WHY ANTI-LUCK VIRTUE EPISTEMOLOGY HAS NO LUCK WITH CLOSURE

Maura PRIEST

ABSTRACT: In Part I, this paper argues that Duncan Pritchard's version of safety is incompatible with closure. In Part II I argue for an alternative theory that fares much better. Part I begins by reviewing past arguments concerning safety's problems with closure. After discussing both their inadequacies and Pritchard's response to them, I offer a modified criticism immune to previous shortcomings. I conclude Part I by explaining how Pritchard's own arguments make my critique possible. Part II argues that most modal theories of knowledge will run into problems similar to those found in Pritchard's Anti-Luck Virtue Epistemology. I hence offer my own theory grounded in risk assessment and explain why and how it does much better.

KEYWORDS: safety, closure, barns, risk. Edmund Gettier

Preliminary Remarks

"Anti-Luck Virtue Epistemology" is Duncan Pritchard's response to what he perceived as an inability of a pure anti-luck theory to accommodate the widespread intuition that knowledge is a product of the knower's cognitive abilities. Pritchard hence modifies his account by incorporating a virtue component; therein lies his move from an anti-luck epistemology to Anti-luck Virtue Epistemology (ALVE).[1] According to ALVE, knowledge consists of two related epistemic criteria to satisfy two compelling intuitions. The anti-luck criterion, the intuition that knowledge is incompatible with luck; the virtue criterion, that knowledge is a product of the knower's cognitive abilities.[2]

This paper argues that safety, Pritchard's *anti-luck* criterion, is incompatible with closure and then offers an alternative solution which does not run into the same problem. I divide the paper into two parts, a negative part and a positive one. Part one is my negative argument against Pritchard's anti-luck virtue epistemology (ALVE). It begins by reviewing past arguments concerning safety's

[1] Duncan Pritchard, "Anti-Luck Virtue Epistemology," *The Journal of Philosophy* 109 (2012): 247-248.

[2] Pritchard, "Anti-Luck," 247-248.

problems with closure. I then explain both the inadequacies of such arguments and Pritchard's response to them. Third, I offer a modified criticism immune to the previously mentioned shortcomings. I conclude my negative argument with an explanation of how Pritchard's arguments for ALVE push him into this predicament.

The positive part of my paper argues in favor of an alternative theory of knowledge that avoids the closure dilemma. I suggest the modal aspect of Pritchard's theory forces him to deny closure. I hence argue that a risk grounded account fairs much better than a modal one, at least in this respect. I will not offer an 'all things considered' argument in favor of my theory. Rather, I hope to get epistemologists interested in risk-centered theories by demonstrating their superiority in this small but important aspect of epistemology. In other words, a theory's ability to explain fake barn examples while also accommodating closure gives us reason to take it seriously.

Part I: No Luck with Closure

1. Kripke's Farm

> One version of Pritchard's safety principle is below.
>
> SAFETY: SP**) S's true belief is safe iff in most near-by possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, and in all very close near-by possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, the belief continues to be true.[3]

The gist of the above is that safe beliefs could not have easily been false, when 'easily' is understood modally. If S's belief that p is safe, there is no close world in which S believes p but p is false.[4]

As mentioned, this safety condition, allegedly, has problems with closure. Let us begin by noting that there are many variations of the closure principle. We can first look at the 'naïve' closure principle, which at first glance seems common sense but upon closer inspection seems implausible:

> *Naïve Closure Principle:* If S knows P, and if P necessitates Q, then S knows Q.

---

[3] Pritchard, "Safety Based Epistemology," *Journal of Philosophical Research* 34 (2009): 34.

[4] As seen in the above quotation, Pritchard makes nuanced distinctions between "close worlds," "very close worlds," and "near-by worlds." I think there are many instances in which such nuance is important. For the purposes of this paper, however, it is unnecessary. For simplicity, we will only distinguish between worlds that are close and those that are not. If a believer would falsely believe p in a 'close' world, his belief p is unsafe.

At first glance this principle appears plausible. We think that if a subject knows some proposition, and that proposition necessitates a second proposition, then he or she will know that second proposition. We can even point to a few cases in which this Naïve Closure Principle seems to hold. Suppose, for instance, that Steve knows that Mary is a women. Steve also knows that being a women implies (necessarily) that one is not a bachelor. It would seem to obviously follow that Steve knows that Mary is not a bachelor.

In spite of some intuitive plausibility, there are various problems with naïve closure. The most obvious is this: one can know P while lacking knowledge, or even awareness of, P's entailments. S might know P and P might imply Q, but if S is unaware that P implies Q, then clearly S will not know Q. It is possible, however, to amend the closure principle to account for this issue. Let us call this modified version "Less Naïve Closure Principle"

> *Less Naïve Closure Principle* If S knows P, and if P necessitates Q, and If S knows that P necessitates Q, then S knows Q

The above closure principle is more plausible than naïve closure, but it still has its problems. The most obvious is this: S can know P, and also know that P implies Q, and also believe Q, but nonetheless believe Q *for the wrong reasons.* Philosophers are capable of dreaming up very strange scenarios. And they may easily dream up one in which, for instance, (1) Sam knows that Paul runs slowly, and, (2) also knows that Paul running slowly implies that Paul does not run quickly. In addition, in this odd world, Sam believes (3) that Paul does not run quickly. He believes this entailment, however*, not because* he has inferred (2) from (1), but rather because he trusts tea leaf readings. In this instance the Less Naïve Closure Principle holds, but because Paul believes for the wrong reasons, it appears he lacks knowledge. Tea leaf readings provide no justificatory grounds. We are once again left with an unsatisfactory closure principle.[5]

In the first chapter of his book, *Epistemic Angst,* Pritchard discusses the problems with the naive versions of the closure principle I just explained (although he does not use my terms). Pritchard argues that the true intuitive aspect of closure is that, "…such principles attempt to codify how one might legitimately extend one's knowledge via competent deduction from what one already knows."[6] Pritchard goes on to formulate this preferred version of the

[5] Pritchard himself has an extensive discussion of the different variations of the closure principle (including the types I am calling the "Naive Closure Principle" and the "Less Naive Closure Principle") in the first chapter of his book, *Epistemic Angst*

[6] Pritchard, *Epistemic Angst: Radical Skepticism and the Groundlessness of Our Believing* (Princeton: Princeton University Press, 2016), 13.

closure principle that accounts for the 'competent deduction' intuition as follows, "If S knows that p, and S competently deduces from p that q, thereby forming a belief that q on this basis while retaining her knowledge that p, then S knows that q."[7]

I agree with Pritchard that the intuitive version of closure is something very close to the above; throughout the rest of the paper when I reference 'closure,' I mean something just along these lines. Or, to use Pritchard's own words, "henceforth when we refer without qualification to the "closure principle" we will have this highly compelling articulation of the principle in mind."[8]

Now that we understand our terminology, let us go back to safety and the potential problems it runs into with closure. Here is one problematic scheme discussed in the literature:

(1) An agent forms a belief about an object and a quality of that object

(2) The agent forms a general belief about that object because it is entailed by (1)

(3) (1) is safe but (2) is unsafe

(4) The agent thereby knows a proposition but not the entailment

The most well-known example comes from Saul Kripke in a criticism of Robert Nozick's sensitivity condition. Kripke alters the traditional fake barn Gettier case along the following lines:

> RED BARN, GREEN BARN: Henry* is driving past a farm with one real green barn and many red fakes. His eyes fall upon the green barn and he believes "That is a green barn." From this he forms a belief in the entailment, "That is a barn."[9]

Although Kripke's case was aimed against sensitivity, as others noticed, it also appears applicable to safety. According to ALVE, Henry* knows 'that is a

---

[7] Pritchard, *Epistemic Angst,* 13. After offering this formulation, Pritchard gives credit in the following footnote, "This is essentially the formulation of the closure principle put forward by Williamson (2000a, 117) and Hawthorne (2005, 29). See also David & Warfield (2008)" (*Epistemic Angst*, 191). In addition to these authors cited by Pritchard, others places readers can find an extensive discussion of closure include Peter Bauman, "Epistemic Closure," in Sven Bernecker and Duncan Pritchard's, *The Routledge Companion to Epistemology* (New York: Routledge, 2011), 597-608 and Sven Bernecker, "Sensitivity, Safety, and Closure," *Acta Analytica* 27 (2012): 367-381.

[8] Pritchard, *Epistemic Angst*, 14.

[9] Kripke's example is a modification of one described by Alvin Goldman in "Discrimination and Perceptual Knowledge," *The Journal of Philosophy* (1976): 771-791. Goldman credits the case to Carl Ginet. See Saul Kripke "Nozick on Knowledge," in *Philosophical Troubles: Collected Papers* (New York: Oxford University Press, 2011), ch. 7.

green barn,' but not 'that is a barn.' He forms his green barn belief via his properly functioning visual abilities, thereby meeting the ability criterion, and because the belief is safe he meets the anti-luck criterion. Regarding the entailment barn belief, however, the ability criterion is meet but *not* the anti-luck criterion. In a close world Henry* falsely believes he is looking at a red barn and so falsely believes the entailment. Hence the belief in the proposition, 'that is a barn,' is unsafe and fails to qualify as knowledge. This is a troubling closure violation. How can you know that there is a green barn but not that there is a barn? Recognizing the unfortunate consequences, Pritchard construes a response to preserve both safety and closure. Contrary to first appearances, he argues, Henry*'s green barn belief is *unsafe*. Here is his reply to a 'Kripke barn' challenge posed by Jonathan Kvanvig.[10]

> The trouble with examples such as this is that it is far from plausible that the agent has knowledge of the antecedent proposition-in this case that this is a green barn-in the first place... it seems that the agent in this example does not have a safe belief in the target proposition, since in an environment where there is barn-deception going on there will be a wide class of nearby possible worlds where, for example, the agent is looking at a green barn facade and yet is nevertheless forming a belief that she is looking at a green barn (it could be, for instance, that this is one of the barn facades that the townsfolk haven't got around to painting red yet).[11]

This reply is puzzling; it is unclear why there *must* be a close world with green fakes. Surely this depends on the details of the example. And even if Kripke's or Kvanvig's particular construction doesn't apply, surely we might imagine a case in which no close world has green fakes. Consider this one:

> RED BARN, GREEN BARN 2: As Henry** drives through fake barn country, he looks at the one real (green) barn, believing, "That is a green barn." Unbeknownst to him, the surrounding barns are fakes, some red, some green. Neither is he aware that this particular fake barn country is managed by Seuss, a demon epistemologist. Bored with the usual Gettier problems, Seuss behaves as follows: If Henry**'s eyes veer toward a green fake, Seuss magically erects a real green barn in front; he does nothing when it comes to red fakes.

The point, of course, is just that of RED BARN, GREEN BARN. Henry**'s green barn belief is safe, his entailment belief unsafe: He knows that there is a green barn but not that there is a barn. The green barn belief is safe, for the demon guarantees that Henry can only view real green barns. However, since the

[10] Jonathan Kvanvig, *The Knowability Paradox* (New York: Oxford University Press, 2006), ch.4.
[11] Pritchard, *Epistemic Luck* (New York: Oxford University Press, 2005), 168.

demon does nothing with red fakes, there is a close world in which Henry** falsely believes a fake red barn is real, and from this he forms a false belief in the entailment. Notice that RED BARN, GREEN BARN 2 leaves little room for Pritchard's already suspect rejoinder. In response to the original RED BARN, GREEN BARN, he argued that there must be a close world with green fakes. It seemed a strange retort, because it seems we can stipulate otherwise. This is displayed in RED BARN, GREEN BARN 2. Thanks to Seuss, in no close world does Henry** have false green barn beliefs. Our demon ensures truth. In Section 4, we see that to insist otherwise, to argue against the stipulation of such demons, is bound to undermine Pritchard's own methodology.

2. Methods Rejoinder

One alternative line of response open to Pritchard involves an appeal to method relativization. Most versions of safety are defined in terms of belief acquisition method: to determine whether a belief is safe, one must consider the method that the agent used to acquire it. Remember that Pritchard has defined his safety condition as follows:

> SAFETY: SP**) S's true belief is safe iff in most near-by possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, and in all very close near-by possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, the belief continues to be true.[12]

Notice that Pritchard stipulates that S's belief is safe if there are no very close worlds in which S forms a false belief in the target proposition, '*in the same way as in the actual world*.' We can think of 'the same way' as referring to a method of belief formation. In order for S's belief that p to be safe, it is *not* necessary that whenever S believes p in very close nearby worlds, she believes truly. Rather, S cannot falsely believe p in a very close world *via the same method.* I can know that my brother is home when I see him sitting on the couch. I can know this even though there is a close world in which I would falsely believe as much. For although I hold a false belief in a close world, it would be acquired via a method that is distinct from the one used in the real world. I might, for example, falsely believe my brother is home via my mother's lying testimony. However, the method by which I would acquire this false belief (testimony) is distinct from the method via which I acquire my true belief in the actual world (vision).[13] Some

[12] Pritchard, "Safety-Based Epistemology," 34.

[13] This example is based off of a similar example provided by Robert Nozick, *Philosophical*

might argue that Pritchard's theory can be saved via method relativization. We just need to suppose that Henry acquires his barn belief via the following method (M1):

> M1: Henry deduces the belief, "I am looking at a barn," from his other belief, "I am looking at a *green* barn."

Given M1, Henry can indeed know he is looking at a barn. Because we stipulated that there are no close worlds with green fakes, there are no close worlds where Henry, *via M1*, falsely believes he is looking at a barn. If Henry indeed has knowledge, there is no longer a problem with closure. Consider: Henry has a safe, true, green barn belief acquired via his own abilities. If Pritchard's ALVE is right, then it follows that Henry also has a safe belief in the entailment which amounts to knowledge. The belief qualifies not only as safe but also as what Pritchard would call a 'safe cognitive success.' In order for the belief to be a 'safe cognitive success,' we must be able to credit the success (true belief) to the agent's cognitive abilities. And in this case we can. First, Henry used his visual abilities to form his belief about the green barn and also his competent deductive abilities to form his true belief in the entailment, "I am looking at a barn." This entailment belief is also safe, because there is no close world in which, *via M1* (deduction from 'green barn belief), Henry holds a false belief in the target proposition ('I am looking at a barn'). Hence Henry *does know* that he is looking at a barn, and there is no longer a problem with closure.

Before evaluating the counter reply at hand, we should note that method relativization has faced criticism that it falls victim to a generality problem. Generality worries amount to the following: If we define a method too broadly, method relativization will not work as desired. (A broadly defined method might not be much better than no method at all). However, if we define a method too narrowly, satisfaction of the safety condition becomes trivial.

Suppose that I form a true belief about the number of grains of sand on a beach. My method, *prima facie*, is 'guessing.' But we might define the method more narrowly as: 'guessing while walking a Dalmatian on a cold winter day in December approximately 7 minutes after 5pm.' This method characterization makes my true sand belief trivially safe. It is safe because I will never have the relevant false belief via that uniquely defined method. Yet it is trivially safe because the method is so obscure that there is no close world in which I would ever use it again.

A similar triviality objection is applicable to Henry if we define his method

---

Explanations (Cambridge: Harvard University Press, 1981).

of belief formation as "deriving the belief that one is looking at a barn from the belief that one is looking at a green barn." The belief is safe, but only trivially so. Henry will never form a false belief via the method in question, because by stipulation, there is no close world with green barns and so no close world in which he could possibly use the same method. Even more, if we get into the habit of relativizing methods this narrowly, it is bound to set safety up for many accusations of triviality in similar cases.

Generality worries, although problematic, are not the biggest problem with the method relativization response. In the next section, we will discuss Pritchard's move to ALVE, which is at least in part a response to safety and triviality worries, and might have some potential to help Pritchard out with the generality problem. The biggest problem for Pritchard is not method generality, but that method relativization will get Pritchard out of the closure dilemma only by leaving him with a completely different problem. Because of method relativization, we are supposed to admit that Henry *does* know that he is looking at a barn. Yet, intuitively, Henry does *not* know this. Recall that Pritchard's initial response to the Kripke barn challenge was to argue that Henry lacked knowledge both about the green barn and the barn itself. Pritchard argued as much because he wanted to show that ALVE aligns with the widespread intuition that Henry lacks knowledge about the barn. The safety condition is supposed to be a preferred epistemic criterion specifically because it gives us the intuitive result in 'Fake barn Gettier cases' (i.e. the result that Henry does not know that he is looking at a barn).

To conclude this section, method relativization cannot save Pritchard from the dilemma at hand. It cannot do this for at least two reasons. First, we must describe Henry's method of belief formation in an unnaturally narrow way if it is to be of any help with the closure dilemma. This excessive narrowness risks subjecting safety to further generality accusations. And even more importantly, method relativization will only get Pritchard out of the dilemma at the cost of giving us the wrong result in fake barn Gettier cases. If one admits that Henry knows he is looking at a barn, one simply bites the bullet on the Gettier case. A purported advantage of ALVE, however, was that it *does not* bite the bullet in this way.

3. Safety & Ability

When Pritchard argued for his change from a mere anti-luck theory of knowledge to an anti-luck *virtue* epistemology, he motivated the switch with the following case:

> TEMP: Temp forms his beliefs about the temperature in the room by consulting a

> thermometer. His beliefs, so formed, are highly reliable, in that any belief he forms on this basis will always be correct. Moreover, he has no reason for thinking that there is anything amiss with his thermometer. But the thermometer is in fact broken, and is fluctuating randomly within a given range. Unbeknownst to Temp, there is an agent hidden in the room who is in control of the thermostat whose job it is to ensure that every time Temp consults the thermometer the "reading" on the thermometer corresponds to the temperature in the room.[14]

Pritchard claims that Temp lacks knowledge, even though his belief is safe. The missing ingredient is ability, "[W]hat is wrong with Temp's beliefs is that… their correctness has nothing to do with Temp's abilities and everything to do with some feature external to his cognitive agency."[15] Because what explains Temp's success is not ability but the hidden agent, we are disinclined to attribute knowledge. Safety is too weak on its own and must be supplemented with an ability criterion.[16]

Notice that for the Temp example to work, the hidden helper must *guarantee* Temp's safe beliefs. In Pritchard's words, "[W]hatever one wishes to say about what is epistemically deficient in Temp's beliefs, it does not seem that his beliefs fail to satisfy the anti-luck intuition. After all, his beliefs are *guaranteed* to be true…"[17] (my emphasis). Let us review the structure of the Temp case:

> Temp has true temperature beliefs
>
> The beliefs are safe because a hidden agent ensures their truth
>
> Intuitively Temp lacks knowledge
>
> Hence safe belief is insufficient for knowledge
>
> Hence we must amend our theory of knowledge (with an ability condition)

The Temp case is critical for Pritchard's move from an anti-luck theory of knowledge to a theory that incorporates a virtue component. It also guarantees problems in RED BARN, GREEN BARN 2. The Temp case needs a hidden helper to ensure safe beliefs. Again, in Pritchard's words, "…Temp's belief *satisfies the safety principle*. This is ensured by the fact that the manner in which Temp is forming his beliefs, such that success is guaranteed, means that *it can hardly be the*

---

[14] Pritchard, "Anti-Luck," 260.

[15] Pritchard, "Anti-Luck," 260.

[16] "Anti-Luck Virtue Epistemology (ALVE): S knows that p if and only if S 's safe true belief that p is the product of her relevant cognitive abilities (such that her safe cognitive success is to a significant degree creditable to her cognitive agency)" (Pritchard, "Anti-Luck," 260).

[17] Pritchard, "Anti-Luck," 261.

case that he could easily have formed a false belief"[18](my emphasis). If Pritchard can stipulate a helper that ensures Temp's safe temperature beliefs, others can stipulate demons who ensures that fellows named Henry have safe green barn beliefs. Pritchard cannot, then, dismiss our objection on the grounds that "…there will be a wide class of nearby possible worlds where… (Henry**) is looking at a green barn facade and yet…forming a belief that (he) is looking at a green barn."[19] Thanks to Seuss, in RED BARN, GREEEN BARN 2, there is no such close world. Henry**'s belief is safe according to Pritchard's own standards.

At this point some are probably thinking, 'Wait a minute, doesn't TEMP *vindicate* Pritchard'? Seuss, just like Temp, ensures safe belief. The demon's help, some might argue, disqualifies Henry** from meeting the ability criterion and hence the complete demands of knowledge. And if Henry** thereby lacks knowledge of the green barn, it is irrelevant that he also lacks knowledge of the entailment. Here are the problems with that claim. In the original fake barn case, Pritchard himself argues that Henry's safe cognitive success is indeed explained by his own abilities. In Pritchard's words, "…given that (Henry) does undertake, using his cognitive abilities, a genuine perception of the barn, it seems that his cognitive success is explained by his cognitive abilities…"[20] (original emphasis). Notice that just like Henry from the original fake barn case, Henry** "undertakes, using his cognitive abilities, a genuine perception of the barn." There is, then, no grounds to say that Henry does not meet the standards of ALVE which demand a 'safe cognitive success.' Henry's belief, after all, is safe. And, as Pritchard himself admits, Henry uses his cognitive abilities in acquiring this safe belief. Moreover, Henry and Henry** exercise the very same abilities. We must then conclude that Henry**'s barn belief qualifies as a safe cognitive success (just like Henry) thereby meeting the standards of Pritchard's very own ALVE.

Given Pritchard's decision to prescribe to a *weak* virtue theory, it is especially difficult for him to reply to the above criticism. Describing ALVE, he argues, "…the ability condition in play here is that proposed by a weak virtue epistemology rather than a strong virtue epistemology…the agent's *safe* cognitive success should be to a significant degree creditable to her cognitive agency"[21] (original emphasis). Pritchard rejects that safe belief must be *because of* or *primarily* creditable to the believer's cognitive ability. What matters is that the agent's abilities are involved to 'a significant degree.' Henry** passes the bar in this

[18] Pritchard, "Anti-Luck," 259-260.
[19] Pritchard, "Anti-Luck," 260.
[20] Pritchard, "Anti-Luck," 272-273.
[21] Pritchard, "Anti-Luck," 274.

regard; his reliable vision is indispensably involved in acquiring his safe belief. We might even stipulate that Henry** has unusually excellent visual abilities, thus highlighting his contribution even further. If Henry of the original fake barn case utilizes his abilities in a way that satisfies Pritchard's self-described standards (i.e. so that 'his cognitive success is explained by his cognitive abilities'), then so must Henry**.

What if Pritchard contends that the demon's help diminishes Henry**'s import to an insignificant level? I find this unconvincing, given that (1) 'significant' involvement of the agent's cognitive abilities is enough to secure the virtue component, and (2) as mentioned, we can stipulate that Henry** had especially exquisite vision. Exquisite vision, it seems, should qualify as 'significant' involvement in the acquisition of a safe belief. Notwithstanding, although I am not convinced of Pritchard's imagined retort, there is indeed a way to counter it. Let us consider one last barn example.

> RED BARN, GREEN BARN 3 A demon named Henry*** is driving through fake barn country, and like the other Henrys, believes he is in regular barn country. Henry*** views the one real (green) barn surrounded by red fakes, forming the belief, "That is a green barn." Connecting the logical dots, he then believes the entailment, "That is a barn." Henry*** is especially confident in his green barn belief, because many years ago he cast a special spell. (Demons can cast one personally enhancing spell as soon as they are of age to exercise demonic powers.) The spell was as follows: *Every belief I form about green objects will qualify as safe, according to Duncan Pritchard's safety condition.*

We can see that Henry***'s green barn belief is safe; the spell guarantees as much. Casting the spell himself, moreover, allows him to fulfill the ability criterion. (Let us imagine that spell casting requires advanced analytic ability. This is why demons cannot cast spells upon birth, but must wait until their cognitive capacities are more fully developed.) Henry*** meets all the requirements of ALVE and so his green barn belief qualifies as knowledge. His entailment belief, however, is unsafe and thereby not knowledge. The following criticism stands: According to ALVE, Henry*** knows there is a green barn but not that there is a barn. This, of course, is a glaring closure violation. And there are likely to be other cases with the same structure.

4. Hard Choices

While it may be possible to argue something stronger, clearly we must abandon one of the following:

(1) Safety

(2) Closure

(3) The stipulation of creatures that guarantee safety.

We see that (3) from above allows us to modify Kripke's original criticism against sensitivity as to make it equally forceful against safety. Pritchard cannot have it all. Given the foundational import to his theory, rejecting safety itself is unlikely. Abandoning closure is an option, one that others have taken before. But closure's compelling intuitive force makes this less than ideal. Pritchard, moreover, argues that an *advantage* of safety is its' compatibility with closure.[22] Rejecting (3), then, may seem the most palatable. But this too comes with undesirable consequences. If Pritchard loses (3), he loses the Temp case. If he loses the Temp case, he loses what grounds his argument for an ability criterion. There are also problems apart from ALVE. Without justification, a prohibition on safety ensuring creatures seems arbitrary. On the other hand, arguments against such magic may cover too much. If, for example, demons cannot ensure safe belief, what else? Epistemologists who limit demonic powers run the risk of biting the hand that feeds them. The next thing we know epistemic villains are unable to deceive the senses. ALVE then runs into trouble not due to any specific flaw, but because by first putting demons out of business, and thereby the skeptic, epistemologists may unintentionally do the same to themselves.

PART II: A Non-Modal Solution

This second-half of the paper is devoted to showing how a non-modal theory of knowledge avoids the problems with closure that troubles ALVE. Indeed, I think that most modal theories will run into the same problems as ALVE. The problem is it seems ever possible to design an entitlement counterexample that simply does not mesh with modal accounts of knowledge. If I can know a proposition about an object O and a quality of that object Q, then it seems intuitively plausible (in the usual cases) that I can derive a belief in the entailment that is propositional knowledge about O itself. However, as long as the theory is a modal one, the skeptic is there awaiting with a counterexample to deny the epistemic agent his entailment belief. The scope of possible worlds, even close ones, is quite expansive, especially when we include demon worlds. Hence there will be cases in which an agent cannot fulfill the modal criterion for the entailment belief (even though she can fulfill the modal criteria for the entailing belief).

[22] Pritchard, *Epistemic Luck*, 94. See also, Pritchard "Sensitivity, Safety, and Anti-luck Epistemology," in John Greco's *The Oxford Handbook of Skepticism* (Oxford: Oxford University Press, 2008), 447.

My solution to the problem of skeptics and closure, explained in what follows, is to bypass the modal criterion all together. Without a modal requirement, the skeptic's challenges will prove irrelevant. Closure comes out alive (that is, it comes out alive in the most important world, the actual one). In my account a *risk* criterion will in some sense replace the modal criterion. Section 2.1 will detail my risk criterion. Section 2.2 will explain how the risk criterion addresses fake barn Gettier cases. Section 2.3 addresses potential confusions. Lastly, Section 2.4 explains specific advantages of a risk grounded theory compared to modal ones.

Before moving on, I want to clarify a potential confusion. I am about to argue for a 'risk-centered' approach I call, 'risk sensitive credit. I should note that Pritchard's most recent work now describes his theory as 'anti-risk' rather than 'anti-luck.' The risk that Pritchard's theory is now committed to, however, is not the type of risk that is relevant for my own 'risk-sensitive credit.' I agree with Pritchard that "risk assessments seem to be essentially forwards-looking…Luck assessments, in contrast, seem to be essentially backwards-looking…" (Pritchard: forthcoming). It is true that whether we are understanding risk in my probabilistic sense or in Pritchard's modal one, risk assessments are forward looking. According to my own theory, for instance, an agent looks forward toward the probability that her future belief will or will not hold true. This agreement on the forward looking characteristic of risk, however, is where the similarities between my own view of risk and that of Pritchard's grinds to a halt.

Pritchard makes clear that the type of risk which he is concerned with is a modal account. My account of risk is explicitly *not* a modal one. I am understanding risk, rather, as it is being used by cognitive scientists, i.e., as a Bayesian type of probabilistic risk. This is also the type of risk, for instance, that David Henderson and Terry Horgan defend in their own risk-centered theory. This type of risk, rather than concern itself with modal possibilities, is concerned with 'chance' understood in terms of Bayesian probabilities. If believing p is epistemically risky (in my sense), it follows that there is a high probability that 'p' is not true. If belief in p is not risky, then there is a high probability that S's belief in p is accurate. When a theory of knowledge is founded in this type of probabilistic risk-sensitivity, we will see that Henry is able to completely evade the closure related problems that pop-up with fake-barn Gettier cases.[23]

[23] Pritchard, "Epistemic Risk," *The Journal of Philosophy* 113 (2016): 550-571, admits that modifying his account of knowledge from an "anti-luck" account to an "anti-risk" account may seem like a minor change, stating that, "the differences between the two views will not be radical" He nonetheless defends the switch by arguing that an anti-risk account has at least two

2.1. Risk Sensitive Credit

For any proposition p, an agent might believe p, believe not-p, or withhold belief. Believing, however, can be epistemically risky, and at times the risk of false belief is not be worth the potential reward. Along these lines, I will argue that in order for a belief to qualify as knowledge, it must *be risk sensitive*. S's belief is what I call risk sensitive only if the likelihood of false belief is low enough that belief (as opposed to disbelief or withholding) is the best epistemic option. Ernest Sosa, for instance, has discussed ideas along these lines, arguing that "[One's] meta-competence governs whether or not one should form a belief at all on the question at issue, or should rather withhold belief altogether."[24] Elsewhere Sosa argues that "A performance can thus easily fail to be 'meta- apt,' because the agent handles risk poorly, either by taking too much or by taking too little. The agent may fail to perceive the risk, when he should be more perceptive; or he may respond to the perceived risk with either foolhardiness or cowardice…"[25]

What I call risk sensitivity is similar to Sosa's meta-aptness. S's belief lacks risk sensitivity if she takes too much risk or too little. What exactly I mean by 'risk assessment' and 'risk sensitivity' is obviously important. I am *not* thinking about risk assessment as a highly reflective cognitive process. Under my account, risk assessment need not include reflection, higher order beliefs, or even the possibility of either. (If 'assessment' sounds too reflective, you may prefer to think of risk 'accommodation')[26] To explain further, we can helpfully turn to the work of David Henderson and Terry Horgan:

> We ourselves find very plausible the idea that competent risk assessment, as an aspect of the process of forming a belief, is required in order for that belief to

---

significant advantages. First, he argues that between risk and luck, it is the former that has a better claim to the status of what we might call the more 'fundamental' concept. He argues that, "…we naturally explain a concern to eliminate (luck) in terms of a concern to eliminate (risk) rather than vice versa…" Lastly, although Pritchard himself if not working with my own probabilistic conception of risk, he acknowledges that "most contemporary treatments of risk" utilize "a probabilistic conception."

[24] Ernest Sosa, "Knowing Full Well: the Normativity of Beliefs as Performance," *Philosophical Studies* 142 (2009): 14.

[25] Sosa, "Knowing Full Well," 12.

[26] In Sosa's own words, "We can now see that knowing something full well requires that one have animal and reflective knowledge of it, but also that one know it with full aptness. It requires, that is to say, that the correctness of one's first- order belief manifest not only the animal, first-order competences that reliably enough yield the correctness of the beliefs that they produce. One's first-order belief falls short if it is not appropriately *guided* by one's relevant meta-competence" Sosa, "Knowing Full Well," 16.

> constitute fully human knowledge. But we doubt whether such competence needs to take the form of a higher-order belief; and we also doubt whether a first-order belief can qualify as any kind of knowledge if it is formed in a way that *utterly lack*s the aspect of competent risk assessment (original emphasis).[27]

The risk sensitivity I advocate aligns with Henderson and Horgan on both counts: S cannot know p unless S (or S's abilities or S's cognitive system) assessed (or accommodated) p's risk, but this can take place without higher order belief. Moreover, risk assessment is necessary for knowledge of any kind. H&H further suggest that, "[We] might have a trained capacity that manages to accommodate [risk] without articulation, automatically and quickly…"[28] I agree, but would add that we might also have innate cognitive capacities that *evolved* to accommodate risk. I suspect that H&H were thinking of 'trained' loosely, and this was what they meant. In any case, visual studies confirm that automated cognitive processes can classify sensory data according to a risk sensitive framework. Consider the following commentary on a recent study,

> …Bayesian concepts are transforming perception research by providing a rigorous mathematical framework for representing the physical and statistical properties of the environment… describing the tasks that perceptual systems are trying to perform, and deriving appropriate computational theories of how to perform those tasks, given the properties of the environment and the costs and benefits associated with different perceptual decisions.[29]

The above suggests that perception works within a cost benefit framework that balances the benefits of perceptual belief versus the risks. Further studies provide evidence that we update statistical frameworks according to perceived environment. In short, there is much more to perception than sensory data. To ensure accuracy, our perceptual system first receives sensory information, and then second and separately, accommodates this data in accordance with the environment and other circumstantial contingencies. Environmental awareness, combined with sensory input, leads to risk assessment. This again is supported with research in cognitive science:

> [T]he objects that are likely to occur in a scene can be predicted probabilistically from natural scene categories that are encoded in human brain activity. This

---

[27] David Henderson and Terry Horgan, "Risk Sensitive Animal Knowledge," *Philosophical Studies* 166 (2013): 601. This quote was aimed at Sosa. Sosa since responded to the criticism and argues that his own account does not demand as much reflection as H&H may have assumed.

[28] Henderson and Horgan, "Risk Sensitive," 603.

[29] Wilson Geisler and Daniel Kersten, "Illusions, Perception and Bayes," *Nature Neuroscience* 5 (2002): 508.

> suggests that humans might use a probabilistic strategy to help infer the likely objects in a scene from fragmentary information available at any point in time.[30]

Our perceptual system matches visual sensations to familiar objects given other information about the environment and contextual circumstance. Suppose you experience a visual stimulus of a small furry animal. If you believe you are in the forest, this stimuli might indicate a squirrel. Contrastingly, if you were at home, your unconscious cognitive processes might suggest that the animal is a cat. In order to acquire perceptual knowledge, your sensory data must first accurately reflect the perceptual object. In other words, your vision is not blurry, you are an appropriate distance from the object, and you are not under the influence of hallucinogens. If this holds, you have data to make a probability assessment in accordance with the environment and other relevant conditions. Back to our visual studies:

> [A]n ideal observer convolves the posterior distribution with a utility function (or loss function), which specifies the costs and benefits associated with the different possible errors in the perceptual decision. The result of this operation is the expected utility (or Bayes' risk) associated with each possible interpretation of the stimulus. Finally, the ideal observer picks the interpretation that has the maximum expected utility.[31]

The above quote nicely explains how sensory input prompts the following evaluation: What are the chances that this stimulus comes from object O given environment E and circumstances C? The answer determines whether it is best to believe p, withhold belief, or believe not-p. Assume that a true belief is an epistemic benefit and a false belief a cost. Ideal agents, we might say, believe p only if belief has the highest expected epistemic value. I do not think, however, that in order to acquire perceptual knowledge one needs to be an 'ideal observer.' Indeed, in order to acquire any type of knowledge one need not be epistemically 'ideal' in any sense at all. Yet I do want to argue that knowledge demands a type of 'creditworthiness.' Hence the account I am arguing for falls under the umbrella of 'credit theories' of knowledge.

While there is all sorts of disagreements between credit theorists, most agree that an agent acquires knowledge when she forms her belief through a process which is 'epistemically creditworthy.' Credit theorists further argue that their accounts are especially well fit to explain the value of knowledge. For while both true belief and knowledge are in some sense epistemically desirable, the

[30] Dustin Stansbury, Thomas Naselaris, and Jack Gallant "Natural scene statistics account for the representation of scene categories in human visual cortex," *Neuron* 79 (2013): 1031.
[31] Geisler and Kersten, "Illusions, Perception," 508.

latter is preferable for it is an 'achievement' or an act of 'creditworthiness.' It is these creditworthy beliefs alone that count as 'knowledge.' The crux of the issue, of course, is just what belief forming mechanisms count as creditworthy. I am arguing that beliefs formed through a process that assesses epistemic risk are those special sort of beliefs that we might deem creditworthy. The creditworthy agent believes p only if believing presents minimal epistemic risk. We can call this Risk Sensitive Credit (RSC). More formally,

> RSC: An agent's belief p is risk sensitive and hence creditworthy if (1) her own abilities assess belief risk, and (2) she correctly believes p because (1) indicates a reasonably low chance of p's falsity.

Some might object to the vagueness of 'reasonably low.' It is used for two reasons. First, it seems a fruitless effort to determine whether the risk of falsehood must be below 15, 10, or 5 percent. Second, philosophers who disagree about justificatory *degree* might still agree on justificatory *kind*. But if we agree that risk sensitive belief is belief in accordance with reasonable risk assessment. What is risk assessment? Briefly, it is a means of analyzing and interpreting relevant data within an environment and set of conditions. Assessment goes about as follows: an agent's cognitive system, consciously or unconsciously, assesses the chances of p given what I call her *total information*. Total information consists of certain epistemic data D and epistemically relevant conditions C. That is, P(P/D&C). Risk assessment can go awry in at least three ways:

> Risk Assessment Errors
>
> (1) Inaccurate data
>
> (2) Inaccuracy regarding the conditions
>
> (3) Misinterpreting the meaning of the data given the conditions

Imagine a risk management company, SECURE, that is hired to assess the safety of a mansion hosting a prestigious fundraiser. SECURE might blunder through inaccurate data gathering, inaccurate conditional assessment, or misinterpretation of the data given the conditions. Examples of the first could include miscounting the fire alarms or misreading the thermostat. Either error would skew total assessment. But maybe there is no data inaccuracy. Problems ensure, however, because there is failure to consider a tornado warning. (A failure of conditional assessment).

A third possibility is that SECURE makes no error in data collection nor conditional assessment, yet still goes wrong in interpretation. They might judge that 7 fire alarms is appropriate when 15 are needed. To do their job, SECURE

must collect good data, carefully apprise conditions, and then use both of the aforementioned to arrive at an all things considered risk assessment. Note that a safe event is not enough to fend off criticism. SECURE'S customers can demand a refund upon discovering the event unknowingly presented a high safety risk, even if no risk actualized. Each of us, when making an epistemic risk assessment, functions in a manner similar to SECURE. In other words, we attempt to make an accurate risk assessment (that is, an assessment of the chance of p's truth) given all relevant information. Things can go wrong when we either misinterpret the meaning of information or receive misinformation from the start. In the next section, I will explain the sad story of a misfired risk assessment by an innocent fellow named Henry deep within the land of barn facades.

2.2 Resolving The Fake Barn Dilemma

With the risk sensitive framework just described, we can now explain what goes wrong when Henry views the one real red barn. Although Henry has a true belief, he does not have a risk sensitive belief. Because risk sensitivity is required for knowledge, Henry's true belief fails to qualify.

Let us describe the process that Henry engages in in more detail. We can then see exactly where things go wrong. Henry, through his visual experience of the barn, receives data in need of epistemic analysis. Shortly after receiving this data his perceptual system gauges epistemic risk. Henry, however, assumes he is in a traditional barn environment; this skews assessment. We should think of epistemic evaluation in terms of 'total risk assessment.' In other words, creditworthy epistemic endeavors demand the proper processing of all relevant epistemic information. An agent might receive various information from many sources and over a long time period. Some of this information might be consciously accessible, while other information is not. An agent deserves credit (and so acquires knowledge) when she first accurately processes this data, second comes to the (correct) conclusion that not-p is improbable and therefore truly believes p. With this in mind we can recognize what goes wrong in fake barn country: Henry misinterprets a critical portion of epistemic information when he misjudges his environment 'traditional barn country.' This misinterpretation is critical to his misfired risk assessment and the ensuing failure to obtain knowledge.

Prima facie, we might be tempted to think that Henry's belief forming mechanism is nothing more than visual perception, and this would lead us to conclude he forms his belief via epistemically acceptable means. But things are not so simple. For instance, in challenging Fred Dretske's argument against closure,

Pritchard himself has pointed out that beliefs ostensibly formed, 'just by looking,' are in reality much more complex. Suppose, for instance, that Zula looks at a zebra and forms the true belief that what she sees is a zebra. It may be tempting to say she forms her belief, 'just by looking.' But as Pritchard explains, this isn't quite right.

> I think that while there is a sense in which it is obviously true that Zula gains her knowledge just by looking…perceptual knowledge can…involve a wide range of specialist expertise and background knowledge…such expertise and background knowledge would surely have ramifications for the total evidence that you possess in support of your belief… to know a proposition just by looking need not entail that the only evidence you possess for your belief is the evidence you gained from the bare visual scene before you.[32]

Like Zula, Henry's 'evidence' (what I prefer to call 'information') consists in much more than just the bare visual scene before him. Background knowledge plays an important role; only from past experience does Henry know his percept has the appearance of an object called a 'barn,' and that open grassy areas are the types of places where barns are commonly found. Yet unfortunately for Henry, some of his background information misleads. If we assume Henry an ordinary fellow, he hasn't any reason to think that objects that appear like barns are actually barn facades. As far as he knows, it would be pointless to have a town full of barn facades, he has never heard of such things, and he would be prone to suspect (quite reasonably) that those who believe in fake barn country are conspiratorial loons. While these are all reasonable assumptions on his part, they have distorting consequences on his epistemic evaluation.

Total risk assessment is derived from various sources of epistemic information which are first individually interpreted and then collectively assessed. Going too far off the mark when interpreting information will corrupt the collective assessment. This is what happens with Henry. He misinterpreted his environment and unfortunately for him, this misinterpretation played a key role in his total risk assessment. Epistemic creditworthiness does not allow for these types of mistakes. In line with previous credit theorists emphasis on 'credit for *success*,' an understandable epistemic mistake is still a mistake. The idea is similar to the common externalist/reliabilist notion that justification goes beyond that which is internal to the believer. Even if an agent has good reason to think her method is reliable, she cannot be justified if it is unreliable. Similarly, even if we can understand why Henry made the risk assessment that he did (that the object

[32] Pritchard, "Relevant Alternatives, Perceptual Knowledge and Discrimination," *Nous* 44 (2010): 256-257.

before him was likely a barn), it was inaccurate (the object before him was *not* likely to be a barn) and therefore not creditworthy.

2.3 Clarifications

Let me make clear that RSC is not a variant of the so called 'no false lemmas' theory. As some may recall, shortly after Gettier introduced his Problem, a view often referred to as the 'no false lemmas' approach (NFL) suggested a simple solution.[33] According to NFL, Gettier's examples of troublesome beliefs are, in actuality, illegitimate (or unjustified) because they rely on false premises: Smith's true belief that 'the man who will get the job has 10 coins in his pocket' is acquired by reasoning through the false premise that 'Jones will get the job.' Similarly, Smith's true belief that "Either Jones owns a ford or Brown is in Barcelona,' is acquired only via reasoning through the false premise that 'Brown is in Barcelona.' NFL proponents argued that a necessary condition of knowledge was that the 'belief' in 'justified true belief' could not be acquired by reasoning through false premises. With this requirement, we see that the heroes of Gettier's puzzles rely on false premises and this therefore prevents them from acquiring knowledge.

Many problems with NFL soon came to light. First, with some imaginative effort, it is possible to come up with examples similar to those in Gettier's original paper that *do not* rely on false premises.[34] And second, a new breed of Gettier cases, those of the fake barn variety, were introduced onto the epistemological stage.[35] It seemed to many that simple visual beliefs (like the barn façade belief) do not rely on any premises at all, and hence even more so do not rely on false premises.

Because I emphasize the role false information plays in skewing risk assessment, some might confuse RSC with NFL. I want to be clear that RSC is entirely distinct from, and bears very little relation to any variant of the no false premise approach and *does not* suggest that NFL is necessary for knowledge. Let us return to Henry. I argued that his true barn belief, which might appear to arise

---

[33] See D.M. Armstrong, *Belief Truth and Knowledge* (Cambridge: Cambridge University Press, 1974).

[34] See Keith Lehrer, "Knowledge, Truth and Evidence," *Analysis* 25 (1965):170, and Richard Feldman, "An Alleged Defect in Gettier Counter-Examples," *Australasian Journal of Philosophy* 52 (1974): 68-69. For a challenge to Lehrer and Feldman, See Michael Levine, "Gettier Cases without False Lemmas?" *Erkenntnis* 64 (2006): 381-392.

[35] See Alvin Goldman, "Discrimination and Perceptual Knowledge," *The Journal of Philosophy* (1976): 771-791.

spontaneously, is actually dependent on a vast array of background information, much of which is really misinformation. Such misinformation plays a critical role in tipping Henry's risk assessment scales in the wrong way. However, we should not understand Henry's risk assessment failure in terms of false premises. First off, this would make knowledge requirements unreasonably strict. After all, much everyday knowledge is partly based on false or misleading background information.

Not only is false background information compatible with knowledge, it is unclear that background information necessarily consists of beliefs (beliefs to potentially serve the role of a false lemma). Our cognitive system can register information that never makes its way into the realm of explicit beliefs, and might not even rise to the level of implicit belief. But background information contributes to assessment of epistemic risk nonetheless. It is this *failure to accurately assess epistemic risk* which accounts for Henry's failure to obtain knowledge. Of course, there are many cases in which misleading background information (which may or may not consist of false beliefs) is not enough to prevent a reasonable assessment of epistemic risk. In such cases, one might have knowledge partly based on inaccurate information. However, in other instances, (like with Henry) inaccurate information *does* interfere with a reasonable risk assessment, and thus does prevent one from attaining knowledge.

Let us return to our analogy of the risk assessment company. Imagine that SECURE concludes that there is minimal safety risk at the mansion, but only because the company is unaware of the man-eating grizzly bears who reside in the courtyard. Clearly any valuation made without awareness of this environmental feature will interfere with a successful assessment. Similarly, Henry's ignorance of fake barn country prevents him from accurately assessing the riskiness of his situation.

2.4 Risk Sensitivity & Fake Barns

We can now see the benefits of a risk grounded theory as opposed to a modal one, at least in respect to fake barn Gettier cases. If one is committed to a modal theory and also wants to preserve intuitions in fake barn examples, Henry's lack of knowledge must be explained in terms of false beliefs in close worlds. Indeed, we all admit that Henry has a true barn belief formed through his reliable vision. The challenge is to explain *why* this seemingly true justified belief does not qualify as knowledge. Modal theories must turn to close worlds for an explanation. Pritchard in particular argues that knowledge demands 'safety'. A belief is unsafe if false in a nearby world. Henry, in turn, lacks knowledge because in a close world he falsely

believes that a fake barn is real. But this is exactly where Pritchard leaves the door open to demons and skeptics. These challenging critics will quickly create a world in which an entailing belief is safe but the entailed belief is unsafe. And this is when Pritchard (and some other modal theorists) must come face to face with closure denial.

My risk sensitive theory avoids the problems just explained above, because it never gives the skeptic cause to dream about strange and troublesome close worlds. Rather, a risk sensitive theory explains that Henry lacks knowledge without any appeal to modal conditions. All of Henry's epistemic failures can be explained by reference to Henry's epistemic practices in the actual world. In order to acquire knowledge in the actual world, an agent must gather information and make a reasonably accurate epistemic risk assessment. Henry gathers information and makes a risk assessment. Unfortunately for Henry, it is *not* a reasonably accurate assessment.

Risk sensitivity demands reasonable accuracy regarding data, environment, and other relevant conditions. Mistakes about any of these can result in an assessment that either (1) misrepresents epistemic risk, or (2) makes an accurate assessment but only by luck. Both (1) and (2) are incompatible with creditworthiness and thereby knowledge. In the former case inaccuracy is the problem; in the latter accuracy is powerless because it does not derive from the agent's abilities. Henry's problem is with (1). His mistaken environmental assumption that he is in traditional barn country give rise to an inaccurate assessment and he gravely misrepresents epistemic risk. Because accurate risk assessment is required for knowledge, Henry lacks knowledge both about the barn and its color.

Conclusion

At first glance, Kripke's 'green barn challenge' to Nozickian sensitivity appears applicable to safety. Pritchard replied to the challenge, arguing that safety and closure get along just fine. This paper argued that his response works only for certain constructions of the green barn challenge; we have seen that an alternative version leaves no room for Pritchard's counterargument. Thanks to the help of demons, a subject can have a safe green barn belief while her belief in the entailment remains unsafe. Pritchard opened the door to this possibility in his argument for ALVE, which stipulated a safety ensuring hidden helper. Either safety and closure are incompatible, or Pritchard's argument for ALVE falls flat. If it was up to me, I would choose the former. I argued that we can keep both safety and closure if we replace ALVE with an alternative epistemic theory that is

grounded in risk assessment rather than modality. My alternative theory, "Risk Sensitive Credit,' preserves safety and closure while also explaining Henry's lack of knowledge in fake barn Gettier cases.