

The causal theory of knowledge revisited:

An interventionist approach*

Job de Grefte · Alexander Gebharter†

Abstract: Goldman (1967) proposed that a subject s knows p if and only if p is appropriately causally connected to s 's believing p . He later on abandoned this theory (Goldman, 1976). The main objection to the theory is that the causal connection required by Goldman is compatible with certain problematic forms of luck. In this paper we argue that Goldman's causal theory of knowledge can overcome the luck problem if causation is understood along interventionist lines. We also show that the modified theory leads to the correct results in contexts involving other prominent forms of epistemic luck and compare it with other accounts on the market.

1 Introduction

Goldman's (1967) causal theory of knowledge faces notorious problems in accounting for certain kinds of luck. Partly because of this, Goldman (1976)

*This is a draft paper. The final version of this paper is published under the following bibliographical data: de Grefte, J., & Gebharter, A. (2021). The causal theory of knowledge revisited: An interventionist approach. *Ratio*, 34(3), 193–202. doi:10.1111/rati.12304

†The order of authorship is alphabetical; both authors contributed equally to this paper.

abandoned his theory. In this paper, we suggest to salvage the causal theory of knowledge by incorporating elements from the recent causation debate. In particular, we draw on basic ideas underlying the interventionist notion of *stability* (cf. [Woodward, 2010](#)) in order to distinguish between lucky and non-lucky cases of causation. This allows us to save the causal theory of knowledge by adding the requirement that knowledge must be based on cases of non-lucky causation.¹

The paper is structured as follows. In [section 2](#), we spell out the causal theory of knowledge and the luck problem. In [section 3](#), we introduce the interventionist framework. In [section 4](#), we define the notion of lucky causation and explain how it helps the causal theory of knowledge to solve the luck problem. In [section 5](#), we argue that our interventionist version of the causal theory of knowledge correctly distinguishes between kinds of luck compatible and incompatible with knowledge. This shows that our theory does not only solve the particular problem discovered by [Goldman \(1976\)](#), but also provides a unified explanation of the compatibilities and incompatibilities of knowledge with different kinds of luck. In [section 6](#), we briefly compare our account with competing accounts. We conclude in [section 7](#).

2 The causal theory of knowledge and problems with luck

[Gettier \(1963\)](#) famously showed that not every justified, true belief amounts to knowledge. Consider the following case, adapted from Gettier's paper:

¹We do not claim that interventionism is the only theory of causation suitable to overcome the mentioned problems. However, since interventionism is one of the more powerful, flexible, easily accessible, and rather intuitive accounts currently on the market, it seems to be a good first choice for our project.

Smith has strong reasons to believe Jones will get the job, and has seen Jones put ten coins in her pocket. She infers from this that the person who will get the job has ten coins in her pocket. Alas, Smith's evidence is misleading. Contrary to what Smith believes, she herself will get the job. As it happens, Smith also has ten coins in her pocket.

Uncontroversially, Smith has a true, justified belief that the person who will get the job has ten coins in her pocket. Still, this belief does not amount to knowledge. The reason seems to be that her belief is objectionably *lucky* (Unger, 1968; Engel, 1992; Pritchard, 2005). This kind of luck is called veritic luck, which roughly amounts to the following (cf. de Grefte, 2018, p. 3824): *s*'s belief that *p* is true, but the belief-forming method that generated *s*'s belief that *p* could very easily have produced a false belief.

If Gettier cases show that the justified true belief analysis of knowledge fails, then what should replace it? Goldman's causal theory of knowledge attempts to provide an alternative:

(CTK) *s* knows that *p* if and only if the fact *p* is causally connected in an "appropriate" way with *s*'s believing *p*. (Goldman, 1967, p. 369)

Since Smith's belief in the above case is exclusively based on her misleading evidence about Jones, Goldman (1967) can explain the absence of knowledge in the case above by referring to the absence of an appropriate causal connection between the fact that the person who will get the job has ten coins in her pocket and Smith's belief. The causal connection required is that either (i) *p* causes *s*'s believing *p* or (ii) *p* and *s*'s believing *p* are both consequences of a common cause. Much hangs on whether and how we can satisfactorily explain the notion of appropriateness in (CTK). Goldman does not further specify conditions for

appropriate causation in this context. Instead, he lists examples of appropriate, knowledge-producing causal processes, such as perception, memory, and inference (Goldman, 1967, p. 369). In section 4, we offer our own interpretation. For now, we assume that (CTK) adequately captures what goes on in the Gettier case above. Unfortunately, Goldman's (1967) analysis is unable to adequately capture *all* Gettier cases. As noted by Goldman (1976) himself, the approach struggles to explain why knowledge is lacking in so-called fake barn cases:

Henry is driving in the country-side and sees a barn. He forms the belief that there is a barn over yonder. Unbeknownst to Henry, however, there are numerous fake barns around, which are constructed in such a way as to make them indistinguishable from real barns from Henry's position. Luckily, Henry is presently looking at a real barn.

As Goldman admits,

[the] causal analysis cannot handle the problem [...]. Henry's belief that the object is a barn is caused by the presence of the barn; indeed, the causal process is a perceptual one. Nonetheless, we are not prepared to say [...] that Henry knows. (Goldman, 1976, p. 773)

Henry's belief that there is a barn over yonder is appropriately caused by the fact that there is a barn over yonder on Goldman's view, because it is the result of perception, one of the knowledge-producing processes on Goldman's list. Yet, it fails to amount to knowledge. Goldman's notion of appropriate causation thus wrongly classifies Henry's belief as knowledge.

We may diagnose what goes wrong with Goldman's analysis by distinguishing between two kinds of veritic luck: intervening and environmental luck (Carter & Pritchard, 2015, sec. 2). Intervening luck, as the name suggests,

intervenes between an agent's belief and the facts that make her belief true. Standard Gettier cases like the Smith case above involve this kind of luck: The truth of Smith's belief is not caused by the fact that she herself has ten coins in her pocket, but rather by the misleading evidence that Jones has. Environmental luck, on the other hand, involves no such causal disconnection. In fake barn cases, Henry's belief *is* caused by the real barn he is looking at. Rather, in the case of environmental luck, the luckiness of one's belief depends on the environment. Henry ends up with a veritically lucky belief, precisely because he could have easily looked at one of the fake barns, in which case he would have formed the false belief that there was a barn over yonder.

Summarizing, the causal account accurately rules out intervening veritic luck, simply because in these cases the beliefs are causally disconnected from the facts. However, because cases of environmental luck may involve appropriate causal connections, (CTK) fails to correctly rule out these cases. In the remainder of this paper we draw on interventionist resources to save the causal account from this problem by further specifying what counts as an appropriate causal connection.

3 The interventionist theory of causation

According to interventionist theories of causation, causal relations can be analyzed in terms of systematic interventions. Since it is straightforward, intuitively comprehensible, and probably the most prominent interventionist theory on the market, we will mainly focus on Woodward's (2003) version.² We will only introduce the basic ideas behind interventionism while pointing the reader to the

²For more sophisticated versions of interventionist approaches see, for example, (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). For how these approaches relate to Woodward's (2003) see, for example, (Gebharder, 2017; Gebharder & Schurz, 2014; Zhang & Spirtes, 2011).

relevant passages in (Woodward, 2003) for details.

Interventionism is first and foremost a theory about causation at the type-level, meaning that it is more about the causal relations between random variables (e.g., smoking behavior is causally relevant for whether lung cancer occurs) than the causal relations between specific token-level events (e.g., Pete's smoking 20 cigarettes a day from 1990 to 2000 caused his lung cancer in 2002). At the core of the interventionist framework lie three interdependent concepts: direct causation, contributing causation, and intervention. The first two of these concepts are defined w.r.t. specific sets of variables \mathbf{V} . In a nutshell, direct causation could be defined as follows (cf. Woodward, 2003, p. 59):

(DC) X is a direct cause of Y w.r.t. \mathbf{V} if and only if (iff) there are interventions on X that are associated with changes in Y if the values of all other variables Z_i in \mathbf{V} are fixed by additional interventions.

Based on **(DC)**, the basic idea underlying contributing causation can be expressed as follows (cf. Woodward, 2003, p. 59):

(CC) X is a contributing cause of Y w.r.t. \mathbf{V} iff there is a chain (or path) of direct causal relations from X to Y and an intervention on X that is associated with changes in Y if the values of all variables Z_i in \mathbf{V} not lying on this particular path are fixed by additional interventions.

Note that both **(DC)** and **(CC)** speak about interventions. An intervention is a causal notion. The basic idea (for details, see Woodward, 2003, p. 98) is that the interventions on X in both **(DC)** and **(CC)** are causes of X that can influence Y only over chains of directed causal arrows going through X . If such an intervention X then leads to a change in Y while the values of relevant other variables are kept constant, this change in Y can only be due to the fact that

X is a (direct or contributing) cause of Y .

In this paper, we are more interested in causal relations at the token-level than at the type-level. Token-level causation is often discussed under the label of “actual causation”. The literature features different proposals for how to best analyze actual causation. Here, a simple version of actual causation (cf. [Woodward, 2003](#), p. 77) suffices:³

(AC) $X = x$ is an actual cause of $Y = y$ iff

- (i) the actual value of $X = x$ and the actual value of $Y = y$, and
- (ii) there is at least one route from X to Y for which an intervention on X will change the value of Y , given that other direct causes Z_i of Y that are not on this route have been fixed at their actual values.

4 The causal theory of knowledge revisited

In this section, we reconstruct (CTK) and the luck problem using interventionist resources. We then use these resources to add a requirement to (CTK) that allows us to overcome the luck problem. We will use the fake barn case for illustration.

First, we translate (CTK) into the interventionist framework:

(CTK_I) s knows that $X = x$ iff

- (i) $X = x$ is an actual cause of $B_s(X = x)$, or

³This basic version is subject to a range of different problems arising in cases involving preemption and overdetermination. A more sophisticated account of actual causation is, for example, ([Halpern & Pearl, 2000](#)).

- (ii) there is an actual common cause $Z_i = z_i$ of $X = x$ and $B_s(X = x)$.

Conditions (i) and (ii) reflect the requirement that $X = x$ and $B_s(X = x)$ have to be causally connected in an appropriate way. According to Goldman (1967), this requires that (i) $X = x$ causes s to believe $X = x$ or if (ii) both $X = x$ and s 's belief in $X = x$ have a common cause.

Some clarification is in order. Requiring $X = x$ to be an actual cause of $B_s(X = x)$ implies that $X = x$ and $B_s(X = x)$ are true; and likewise for $Z_i = z_i$ if condition (ii) applies. Also note that **(CTK_I)** provides truth conditions for knowledge about particular token-level events or facts only. It does not provide truth conditions for knowledge about causal relations themselves, type-level generalizations (such as laws of nature), or mathematical facts. Causal theories—ours as well as Goldman's (1967)—apply only to entities that can stand in causal relations. We do not claim that one cannot have knowledge of other kinds of entities, but rather that such knowledge will require a different explanation. Here, we focus on empirical knowledge. How causal knowledge and knowledge that a generalization holds relate to empirical knowledge, are central topics investigated in general philosophy of science (see, e.g., Schurz, 2013).

Let us restate the luck problem in terms of **(CTK_I)** and interventionism. Assume Henry (s) is entering fake barn village. There is one real barn and there are 79 fake barns in fake barn village. Henry sees an object at a certain location l_1 and as it happens, this object is a real barn and the presence of this object at l_1 ($X = x$) causes Henry to believe that there is a barn at location l_1 ($B_s(X = x)$). As it happens, this belief is true. In our interventionist framework the presence of a barn at location l_1 ($X = x$) is an actual cause of Henry's belief ($B_s(X = x)$) because putting a barn there by intervention vs. not putting a barn there or putting a completely different object there makes a difference for

whether Henry holds the particular belief $B_s(X = x)$. However, we intuitively would not want to say that Henry knows that there is a barn at location l_1 . The reason is that given the circumstances—i.e., that Henry is in fake barn village—he holds the true belief $B_s(X = x)$ only luckily. In interventionist terms this means that $X = x$ did not make that much of a difference for Henry’s belief given the specific circumstances. If the barn and the fake barns were only slightly differently arranged in fake barn village, then one of the 79 fake barns would have been at location l_1 and the presence of a fake barn ($X = x'$) rather than the presence of a real barn ($X = x$) would have caused Henry’s belief ($B_s(X = x)$). In that case, Henry would have hold a false belief. So although $X = x$ is the actual cause of Henry’s true belief $B_s(X = x)$, Henry holds a true belief only luckily. Slight variations in the actual circumstances would have resulted in a different cause $X = x'$ of Henry’s (now false) belief $B_s(X = x)$.

How can we fix the luck problem? Interventionism allows for two kinds of questions about causal relations. The first is answered by the core theory presented in [section 3](#). It is about whether a particular causal relation between two variables holds or does not hold. The second asks for further features of causal relations such as how specific, proportional, or stable the cause is w.r.t. its effect.⁴ These characteristics are typically defined for type-level causal relations. However, modified versions may apply to cases of actual causation too.

In particular, we take inspiration from the concept of stability for type-level causal relations and introduce notions of *lucky and non-lucky causation*. According to [Woodward \(2010\)](#), stability is about whether and to what extent the pattern of counterfactual dependence of the effect variable’s values on the cause variable’s values would change in different background circumstances. So

⁴For a detailed discussion of these features of causal relations see, for example, ([Woodward, 2010](#)).

stability requires that something—the pattern of counterfactual dependencies—keeps relatively stable while something else—the background circumstances—changes. Note that stability comes in degrees; a causal relation might be more or less stable depending on how much the counterfactual pattern between the cause and effect variable’s values would change when changing background conditions. It also depends on the range of such changes that would have an influence on this pattern. It is possible to introduce a yes-or-no version of stability: A causal relation is stable if the pattern of counterfactual dependencies between cause and effect values does not or only minimally change in a range of changes in background conditions that exceeds a threshold, and unstable otherwise.⁵

Inspired by stability, we propose the following yes-or-no token-level notion of *lucky causation*:

(LC) $Y = y$ is luckily caused by $X = x$ iff

- (i) $X = x$ is an actual cause of $Y = y$, and
- (ii) under most small variations of the actual circumstances a different value x' of X would be an actual cause of $Y = y$.

First of all, $X = x$ must actually have caused $Y = y$ in order to have caused $Y = y$ luckily (i), which implies, among other things, that both $X = x$ and $Y = y$ actually happened. But according to condition (ii) there are other X -values x' that would also have caused $Y = y$ if X had taken one of these values. However, given the actual circumstances, x and none of these other values x' caused $Y = y$. $X = x$ caused $Y = y$ luckily if most slight changes in the actual circumstances

⁵Note that the threshold may be context-dependent. Because of this, one might doubt that this version of stability is really a yes-or-no concept. We would like to thank an anonymous reviewer for this point. However, once a given context is fixed, the causal relation in question will either be stable or unstable, depending on whether the relevant threshold for that context is exceeded.

would also have lead to $Y = y$, but not because $Y = y$ would have been caused by x , but rather by one of these other x' . In this sense the specific actual cause $X = x$ was quite lucky— $Y = y$ could easily have been caused by other X -values x' .

Based on **(LC)** we can now say when $Y = y$ has been caused by $X = x$ non-luckily:

(NLC) $Y = y$ is non-luckily caused by $X = x$ iff $X = x$ is an actual cause of $Y = y$, but $Y = y$ is not caused luckily by $X = x$.

Here the similarities between non-lucky causation and stability become clear: If $X = x$ is a non-lucky cause of $Y = y$, then small variations in the actual circumstances would not easily have led to a different cause $X = x'$ of $Y = y$. As in the case of a stable causal relation, the background circumstances vary while something else stays fixed: In the case of stability, the pattern of counterfactual dependencies stays the same, and in the case of non-lucky causation, the X -value x that causes $Y = y$ stays the same.

Next, we use the concept of non-lucky causation to modify **(CTK_I)**:

(CTK_I^{*}) s knows that $X = x$ iff

- (i) $B_s(X = x)$ is non-luckily caused by $X = x$, or
- (ii) there is a common cause $Z_i = z_i$ of $X = x$ and $B_s(X = x)$ such that $Z_i = z_i$ causes $X = x$ as well as $B_s(X = x)$ non-luckily.

Recall that conditions (i) and (ii) in **(CTK_I)** were intended to capture the “appropriate” causal connections in the original **(CTK)**. The new conditions (i) and (ii) in **(CTK_I^{*})** can be understood to further specify the required appropriateness.

(CTK_I^{*}) fixes the luck problem by requiring that s 's belief is caused non-luckily by the relevant facts. To illustrate, consider again that Henry (s) enters

fake barn village, which features 1 real barn and 79 fake barns. Henry looks at location l_1 , where the only real barn is located, and this barn at l_1 ($X = x$) causes Henry to hold the belief that there is a barn at l_1 ($B_s(X = x)$). Thus, $X = x$ is an actual cause of $B_s(X = x)$. However, since there are 79 fake barns in fake barn village, $X = x$ is a lucky cause for $B_s(X = x)$. Under most slight changes of actual circumstances—i.e., if the barns and fake barns would be arranged slightly differently—not $X = x$, but the fact that there is a fake barn at location l_1 ($X = x'$) would have caused $B_s(X = x)$. Since $X = x$ is a lucky cause of $B_s(X = x)$ in the fake barn case, **(CTK_f^{*})** implies that Henry does not know that there is a barn at location l_1 ($X = x$), which is the desired result.

Contrast now the fake barn case to a case in which Henry enters a village featuring 79 real barns and only 1 fake barn. Again, there is a real barn located at l_1 ($X = x$), Henry looks at l_1 and the fact that there is a barn at l_1 causes his belief that there is a barn at l_1 ($B_s(X = x)$). The only difference to fake barn village is that under almost all variations of the actual circumstances it would be $X = x$ that causes $B_s(X = x)$. This means that the role $X = x$ has in causing $B_s(X = x)$ could not have been replaced easily by another potential cause $X = x'$. $X = x$ is a non-lucky cause of $B_s(X = x)$ and, thus, in this case **(CTK_f^{*})** leads to the desired consequence that Henry knows that there is a barn at location l_1 .

One may object that our interventionist account is not doing much work; lucky and non-lucky causation could potentially be defined (i) on the basis of an informal notion of causation or (ii) to the background of another available theory of causation. The problem with (i) is that an informal notion of causation does not provide us with clear criteria for whether the relevant causal relation holds in a given situation. Consequently, we would be unable to assess whether an agent possesses knowledge in any given situation. Our approach draws on the

well-defined interventionist concept of causal stability to produce a clear verdict about whether Henry knows. The problem with (ii) is that there are many theories of causation on the market, none of which currently features a clear analogue of the notion of lucky causation. Our point is not that such analogues cannot be developed, but rather that more work needs to be done to carefully investigate each candidate theory. One needs to check how lucky and non-lucky causation can be spelled out within each theory of causation and which results it would give us for knowledge. The present paper may be seen as one part of this much larger project.

5 Classifying various forms of epistemic luck

In [section 4](#) we argued that (CTK_7^*) correctly handles fake barn cases. This is already a significant advantage of our approach over Goldman's (1967) original account. One may worry, however, that our account might be *ad hoc* in so far as it is designed specifically to solve this problem. In this section, we defend our account from this charge by showing how it provides a general explanation of the relation between knowledge and luck. As suggested in [section 2](#), our account provides a principled interpretation of the notion of causation appropriate for knowledge, one that correctly rules out problematic forms of luck but that can also explain why certain other forms of luck are compatible with knowledge.

Forms of luck compatible with knowledge are standardly called *benign*, and forms incompatible with knowledge *malicious* forms of luck. The principal benign form of luck is called evidential luck.⁶ Evidential luck is the kind of luck at issue when an agent luckily comes into possession of a given piece of evidence, say, a reliable encyclopedia, on the basis of reading of which she forms the belief

⁶The standard locus for the terms evidential and veritic luck is [Engel \(1992\)](#). See also [Pritchard \(2005\)](#).

that World War II ended in 1945. There is a sense in which this belief is lucky, since easily the agent could have failed to find the evidence and thus easily she could have failed to form the relevant belief. However, as long as she appropriately responds to her reliable evidence, many maintain that she should be able to acquire knowledge in such cases (Engel, 1992).

Crucially, our account accommodates this verdict. Let $(B_s(X = x))$ stand for our subject's belief that World War II ended in 1945. Now the fact that World War II ended in 1945 ($X = x$) was an actual cause of someone writing the corresponding encyclopedia entry, which was, in turn, an actual cause of our subject forming her belief $(B_s(X = x))$ when reading it. Note that $X = x$ is, again, a non-lucky cause of $B_s(X = x)$ since under most small variations of the actual circumstances no other X -value x' (such as World War II ending in 1946, in 1947, etc.) would have been an actual cause of our agent's belief $B_s(X = x)$. Under most small variations of the actual circumstances she would not have found the encyclopedia at all and, thus, none of these other X -values x' would have caused her to believe that World War II ended in 1945. But even in those circumstances in which she would have found the encyclopedia, none of these other X -value x' would have caused her belief that World War II ended in 1945. In that case she would not believe that World War II ended in 1945 at all, but rather that it ended in 1946, in 1947, etc. As a result, our account nicely accommodates the idea that evidential luck is compatible with knowledge.

We now turn to two forms of malicious luck: environmental and intervening luck. These are sub-species of veritic luck in that in both cases it is true that the agent's belief-forming method could have easily generated a false belief. But in the case of environmental luck, the belief is causally connected to the fact believed, whereas in the case of intervening luck, no such causal connection is present. Fake barn cases are cases of environmental luck: The agent's belief

$B_s(X = x)$ is actually caused by the relevant fact $(X = x)$ (e.g., there being a barn over yonder), but the causation in question is lucky causation because slight differences to the environment would result in a different fact $X = x'$ (e.g., there being a fake barn over yonder) causing $B_s(X = x)$. (\mathbf{CTK}_I^*) entails that $B_s(X = x)$ fails to qualify as knowledge in such cases, which is as it should be.

What about intervening luck, the kind of luck present in original Gettier cases? Take the case of Jones, who pretends to own a Ford in such a convincing way that Smith comes to believe she has a Ford. As it happens, Jones did not own a Ford and only inherited one right before Smith formed her belief. As the overwhelming majority of epistemologists agrees, this is not a case of knowledge.⁷ (\mathbf{CTK}_I^*) correctly yields this result because this is not a case of non-lucky causation simply because it is not a case of causation at all. The fact that Jones owns a Ford $(X = x)$ that makes Smith's belief to this effect $(B_s(X = x))$ true is not causally connected at all to her belief.

Let us briefly summarise: Our account correctly classifies the main forms of luck in the literature. This suggests our account is not an *ad hoc* solution to a particular problem, but that it captures the relation between knowledge and luck generally. Moreover, our account provides a unified answer to the questions *why* some forms of luck are compatible with knowledge whereas others are not: Knowledge is non-luckily caused belief. This is a clear advantage over extant discussions of epistemic luck, which tend to rest on brute intuitive verdicts about the (in)compatibility of knowledge and various forms of luck.

6 Similar accounts of knowledge

We have shown so far how our interventionist account solves the fake barn problem and correctly classifies various forms of epistemic luck. In this section,

⁷For some exceptions, see (Weatherson, 2003; Hetherington, 2011).

we contrast our account to competitors. In particular, we discuss Goldman's (1976) own *perceptual equivalence account* (**PEA**), as well as recent safety-based accounts and highlight some advantages of our approach.

(PEA) can be summarized as follows:

(PEA) *s* noninferentially perceptually knows of object *b* that it has property *F* iff

- (1) *b* has property *F*, and
- (2a) *b*'s having *F* noninferentially causes *s* to believe of object *b* that it has property *F*, and
- (2b) there is no alternative state of affairs featuring object *c* that is a relevant perceptual equivalent for *s* relative to property *F* where *c* does not have property *F*.

Goldman's (1976) claim is that in fake barn cases, condition (2b) is violated because the barn facades count as relevant perceptual equivalents for the believing agent relative to the property of being a barn. Since these facades do not have that property, condition (2b) fails. Like **(CTK₇^{*})**, **(PEA)** recognizes that alternative states of affairs where different facts cause *s* to have the same (but now false) belief undermine knowledge.

(CTK₇^{*}) is preferable, however, for several reasons. First, Goldman's (1976) addition of (2b) is somewhat *ad hoc* and disconnected from his causal condition (2a). Alternatively, **(CTK₇^{*})** handles fake barn cases quite naturally by utilising resources drawn directly from a well-established theory of causation. As such, our account stays closer to the core of the original causal account. Second, Goldman's account is only meant to explain non-inferential *perceptual* knowledge. Our account is more general. For example, we can easily explain testimonial knowledge. Suppose Mr. X tells you that you have left the stove

on, and that you believe him. Our account explains that this belief amounts to knowledge if your belief is non-luckily caused by the stove being on. Perhaps (PEA) can be extended to accommodate such cases, but as it stands, our account has an advantage here. Third, Goldman assumes that only perceptual equivalents undermine knowledge. We allow cases in which non-perceptual equivalents undermine knowledge. Consider Cathy, who strangely thinks all cats are dogs. Suppose Cathy is in an environment where there are many cats but only one dog, Dach the Dachshund. She happens to look at Dach and forms the belief that there is a dog there. For Goldman, this belief would constitute knowledge since there are no perceptual equivalents that would cause Cathy to form the same belief. On the other hand, we maintain this belief is disqualified from knowledge precisely because in most variations of the actual circumstances a cat would have caught Cathy's eye and produced the same belief that there is a dog. We submit people who think cats are dogs simply do not know what a dog is, and therefore are precluded from knowing that Dach the Dachshund is a dog.

While we agree that taken separately, these considerations may not provide knock-down arguments against Goldman (1976), we think together they make a compelling case for the superiority of our approach. In any case, they serve to contrast Goldman's way of addressing the fake barn problem with ours.

A second contrast may be drawn between our approach and recent safety based approaches to knowledge (Sosa, 1999; Williamson, 2000; Pritchard, 2005). Safety theorists maintain that knowledge is incompatible with false beliefs in nearby possible worlds, provided those beliefs are produced in the same way as the relevant belief is in the actual world. Thus, standard Gettier-cases are cases of unsafe belief, since in those cases the subject could have easily formed a false belief in the same way as she actually formed her belief. In our example, Henry

could have easily formed a false belief about there being a barn over yonder by looking at one of the fake barns. So, safety-based approaches have no problem with accounting for the kind of luck at issue in Gettier cases, including those involving environmental luck.

However, safety-based approaches do have a problem with accommodating the idea that knowledge requires a certain direction of fit between one's belief and the world (Vogel, 2017). Suppose you believe that the number of stars is even, purely on the basis of guessing. Even if true, this belief should not amount to knowledge. In normal cases, safety-based approaches can accommodate this verdict, since in normal cases such guesses may easily produce a false belief; we assume that you may just as easily have guessed that the number of stars is uneven. Now suppose, however, that unbeknownst to you, a guiding angel is helping you, such that she creates a star whenever your guess is wrong, such that after her intervention, your belief is bound to be true. If such an angel exists not only in the actual world, but also in nearby worlds, then your belief will be safe, since not easily will you form a false belief in this way. But such beliefs should not qualify as knowledge. While safety theorists thus need to add conditions to their account of knowledge to maintain proper direction of fit (Pritchard, 2012), it is an advantage of causal accounts generally, the present one included, that they are able to capture the required direction of fit quite naturally. Both variations of the above example fail to satisfy (CTK_I^*) . So whereas safety alone turns out to be insufficient for knowledge, a non-lucky causal condition seems to do better.

7 Conclusion

While Goldman's (1967) causal theory of knowledge accommodates Gettier cases involving intervening luck, it fails to rule out cases of environmental luck. In

this paper, we have proposed a fix. In particular, we reformulated the causal theory of knowledge within an interventionist framework and drew a distinction between lucky and non-lucky causation inspired by the interventionist notion of causal stability. We then proposed to read the “appropriate” causal connection in Goldman’s original account as non-lucky causal connection, and showed that this appropriately rules out cases of environmental luck. We further argued that our interventionist causal theory of knowledge is not simply an *ad hoc* solution to fake barn cases, but that it allows for a unified explanation of the relation between knowledge and different types of epistemic luck. Finally, we compared our account to two other prominent proposals that have been made in order to avoid problems with fake barn cases.

There remain, of course, some limitations and open questions. A causal theory of knowledge only applies to entities that can stand in causal relations. In the future it needs to be shown how our account relates to knowledge of causal relations and type-level generalizations such as laws of nature or scientific hypotheses and theories. Also, in this paper we specifically endorsed an interventionist theory of causation. As we saw in [section 4](#), it is still an open question whether other recent accounts of causation might also be able to do the job. In the future, we hope to provide a more thorough defence of (CTK_I^*) . Here, we rest content with motivating a cautious reconsideration of the causal approach to knowledge.

Acknowledgements: We would like to thank an anonymous reviewer for helpful comments and suggestions.

References

- Carter, J. A., & Pritchard, D. (2015). Knowledge How and Epistemic Luck. *Noûs*, 49(3), 440–453.
- de Grefte, J. (2018). Epistemic justification and epistemic luck. *Synthese*, 195(9), 3821–3836.
- Engel, M. (1992). Is Epistemic Luck Compatible with Knowledge? *The Southern Journal of Philosophy*, 30(2), 59–75.
- Gebharder, A. (2017). *Causal nets, interventionism, and mechanisms*. Cham: Springer.
- Gebharder, A., & Schurz, G. (2014). How Occam’s razor provides a neat definition of direct causation. In J. M. Mooij, D. Janzing, J. Peters, T. Claassen, & A. Hyttinen (Eds.), *Proceedings of the UAI workshop Causal Inference: Learning and Prediction*. Aachen.
- Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, 23(6), 121–123.
- Goldman, A. I. (1967). A causal theory of knowing. *Journal of Philosophy*, 64(12), 357–372.
- Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, 73(20), 771.
- Halpern, J. Y., & Pearl, J. (2000). Causes and explanations: A structural-model approach, Part I: Causes. *arXiv.org*.
- Hetherington, S. C. (2011). *How to Know: A Practicalist Conception of Knowledge*. John Wiley & Sons.
- Pearl, J. (2000). *Causality* (1st ed.). Cambridge: Cambridge University Press.
- Pritchard, D. (2005). *Epistemic luck*. New York, NY: Oxford University Press.
- Pritchard, D. (2012). Anti-Luck Virtue Epistemology. *Journal of Philosophy*, 109(3), 247–279.

- Schurz, G. (2013). *Philosophy of science: A unified approach*. New York: Routledge.
- Sosa, E. (1999). How to Defeat Opposition to Moore. *Nous*, 33(s13), 141–153. doi: 10.1111/0029-4624.33.s13.7
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* (1st ed.). Dordrecht: Springer.
- Unger, P. (1968). An analysis of Factual Knowledge. *The Journal of Philosophy*, 65(6), 157–170.
- Vogel, J. (2017). Accident, Evidence, and Knowledge. In *Explaining knowledge*. Oxford: Oxford University Press. Retrieved from <http://www.oxfordscholarship.com/10.1093/oso/9780198724551.001.0001/oso-9780198724551-chapter-7> doi: 10.1093/oso/9780198724551.003.0007
- Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, 115(1), 1–31.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford, NY: Oxford University Press.
- Woodward, J. (2003). *Making things happen*. Oxford: Oxford University Press.
- Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy*, 25(3), 287–318.
- Zhang, J., & Spirtes, P. (2011). Intervention, determinism, and the causal minimality condition. *Synthese*, 182(3), 335–347.