Synthese (2025) 206:296
https://doi.org/10.1007/s11229-025-05395-8

ORIGINAL RESEARCH

Sturdy beliefs

Jack Madock[1]

Received: 14 July 2025 / Accepted: 26 November 2025

Abstract
This paper explores and defends sturdiness, that p would be false were one not to believe it (~Bp→~p), as a plausible necessary condition on knowledge. Sections one and two review previously offered modal accounts of knowledge, namely sensitivity and safety. Section three outlines sturdiness in full and section four tests sturdiness against Gettier-style cases that gave trouble to other modal accounts. Sturdiness is then defended against objections from closure and knowledge of logical truths. Finally, the paper considers whether sturdiness may deliver on intuition in skeptical scenarios. The arguments here show that, given a certain reading, sturdiness is a plausible necessary condition on knowledge and motivates further exploration of the view.

Keywords Modal epistemology · Safety · Sensitivity · Knowledge · Counterfactuals

1 Introduction

Many of the well-received necessary conditions on knowledge come in the form of counterfactual conditionals. One such example is Robert Nozick's sensitivity condition:

Sensitivity: S knows p only if S's true belief in p is sensitive.

-If p were false, s would not have believed p.

✉ Jack Madock
johnmadock@ufl.edu

[1] Department of Philosophy, University of Florida, 330 Griffin-Floyd Hall, Gainesville, FL 32611-8545, USA

Published online: 08 December 2025

Sensitivity, however, took some heavy criticism. Philosophers discovered that sensitivity had its own host of issues that may or may not outweigh its value as a theory of knowledge. A more recent and popular proposal is safety, also defined in counterfactual terms.

Safety: S knows p only if S's true belief in p is safe.

-If S were to believe p, then p would be true.

Recent years has seen lively debate over which (if any) of these options has greater merit. The majority of these debates focus on whether safety is superior to sensitivity or to any number of non-modal approaches. Additionally, safety has had to be defended from various criticisms itself.

But are these the only two forms that a counterfactual theory of knowledge might take? Taking at a closer look at sensitivity reveals that they are not. One view that was defended by Nozick but has received scant attention is adherence.

Adherence: S knows p only if S's true belief in p is adherent.[1]

-If p were true, then S would believe p.

Rightly or wrongly, adherence has received nowhere near the attention of sensitivity or safety.

The careful reader may notice that there is yet a fourth counterfactual option available in the relationship between some proposition (p), and some subject's belief in that proposition (Bp). Another candidate counterfactual is: *if S were not to believe p, then p would be false.* Surprisingly, this option has not been widely discussed in the debate over candidate theories of knowledge.[2] For want of a label let us call beliefs which satisfy this condition "sturdy." As a theory of knowledge, we can define sturdiness:

Sturdiness: S knows p only if S's true belief in p is sturdy.

-If S were not to believe p, then p would be false.

Sturdiness is undeniably a genuine property of some beliefs. The topic of the remainder of this paper is: Is sturdiness a plausible condition on knowledge? The paper argues that that, while sturdiness may initially look implausible, it has significant

[1] Nozick's proposal included variance and adherence as two necessary conditions. I am here referring a sensitivity as condition three of Nozick following Pritchard (2012). Some refer to sensitivity as the variance condition and adherence as the adherence condition (Becker, 2018).

[2] Strictly speaking, this conditional was proposed as part of a different theory of knowledge put forth by Carrier (1971). Carrier offered it in addition to Nozick's sensitivity condition (3) and as an analysis of when we can correctly say of someone that he knows some proposition. He further thinks of this a material conditional whereas I am interpreting this as a counterfactual conditional. There is also a brief discussion in Shope (1983, 119).

advantages over the more familiar counterfactual conditions such as sensitivity and safety. If this proves unconvincing, the hope is to motivate the condition for further exploration as it may have downstream implications for other areas in the theory of knowledge.

The plan for the paper is as follows. Section two (2) lays out sensitivity and safety in more detail and discusses some counterexamples that give each of them trouble. Section three (3) offers a thorough defense of sturdiness and explains how it is to be interpreted. Section four (4) tests sturdiness against the very same Gettier-style cases that pose problems for sensitivity and safety. In so doing, it will address the motivation for this interpretation, as well as demonstrate how sturdiness deals with epistemic closure, logical truths, and skeptical scenarios. Section five (5) concludes.

2 Sensitivity

Recall sensitivity:

Sensitivity:

If p were false, s would not have believed p. (Nozick, 1981, 159) (~p→~Bp).

Nozick's account is often referred to as the tracking or truth tracking theory. On this account S knows a proposition p only if S's belief tracks the truth. As has been argued previously, sensitivity has a strong record of success in handling many of the Gettier-style cases that proliferate the literature (Pritchard, 2012).

Like many instances of counterfactuals, the truth of sensitivity propositions are interpreted using the Lewis-Stalnaker possible world semantics (Lewis, 1973; Stalnaker and Thomason 1970). For a belief to be sensitive then, it must satisfy the following test: S's belief in p is sensitive only if in the nearby possible worlds where p is not the case, S does not believe p.

Imagine the set of nearby possible worlds in which p is false. Then check those worlds to see if S believes p. If S does believe p, the belief is not sensitive and thus not a candidate for knowledge. Alternatively, if in the worlds where ~p, s does not believe p, then the belief is sensitive and is a candidate for knowledge. In other words, you will have a counter example to sensitivity if you can easily craft a nearby possible world where (Bp^~p).

This condition handles a wide variety of cases but is not without its problems. The rest of this section will be dedicated to examining some of these issues.

2.1 Tough cases

First, consider a modified fake barns case from Kripke (2011). In *Barns* the subject has two intuitively related beliefs. Sensitivity predicts that only one of these beliefs is a candidate for knowledge. *Barns* objects to the necessity of sensitivity and raises issues for closure.

Barns:

> Sally is driving through fake barn county which is full of barn façades. She happens to be looking at the one real barn in the area. The real barn she is looking at is red while all the façades are green. Sally thus comes to believe the following two propositions. There is a barn in front of me (p). There is a red barn in front of me (q).

Let us first evaluate q: there is a red barn in front of me. Consider the nearest worlds where q is false, that is where Sally is not in front of the red barn. We want to be sure that in these worlds, Sally does not believe there is a red barn in front of her. In some of these worlds, Sally is still looking at a barn façade. But that is okay, because all the barn façades are green, she would thus not believe that there is a red barn in front of her. Her belief in q is sensitive.

Now consider p: there is a barn in front of me. Consider the nearest worlds where p is false. At some of these worlds Sally may still be looking at a barn façade, because they are indistinguishable by hypothesis, and would thus still form the belief in p. Her belief in p is not sensitive.

Regardless of one's intuition about whether p or q should be candidates for knowledge, it seems odd that the two beliefs should not have the same result on the application of sensitivity. After all, seeing a red barn entails seeing a barn, yet somehow the result is that q is a candidate for knowledge while p is not. This amounts to an egregious violation of epistemic closure. A simple formulation of epistemic closure holds that if s knows p, and s knows that p entails q, then s also knows q. In this case, because seeing a red barn entails seeing a barn, when Sally knows she sees a red barn she should also know that she sees a barn. Nozick embraced this violation of closure, while others found this disturbing enough to abandon sensitivity.

2.2 Sensitivity, closure and skepticism

Sensitivity was also forced to adopt some overly skeptical results. Consider the possibility of knowing that one is not a brain in vat (BIV). Will the belief that one is not a brain in a vat satisfy sensitivity? Imagine the nearest possible worlds in which p is false (in which one is a brain in a vat). At those worlds one will still believe they are not brains in vats, because the demon or vat operators will have made it appear this way (Nozick, 1981, 201). The belief is not sensitive.

Moreover, Nozick is steadfast that common sense beliefs, such as the fact that one has hands, is sensitive. Imagine the nearest worlds where one does not have hands, in these worlds one would not believe that they have hands. This may seem like a good result at first. Consider, however, an argument from closure which purports to rule out a skeptical hypothesis.

1. I know I have hands (p).
2. If I know I have hands, then I know I am not a handless brain in a vat (p→q).
3. I know I am not a handless brain in vat (q).

Nozick holds that 1 is sensitive, but that 3 is not. This is a bigger problem than it may seem because 1 entails 3. After all, having hands entails that one is not a handless BIV. Faced with these problems Nozick was forced to accept skeptical arguments under certain conditions and deny closure (Nozick, 1981, 204–10). These results led to consideration of alternate counterfactual conditionals in hopes of avoiding these pitfalls.

2.3 Safety

More recently, safety has been proposed as an improvement over sensitivity.

> **Safety**:
>
> If S were to believe p, p would be true (Bp→p) (Sosa, 1999) (Pritchard, 2012).

Similarly to sensitivity, the safety of beliefs is assessed by checking whether at the nearby worlds where S believes p, p is true. Safety also offered effective remedies to the original Gettier cases as well as many of those that came in its wake. It should be noted that just as sensitivity has difficulty with *Barns*, so too does safety. That is so because there are very nearby worlds where S forms the belief in the presence of a barn, but she is not looking at an actual barn. This would render her belief in seeing a barn unsafe, while her belief in seeing a red barn is safe. This presents another problem for safety which is shared with sensitivity, namely that neither account validates closure in this instance.

2.4 Challenges and tough cases

In addition to cases like *Barns*, there are a set of novel objections to the necessity of safety. One notable counterexample comes from Comesaña (2005).

> *Halloween Party*:
>
> There is a Halloween party at Andy's house, and Sam is invited. Andy hires Judy to stand at a crossroads and tell people that the party is at the house down the left road. Unbeknownst to Sam, Andy does not want Michael to go to the party, so he tells Judy that if she sees Michael she should tell him the same thing she tells everybody else (that the party is at the house down the left road), but she should immediately phone Andy so that the party can be moved to Adam's house, which is down the right road. Sam seriously considers disguising himself as Michael, but at the last moment he does not. When Sam gets to the crossroads, he asks Judy where the party is, and she tells him that it is down the left road (Comesaña, 2005).

According to Comesaña, *Halloween Party* is an intuitive case of knowledge that is also unsafe. Consider the nearest possible worlds where S believes p, in any of these worlds is p false? There is a nearby possibility where S forms the belief in the party's

being at Andy's (Bp) when this is not in fact the case (~p). Namely, it is the possible world wherein S comes dressed as Michael, which recall that he seriously considered, thus increasing the proximity of this world. In that world S forms the belief in p while p is not the case. P is an unsafe belief but is an intuitive example of knowledge.

Another well-known counterexample to safety comes from Neta and Rohrbaugh (2004). This case involves a scientific trial and a chemical substance.

> *Chemical*:
>
> Sandra is in a psychological experiment which asks her to report the number of light flashes she recalls seeing. Before being shown the lights she drinks a glass of juice. Unbeknownst to her, she is in a control group and her glass is regular juice. Other groups received memory hindering chemicals in their juice. Sandra sees seven flashes and reports this to the experimenter. Had Sandra been in another group, she would still have reported seven flashes although she would have seen only six. Sandra forms the true belief that she has seen seven flashes (p) (Neta and Rohrbaugh 2004, 400).

Chemical is another instance of intuitive knowledge which is unsafe. The belief of Sandra in *Chemical* is unsafe because she very easily could have been in the experimental group wherein the members received memory altering chemicals. Because of the presence of the experimental group, Sandra's belief in p could have easily been false. In possible worlds terms, there is a nearby possible world where Sandra still believes she was shown seven flashes (namely the one in which she is in the experimental group) but that she was not in fact shown seven flashes, because she was shown six flashes.

2.5 Safety and skepticism

Safety also offered an improvement over sensitivity regarding skepticism. According to a bases relativized version of safety, one's belief that they are not a BIV is safe. This is so because the worlds where one is a brain in a vat take completely different evidence as their inputs for forming their beliefs (Williamson, 2000; 2009). According to Williamson, "one's total evidence simply is one's total knowledge" (Williamson, 2000, 9). This is Williamson's well know E=K thesis which informs his base relative formulation of safety.

Recall the anti-skeptical argument from above:

1. I know I have hands (p).
2. If I know I have hands, then I know I am not a handless brain in a vat (p→q).
3. I know I am not a handless brain in vat (q).

Because the subject in the non-skeptical case has good evidence, they have seen their own hands using reliable perception; and the subject who is in the skeptical case has bad evidence; they are fed hand precepts by a machine; these two cases are much more radically different than they seem at first glance. By highlighting the difference

in evidence that one has between the good and bad case, the BIV world is driven further away and thus made irrelevant. Because the world where one is a BIV is more distant, safety can predict that one knows 3 on the basis of their evidence and the competent deduction from 1 to 2.

3 Sturdy beliefs

In study of sensitivity and safety, one may notice that there is one counterfactual left in the logical space of the relationship between a subject believing a proposition (Bp) and that proposition itself (p).

Sturdiness: S knows p only if S's true belief is sturdy.

-If S were not to believe p, then p would be false ($\sim Bp \rightarrow \sim p$).

Sensitivity ensures that our beliefs track the truth. Safety ensures that our beliefs are not false in sufficiently similar circumstances. What does sturdiness ensure? In some ways, sturdiness itself 'tracks the truth,' with the added benefit that the relevant worlds, not-belief-in-p worlds, will typically not be as far away as potential not-p worlds. When we evaluate a belief against the conditional ($\sim Bp \rightarrow \sim p$), we are asking what a world which is maximally consistent with the actual world and actual belief forming method would have to be like if we were to make the antecedent true.

Like sensitivity and safety, sturdiness is tested in possible worlds terms. In these possible worlds one should hold maximally fixed the conditions of the actual world. This is all the facts and laws about the actual world as well as the method used in belief formation. The nearby possible worlds in question are those in which the belief in p is absent from the subject's mind ($\sim Bp$). In good cases, this will automatically entail that the fact making the belief true is absent from those worlds as well ($\sim p$). If this does not happen, if a relevant possible world which is a $\sim Bp$ world can still have p be true, then the belief has been exposed as poorly constructed. Beliefs that pass this test are sturdy.

It is natural to wonder why, in such an otherwise focused and crowded area of epistemology, sturdiness was so totally overlooked. As outlined above, there is not even an attempt at a treatment of this condition but for the brief mention by Carrier (1971). A reasonable story that one might tell in explanation of this is that, at first glance, sturdiness looks plainly false.[3] On a somewhat flat-footed view of our world, and the epistemic agents that interact with it, it is easy to imagine a picture wherein beliefs are typically frailer than the facts about which they are beliefs. On a view like this, sturdiness looks to get even the most basic cases wrong. If one had failed to look at just the right moment, or glanced at a slightly different object, then surely the fact in question would remain the same. In other words, it is easy to think straight out of the gate that one's believing or not believing p would have no effect whatsoever on

[3] I am exceptionally grateful to an anonymous reviewer for pointing this out as a way to highlight the upshot of the view.

the truth of p. The aim of the remainder of this paper is to look beyond this simple view and offer a story that presents sturdiness as a nuanced and plausible alternative to the received modal accounts of knowledge.

3.1 Sturdiness in action

Before testing sturdiness against *Barns*, *Halloween Party*, and *Chemical*, sturdiness can prove its worth on another Gettier-style case from Chisholm (1977). This section will also motivate a backwards reading of the sturdiness counterfactual.

> *Sheep in the Field*:
>
> Sally is driving through a field. Looking out the window, she sees what appears to be a sheep in the field. She forms the belief, there is a sheep in the field (p). Unbeknownst to Sally, the object she saw is actually a rock that happens to perfectly resemble a sheep from a distance. But, unbeknownst to her, there is also a sheep just out of view in the very same field.

Sheep in the field is an intuitive case of justified true belief which fails to be knowledge. A theory of knowledge is thus working properly if it predicts that Sally does not know the proposition in question. What does sturdiness predict? Consider the nearest worlds in which Sally does not form the belief in p, in these worlds p is still true. This is so because Sally never saw the real sheep, her belief's content is of the sheep-rock and is thus unconnected to the real sheep in the field. In the ~Bp worlds Sally looks at something else, but the sheep is still just out of view. Sally's belief is not sturdy, which matches intuition.[4]

As mentioned above, the sturdiness counterfactual should receive a backward interpretation. This structure of conditional counterfactual asks us to reconstruct the world to maintain consistency with a true antecedent. This allows sturdiness some freedom to work backwards from the truth of the antecedent to the truth of the consequent.

One thing that might be said in response to this reading is that it may imply a view of beliefs which always temporally follow some fact. Experience however tells us that beliefs are not always about some fact that occurs just before their formation. Indeed, we can and often do have beliefs about the future. In response, two things come to mind about this backward reading. The first is that whether or not beliefs about the future can come out as known will depend on quite a bit outside of the scope of this argument. I will thus bracket that concern here. Additionally, if indeed the backwards reading is plausible, then this might also inspire novel resources for safety. This too is outside of the scope of the argument, but it is an interesting research program to examine the potential of safety given a sturdiness style interpretation.

Frank Jackson addresses this method of interpretation by using an example of two friends discussing a third friend visible on a rooftop.

[4] One may worry about subtle variations to this (or any) case that could spell trouble for sturdiness. I will address just such modifications in Sect. 4.3.

Roof Jumper:

> Smith is outside a building and sees his friend Jones on the roof. He gets nervous as Smith looks over the edge. To his relief, Jones steps away from the edge. Smith thus says aloud "Phew, if Jones had jumped, he would have died."
>
> Smith is overheard by Betty, who is right behind him, and who is also a friend of Jones. Betty replies: "Well we know Jones is rational and has no desire to die. He only would have jumped for good reason. So actually, if Jones had jumped, he would have lived" (Jackson, 1977; Khoo, 2017).

It is arguable that Betty has made a good case. What is not arguable is that she has held different things fixed in her possible world than Smith has. Betty has interpreted causality consistently with the facts and laws of the world up until the antecedent, the moment Jones jumps. She has then done some clever rearranging to make the antecedent true while maintaining consistency with those facts and laws. She then checks how the consequent comes out as a result. Smith on the other hand has reasoned from the truth of the consequent, Jones' death, to determine what would have made him jump (i.e. a desire to die).

The process here is to imagine a causal chain which is rationally consistent with the facts known up until the moment of belief formation and the physical laws that would have resulted in Jones jumping. First, examine what facts are known, these include that Jones is rational, Jones has a will to live, and Jones is not impulsive. Then imagine a causal chain that would have led to Jones jumping, such as his going to the roof and seeing a giant net below. This interpretation gives the result that the proposition: 'If Jones had jumped, then Jones would have lived', is true.

The example is provided to show that there are at least two ways that one can read these counterfactuals. One need not be convinced that the first reading is a misreading, but only that the second sort of reading (Betty's) is plausible. If this is right, then it may be that philosophers have overlooked this second possible reading for sturdiness, which might constitute a plausible necessary condition on knowledge.

We can see the backwards method of reading these counterfactuals in action with reference to a simple case of everyday knowledge. Imagine that one opens the fridge door, sees a case of beer, and comes to know p: "There is beer in the fridge." Sturdiness about this proposition p should now be formulated as follows:

> If I didn't believe there was beer in the fridge, then there would be no beer in the fridge.

In order for p to come out sturdy, we need to adopt a reading like Betty's in *Roof Jumper*. Motivating this interpretation of the conditional is part of the goal of this paper. In this interpretation we hold everything fixed up until the moment of the belief formation including the method used. On might naturally worry that in situations where no belief is formed, in what sense is the method a 'belief-forming method'? The most natural response is to say that belief-forming methods should be understood as the sorts of things that could issue in different beliefs (or no belief at

all) with regard to some topic and depending on the circumstances. In this example, the method could be forming a belief about whether there is beer in the fridge on the basis of clear visual perception. This then might issue in different beliefs (There is beer, there is not beer) depending on the content of that visual perception.[5]

The subject then will still look in the fridge using her visual faculties, observe what is inside, and either form or not form a belief about the presence of beer. The most likely possibility, this interpretation contends, in which S does not form the belief that there is beer in the fridge is one in which the fridge contains no beer. Given that in the actual world S gets a clear look at the beer, this too is something to hold fixed across the range of relevant not-belief possibilities. This is what makes it the case that the nearest not-belief worlds are no-beer-in-the-fridge worlds as opposed to, for example, beer-behind-milk worlds.[6] This interpretation gives the verdict that p is a sturdy belief, formed via an appropriate method with veridical visual perception.

Outside of the benefits sturdiness confers as a theory of knowledge, there are at least two independently motivated reasons to evaluate beliefs in this way. First, There is precedent in the field of empirical psychology for this interpretation of counterfactuals (Khoo, 2017; Jackson, 1977; von Kügelgen et al., 2023). Studies suggest that when asked to interpret counterfactuals, participants are more likely to employ a backwards than a forwards method (von Kügelgen et al., 2023, 12). Moreover, sturdiness obeys what historians refer to as the minimal rewrite rule, wherein the subject makes: "[…] minimal changes to the actual world to create the necessary conditions that would have led to a particular counterfactual antecedent" (Ibid).

There is also a growing support in the philosophical literature that the semantics of backward interpretations are not as dicey as may have once been thought (Jackson, 1977; Khoo, 2017).[7] Justin Khoo has even outlined a formal semantics for backwards interpretations of conditionals (Khoo, 2017). A full discussion of those semantics is beyond the scope of this argument. The hope is that those readers who are curious about the formal semantics can see the benefits of sturdiness as a theory of knowledge and consult the technical aspects as necessary. The following section will show how sturdiness stacks up against the prominent counterexamples to sensitivity and safety presented above.

4 The benefits of sturdiness

As with sensitivity and safety, sturdiness has succeeded if it matches our intuitions about various Gettier-style cases and delivers the intuitive verdict in cases of everyday knowledge. We have seen sturdiness deliver the intuitive result on cases of every-

[5] There is always some trouble about just how fine-grained our methods description should be. A full discussion of precisely how to individuate methods would burden this argument but is well worth taking up in the wake of sturdiness. It should also be noted that these issues are shared equally by sensitivity and safety theories, which also appeal to belief-forming methods, and by process reliabilist theories, which appeal to belief-forming processes.

[6] This interpretation of the method employed follows Nozick's method relative sensitivity which he proposed in response to objections about the adherence condition.

[7] The foremost opponent of these interpretations was of course David Lewis (Lewis, 1979).

day knowledge with reference to the beer in the fridge case presented above. But how does sturdiness fare against some of the more challenging game in town?

4.1 Counterexample cases

4.1.1 Barns

Recall that *Barns* had two problems. First it delivered divergent verdicts on propositions that are in an entailment relationship. If one is looking at a red barn (q), then it follows that one is looking at a barn (p). On sensitivity, p was sensitive while q was not. Safety has the same result. In the nearby worlds in which one believes they are looking at a red barn they are safe. However, there is a nearby world wherein one believes themselves to be looking at a barn but are in fact looking at a barn façade. This amounts to a denial of closure, at least in this case.

Is q sturdy? Imagine the nearest worlds where S does not believe q. In these worlds S will not have the belief of there being a red barn, so she must have looked at something else including a barn façade. In these worlds S would also not have a red barn in front of her because she is in front of something else, namely a façade. Q is thus sturdy. So far sturdiness has delivered the same results as sensitivity and safety.

Now consider p. Imagine the nearest possible worlds where S does not believe there is a barn in front of her. These cannot be worlds wherein S looks at one of the many barn façades. This is so because in those worlds S would still form the belief that there is a barn in front of her, despite the phenomenon being a mere façade. In these cases, S does not form the belief that she is in front of a barn ($\sim Bp$). But notice here that in these worlds, S is not in front of a barn or a barn façade. In other words, p is not the case ($\sim p$). S's belief in p is sturdy.

This is a different result than the result arrived at by applying safety or sensitivity. In this case the two verdicts match, which preserves the notion of knowledge being closed under entailment. Sturdiness avoids the egregious violations of closure that sensitivity and safety were forced to accept.

It has been received wisdom for some time that fake barns were not to be understood as intuitive cases of knowledge. Accepting that wisdom would make the result that p is sturdy to be counter intuitive. There is some precedent for the idea that fake barn cases are instances of knowledge as discussed by Sosa among others (Biro, 2017; Hetherington, 1999; Lycan, 1977; Sosa, 2009, 2011; Turri, 2017). The central line of thought is that, while certainly lucky, S's true belief is a result of her epistemic abilities and caused by a genuine barn. Additionally, recent experimental philosophy seems to concur with this result. The average person considers traditional fake barn cases to be intuitive cases of knowledge (Colaço et al., 2014).[8] Sturdiness has predicted that one can have knowledge despite the lurking danger of barn facades in the area.

P is sturdy because despite the possibility of error, it is formed in the most appropriate causal way. Recall that sensitivity and safety both predicted that when Sally

[8] Even if one remains unconvinced that the propositions in Barns should be considered knowledge, it is important to remember that sturdiness is proposed as a necessary condition.

looked at a red barn, she had a belief which was a candidate for knowledge. This is yet another odd feature of sensitivity and safety. If the belief is made just slightly more specific in fake barn cases, then the danger of the fake barns is somehow avoided. Regardless of what one may think about the status of the beliefs in *Barn*, this case also presented one of the more serious violations of epistemic closure. If one is dissatisfied with raising the beliefs in *Barns* to the level of knowledge, they could still be glad if closure were not violated in this way.

4.1.2 Halloween party

Recall that *Halloween party* was an intuitive case of knowledge that was unsafe. Is the belief in question sturdy?

Imagine the nearby worlds wherein S does not form the belief in the party being at Andy's house. If S does not form this belief, then this also cannot be a world where S arrives dressed as Michael. This is so because in that world, S does indeed form the belief that (p). In fact, the relevant (~Bp) world needs to be a world where S does not arrive at the fork in the road at all. That is so because in any world wherein S arrives at the fork, he will form the belief in p. This is because *Halloween Party* is an inevitable causation case. Sturdiness however allows for different facts to be held fixed; we thus look for the nearest world where S does not arrive at the fork and thus forms no belief about the party's location.

What is the nearest world in which S does not arrive at the fork in the road at all? It seems that the most likely possibility is one where there simply is not a party to which S is invited. If there is no party, then there is no party down the left road (~p). This world is more nearby than a world where S has all the same friends and acquaintances, but they do not invite him to the party. This is so because S is invited in the original case. Halloween party thus satisfies sturdiness and is a candidate for knowledge.

4.1.3 Chemical

Chemical is another case of inevitable causation. In all of the worlds in the case, S will form the belief in seeing seven flashes. This belief will be true in the actual world, but in nearby worlds in which S is in the experimental group, and has thus consumed memory altering chemicals, it will be false. Sturdiness defines the relevant possible worlds as worlds where S does not form the belief in p. These worlds cannot be worlds where S is in the experimental group or where she is in the regular group. There is some freedom here in what world is the most nearby, say it is a world where S opts out of the study at the last minute. Of course, then, due to not being in the study, S will not see seven flashes. This satisfies sturdiness and matches the intuition that *Chemical* is a case of knowledge.

It may at first seem odd that the most relevant possibility is one where S opts out of the test. What this shows, however, is that the belief is well-formed and that the relevant possibility of error is very far-off. The focus is on the way S arrived at her belief in the actual world, via visual perception, and not the nearby threat of the experimental group.

So far it has been demonstrated that sturdiness can predict the correct result on some of the tough cases for safety and sensitivity. Sturdiness is thus far showing itself to be a relevant and plausible modal account of knowledge. One way sturdiness proved its worth was by not requiring a blatant violation of closure. Another issue for sensitivity was that it was overly skeptical in its results. The following section will further discuss sturdiness and how it performs regarding necessary truths, skepticism, and closure.

4.2 Objections and replies

The previous sections demonstrated some of the advantages sturdiness has as a theory of knowledge. At the same time, it seems that sturdiness may face some serious problems of its own. Much like safety and sensitivity, the issues are validating closure and handling necessary truths. This section will examine these objections and offer a unified solution.

4.3 Closure

First, consider whether sturdiness preserves closure in all relevant cases. One of the benefits of sturdiness as a theory of knowledge was that it preserved closure in difficult cases such as *Barns.* Consider cases such as the following:

> *Clouds*
>
> Steve looks out of his window in Miami, Florida and sees that it is cloudy. Based on his seeing the clouds he forms two beliefs; It is cloudy in Miami, Florida (p), and it is cloudy somewhere in Florida (q).

First, p entails q. Miami is in Florida, so if it is cloudy there then it is cloudy somewhere in Florida. At first glance, sturdiness will have trouble preserving closure in this instance. P is straightforwardly sturdy, in the nearest world where S does not believe p and S uses the method of visual perception in Miami then it would not be cloudy there. Is q sturdy? In the nearest possible worlds where S does not believe it is cloudy somewhere in Florida then it should not be the case that it is cloudy somewhere in Florida. Considering S is in Miami and using his perception there, the nearest possible world is one where it is not cloudy in Miami. Unfortunately, there is nothing precluding the weather from being cloudy at some other location in Florida. Then there would be a nearby possible world in which S does not believe it is cloudy in Florida, but it still is cloudy, in Tampa for example. Q is not sturdy, and this is an uncomfortable result.

The intuitive response leads to a solution and adds benefits for sturdiness. It seems odd that the possibility of cloudy weather at some unrelated location such as Tampa should count against one's true sturdy belief in the weather at the location in which they do the observing in the actual world. This seems to be a failure of the formulation of the sturdiness condition as opposed to its guiding principle. That principle put

simply is: In the relevant world's where S does not believe p (~Bp), then the relevant fact on which the belief is based would not be true (~p).

Sturdiness can be reformulated in order to better accommodate the guiding intuition and see if that offers a solution to *Cloudy.* A belief is sturdy just in case:

> If S were not to have believed p, the fact on which the belief is based would not be true.

In the case described above, the fact that is making S's belief true is the cloudy weather in Miami, the fact that it might be made true by some cloudy weather elsewhere should not count against it. In the actual world, the belief is well-formed and well-connected to the fact that makes it true.

P remains straightforwardly sturdy on the new formulation; the belief is being made true and is about the clouds in Miami so in the nearest worlds in which S does not believe it is cloudy in Miami; it would not be cloudy there. Likewise, q concerns the clouds in Miami, despite it being a generalization. For q to be sturdy, it must be that in the relevant worlds where S does not form the belief in q with p as its doxastic basis, then p must not be true. This is precisely the case that S is in. In the nearest worlds where he does not believe q (that it is cloudy somewhere in Florida) that is because he never saw clouds in Miami, that is because there were no clouds in Miami for him to see (~p). Q is now sturdy[9].

With the above formulation of sturdiness in mind, we are now in a position to recursively define sturdiness for any beliefs p and q.

> Modified Sturdiness:
>
> Base Clause: If both (i) S's true belief that p is non-inferential, and (ii) if S were not to believe p, p wouldn't be true (~Bp → ~p), then S's belief that p is sturdy.
>
> Recursive Clause: If both (i) S forms the belief that q via competent inference from a sturdy belief that p, and (ii) if S were not to believe q, then the fact on which the belief is based would not be true (~Bq →~p), then S's belief that q is sturdy.
>
> Closure Clause: No other beliefs are sturdy.

This recursive definition is merely the modified sturdiness from above unpacked into formal clauses. Modified sturdiness is importantly related to early attempts to solve Gettier's cases. These include the `No False Lemmas' theory from Clark (Clark, 1963) as well as the `Causal Theory' and `Theory of Discrimination' from Goldman (Goldman, 1967, 1976). Perhaps the most relevant of these theories is Clark's which argues that S knows p just in case S's belief in p is fully grounded or based on no false beliefs. The notion of grounding a belief in another, has played a significant role

[9] It will not matter now if it is cloudy somewhere else in Florida. This is so because the relevant counterfactual that needs to be satisfied just states that it is not cloudy in Miami when S does not believe q.

in influencing the modified sturdiness presented above.[10] These theories all involved drawing a rudimentary basic vs non-basic distinction, and that is part of the procedure for modified sturdiness as well.[11] The causal theory was abandoned because precisely what was the right sort of causal relationship was deemed ambiguous, and the no false lemmas theory was abandoned because of clear objections to its necessity. Sturdiness avoids these worries by using a formal definition to make precise the conditions under which it is satisfied as well as handling a range of objections to its necessity.

One may think that modified sturdiness is quite a bit of extra machinery. It turns out, however, that applying the modified definition has added benefits in that it provides tools for responding to additional difficult cases. Consider the following variation on Chisholm's *Sheep in the Field* case:

Farmer Ed:

> Suppose farmer Ed flips a fair coin. If it lands heads, he puts the rock that resembles the sheep out of sight in the back of the field and moves the genuine sheep out of his field to farmer Ted's field across town. If it lands tails, he puts the rock right where Sally sees it and leaves the genuine sheep in the back corner of the field out of sight. Things are otherwise just like Chisholm's original case–Sally sees the rock that looks like a sheep and forms her belief that there's a sheep in the field.[12]

The initial reaction to this case is that sturdiness gets the wrong result. Given the relevant possibility of the 'heads' world, the most nearby world where Sally does not believe p will be one where p is not the case. But since this is an intuitive example of non-knowledge, sturdiness has failed to deliver. Can modified sturdiness save the day?

To assess modified sturdiness the belief must first be re-interpreted as a belief p and an inference to q arrived at via the belief p. Those beliefs seem most naturally to be:

p: That is a sheep; and

q: There is a sheep in the field.

The problem that presents itself straight away is that p will fail to be sturdy. This is so because the base clause writes that in order to be sturdy a belief must satisfy two conditions: (i) S's true belief that p is non-inferential, and (ii) if S were not to believe

[10] Many thanks to an anonymous referee for situating sturdiness among these more familiar conditions.

[11] There is always some trouble about where precisely to draw the distinction between what is and is not an inferential belief. It is, however, obvious that the distinction is a genuine one. A thorough discussion of that distinction is beyond the scope of this argument. This is at least one way of applying the distinction that delivers on the intuitions in the cases presented.

[12] Many thanks to an anonymous referee for bringing this case and others like it to my attention.

p, p wouldn't be true (~Bp → ~p), then S's belief that p is sturdy (s knows p). But p fails the first condition. While the belief is non-inferential, it is not true and thus cannot be sturdy.

The task then is to look at whether or not q satisfies the recursive clause, (~Bq→~p). The trouble here is that (i) of the recursive clause states that q be formed via competent inference from a sturdy belief that p, and p is not a sturdy belief. Given that both elements of the recursive clause need to be satisfied, q fails and is thus not a sturdy belief. By applying modified sturdiness, the intuitive result has been achieved. Because *Farmer Ed* is an intuitive case of non-knowledge, we are glad that the beliefs in question are not sturdy.[13]

Consider a further case which could spell trouble for sturdiness:

Water Bottle:

Suppose S has a water bottle on the table. He sees X put what S knows to be a drug into the water bottle. He comes to believe that his water bottle is drugged. Intuitively, S knows that it is drugged. After all, he saw X put what he knows to be a drug into his water bottle. But what S doesn't know is that X has drugged every single one of the water bottles in S's town. No one, including S, can drink from a water bottle in town and not be drugged. In some cases, X drugs water bottles discreetly, and no one knows, but in other cases X does it out in the open so that the person can see that the water bottle is having drugs placed in it.

This case objects to the necessity of sturdiness by showing an intuitive case of knowledge that fails to be sturdy. That is so because X is drugging every bottle in town, sometimes discreetly and sometimes not discreetly. So, whether or not S believes his bottle is drugged, it is going to be drugged. Can modified sturdiness deliver the right result? We again begin by reinterpreting the case as two inferentially related beliefs:

p: I see X drugging my bottle; and

q: my bottle is drugged.

P is straightforwardly sturdy by visual perception. Because the perception is veridical in the actual world, and everything else is held fixed, the most relevant possibility where S does not believe p is one where p is false, namely where S did not see X drug his bottle. Now take q, my bottle is drugged. This needs to satisfy the conditional (~Bq→~p). We then check that the nearby worlds where S does not believe his bottle is drugged are also worlds where S did not see X drug his bottle. Because S forms his belief in q based on the veridical perception of p, the most likely explanation for S not believing q is one where p is false. On this interpretation q is sturdy.

[13] It is worth noting that this is an extremely difficult case to handle for any of the extant modal accounts of knowledge. With that said, in case one is not satisfied with sturdiness's result, the safety or sensitivity theorist will also have to wrestle with this case.

Modified sturdiness has hopefully already carried its weight. In the following section we can see some even more challenging objections to sturdiness.

4.4 Necessary truths

Another objection to sturdiness might be its ability to handle necessary truths. The idea is that satisfying sturdiness requires a relevant fact to be false in a nearby world. But because necessary truths are never false, it looks impossible for them to satisfy sturdiness. Call this the problem of logical ignorance, as it is related to the more well-known problem for safety of logical omniscience. Sturdiness can solve this problem by appealing to the recursive formulation. That is, by conceiving of knowledge of logical truths as inferences based on some other contingent fact about the world. For example, one might see a proof of a logical truth and based on that proof infer that the given formula is indeed a logical truth. Consider the following case:

Logic:

Sam is shown the proof of a logical truth. Based on this he forms two beliefs. I see the proof of theorem x (p), and theorem x is a logical truth (q).

Based on the modified version of sturdiness above, S's belief in p is sturdy. He forms this belief based on visual perception so in the nearby relevant not-belief-worlds he also did not see the proof in question. Q will trigger the recursive clause, which states that a in the worlds where S does not believe q, p would be false ($\sim Bq \rightarrow \sim p$). In this case the relevant nearby worlds where S doesn't believe q need to be worlds where he does not see the proof in the first place.

With this formulation q also turns out sturdy. In the nearest world's where S does not believe that theorem x is a logical truth, p is not the case i.e. S did not see the proof that theorem x is a logical truth. There is an intuitive pull to this solution as it mirrors the method by which many people come to believe in logical or mathematical truths. Even if we consider axioms, which do not readily admit of proofs, we could think of beliefs that these are axioms as being importantly related to the original place that this fact was learned, for example by a logic professor or a textbook.

4.5 Skepticism

A final objection wonders whether or not sturdiness can handle skeptical scenarios. Recall the simple anti-skeptical argument which gave sensitivity trouble:

1. I know I have hands.
2. If I know I have hands, then I know I am not a handless BIV.
3. I know I am not a handless BIV.

The challenge of any theory is to predict that three is known. How does sturdiness fare? There is an extensive literature that deals with skepticism in epistemology. Within that literature, what has been written on closure and how it may ameliorate

or intensify our skeptical worries is particularly strong. It is beyond the scope of this paper to deal with skeptical worries in full. But it is possible that sturdiness could have good results in this area as well. Those good results could look something like the following.

Premise one is straightforwardly sturdy; it is a simple uncomplicated case of visual perception. Assuming S knows the implication that makes up premise two, and makes a competent deduction, then S should be able to know three as well.

Because the belief in premise three is formed on the basis a competent deduction, we should apply sturdiness according to the recursive formulation. This entails holding fixed the method of visual perception used to form the belief in premise one. S will thus know three if she satisfies this conditional: If S did not believe she was not a handless BIV (via perception), then she would not have hands.

The fixed method of visual perception makes it such that the closest world in which S does not believe that she is not a BIV is one where she believes she *is* a BIV. If the belief must be formed via visual perception, then this would also necessarily be a world where she was in fact a BIV and somehow came to discover this. The good news is that this means that (3) is sturdy, because if S is a BIV then she does not have hands.

Why is the closest possible world not the world where S has just taken an epistemology course, or has taken LSD and is hallucinating? As mentioned above, sturdiness holds the belief-forming method fixed. Because of this the relevant possibilities need to be those that share the method used in (1), visual perception. Premise one is sturdy because the most nearby world where S does not believe she has hands is a world where she is looking at her empty hand spots. This makes it so that the only way (3) can be interpreted is to use visual perception in the assessment. While this requires some imagination, it delivers the intuitive result that premise three is sturdy.

Being a brain in vat is a remote possibility. What sturdiness tells us here is that a possibility can be simultaneously remote and the most nearby. In fact, it is possible that admitting that being in the skeptical scenario is the most nearby, while still maintaining it is not *ceteris paribus* nearby at all, may open up a new response to the skeptic more generally. What applying sturdiness in this case shows is that the belief in three, that S is not a BIV, is formed in the most appropriate causal way, by looking around and seeing their hands and all the other vivid non-BIV phenomena of the world.

5 Conclusion

This paper has argued that there is an underexplored option in the debate over modal conditions on knowledge. Despite some obvious initial worries about the appearance of the counterfactual, the sturdiness condition was outlined in full and was shown to achieve intuitive results on a wide variety of Gettier-style cases. Some objections concerning closure, logical ignorance, and skepticism were considered. The hope is

that this argument has laid the ground for research into this modal conditional which should be given adequate attention as a *bona fide* condition on knowledge.[14]

Acknowledgements No.

Funding The authors have no relevant financial or non-financial interests to disclose.

Declarations

Competing interests Authors are required to disclose financial or non-financial interests that are directly or indirectly related to the work submitted for publication. Please refer to "Competing Interests and Funding" below for more information on how to complete this section."

References

Becker, K. (2018). The sensitivity response to the gettier problem. In: *The Gettier Problem*, edited by Stephen Hetherington. Classic Philosophical Arguments. Cambridge University Press. https://doi.org/10.1017/9781316827413.009

Biro, J. (2017). Non-Pickwickian belief and 'the Gettier problem'. *Logos & Episteme*, *8*(1), 47–69. https://doi.org/10.5840/logos-episteme2017813

Carrier, L. S. (1971). An analysis of empirical knowledge. *The Southern Journal of Philosophy*, *9*(1), 3–11. https://doi.org/10.1111/j.2041-6962.1971.tb02131.x

Chisholm, R. M. (1977). *Theory of Knowledge*. 2nd edition. Prentice-Hall, inc.

Clark, M. (1963). Knowledge and grounds: A comment on Mr. Gettier's Paper. *Analysis*, *24*(2), 46–48. https://doi.org/10.1093/analys/24.2.46

Colaço, D., Buckwalter, W., Stich, S., & Edouard M. (2014). Epistemic intuitions in fake-barn thought experiments. *Episteme*, *11*(2), 199–212. https://doi.org/10.1017/epi.2014.7

Comesaña, J. (2005). Unsafe knowledge. *Synthese*, *146*(3), 395–404. https://doi.org/10.1007/s11229-004-6213-7

Goldman, A. (1967). A causal theory of knowing. *The Journal of Philosophy*, *64*(12), 357–372.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, *73*(20), 771–791. https://doi.org/10.2307/2025679

Hetherington, S. (1999). Knowing failably. *The Journal of Philosophy*, *96*(11), 565–587. https://doi.org/10.2307/2564624

Jackson, F. (1977). A causal theory of counterfactuals. *Australasian Journal of Philosophy*, *55*(1). https://doi.org/10.1080/00048407712341001

Khoo, J. (2017). Backtracking counterfactuals revisited. *Mind*, *126*(503), 841–910. https://doi.org/10.1093/mind/fzw005

Kripke, S. A. (2011). *Philosophical troubles: Collected papers*. (Volume 1). Oxford University Press.

Lewis, D. K. (1973). *Counterfactuals*. Harvard University Press.

Lewis, D. K. (1979). Counterfactual dependence and time's arrow. *Nous*, *13*(4), 455–476.

Lycan, W. G. (1977). Evidence one does not possess. *Australasian Journal of Philosophy*, *55*(2), 114–126. https://doi.org/10.1080/00048407712341141

Neta, R., & Guy R. (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly*, *85*(4), 396–406. https://doi.org/10.1111/j.1468-0114.2004.00207.x

Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.

Pritchard, D. (2012). Anti-Luck virtue epistemology. *Journal of Philosophy*, *109*(3), 247–279. https://doi.org/10.5840/jphil201210939

Shope, R. K. (1983). *The analysis of knowing: A decade of research*. Princeton University Press.

[14] Many thanks to Bob Beddor for helpful comments and discussion. Thanks also to Arjun Krishna Kumar, Marcelino Hudgson Steele, Marcus Davis, and two external reviewers for *Synthese;* as well as audiences at the University of Connecticut, the University of Florida, and the Portuguese Philosophy Society.

Sosa, E. (1999). How must knowledge be modally related to what is known? *Philosophical Topics*, *26*(1/2), 373–384.
Sosa, E. (2009). A virtue epistemology: apt belief and reflective knowledge, Volume 1. 1st edition. Oxford University Press.
Sosa, E. (2011). Knowing full well. 1st edition. Princeton University Press.
Stalnaker, R. C., Richmond, H., & Thomason (1970). A semantic analysis of conditional logic. *Theoria*, *36*(1), 23–42. https://doi.org/10.1111/j.1755-2567.1970.tb00408.x
Turri, J. (2017). Knowledge attributions in iterated fake barn cases. *Analysis*, *77*(1), 104–115. https://doi.org/10.1093/analys/anx036
von Kügelgen, J., & Mohamed, A., Sander B. (2023). Backtracking counterfactuals. *Proceedings of Machine Learning Research, 213* (May), 1–21. http://arxiv.org/abs/2211.00472
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.
Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy*, 4. https://www.amherstlecture.org/williamson2009/