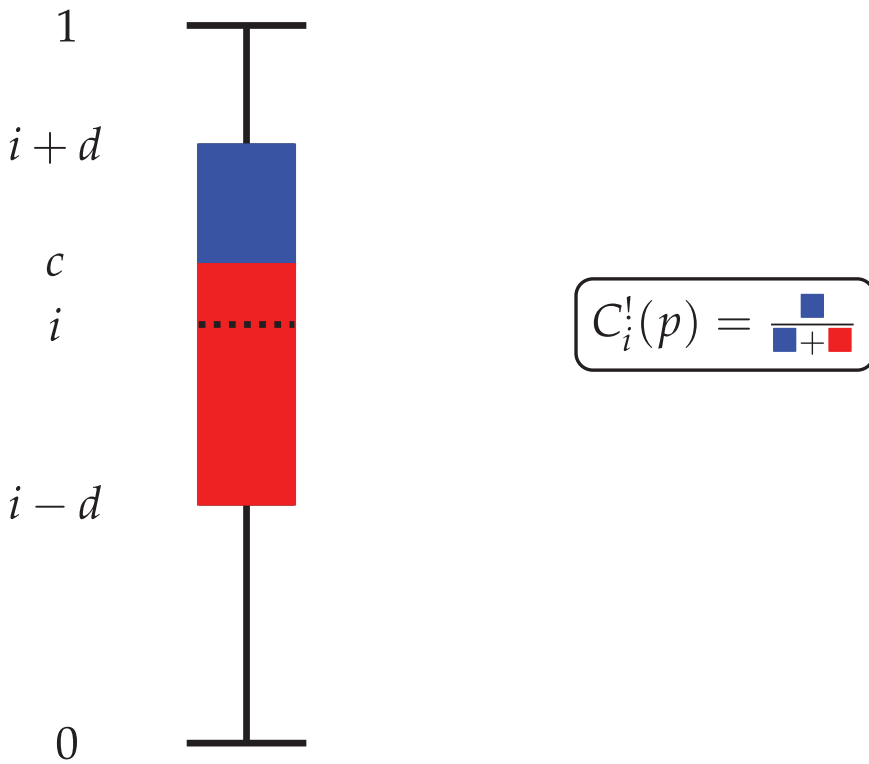


FIGURE 1 Ideal confidence [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Objectivism differs markedly from the rival views in the literature. Compare Vindication first. Given Dispositionalism, Vindication amounts to the claim that if  $p$  is actually true, then the chance that an ideal agent relies on  $p$  in nearby worlds is 1. But as we have argued, this is non-ideal: when the agent just barely feels cold, Vindication requires them to rely on  $p$  in all nearby worlds, including some where  $p$  is false.

Calibration fares similarly. It amounts to the claim that in  $w_i$ , the proportion of nearby worlds in which the ideal agent relies on  $p$  is  $i$  itself. But by considering worlds where the agent just barely feels cold, we can see that Calibration also requires them to rely on feeling cold in some worlds where they do not feel cold. Again, by our lights, this is non-ideal.

Objectivism is a novel and plausible account of ideally reliable confidence. What is its bearing on the anti-luminosity argument? We turn to that question now.

#### 4 | ANTI-LUMINOSITY REGAINED

In this section, we develop a new anti-luminosity argument. The details are complex, so we provide an informal summary of the argument here. Knowledge requires high confidence. So the ideally reliable agent knows that she is cold only if she is highly confident that she is cold. But our theory of ideal confidence predicts that there are cases where the ideally reliable agent is cold and yet is not highly confident that she is cold. This occurs for example when the agent is just barely over the threshold for being cold. In this case, about half of the nearby worlds are ones in which

the agent is not cold; and so her degree of confidence that she is cold is only about 0.5, below the threshold required by knowledge.

We now develop this argument more precisely. First, before getting into the details, we establish some basic properties of our theory, connecting confidence with truth and with knowledge. Throughout, we hold fixed two previous simplifying assumptions: Cutoff (that an agent is cold at  $w_i$  iff  $i$  exceeds the cutoff  $c$ ) and Symmetry.

In the presence of these assumptions, Objectivism implies that  $p$  is true just in case the ideally reliable agent's confidence in  $p$  is greater than  $\frac{1}{2}$ :

*Observation 3.* Suppose Cutoff and Symmetry. Then Objectivism implies that  $p$  is true at  $w_i$  iff  $C_i^1(p) > \frac{1}{2}$ .

Observation 3 connects ideal confidence and truth; our next observation connects ideal confidence and knowledge. The ideally reliable agent knows  $p$  at  $w_i$  just in case  $i > 2dc + c - d$ . To see why, first note that Observation 3 implies that ideal confidence is factive.

(21) **Factivity.**  $C^1 p \rightarrow p$

Next, we make a crucial assumption: that knowledge implies all-or-nothing confidence.

(22) **KC.**  $Kp \rightarrow C^1 p$

The motivation for *KC* comes from the facts that all-or-nothing confidence can be viewed as a form of all-or-nothing belief, and the principle that knowledge requires belief.<sup>20</sup>

Given *KC* and Ideal Reliability, the ideally reliable agent knows  $p$  iff she is confident in  $p$ :

(23) **Collapse.**  $C^1 p \leftrightarrow K^1 p$

Collapse is not an analysis of knowledge, since it only concerns ideally reliable knowledge. But Collapse lets us test whether an ideally reliable agent knows  $p$ . Summarizing:

*Observation 4.*

4.1 Suppose Cutoff and Symmetry. Then Objectivism and Lockeanism imply Factivity.

4.2 Factivity, *KC*, and Ideal Reliability imply Collapse.

Collapse characterizes ideal knowledge in terms of ideal confidence. Objectivism characterized ideal confidence in terms of objective chance. When we chain these characterizations together, we reach a simple account of ideal knowledge:

*Observation 5.* Suppose Cutoff, Cutoff Identification, and Symmetry. Then Objectivism, Lockeanism, *KC*, and Ideal Reliability imply  $K^1 p$  is true at  $w_i$  iff  $i > 2dc + c - d$ .

(We summarize our results so far in Figure 2.)

With these results in place, we can return to the anti-luminosity argument. It turns out that our theory rules out Luminosity, even without appeal to Ideal Reliability, puts severe pressure

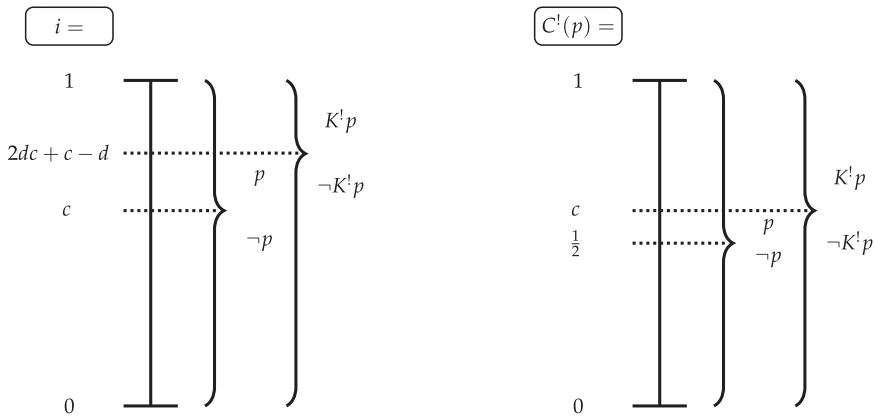


FIGURE 2 Truth and knowledge

on Confidence Safety, and is consistent with ordinary coarse-grained Safety. Thus, our theory generates a new anti-luminosity argument and provides a philosophical framework for assessing a family of safety principles.

### 4.1 | Luminosity

We now argue that Luminosity is false. Consider a weakened form of Luminosity concerning ideally reliable agents:

(24) **Ideal Luminosity.**  $p \rightarrow K^1 p$

To show that Ideal Luminosity fails, we rely on our previous observation that the ideally reliable agent knows  $p$  at  $w_i$  iff  $i > 2dc + c - d$ . It follows that there is a range of degrees of coldness where the ideally reliable agent feels cold but is not confident that she feels cold, and hence by KC does not know. Say that Ideal Luminosity fails at  $w_i$  just in case  $p$  is true there while  $K^1 p$  is false. Then:

*Observation 6.* Suppose Cutoff, Cutoff Identification, and Symmetry. Then Objectivism, Lockeanism, KC, and Ideal Reliability imply that:

1. Ideal Luminosity fails at  $w_i$  iff  $c < i < 2dc + c - d$ .
2. Ideal Luminosity fails at some  $w_i$  iff  $c > 0.5$ .

The failure of Ideal Luminosity sharply distinguishes Objectivism from Calibration and Vindication. As we showed in §3, these rival views both entail that ideal agents are luminous.

This discussion provides a new kind of anti-luminosity argument. On our view, Luminosity fails in a wide range of cases, even for ideally reliable agents, and therefore does not hold in general. This argument doesn't appeal in any way to Confidence Safety. This is no accident; Confidence Safety is no part of our view. We therefore have an anti-luminosity argument that avoids the most controversial assumptions of previous arguments to the same effect.



In fact, our argument against Luminosity can be strengthened significantly. The observation above relies on Ideal Reliability to establish necessary and sufficient conditions for Luminosity violations. But the right to left directions of Observation 6 can be established without appeal to Ideal Reliability. Those directions merely require a case where  $p$  is true and  $K^l p$  false, which given  $KC$  requires only that  $C^l p$  is false. We can use the assumptions above without Ideal Reliability to show that this occurs if  $c < i < 2dc + c - d$ , which holds for some  $i$  iff  $c > 0.5$ .

It is true that our model includes simplifying assumptions such as Symmetry and Cutoff Identification. But the argument is robust; it could be made without them, given any other reasonable conception of the objective chances. Assuming Objectivism, Luminosity holds only if the chances conspired precisely so that the chance at  $i$  of  $p$  in nearby worlds is greater than  $l$  whenever  $i > c$ . For Luminosity to hold in general, this constraint on the chances would have to hold in general, but we see no plausible way to motivate such a requirement. It certainly does not correspond to any of the distinctive theses in epistemology or the philosophy of mind to which defenders of luminosity typically appeal. Discussion of models without these simplifying assumptions is deferred, for the sake of readability, to an appendix.<sup>21</sup>

## 4.2 | Confidence safety

We now show that our theory is not only not committed to Confidence Safety, but also predicts that it fails in a wide range of cases. Again, we consider a version of Confidence Safety for ideal agents.

(25) **Ideal Confidence Safety.** If  $K^l p$  is true at  $w_i$  and  $C_i^l(p) \approx C_j^l(p)$ , then  $p$  is true at  $w_j$ .

Confidence Safety implies Ideal Confidence Safety. We will characterize the conditions under which Ideal Confidence Safety fails.

Confidence Safety is formulated in terms of a similarity relation  $\approx$  between two degrees of confidence  $C_i(p)$  and  $C_j(p)$ . It is helpful to introduce a similarity parameter, so that  $C_i(p)$  and  $C_j(p)$  count as sufficiently similar just in case the distance between them is less than  $s$ :

(26) **Similarity.**  $C_i(p) \approx C_j(p)$  iff  $|C_i(p) - C_j(p)| < s$

For simplicity, we assumed earlier that the thresholds for coldness and confidence are identical. In the same vein we assume that two degrees of confidence are sufficiently similar when they are indiscernible, and that the threshold of indiscernibility for degrees of confidence is identical to the threshold of indiscernibility for degrees of coldness:

(27) **Distance Identification.**  $d = s$

In our theory, the validity of Ideal Confidence Safety amounts to a constraint on the values of  $d$  and  $c$ . Say that Ideal Confidence Safety fails at  $w_i$  and  $w_j$  just in case  $K^l p$  is true at  $w_i$ ,  $C_i^l(p) \approx C_j^l(p)$ , and yet  $p$  is false at  $w_j$ . Then:

*Observation 7.* Suppose Cutoff, Cutoff Identification, Distance Identification, and Symmetry. Then Objectivism, Lockeanism,  $KC$ , and Ideal Reliability imply that:

1. Ideal Confidence Safety fails at  $w_i$  and  $w_j$  iff (a)  $i > 2dc + c - d$ , (b)  $|i - j| < 2d^2$ , (c)  $j \leq c$ , and (d)  $c - d < i < c + d$  and  $c - d < j < c + d$ .
2. Ideal Confidence Safety fails at some  $w_i, w_j$  iff  $c - d < \frac{1}{2}$ .

The way in which our theory undermines Confidence Safety is not as straightforward as its treatment of Luminosity. As we saw, Luminosity fails immediately, since given our background assumptions it is provably violated by ideally reliable agents. Confidence Safety, by contrast, is consistent with our view provided that the  $c, l, d$  and  $s$  parameters are finely-tuned in exactly the right way. But once the thesis is seen to be so sensitive to the particular parameters in play, its plausibility diminishes considerably. Confidence Safety is offered as a general constraint on knowledge. Presumably then, if it is true, it is generally applicable, i.e. it is necessarily true and constrains not just knowledge of whether one feels cold but knowledge in general. But even if it happens to be that the relevant parameters for the mental state of feeling cold line up in the way required for Confidence Safety to hold, it is beyond the bounds of plausibility to hold that they do so for every mental state capable of undergoing gradual change. Furthermore, even if the parameters do line up, they presumably do so only contingently; so even if there are no actual failures of Confidence Safety, there are at least possible ones.<sup>22</sup> Thus, in light of our discussion, we see no reasonable way to preserve Confidence Safety as a general epistemological principle.<sup>23</sup>

### 4.3 | Safety

Our theory provides a principled way to reject Confidence Safety. It is natural to wonder whether our framework is generally hostile to safety principles. As a test case, we consider a coarse-grained safety principle according to which knowledge requires safe all-or-nothing confidence.

Let Safety be the principle that if an agent knows that  $p$ , then in any nearby cases in which they are confident in  $p$ , it must be the case that  $p$  is true.<sup>24,25</sup>

(28) **Safety.** If  $Kp$  is true at  $w_i$  and  $Cp$  is true at  $w_j \in N_i$ , then  $p$  is true at  $w_j$ .

We now consider whether Objectivism is compatible with Safety, focusing again on ideal agents:

(29) **Ideal Safety.** If  $K^1p$  is true at  $w_i$  and  $C^1p$  is true at  $w_j \in N_i$ , then  $p$  is true at  $w_j$ .

We saw above that our theory implies that ideal all-or-nothing confidence is factive. This immediately implies that Ideal Safety is valid. Whenever an ideal agent knows  $p$ ,  $p$  is bound by Factivity to be true in any nearby case where she is confident in  $p$ .

*Observation 8.* Suppose Cutoff and Symmetry. Then Objectivism and Lockeanism imply Ideal Safety.

Ideally reliable agents thus satisfy Safety. Since the degrees of confidence of non-ideal agents are not constrained in any further way by our theory, it follows that it is consistent with Safety. This is a pleasing result: coarse-grained safety principles possess considerable plausibility; indeed, they are accepted by many proponents and opponents of luminosity alike. It would have been a mark

against our theory if the most natural confidence-theoretic principles of this kind were ruled out. Fortunately, that is not the case. In this way, our theory provides the tools to distinguish between valid and invalid safety conditions on knowledge.

## 5 | CONCLUSION: FUTURE DIRECTIONS FOR CONFIDENCE

The primary aim of this paper has been to develop a theory of confidence, explore some of its connections to knowledge, and assess the anti-luminosity argument in its light. In this concluding section, we briefly explore the extent to which the Dispositionalist view of confidence introduced the paper motivates various structural constraints on degrees of confidence, address some possible concerns about the notion, and gesture towards some further philosophical applications.

### 5.1 | Structural constraints on confidence

A natural question is whether the account of confidence we have offered gives rise to any substantive structural constraints on degrees of confidence—the most salient being whether an agent's (or a *rational* agent's) degrees of confidence must satisfy the axioms of probability.<sup>26</sup>

Probabilism imposes three synchronic norms on degrees of belief: that they can be represented by real numbers in the interval  $[0, 1]$ , that every tautology is assigned degree 1, and that they are additive, so that for any exclusive  $p$  and  $q$ , the agent's degree of belief in  $p$  or  $q$  is the sum of her degree of belief in  $p$  and in  $q$ .

The Dispositionalist theory of confidence immediately satisfies the first requirement of probabilism, since it views degrees of confidence as chances of a certain kind, which therefore fall in the interval  $[0, 1]$ .

Whether our theory satisfies the second requirement depends on the theory of reliance. The constraint corresponds to the requirement that (rational) agents always rely on tautologies in practical reasoning. This seems an innocent initial assumption, and one that is especially plausible on the pragmatic theory of reliance discussed in §3.

Crucially, however, our theory rejects the additivity requirement on confidence. Our theory is superadditive: an agent's confidence in a disjunction can exceed the sum of her confidence in each disjunct. To see why, return to the case of a fair lottery. An agent's degree of confidence that her ticket will either win or lose is 1, since the chance that she relies on this in practical reasoning is 1. However, her degree of confidence in each disjunct may be 0: after all, she may not rely at all on her ticket losing while simultaneously not relying at all on her ticket winning. Examples of this sort are no mere edge cases, but are important to the way in which confidence is initially motivated, by Williamson and those following him, as a state distinct from credence. Furthermore, at least *prima facie*, it is plausible that a *rational* agent's dispositions to rely can have this structure, in which case additivity is neither a constraint on agents in general nor on rational agents. Here the key point is that on our theory, degrees of confidence are understood *not* in terms of dispositions to take bets (as credences are, on the traditional conception), but rather in terms of dispositions to rely. As a result, interestingly different structural constraints arise for degrees of confidence than for credences, classically conceived.

Our denial of additivity raises the question of whether other rational requirements on confidence can be motivated. A natural weakening to consider is a rule of comparative entailment: that whenever  $p$  entails  $q$ , an agent's degree of confidence in  $p$  should not exceed her confidence

in  $q$ . One way to motivate this constraint would be to argue that it is necessary for all agents, in virtue of the fact that reliance is itself closed under logical entailment. Such a claim would be roughly analogous to the well-known thesis of Stalnaker (1984) that *belief* is closed under logical entailment.<sup>27</sup> A different version of this constraint might be motivated by showing that *rational* agents must conform to it. One argument to this effect can be sketched as follows: if an agent is rational in relying on  $p$ , then they are rational in acting in such a way that presupposes that  $p$  is true. But if  $p$  is true, and  $p$  entails  $q$ , then  $q$  must also be true. So—at least to the extent that the truth of  $q$  is relevant to the decisions facing the agent—failing to rely on  $q$  seems to give rise to pragmatic incoherence, and hence, arguably, irrationality.

Our discussion here is far from conclusive; much work remains to be done in exploring constraints on confidence. The examples above suggest a general strategy: by analogy with Dutch Book arguments in the case of credence, we can evaluate whether a pattern of confidence is rational by considering whether its associated dispositions are pragmatically appropriate.<sup>28</sup>

A final observation to make is that, on natural background assumptions, all three theories of *ideally reliable* confidence considered in §3 imply that the degrees of confidence of an ideally reliable agent are indeed probabilistic. Since ideal reliability in the sense discussed is an idealization more akin to assuming *full information* than *perfect rationality*, we do not take this observation to conflict with our claim above that probabilism about confidence is not rationally required. Nevertheless, we think it is potentially interesting that, on all of the natural views, agents who are idealized along a salient epistemic dimension satisfy probabilism. The confidence profile of ideally reliable agents may satisfy additional constraints: for instance, Objectivism together with the Principal Principle (Lewis, 1980) imply that an ideal agent who knows the chances (and no inadmissible information) will identify them with both her credence and her degrees of confidence. Again, we leave further investigation for future work.

## 5.2 | The theoretical role of confidence

Some readers might reasonably wonder about the acceptability of the concept of confidence. Is it in good standing, and even if so, is it capable of earning its philosophical keep? If confidence does not have much theoretical utility, one might worry that the important epistemological role that we assign it is undermotivated. In that case, our results above would be of limited interest.

An immediate point to make is that a certain kind of scepticism about the notion is misplaced. Our account of confidence explains what it is for an agent to be confident in a proposition using the notions of (i) objective chance and (ii) reliance by an agent on a proposition. Anyone who accepts these resources ought to accept that confidence, so defined, exists—it is not simply a philosophical fiction.

But even granted that our notion of confidence is in good standing, it does not follow that it plays a significant theoretical role. In particular, we've relied on the principle that knowledge requires confidence; but this might be disputed if confidence turned out to be an artificial notion tailor-made to a particular argument. We've already discussed several links between confidence and the theory of knowledge. In this final subsection, we consider further applications of confidence in epistemology and the theory of action. None of these applications are conclusive; indeed, we hope to explore some further in future work. We include them here because we think that they potentially strengthen the case for confidence: if it really can play some or all of these additional roles, it is more plausible that it is linked to knowledge in the ways discussed above.

A first potential application of confidence concerns puzzles about the epistemic significance of merely statistical evidence. In cases of merely statistical evidence, an agent's only information about whether  $p$  concerns the objective chance of  $p$ . In this way, cases of merely statistical evidence resemble Lottery Bag. Consider the following case from the legal realm:

- (30) **Blue Bus.** Mary's car is hit by a blue bus. Mary knows that 80% of blue buses are owned by the Blue Bus Company (BBC), and 20% by the Red Bus Company. Mary has no other evidence about which company owned the bus.

In Blue Bus, Mary has merely statistical evidence for  $p$ . Buchak (2014) observes that merely statistical evidence licenses quite different behavior than other kinds of evidence:

- (31) **Green Bus.** Mary's car is hit by a green bus. Mary knows that Half of all green buses are owned by the Green Bus Company (GBC), and half by the Yellow Bus Company. Mary also knows that an eyewitness identifies the bus as belonging to the Green Bus company. Eyewitness reports are 80% reliable in general.

In Blue Bus, Mary's credence that BBC is responsible is 0.8. This is the same as her credence that GBC is responsible in Green Bus. Nonetheless, Buchak (2014) argues that Mary should behave differently in the two cases. In Green Bus, Mary can permissibly blame GBC for the accident; but she cannot blame BBC in Blue Bus. Buchak (2014) suggests that this difference in permissibility tracks a corresponding epistemic difference: Mary believes that BBC is responsible in Blue Bus; but she does not believe that GBC is responsible in Green Bus. In this way, Buchak (2014) uses this case to argue against a simple form of Lockeanism, on which belief is simply high credence.

A natural explanation of these examples appeals to confidence. In Green Bus, Mary is disposed to rely on GBC being responsible. She manifests this disposition by blaming GBC, taking them to court, etc. In Blue Bus, Mary has a low degree of disposition to rely on BBC being responsible. Thus we can explicate Buchak (2014)'s distinction by suggesting that belief requires a high degree of *confidence*, and not merely a high degree of *credence*. This in turn strengthens our case for the assumption that knowledge requires high confidence.

To defend this proposal in detail, we need to explain why Mary can rationally have different degrees of confidence in the two cases. Our earlier account of ideal reliability does not make the necessary prediction. The objective chance of the relevant company being responsible is not higher in Blue Bus than in Green Bus. So the ideally reliable agent's confidence is no higher in Green Bus.

On the other hand, ordinary rational agents are not ideally reliable. A natural next step is then to give a theory of the rational norms on confidence for ordinary agents, once we have left the realm of ideal reliability behind. For example, perhaps there is a further kind of safety norm on degrees of confidence, which forbids agents to foresee that they will rely on  $p$  at a nearby world where  $p$  is false.

- (32) **Safety\*.**  $Ch(Rp \wedge \neg p) = 0$

Imagine that in Blue Bus Mary were to form a high degree of confidence that BBC is responsible for the action. Then in most nearby worlds, she would rely on BBC being responsible. Since Mary's evidence in Blue Bus is merely statistical, there are worlds consistent with her evidence in which the Red Bus Company committed the crime. But Mary can foresee that if she relies on BBC being

responsible at most nearby worlds, she will rely on BBC being responsible in a world where they are not responsible. That is, Mary would know that there is a chance that she relies on  $p$  when  $p$  is false. Perhaps this is irrational.

By contrast, one might think that in cases of eyewitness testimony, there can be uncertainty about whether the nearby worlds include ones in which Yellow Bus Company is responsible. If Mary doesn't know that the nearby worlds have this structure, then she doesn't know that she violates Safety\*. On this proposal, the difference between merely statistical evidence and other kinds of evidence concerns our knowledge about nearby worlds.

Safety\* is interestingly weaker than Objectivism, because it allows that there are cases in which  $p$  is true and an agent does not rely on  $p$ , because she relies on neither  $p$  nor not  $p$ . Safety\* encodes a norm to avoid relying on falsehoods; it does not require relying on every truth.

The Blue Bus case is structurally similar to Lottery Bag. In Lottery Bag, the agent also has merely statistical evidence about the ball. In Lottery Bag as in Blue Bus, the agent intuitively should not rely on the relevant claim in practical reasoning. There is thus hope that Safety\* can also explain our judgments about Lottery Bag.

This isn't the only way one might appeal to confidence to explain the puzzle of merely statistical evidence. Another option would be to appeal to the principle that if  $S$  knows that  $S$  doesn't know that  $p$ , then  $S$ 's degree of confidence that  $p$  ought to be 0.

(33) **Safety\*\***.  $K\neg Kp \rightarrow C(p) = 0$

This might follow from a more general requirement: if you don't know  $p$  at  $w$ , then you shouldn't rely on  $p$  at  $w$ . Then one could argue that in Blue Bus and Lottery Bag, the relevant agent knows that she doesn't know the relevant proposition. A natural next question is whether one could motivate Safety\*\* from facts about the nature of reliance or confidence. (For example, Safety\*\* might follow from the theory that an agent's degree of confidence in  $p$  is simply her credence that she knows  $p$ .)

A second potential application for the notion of confidence is in making sense of pragmatic encroachment on knowledge. Consider the following pair of cases.<sup>29</sup>

(34) **Low**. Hannah made three sandwiches and placed them in the refrigerator. She told Sarah that she placed the peanut butter sandwich on the left, the tuna sandwich in the middle, and the almond butter sandwich on the right. Sarah knows that Almira has no allergies. Almira says: 'I'd love an almond butter sandwich.' And so Sarah opens the refrigerator door, points to the sandwich on the right, and says: 'The sandwich on the right is an almond butter sandwich. You can have it.'

(35) **High**. Just like Low, except it is Sarah's nephew Algernon who is visiting for lunch, and he has a severe peanut allergy. He asks Sarah for a sandwich. Sarah knows that the peanut butter sandwich would be fatal to Algernon, but that the almond butter sandwich would be harmless. She also knows that he would slightly prefer the almond butter sandwich to the tuna sandwich. So she gives him the tuna sandwich.

In Low, Sarah knows that the sandwich on the right is made of almond butter; in High she does not. In Low and High, Sarah has the same evidence. But in High, the practical stakes regarding whether the sandwich is made of almond butter are much higher.

Above, we considered a pragmatic account of confidence, on which reliance was understood in terms of preference. That account built on Weatherson (2005)'s theory of belief. For that reason,

our theory gives a natural account of the cases above. In Low, Sarah's degree of confidence in the sandwich containing almonds is higher than her degree of confidence in High. The stakes do not simply affect the threshold of credence required for justified belief; they affect the agent's actual level of confidence.<sup>30</sup>

Why do stakes affect levels of confidence? Our pragmatic account of reliance suggests an answer. When the stakes regarding  $p$  increase, the agent has a lower chance of relying on  $p$  at nearby worlds. After all, Reliance by Action says that an agent relies on  $p$  when she performs the action that maximizes expected utility given  $p$ . Let  $p$  be the proposition that the sandwich on the right contains almond butter, and let  $R$  be the action of giving the relevant person the sandwich on the right. In Low and High,  $R$  is the action that maximizes expected utility given  $p$ . In Low, Sarah is highly disposed to perform  $R$ ; in High, she is not. So our account of confidence immediately predicts that Sarah's degree of confidence in  $p$  is high in Low, and low in High.

This explanation of pragmatic encroachment provides further support for KC. Given KC, Sarah's changing degrees of confidence generate different predictions about knowledge. Since in High, Sarah has a low degree of confidence in  $p$ , Sarah does not count as knowing  $p$ . By contrast, KC provides no barrier to knowing  $p$  in Low.

We suspect that the theory of confidence may also be fruitfully applied to other interesting questions in epistemology, at which we can only gesture here. First, the theory could be extended into a theory of *degrees* of credence. On this picture, an agent's credence in  $p$  of, say,  $\frac{1}{2}$  could itself come in degrees, depending on how robustly she is disposed to accept various odds. The reliability of these degrees of confidence might then be of use in giving a theory of probabilistic knowledge (Moss, 2018). Second, perhaps confidence can be used to distinguish between cases of risk and uncertainty in a satisfying way. Finally, a natural question is whether we can extend our account of ideal reliability to a full fledged analysis of knowledge. One possibility is that  $S$  knows that  $p$  iff the ideally reliable agent knows  $p$ ,  $S$  is confident in  $p$ , and  $S$ 's degree of confidence in  $p$  is sufficiently reliable, understood in terms of sufficient closeness to the ideal degree of confidence in  $p$ .<sup>31</sup>

## ENDNOTES

<sup>1</sup> For responses to Williamson (2000), see Leitgeb (2002), Berker (2008), Ramachandran (2009), and Cohen (2010). Srinivasan (2013) provides a recent defense of anti-luminosity.

<sup>2</sup> Strictly speaking, luminosity is defined in terms of being in a position to know rather than knowing. Williamson (2000, p. 95) glosses the notion as follows: "If one is in a position to know  $p$ , and one has done what one is in a position to do to decide whether  $p$  is true, then one does know  $p$ ." Following the literature, we make the simplifying assumption that our agent does everything they are in a position to do to decide whether the relevant propositions are true.

<sup>3</sup> The existence of a cutoff is (classically) equivalent to Williamson's stipulations that the agent starts off feeling cold, ends up not feeling cold, and warms up over time. The existence of a cutoff is a standard assumption about gradable adjectives like *cold*; it is compatible with the vagueness of 'feels cold', since, for example, indeterminacy in whether an agent feels cold can be understood in terms of indeterminacy about the location of the cutoff. See Kennedy (2007) for a recent discussion of gradable adjectives, cutoffs, and their connection to vagueness.

More generally, some readers might worry that the appeal to cutoffs will lead to soritical arguments in what follows. Here we appeal to something like the response in Williamson (2000, p. 104): if an argument is soritical, then the result of precisifying vague terms in the premises leads to a reduction in their plausibility. But when 'feels cold' and 'is confident' are replaced by precisifications 'feels cold to degree  $m$ ' and 'is confident to degree  $n$ ', our premises become even more plausible. Williamson's original anti-luminosity argument has been criticized (see e.g. Berker, 2008) for ignoring the possibility of penumbral connections; we discuss this issue further in §4.2.



- <sup>4</sup> Nothing turns on the particular value chosen here; for our purposes, we need only that degrees of coldness in the interval  $[0,1]$  can be subdivided sufficiently finely that the agent's confidence in 'neighboring' values is similar.
- <sup>5</sup> See Berker (2008), Schafer (2015), Schoenfield (2015), Srinivasan (2015), Williamson (2000), p. 110-113 and Dummett (1978, p. 131).
- <sup>6</sup> See again Berker (2008) and Weatherson (2004).
- <sup>7</sup> We assume that an agent's degrees of confidence occupy a ratio scale with a distinguished minimum (no confidence) and distinguished maximum (full confidence), allowing their representation by a real-valued function from the set of propositions to the interval  $[0,1]$ . For more on scales of measurement, see Krantz, Luce, Suppes, and Tversky (1971).
- <sup>8</sup> The view in Manley and Wasserman (2008) is more complex, understanding degrees of disposition in terms of cases rather than worlds. Each case is a precise specification of the background conditions relevant to the disposition ascription (how high is the glass? how hard does the wind blow?).
- <sup>9</sup> Before continuing, it's worth clarifying the notion of objective chance at issue. Throughout, we appeal to this notion of chance to understand an agent's degree of confidence in various claims. For this reason, it is important that this chance function is not automatically opinionated about these propositions. So we cannot analyze an agent's present degrees of confidence in terms of the present chance function, since this will lead to opinionation about all past facts, while an agent can have intermediate degrees of confidence about, say, whether it was cold yesterday. In response to this problem, there are a variety of options we might rely on. Manley and Wasserman (2008) offer two suggestions here (p. 80). First, they suggest relying on the Louiville measure on phase space, which will help make sense of the proportional relationship between various claims about the world. Second, they suggest that one could appeal to the kind of Williamsonian *ur* prior that many epistemologists have used to make sense of evidential probability. Finally, departing from the suggestions in Manley and Wasserman (2008), we could appeal to the chance function from the time right before various relevant propositions are intuitively 'resolved'. One problem for this last proposal, however, is that it might lead to trouble when comparing an agent's degrees of confidence about propositions that are resolved at quite different times. Finally, one place where this question affects our argument is when we suggest that any agent who defers their credences to the chances will match their credences to their degrees of confidence. This result is only interesting if the relevant notion of chance in our argument is one to which rational agents defer.
- <sup>10</sup> The analogy is with Lockeanism concerning credence/belief. For more, see Foley (1993, 2009), Christensen (2005), Sturgeon (2008), Beddor and Goldstein (2018), and Dorst (2018). For applications of Lockeanism in a confidence-theoretic setting, see e.g. Srinivasan, 2013, p. 20.
- <sup>11</sup> Williamson, 2000, p. 98-9; see also Williamson, MS, p. 9 for a more recent endorsement. 'Confidence' may be an infelicitous term for the notion, to the extent that it is intuitive to say that one's degrees of confidence should track known objective chances. Perhaps Williamson's 'degree of outright belief' would be better; still, we stick with 'confidence' given its entrenchment in the literature. Thanks to an anonymous referee for pressing us to address these points.
- <sup>12</sup> For discussion, see Wong (2008) and Cohen (2010).
- <sup>13</sup> Treatments of knowledge as safe belief are considered in Beddor and Pavese (forthcoming), Lasonen-Aarnio (2010), Pritchard (2005), and in Williamson (2000) in a non-reductive form.
- <sup>14</sup> In their own words:

A calibrated system simply draws the relevant distinctions appropriately, such that there is no danger of misapprehending a total state in which one does not feel cold by having the simultaneous belief that one does. (Leitgeb, 2002, p. 216)

[If Calibration is satisfied, then], at  $[w_i]$  one just barely feels cold and one just barely believes one feels cold. So how is one's confidence at  $[w_i]$  misplaced? Clearly one should not have more confidence at  $[w_i]$ . (Cohen, 2010, p. 726)

For another potential defense of this view, see Zardini (2012).

- <sup>15</sup> If different thresholds for confidence and coldness are allowed, the statement of Calibration will be more complex. The view would need to be tweaked so that the ideal agent's degree of confidence in  $p$  exceeds the threshold

for outright confidence exactly when  $i$  exceeds the threshold for feeling cold. There is no guarantee in such a scenario that  $C_i^1(p) = i$ , but some analogue of this equation would hold with suitable scaling factors applied.

- <sup>16</sup> The crucial passage in which this suggestion is found is Srinivasan (2013)'s discussion, on p. 20, of the case of Henrietta.
- <sup>17</sup> Several readers have suggested that Ideality makes strange predictions about the degrees of confidence of an omniscient agent. One might think that God should have a degree of confidence of 1 in all the truths at any world. But Ideality seems to predict that this is false. If  $p$  is true at  $w$  but could easily have been false, Ideality predicts that the ideally reliable degree of confidence in  $p$  is less than 1. We don't see this as a serious objection to Ideality, but rather as a helpful way of illustrating the view. Which degrees of confidence are ideal should depend on what confidence is; and on our conception of confidence, there doesn't seem to be any particular reason why God would have maximal degrees of confidence. On the other hand, for those who are moved by this worry, one option is to say that when we assess God's degrees of confidence we use a conception of nearness on which every world is uniquely near itself. On this conception, God would have maximal confidence in all and only the truths.
- <sup>18</sup> Here our models are structurally similar to the distance-based models of epistemic accessibility explored in Williamson (2011), Goodman (2013), and Carter (2018).
- <sup>19</sup> Interpreting 'nearby' is a matter of some delicacy. Another possible account is that a world is nearby to  $w_i$  iff it is consistent with what is known by an agent at  $w_i$ . On this view, Confidence Safety simply says that an agent has a confidence of degree 1 in whatever they know. But there are independent reasons not to accept this view of 'nearby'. In particular, it has a trivializing effect on standard formulations of (coarse-grained) safety, e.g. that knowledge requires no nearby false belief. (For if one knows  $p$ , then on this view there is no nearby case where  $p$  is false, hence trivially no nearby case where one falsely believes  $p$ ).
- <sup>20</sup> Even if all-or-nothing confidence and belief come apart, confidence—a disposition to rely on  $p$ —is very plausibly necessary for belief that  $p$ . Earlier we considered reasons to be wary of a dispositional analysis of belief; but these reasons challenge the sufficiency of the dispositional analysis, not its necessity.
- <sup>21</sup> As we noted earlier, Luminosity has often been formulated in terms of being in a position to know, rather than knowing proper (Williamson, 2000). A defender of Luminosity might exploit this distinction to defend Luminosity from our objection. On the resulting view, even though the ideal agent does not know  $p$ , she is in a position to know  $p$ . Of course, to implement this view requires a substantive conception of being in a position to know. For example, one might think that being in a position to know  $p$  implies that if one has considered the question of whether  $p$  and formed a degree of confidence on this basis, then one knows  $p$ . On this conception of being in a position to know, the ideal agent is not in a position to know, since she has a certain degree of confidence in  $p$  and yet does not know. On the other hand, another conception of being in a position to know, developed in Cohen (1999) and endorsed in Schaffer (2007), suggests that an agent is in a position to know  $p$  iff were she to competently form a belief in  $p$  on the basis of evidence she currently possesses, she would know  $p$ . One might then argue that the ideal agent could raise her confidence in  $p$  on the basis of her current evidence (feeling like *this*) and thereby come to know.
- If an ideal agent who is not confident in  $p$  were to raise her confidence in  $p$  while retaining her previous evidence, she would no longer be ideal in our sense. Thus our current theory does not issue an immediate verdict about whether she would then know  $p$ . This allows some room for a version of Luminosity that appeals to being in a position to know. However, one obstacle to this view is that it requires that non-ideal agents can know more than ideal agents. That is: an agent could shift her degrees of confidence to become less reliable, and thereby gain knowledge. This is difficult to accept. That being said, one way to investigate this view further might be to give an analysis of knowledge where  $S$  knows that  $p$  iff  $S$  is confident in  $p$ ,  $p$  is true, and  $S$ 's degree of confidence in  $p$  is sufficiently close to the ideal agent's degree of confidence in  $p$ . Then one could characterize the range of potential Luminosity violations in terms of this new parameter: sufficient proximity to ideal confidence.
- <sup>22</sup> The situation is even worse for the defender of Confidence Safety in light of the vagueness of 'feels cold' and 'is confident'. For Confidence Safety to hold in all precifications, these terms must be constrained by a penumbral connection that forces the relevant parameters to line up. Since no natural philosophical thesis entails that the parameters line up in this way, we should treat this as extremely implausible.
- <sup>23</sup> Cases where Confidence Safety fail are also plausibly cases where Williamson-style margin-for-error principles fail, for  $K^1 p$  at  $w_i$  but  $\neg p$  at  $w_{i-\epsilon}$ .

- <sup>24</sup> Coarse-grained safety principles are usually formulated in terms of belief. Since we are thinking of confidence as a kind of degreed notion of belief, it is natural to consider versions of coarse-grained safety formulated, as is Safety, in terms of all-or-nothing confidence. For influential defenses of safety principles concerning belief, see Sosa (1999), Williamson (2000), and Pritchard (2005) among others.
- <sup>25</sup> Notice that Safety does not imply Confidence Safety, since at  $w_j$  the agent may fall below the threshold necessary for outright confidence, even though her confidence levels at  $w_i$  and  $w_j$  are similar. But Confidence Safety implies Safety given the plausible assumption that if S knows  $p$  at  $w_i$ , and the agent is confident in  $p$  at  $w_j \in N_i$ , then there is some world  $w_k$  where  $Kp$  is true at  $w_k$ ,  $j \in N_k$ , and  $C_k(p) \approx C_j(p)$ .  $w_k$  might well be  $w_i$ ; but even if the agent's confidence at  $w_i$  is too different from her confidence at  $w_j$ , there will plausibly be some intermediate world  $k$  between  $w_i$  and  $w_k$  where  $Kp$  is still true, and yet the agent's confidence is similar to her confidence at  $w_j$ .
- <sup>26</sup> Thanks to an anonymous referee for pressing us to clarify this point.
- <sup>27</sup> Although such a claim would be relatively natural on the pragmatic account of reliance discussed in the paper, we prefer not to take a firm stand here, since we suspect that the surrounding dialectic would be similar to that arising for Stalnaker's account of belief and would therefore take us far afield.
- <sup>28</sup> Some recent work of possible relevance to this strategy is Goldstein (2016), which investigates the norms on rational intention by exploring the conditions under which the dispositions characteristic of intending are pragmatically appropriate.
- <sup>29</sup> The case is originally from Ross and Schroeder (2014). The presentation below is from Gao (2018). See Cohen (1999), DeRose (1992), Gerken (2017), and Stanley (2005) for similar cases.
- <sup>30</sup> This treatment of degrees of confidence is structurally analogous to Gao (2018)'s analysis of pragmatic encroachment in terms of shifting levels of credence.
- <sup>31</sup> Versions of this paper were presented at Lingnan University, the National University of Singapore, NYU Shanghai, University College London, and Yonsei University. We are grateful to the audiences for their feedback. For helpful suggestions, comments, and discussion, we would like to thank an anonymous referee for this journal, Derek Baker, Bob Beddor, Sam Carter, Dan Greco, Sebastian Liu, Jenny Nado, Lavinia Picollo, Lorenzo Rossi, Miriam Schoenfield, and Jiji Zhang.

## REFERENCES

- Beddor, B., & Goldstein, S. (2018). Believing epistemic contradictions. *Review of Symbolic Logic*, *11*, 87–114.
- Beddor, B., & Pavese, C. (forthcoming). Modal virtue epistemology. *Philosophy and Phenomenological Research*.
- Berker, S. (2008). Luminosity regained. *Philosophers' Imprint*, *8*(2), 1–22.
- Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, *169*, 285–311.
- Carter, S. (2018). Higher order ignorance inside the margins. *Philosophical Studies*, 1–18.
- Christensen, D. (2005). *Putting Logic in its Place*. Oxford University Press.
- Cohen, S. (1999). Contextualism, skepticism, and the structure of reasons. *Philosophical Perspectives*, *13*, 57–89.
- Cohen, S. (2010). Luminosity, reliability, and the sorites. *Philosophy and Phenomenological Research*, *81*(3), 718–730.
- DeRose, K. (1992). Contextualism and knowledge attributions pragmatism. *Philosophy and Phenomenological Research*, *52*, 913–929.
- Dorst, K. (2018). Lockean maximize expected accuracy. *Mind*.
- Dummett, M. (1978). *Truth and Other Enigmas*. Duckworth.
- Foley, R. (1993). *Working Without a Net*. Oxford University Press.
- Foley, R. (2009). Belief, degrees of belief, and the Lockean thesis. In Franz Huber and Christoph Schmidt-Petri (Eds.), *Degrees of Belief* (pp. 37–47). Springer.
- Gao, J. (2018). Credal pragmatism. *Philosophical Studies*, *Forthcoming*.
- Gerken, M. (2017). *On Folk Epistemology*. Oxford University Press.
- Goldstein, S. (2016). A preface paradox for intention. *Philosophers' Imprint*, *16*(14).
- Goodman, J. (2013). Inexact knowledge without improbable knowing. *Inquiry*, *56*, 30–53.
- Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*(1), 1–45.
- Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Vol. I: Additive and Polynomial Representations*. New York Academic Press.

- Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives*, 24, 1–21.
- Leitgeb, H. (2002). Review of ‘Timothy Williamson, *Knowledge and Its Limits*’. *Grazer Philosophische Studien*, 65, 207–217.
- Lewis, D. (1980). A subjectivist’s guide to objective chance. In R. C. Jeffrey (Ed.), *Philosophy of Probability: Contemporary Readings* (pp. 83–132). University of California Press.
- Manley, D., & Wasserman, R. (2008). On linking dispositions and conditionals. *Mind*, 117(465), 59–84.
- Moss, S. (2018). *Probabilistic Knowledge*. Oxford University Press.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford University Press.
- Ramachandran, M. (2009). Luminosity and the safety of knowledge. *Australasian Journal of Philosophy*, 87(4), 659–673.
- Ross, J., & Schroeder, M. (2014). Belief, credence, and pragmatic encroachment. *Philosophy and Phenomenological Research*, 88, 259–288.
- Schafer, K. (2015). Doxastic planning and epistemic internalism. *Synthese*, 191(12), 2571–2591.
- Schaffer, J. (2007). Closure, contrast, and answer. *Philosophical Studies*, 133, 233–255.
- Schoenfield, M. (2015). Internalism without luminosity. *Philosophical Studies*, 25(1), 252–272.
- Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, 13, 141–154.
- Srinivasan, A. (2013). Are we luminous? *Philosophy and Phenomenological Research*, 90(2).
- Srinivasan, A. (2015). Normativity without cartesian privilege. *Philosophical Issues*, 25, 273–299.
- Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.
- Stanley, J. (2005). Fallibilism and concessive knowledge attributions. *Analysis*, 65, 126–131.
- Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, 42(1), 359–396.
- Weatherson, B. (2004). Luminous margins. *Australasian Journal of Philosophy*, 83, 73–83.
- Weatherson, B. (2005). Can we do without pragmatic encroachment. *Philosophical Perspectives*, 19, 417–43.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.
- Williamson, T. (2011). Improbable knowing. In T. Dougherty (Ed.), *Evidentialism and its Discontents*. Oxford University Press.
- Williamson, T. (MS). Knowledge, credence, and the strength of belief. available online at <http://media.philosophy.ox.ac.uk/docs/people/williamson/Strengthofbelief.pdf>.
- Wong, W. H. (2008). What Williamson’s anti-luminosity argument really is. *Pacific Philosophical Quarterly*, 89, 536–543.
- Zardini, E. (2012). Luminosity and vagueness. *Dialectica*, 66, 375–410.

**How to cite this article:** Goldstein S, Waxman D. Losing confidence in luminosity. *Noûs*. 2020;1–30. <https://doi.org/10.1111/nous.12348>

## APPENDIX: PROOFS OF OBSERVATIONS

Here we demonstrate observations not proven in the main text.

*Observation 2.* Suppose Cutoff and Symmetry. Then Objectivism implies that:

$$C_i^!(p) = \begin{cases} 1 & \text{if } c < i - d \\ \frac{i+d-c}{2d} & \text{if } i - d < c < i + d \\ 0 & \text{if } c > i + d \end{cases}$$

*Proof.*

$$\begin{aligned} C_i^!(p) &= Ch(p \mid N_i) && \text{(Objectivism)} \\ &= \frac{Ch(p \wedge N_i)}{Ch(N_i)} \\ &= \frac{Ch(\{w_j \mid j > c \wedge [i - d < j < i + d]\})}{2d} && \text{(Cutoff, Symmetry)} \end{aligned}$$

To evaluate this final term, there are three possible cases.

- $c < i - d$ . Then no nearby worlds are above the cutoff  $c$ , and thus  $C_i^!(p)$  is 1.
- $c > i + d$ . Then no nearby worlds are above the cutoff  $c$ , and thus  $C_i^!(p)$  is 0.
- $i - d < c < i + d$ . Then  $Ch(\{w_j \mid j > c \wedge [i - d < j < i + d]\}) = Ch(\{w_j \mid c < j < i + d\})$ , so that the proportion of the nearby worlds where  $p$  holds is  $i + d - c$ . Thus  $C_i^!(p)$  is  $\frac{i+d-c}{2d}$ .

□

*Observation 3.* Suppose Cutoff and Symmetry. Then Objectivism implies that  $p$  is true at  $w_i$  iff  $C_i^!(p) > \frac{1}{2}$ .

*Proof.* By Symmetry and Cutoff,  $p$  is true at  $w_i$  iff  $Ch(p \mid N_i) > \frac{1}{2}$ . So by Objectivism, this holds iff  $C_i^!(p) > \frac{1}{2}$ . □

*Observation 4.*

4.1 Suppose Cutoff and Symmetry. Then Objectivism and Lockeanism imply Factivity.

4.2 Factivity, KC, and Ideal Reliability imply Collapse.

*Proof.* Part 1: By Lockeanism,  $C^!p$  iff  $C^!(p) > l > \frac{1}{2}$ . By the previous observation,  $C^!(p) > \frac{1}{2}$  implies  $p$ .

Part 2: Suppose  $C^!p$ . By Factivity,  $p$ . By Ideal Reliability,  $K^!p$ . By KC,  $K^!p$  implies  $C^!p$ . □

*Observation 5.* Suppose Cutoff, Cutoff Identification, and Symmetry. Then Objectivism, Lockeanism, KC, and Ideal Reliability imply  $K^!p$  is true at  $w_i$  iff  $i > 2dc + c - d$ .

*Proof.* There are three cases to consider.

Case (a):  $i - d < c < i + d$ , so that  $C^1 p = \frac{i+d-c}{2d}$ .

Suppose that  $K^1 p$  is true at  $w_i$ . By KC, we know that  $C^1 p$ , and so by Symmetry, Cutoff, Cutoff Identification, and Objectivism,  $\frac{i+d-c}{2d} > c$ .

Conversely, suppose  $\frac{i+d-c}{2d} > c$ . By Cutoff, Symmetry, Cutoff Identification, Lockeanism, and Objectivism, we know that  $C^1 p$ . From Collapse, we can infer  $K^1 p$ . Moreover, our previous observation showed that Collapse is implied by our other assumptions. This simplifies to:  $K^1 p$  is true at  $w_i$  iff  $i > 2dc + c - d$ .

Case (b):  $c < i - d$ . Then  $K^1$  is true at  $w_i$ , since  $C^1(p) = 1$ . Since  $c < i - d$ , we can infer  $i > c + d$ . So the following inequality holds, where we substitute  $(c + d)$  for  $i$ :

$$\frac{i + d - c}{2d} > \frac{(c + d) + d - c}{2d} = \frac{2d}{2d} = 1 > c.$$

Thus  $\frac{i+d-c}{2d} > c$ .

Case (c):  $i + d < c$ . Then  $K^1$  fails at  $w_i$  since  $C^1(p) = 0$ . For this reason, the following inequality holds, where we substitute  $c$  for  $(i + d)$ :

$$\frac{i + d - c}{2d} < \frac{c - c}{2d} = \frac{0}{2d} = 0 < c.$$

Thus  $\frac{i+d-c}{2d} < c$ .

□

*Observation 6.* Suppose Cutoff, Cutoff Identification, and Symmetry. Then Objectivism, Lockeanism, KC, and Ideal Reliability imply that:

1. Ideal Luminosity fails at  $w_i$  iff  $c < i < 2dc + c - d$ .
2. Ideal Luminosity fails at some  $w_i$  iff  $c > 0.5$ .

*Proof.* For part 1, we know from Observation 6 that  $K^1 p$  is true at  $w_i$  iff  $i > 2dc + c - d$ . We also know that  $p$  is true at  $w_i$  iff  $i > c$ . Together, these facts imply that Ideal Luminosity fails at  $w_i$  iff  $c < i$  and  $i < 2dc + c - d$ .

For part 2, there exists some  $w_i$  where Ideal Luminosity fails iff for some  $i$ ,  $c < i < 2dc + c - d$ . But this holds iff  $c < 2dc + c - d$  iff  $d < 2dc$  iff  $1 < 2c$  iff  $\frac{1}{2} < c$ . □

*Observation 7.* Suppose Cutoff, Cutoff Identification, Distance Identification, and Symmetry. Then Objectivism, Lockeanism, KC, and Ideal Reliability imply that:

1. Ideal Confidence Safety fails at  $w_i$  and  $w_j$  iff (a)  $i > 2dc + c - d$ , (b)  $|i - j| < 2d^2$ , (c)  $j \leq c$ , and (d)  $c - d < i < c + d$  and  $c - d < j < c + d$ .
2. Ideal Confidence Safety fails at some  $w_i, w_j$  iff  $c - d < \frac{1}{2}$ .

*Proof.* For part 1, notice that there are three possibilities for each of  $i$  and  $j$ :

- (i)  $i$  (respectively,  $j$ ) is in the interval  $(c - d, c + d)$ , in which case  $C_i(p)$  is  $\frac{i+d-c}{2d}$ .  
(ii)  $i$  (respectively,  $j$ ) is above the interval, i.e.  $c + d < i$ , in which case  $C_i(p) = 1$ .  
(iii)  $i$  (respectively,  $j$ ) is below the interval, i.e.  $i < c - d$ , in which case  $C_i(p) = 0$ .

There are thus five cases to consider, without loss of generality. For each case we first calculate the conditions under which  $C_i^!(p) \approx C_j^!(p)$ , and then investigate the conditions under which Confidence Safety fails. Condition 9d) holds only in Case (1); in all other cases, we'll show that Ideal Confidence Safety holds.

Case (1). Both  $i$  and  $j$  are in the interval. Then  $C_i^!(p) \approx C_j^!(p)$  iff  $|\frac{i+d-c}{2d} - \frac{j+d-c}{2d}| < d$  iff  $|\frac{i-j}{2d}| < d$  iff  $|i - j| < 2d^2$ .

Conditions (a) and (c) are trivially satisfied. For condition (b), notice that this holds iff  $2dc + c - d - c < 2d^2$ , which holds iff  $2dc - d < 2d^2$  iff  $2c - 1 < 2d$  iff  $c - \frac{1}{2} < d$  iff  $c - d < \frac{1}{2}$ .

Case (2).  $i$  is above the interval and  $j$  is below. Then  $C_i^!(p)$  is not similar to  $C_j^!(p)$ , since 0 is not similar to 1.

No violations of Ideal Confidence Safety arise, since  $C_i^!(p) \not\approx C_j^!(p)$ .

Case (3).  $i$  is in the interval,  $j$  is above. Then  $C_i^!(p) = \frac{i+d-c}{2d}$  and  $C_j^!(p) = 1$ . So  $C_i^!(p) \approx C_j^!(p)$  iff  $|C_i^!(p) - C_j^!(p)| < d$  iff  $|1 - \frac{i+d-c}{2d}| < d$  iff  $1 - \frac{i+d-c}{2d} < d$  iff  $2d - (i + d - c) < 2d^2$  iff  $d + c - i < 2d^2$ .

No violations of Confidence Safety arise. For  $C_j^!(p) = 1$ , so  $K_j^!$  holds. Thus a violation occurs iff (i)  $C_i^!(p) \approx C_j^!(p)$  and (ii)  $p$  is false at  $w_i$ , i.e.  $i < c$ . But if a violation occurs,  $i < c$ , so  $c - i > 0$ , and so  $d < d + c - i < 2d^2$ . But then  $1 < 2d$  i.e.  $1/2 < d$ . But  $d > 1/2$  is ruled out by Symmetry, since it has the trivializing consequence that the agent's confidence level in  $p$  is  $1/2$  at every  $w_i$ .

Case (4).  $i$  is in the interval,  $j$  is below. Then  $C_i^!(p) = \frac{i+d-c}{2d}$  and  $C_j^!(p) = 0$ . So  $C_i^!(p) \approx C_j^!(p)$  iff  $\frac{i+d-c}{2d} < d$  iff  $i + d - c < 2d^2$ .

No violations of Confidence Safety arise. For  $C_j^!(p) = 0$ , so  $p$  is false at  $w_j$ . Thus a violation occurs iff (i)  $C_i^!(p) \approx C_j^!(p)$  and (b)  $K_i^!$  holds. But suppose a violation occurs. From part 1, (i) holds iff  $i + d - c < 2d^2$  iff  $i < 2d^2 + c - d$ .

Thus a violation occurs iff (i) and (ii) both hold iff  $2dc + c - d < i < 2d^2 + c - d$  iff  $2dc < i + d - c < 2d^2$  iff  $c < (i + d - c)/2d < d$ . This again implies that  $d > \frac{1}{2}$ .

Case (5). Both  $i$  and  $j$  are above (or below) the interval. Then  $C_i^!(p)$  and  $C_j^!(p)$  are both either 0 or 1 and thus  $C_i^!(p) \approx C_j^!(p)$ .

No violations of Confidence Safety arise, since  $p$  is true (or false) at both  $w_i$  and  $w_j$ .

For part 2, note first that whenever Ideal Confidence Safety fails at  $w_i$  and  $w_j$ , Ideal Confidence Safety also fails at  $w_i$  and  $w_k$ , where  $j < k < c$ . Letting  $\epsilon > 0$ , we can thus limit attention to the case of  $j = c - \epsilon$ . Similarly, whenever Ideal Confidence Safety fails at  $w_i$  and  $w_j$ , Ideal Confidence Safety also fails at  $w_k$  and  $w_j$ , where  $2dc + c - d < k < i$ . We can limit attention to the case of  $i = 2dc + c - d + \epsilon$ . In the limit case, then, conditions (a) and (c) become trivial, and condition (b)



collapses to the constraint that  $2dc + c - d - c < 2d^2$ , which holds iff  $2dc - d < 2d^2$  iff  $2c - 1 < 2d$  iff  $c - \frac{1}{2} < d$  iff  $c - d < \frac{1}{2}$ .  $\square$

We now consider relaxing the assumptions of Indifference, Cutoff Identification, and Distance Identification.

Suppose as before Dispositionalism ( $C^1(p) = n$  iff  $Ch_i(R^1 p \mid N_i) = n$ ), Cutoff ( $p$  is true at  $w_i$  iff  $i > c > \frac{1}{2}$ ), Lockeanism ( $C^1 p$  is true at  $w_i$  iff  $C^1(p) > l$  is true at  $w_i$ ), Centering ( $N_i = \{w_j \mid |i - j| < d\}$ ), and Similarity ( $C_i^1(p) \approx C_j^1(p)$  iff  $|C_i(p) - C_j(p)| < s$ ).

*Observation 9.*

1. Luminosity holds iff: whenever  $i > c$ ,  $\frac{Ch_i(\{w_j \mid j > c \wedge |i - j| < d\})}{Ch_i(\{w_j \mid |i - j| < d\})} > l$ .
2. Confidence Safety holds iff: if (i)  $\frac{Ch_i(\{w_j \mid j > c \wedge |i - j| < d\})}{Ch_i(\{w_j \mid |i - j| < d\})} > l$ , (ii)  $i > c$ , and (iii)  $|\frac{Ch_i(\{w_j \mid j > c \wedge |i - j| < d\})}{Ch_i(\{w_j \mid |i - j| < d\})} - \frac{Ch_j(\{w_k \mid k > c \wedge |j - k| < d\})}{Ch_j(\{w_k \mid |j - k| < d\})}| < s$ , then (iv)  $j > c$ .
3. Safety holds iff: if (i)  $\frac{Ch_i(\{w_j \mid j > c \wedge |i - j| < d\})}{Ch_i(\{w_j \mid |i - j| < d\})} > l$ , (ii)  $i > c$ , (iii)  $j \in N_i$ , and (iv)  $\frac{Ch_j(\{w_k \mid k > c \wedge |j - k| < d\})}{Ch_j(\{w_k \mid |j - k| < d\})} > l$ , then (v)  $j > c$ .