Episteme (2026), 1–14
doi:10.1017/epi.2026.10165

episteme

ARTICLE

A Modal Explanationist Solution to the Gettier Problem

Ying Zhang and Haicheng Zhao

Department of Philosophy, Xiamen University, China
Corresponding author: Ying Zhang; Email: 10420250158062@stu.xmu.edu.cn

(Received 28 November 2025 UTC; revised 19 August 2026 UTC; accepted 13 September 2026 UTC)

Abstract

The Gettier problem suggests that the core challenge for theories of knowledge lies in eliminating knowledge-undermining epistemic luck. In response, two major approaches have emerged in contemporary epistemology: modalism and explanationism. However, both approaches have limitations and exhibit an interesting theoretical symmetry: while modalism effectively handles environmental luck, it struggles with certain cases of intervening luck; explanationism faces the reverse difficulty. This symmetrical limitation suggests that any theory of knowledge relying on a single dimension is likely to be inadequate. To address this, this paper presents a new, integrated approach – Modal Explanationism (ME). ME requires that a belief counts as knowledge just in case it is explained by the truth of the believed proposition in the actual world, and that this explanatory connection is preserved across a range of nearby possible worlds. It is argued that ME offers a more successful solution to the Gettier problem than either modalism or explanationism alone.

Keywords: Gettier problem; epistemic luck; modalism; explanationism

1. Introduction

Gettier cases reveal that a justified true belief can fail to constitute knowledge when its truth is a matter of luck. Any adequate theory of knowledge must therefore eliminate epistemic luck that is incompatible with knowledge. Pritchard (2016) characterizes such knowledge-undermining luck as "veritic luck" and divides it into two structurally distinct species: intervening luck and environmental luck.

Intervening luck is the kind of luck that permeates paradigmatic Gettier cases, such as Gettier's (1963) own cases. In such cases, "bad luck severs the connection between justification and true belief, only to be counteracted by compensating good luck."[1] Consider Gettier's (1963) original "job-seeker case." Smith forms the belief that *the man*

[1]This formulation of intervening luck is from Carter and Kallestrup (2026: 10). See also Carter and Pritchard (2015).

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

who will get the job has ten coins in his pocket on the basis of strong evidence suggesting that Jones will get the job and has ten coins. Ultimately, Smith himself gets the job and, unbeknownst to himself, happens to have ten coins in his pocket. Smith's belief is true and justified, but fails to be knowledge. Here, bad luck severs the connection between Smith's justification and true belief, such that it is not by virtue of his justification that he forms a true belief. However, another stroke of good luck – which involves the fact that Smith himself gets the job and has ten coins – intervenes, so that his belief is true anyway.

By contrast, environmental luck places the subject in an epistemically hostile and misleading environment such that, although there is no disruption of the justification-truth connection, the subject could very easily have formed a false belief. The well-known barn facade case is illustrative of this kind of luck (Goldman 1976). Barney forms a true belief that he sees a barn while unknowingly situated in a region filled with barn facades – each of which looks exactly like a real barn but is not. He forms a justified true belief, but given the hostile environment, he acquires it only luckily. Here, nothing intervenes between the subject's justification and his true belief. Unlike the job-seeker case, it is by virtue of Barney's (perceptual) justification that he forms a true belief. The knowledge-undermining luck is purely environmental.

Two major responses to Gettier cases have emerged in the post-Gettier literature. Modalism, exemplified by the safety principle, grounds knowledge in counterfactual stability. The idea is that for a belief to constitute knowledge, that belief (or a set of relevant beliefs) must be stably true in a range of nearby possible worlds. Explanationism, by contrast, demands an appropriate explanatory connection between belief and fact in the actual world. We argue, however, that the two approaches exhibit an interesting theoretical symmetry in their respective strengths and limitations. Modalism effectively screens out environmental luck but faces difficulties with certain cases of intervening luck, since a belief involving intervening luck can nevertheless be highly modally stable. By contrast, explanationism offers a straightforward diagnosis of intervening luck but faces difficulties in cases of environmental luck.

These complementary limitations suggest that neither modal stability nor actual explanatory connection, taken on its own, adequately captures all forms of knowledge-undermining luck. We thus propose an integrated alternative: Modal Explanationism (ME). Its core idea is that a belief counts as knowledge just in case it is explained by the truth of the belief in the actual world, and that this explanatory connection is preserved across a range of nearby possible worlds. The main thesis of this paper is comparative: we argue that ME fares better than modalism and explanationism in addressing Gettier cases, and therefore enjoys key theoretical advantages over both accounts.[2]

The paper proceeds as follows. Section 2 introduces modalism and diagnoses its difficulty in dealing with cases of intervening luck. Section 3 turns to explanationism and argues that it faces trouble in handling cases of environmental luck. Section 4 presents and motivates ME. Section 5 addresses Gettier cases that involve deviant causal chains. Section 6 compares ME with virtue epistemology and argues that ME avoids the credit dilemma for virtue epistemology.

[2]Thus, we do not undertake a full-fledged defense of ME – that goes far beyond the scope of this paper. Notice also that even if ME inherits some problematic features of the accounts it draws on (there are well-documented objections against modalism and explanationism that go beyond Gettier cases), this does not affect the present comparative thesis. Our aim is to show that ME fares better than these accounts in handling Gettier cases, not that it remedies all of their theoretical drawbacks.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

2. The modalist solution to the Gettier problem

The core idea behind modalism is that knowledge cannot be accidentally true belief; there must be a stable modal connection between belief and truth. Within this theoretical family, the safety principle has gained much popularity:

> **Safety Principle:**
>
> If S knows that p via method M, then S could not easily have formed a false belief via M. (see e.g., Sainsbury 1997, Pritchard 2012, Zhao 2024)

At first glance, the **Safety Principle** seems to offer plausible diagnoses of both environmental and intervening luck. With respect to cases of environmental luck, Barney's belief that *he sees a barn* is unsafe: in many nearby possible worlds where his gaze shifts slightly, he sees a fake barn and forms a false belief by the same visual method. Such an easy possibility of error makes his belief unsafe and thus excludes it from knowledge. Turning to cases of intervening luck, Smith's belief that *the man who will get the job has ten coins in his pocket* is also unsafe. We can imagine many nearby possible worlds in which Smith still gets the job but happens not to have ten coins. In these worlds, his belief, formed via the same method, would be false. By attending to these counterfactual possibilities, the **Safety Principle** seems to capture epistemic luck involved in Smith's belief.

There are, however, more recalcitrant cases of intervening luck. Consider the following:

> **Blood Type:**
>
> John lives in a remote community with a strong medical-privacy culture. The local medical center has installed a self-service blood-typing machine housed in a completely enclosed, strictly one-person-at-a-time private booth. Owing to local custom and the layout of the medical center, residents have virtually no way of learning anyone else's blood type from this machine. John has good reason to believe the machine is reliable. Thus, having never known his blood type, he decides to test himself. Unbeknownst to John, however, the machine was programmed to always deliver just one result: blood type B. Coincidentally, John's blood type is B. So he ends up with a true belief. (adapted from Zhao 2026: 465)

On a natural reading of the case, John's belief appears safe. On the one hand, given how the machine has been set up, in all sufficiently close possible worlds it outputs only "Type B." On the other hand, it is a modally stable biological fact that John's blood type is B and, given the local culture, he could not easily have formed beliefs about other people's blood types. Together, these facts ensure that, by consulting this machine, John could not easily have formed a false belief. But intuitively, John does not know that his blood type is B. His true belief results from a coincidence: the machine is programmed to output "Type B" regardless of the user's actual blood type, while John independently happens to have blood type B.

A modalist might argue that the safety principle should be supplemented with a more refined account of which possible worlds are relevant to the safety assessment. One influential way of refining safety appeals to normality. On a normality-based conception, safety is assessed not simply across nearby worlds, but across normal possible worlds:

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

Normality-Based Safety:

S knows that p only if S's belief that p is true not only in the actual world but also in normal possible worlds in which S forms the belief via the same method as in the actual world. (Smith 2010, 2016; Dutant 2016; Beddor and Pavese 2020)

A defender of this view might offer the following reply to the blood type case. There may be normal worlds in which the machine malfunctions in a different way, producing a blood type result other than "Type B." In those worlds, John forms a false belief. His belief then fails to be safe.

However, normality depends at least partly on the actual states of affairs (see Valaris 2022, Zhao 2025). In the present case, the actual states of affairs can be easily stipulated to block the above response. In particular, suppose that the manufacturer of the blood type machine has a strong interest in impressing the hospital director, whose blood type is B, and so deliberately programs the machine to produce only Type B results. Under this stipulation, worlds in which the machine outputs a different result appear abnormal – or at least more abnormal than the actual world in which the machine produces only Type B results. This verdict is well supported by Smith's (2016) influential account of normality, according to which abnormal events are those that call for special explanation. Given the manufacturer's interest and the machine's actual design, a deviation from the "Type B" output would itself require some further explanation. Such worlds may therefore be excluded from the relevant normal worlds, in which case John's belief once again satisfies the safety condition.

Blood type exposes a burden that a purely modal approach must bear.[3] Whether John's belief counts as safe depends on which counterfactual possibilities are considered as relevantly close or normal. If worlds in which the machine outputs something other than "Type B" are included, John's belief may be unsafe; if they are excluded, the belief may remain safe. Modalism therefore requires a principled account of relevant-world selection that delivers the appropriate verdict across such cases. Blood type illustrates how difficult this task can become when the truth of a belief results from two modally stable but explanatorily independent features of the actual situation. In this sense, Blood type reveals a genuine difficulty for a purely modal treatment of intervening epistemic luck.

3. The explanationist solution to the Gettier problem

In recent years, explanationism has gradually developed into a rival to modalism.[4] In contrast to modalism, explanationists reject the move of grounding knowledge in a

[3]A more resolute modalist might simply bite the bullet and accept that sufficiently modally stable coincidence is compatible with knowledge. Clarke-Doane's (2012) response to evolutionary debunking arguments in ethics suggests a position of this general kind: the modal stability of the match between relevant beliefs and corresponding facts may be taken to suffice for non-accidentality. Applied to Blood Type, a modalist of this sort might therefore accept that John knows that his blood type is B. But such a theorist would then owe us a non-question-begging explanation of why John's belief is not knowledge-underminingly lucky, despite the fact that the machine's output is explanatorily independent of John's actual blood type. Merely insisting that the match is modally stable does not seem to suffice. We thank an anonymous reviewer for drawing our attention to this possibility.

[4]The origins of explanationism can be traced back to Alan Goldman's work in the 1980s (see Goldman 1982). After a period of relative dormancy, the approach has been revived and defended in recent years by scholars including Jenkins (2006), Faraci (2019), Bogardus and Perrin (2022, 2025), Zhao (2025).

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

counterfactual relation between belief and truth. Rather, they insist on an appropriate explanatory relation between belief and truth in the actual world. Explanationists capture this thought in the slogan "knowledge is believing something because it's true" (Bogardus and Perrin 2022). More precisely:

Explanationism:

S's belief that p is knowledge if and only if the truth of p constitutes the core explanation of S's holding that belief.[5]

This immediately raises a question: what does it mean for the truth of p (or the fact that makes p true) to constitute a *core* explanation of one's belief that p? In response, explanationists appeal to the *salience* and *indispensability* of the explanation: the truth-making fact must be the irreplaceable, pivotal factor in the explanatory chain.

More specifically, Bogardus and Perrin (2022) adapt the kairetic test developed by Strevens (2011) in the philosophy of science to give this idea precision. The procedure works by progressive abstraction: beginning with a full set of potential explanantia for the belief, one strips away potential explanantia in the set until further abstraction would make the explanation of the belief fail. The remaining explanantia in the set are the difference-makers. Consider a subject who knows it is raining outside through an abundance of perceptual experience: she sees raindrops on the window, hears rain on the roof, and feels the moisture in the air. The abstraction procedure strips away these details one by one. As such, we can abstract away any particular perceptual modality – whether she saw the rain, heard it, or felt it – and a recognizable explanation of her belief still stands. But once we remove the fact that it is actually raining outside, the explanation collapses entirely. That fact is indispensable for the explanation of why she holds the belief. It is precisely this indispensability, revealed through the abstraction process, that qualifies the fact that it is raining as the core explanatory item.

It is also worth noting that the explanatory relation invoked by explanationists is broader than causal explanation. This is important when dealing with non-empirical knowledge. Take mathematical or logical truths. The abstract fact that $2 + 2 = 4$ does not "bring about" our belief through a causal process. So if the conditions for knowledge were strictly confined to causal explanations, it would be impossible to account for how subjects acquire knowledge of this kind. The explanationist response is that even in the absence of causal interaction, mathematical facts can be grasped through the subject's rational intuition, and it is the fact itself – apprehended through that rational intuition – that plays the core explanatory role in belief formation (Faraci 2019, Bogardus and Perrin 2022). The same holds for future facts not yet realized. Although future facts cannot, in any temporal sense, be "causes" of present beliefs, a subject can still apprehend them through rational intuition and/or reliable inference, allowing them to figure as core explanatory items that account for our beliefs about the future (see Bogardus and Perrin 2022, Zhao 2025).

The explanationist framework is well suited to diagnosing Gettier cases that involve intervening luck. In the blood type case, where a modalist diagnosis depends on how the relevant counterfactual possibilities are selected, explanationism's verdict is rather straightforward: the immediate explanation of John's belief that his blood type is B is the machine's output, but – given the setup of the case – this output has nothing to do with

[5]See the references in note 4.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

John's actual blood type. That is, the output is not at all explained by John's actual blood type. As such, there is an explanatory break between the belief and the fact. Consequently, John does not know. Likewise, in the classic job-seeker case, the core explanation for Smith's belief involves his mistaken assumption that Jones will get the job and his observation that Jones has ten coins in his pocket, not the fact that the person who gets the job (Smith himself) has ten coins in his pocket. By requiring the truth-making fact itself to constitute the salient explanation for why the belief is held, explanationism effectively excludes intervening luck.

However, this focus on the actual-world explanatory chain also becomes explanationism's Achilles' heel when it comes to cases of purely environmental luck. In the classic barn case, Barney's belief seems to satisfy the explanationist requirement: what explains why he believes that he sees a barn is exactly the fact that there really is a barn. The explanatory chain is wholly intact, so that the explanationist seems forced to conclude that Barney possesses knowledge.

Faced with this objection, Bogardus and Perrin (2022) argue that the genuine core explanation in the barn facade case should encompass the subject's overall environmental information – namely, that the subject is situated in an environment populated with fake barns. They offer a vivid analogy: if a truck scatters nails across the road and the car behind gets a flat tire, what figures crucially into the explanation is the ubiquity of nails on the road, not the presence of the particular nail that caused the puncture (Ibid.: 193). Similarly, in the barn facade case, the difference-maker is the pervasive presence of barn-like objects, not the real barn in front of Barney. On the difference-maker test, even if we abstract away the presence of the single real barn, the explanation of Barney's belief still stands (thanks to the presence of numerous barn facades). So the real barn is not indispensable as an explanans.

We maintain, however, that the explanationists' diagnosis above is untenable. Consider first the following variant of the barn facade case:[6]

Shielded Barn:

> Baum also drives into a county full of barn facades. The local authorities, however, have erected screens that completely conceal all the fake barns. As Baum drives through, he cannot see any fake barns; he can only observe the one real barn in the county, on the basis of which he forms the belief that he sees a barn.

Intuitively, given that the epistemically unfriendly barn facades have been physically shielded, Baum knows that he sees a barn. To reach this verdict, the explanationist seems compelled to grant that the core explanation for Baum's belief is the fact that there really is a barn there. After all, with the fake barns shielded, if the presence of the real barn were abstracted away, nothing would explain why Baum forms the belief.

Now recall the explanationist reply to the original barn facade case: it is the pervasive presence of the fake barns that constitutes the core explanation of Barney's belief. On this basis, Bogardus and Perrin (2022) deny that Barney knows. This diagnosis, however, sits uneasily with the Shielded Barn case. Both Barney and Baum actually perceive the one real barn in the area; both are actually surrounded by fake barns. The two cases share the same features regarding the actual perceptual situation, differing only in one

[6]See Mortini (2024) for a different objection against explanationism's treatment of the barn case, to which Bogardus and Perrin (2025) reply.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

respect: the presence of the screens. Yet the explanationist's verdicts diverge sharply – Barney lacks knowledge, while Baum knows. The explanationist therefore owes us an account of why the screens make this difference.

Obviously, on an intuitive level, the screens make a significant epistemic difference. But for explanationists, they need to provide an answer to the question of why the presence of screens would make a difference to the *explanation* of the subject's belief. The screens stand in no direct causal or constitutive relation to Baum's perception of the real barn: his visual chain runs from the real barn to his retina to his belief, and the screens figure nowhere in this chain.

Why, then, does the presence of screens affect the explanatory chain? Only one answer seems available: the screens matter because they affect what Baum *could have* perceived. Without the screens, Baum's gaze could easily have fallen on a fake barn, and he would have formed a false belief. With the screens in place, however, this possibility is closed off. The screens, in short, do not alter what Baum actually perceives; they alter what he *could have* perceived in other possibilities. And this is surely a modal property: a fact about possible worlds in which the subject's gaze, position, or perceptual conditions vary.

The explanationist's treatment of the two cases therefore tacitly relies on the modal profile that the screens determine. What looks like a difference in core explanation is, on inspection, a difference in modal accessibility. Strip away this modality, and the asymmetry between Barney and Baum collapses. This leaves explanationism with an uncomfortable dilemma. On the first horn, suppose the explanationist lets the actually present fake barns enter the core explanation in both cases. This would correctly diagnose Barney's lack of knowledge, but would undesirably deny Baum's knowledge as well. On the second horn, suppose the explanationist concedes that modal accessibility to the fake barns – as determined by the presence of the screens – does make a difference to the core explanation of the subjects' beliefs. In that case, explanationism diagnoses these cases only by appeal to modal considerations and thus becomes a modal account in disguise, failing to constitute a genuine alternative to modalism. Either way, explanationism cannot plausibly diagnose cases of environmental luck without importing the very modal considerations it was designed to replace.[7]

4. Modal explanationism

The respective strengths and limitations of modalism and explanationism reveal that each captures an important yet incomplete dimension of epistemic non-accidentality. This motivates the search for a framework that integrates the two dimensions rather than treating them as competing alternatives.

We propose such an alternative; consider:

Modal Explanationism (ME):

S's true belief that p is knowledge if and only if: (1) In the actual world, the truth of p is the core explanation of S's holding that belief; and (2) This explanatory connection is maintained in nearly all nearby possible worlds in which S continues to form the belief that p in the same way as in the actual world.

[7]Recent explanationists do intend their accounts to be non-modal accounts and have motivated explanationism as a rival to modalism (see Faraci 2019, Bogardus and Perrin 2022, Zhao 2025).

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

Condition (1) draws on explanationism's explanatory connection requirement between belief and truth, whereas Condition (2) is a modal reinforcement of this connection. ME incorporates both modal and explanatory aspects. However, unlike extant explanationism, ME requires the explanatory connection to hold not only in the actual world but also in counterfactual possibilities. And unlike the modal safety principle, which requires beliefs to be robustly *true* in counterfactual possibilities, ME instead requires the *explanatory connection* demanded by explanationism to continue to hold in relevant counterfactuals.[8]

These two conditions form coherent and complementary evaluative dimensions, jointly defining knowledge as a non-accidental epistemic success. In particular, Condition (1) examines the belief's etiology in the actual world, ensuring its source is traceable to the fact that renders the belief true. Condition (2) examines the modal reliability of this explanatory connection, requiring that it remains robust in counterfactuals where the subject still forms the belief in the same way as in the actual world. By drawing on insights from both modalism and explanationism, ME inherits the theoretical advantages of each while avoiding their major pitfalls in dealing with Gettier cases.

Starting with the barn facade case, Barney's belief satisfies Condition (1) of ME. Unlike Bogardus and Perrin's treatment of this case, ME can maintain the intuitive observation that the fact making Barney's belief true – that there is a real barn before him – is the salient explanation for his belief. The crux is that the belief does not satisfy Condition (2) of ME. Consider the possible worlds in which Barney continues to form the belief that he sees a barn via perception. Given the abundance of fake barns in his cognitive environment, in the vast majority of such worlds Barney would be perceiving a fake barn and thereby forming a false belief. In those worlds, Barney's belief is no longer explained by the presence of a real barn – his perceptual chain is linked to a fake one – and the explanatory connection between truth and belief no longer exists. Condition (2) thus captures the environmental luck involved in Barney's true belief and correctly excludes it from knowledge.

Second, in the blood type case, ME also delivers the desired verdict. John's belief that his blood type is B may exhibit considerable modal stability, at least in the sense that, across the relevant nearby worlds, the machine continues to output "Type B" and John's blood type remains B. The crux, however, is that the fact that John's actual blood type is B plays no explanatory role in his believing that his blood type is B. He forms the belief because the machine produces the Type B result, but the machine produces that result not because it correctly detects John's actual blood type. The belief therefore fails Condition (1). Nor is Condition (2) satisfied: in nearby possible worlds in which John forms his belief in the same way, namely by using the machine, the explanatory connection between the truth-making fact and his belief still does not obtain. This clarifies a comparative advantage of ME. A purely modal diagnosis of the blood type case depends on whether worlds in which the machine produces a different output count as sufficiently close or normal. ME need not settle that question in order to reach a verdict here. Since the fact that John's blood type is B does not explain his belief in the actual world, this actual epistemic defect already suffices, according to ME, to rule out John's belief as knowledge.

[8]ME does not aim to revise existing explanationist criteria for what counts as a core explanation. It takes those criteria over intact and adds a requirement of modal robustness on top of the actual-world explanatory connection. The modal considerations introduced in Condition (2) therefore play no role in determining coreness under Condition (1). We thank an anonymous reviewer for pressing us to clarify this point.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

Zagzebski (1996) famously argues that the crux of the Gettier problem lies in the decoupling of justification and truth, with the gap between them leaving room for knowledge-undermining luck. Traditional epistemology attempts to bridge this gap by refining the justification condition, or by replacing justification with some alternative epistemic condition, but often with limited success. Our suspicion is that this is largely because most theories locate knowledge-undermining luck either in the actual-world etiology of belief or in its counterfactual profile, but not in both. ME addresses the gap at both levels. At the level of the actual world, Condition (1) anchors the source of the belief to the truth-making fact itself, establishing a direct connection between belief and truth and thereby blocking the familiar form of "believe because of A, true because of B" luck. At the counterfactual level, Condition (2) ensures that this explanatory connection is stable and non-accidental. It is precisely this dual grasp of the actual and the counterfactual that gives ME a more comprehensive framework for addressing the Gettier problem.

5. The challenge from deviant causal chains

Since ME inherits the core resources of explanationism, it also inherits a challenge that explanationist theories commonly face: deviant causal chains.[9] We can construct a Gettier case in which the fact that makes the belief true explains the subject's belief via a deviant causal chain, and in which this explanatory relation remains stable across nearby possible worlds. Then both of ME's conditions appear to be met, forcing ME to deliver a knowledge verdict that intuitively seems incorrect. Here is such a case:

> **Sheep:**
>
> Harold sees what appears to be a sheep in a field, and thereby comes to believe that there is a sheep in the field. What Harold sees is actually a sheepdog disguised as a very convincing fake sheep. The sheepdog is present only to keep a real sheep company. Indeed, whenever the disguised sheepdog is in the vicinity of a real sheep, it is because there is a real sheep nearby, though Harold does not know this. In this particular instance, Harold does not see the real sheep, because it is hidden behind the disguised sheepdog.[10]

Intuitively, Harold does not know there is a sheep in the field. Yet his belief seems to satisfy both conditions of ME. Take Condition (1) first: the fact that there is a real sheep in the field explains why Harold believes that there is a sheep in the field, because it explains why the fake sheep – the disguised sheepdog – is there. And the presence of the fake sheep in turn explains why Harold forms his belief. Now consider Condition (2). In nearby worlds where Harold forms the belief in the same general way, he either sees the fake sheep or the real sheep. Given the case's stipulation, even when he sees the fake sheep, the fake sheep is present only because there is a real sheep nearby. Thus, the explanatory connection between truth and belief remains intact across nearby worlds. If Harold sees the fake sheep, the real sheep explains the fake sheep's presence, and the fake sheep's presence explains Harold's belief. If he sees the real sheep, the real sheep's presence explains his belief directly. The explanatory relation between truth and belief

[9]See Bogardus and Perrin (2022, 2025) and Zhao (2025) for further discussions of this problem.

[10]Thanks to an anonymous reviewer for raising this counterexample. This case is a variation on one that Chisholm gives in his (1966: 23). For a similar case, see Miracchi (2015).

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

therefore appears to be modally stable. Thus, one may argue that ME still struggles to handle the Gettier problem.

In reply, the issue lies in how we interpret the spatial constraint implicit in the "keeping company" stipulation of the case. There are two plausible readings of this constraint, and ME can plausibly accommodate both.

On the weak reading, the fake sheep "keeping company" with a real sheep requires only that the real sheep be spatially close to the fake one. It does not require that the two occupy the same strictly demarcated region. To be sure, Harold believes there is a sheep in the field because he sees a very convincing fake sheep, and the fake sheep is there to accompany a real sheep. But notice that, for this explanation, the precise location of the real sheep – whether in the field, on the hillside, or down by the stream – does not really matter. As Strevens (2011) might put it, the real sheep's precise location is not a difference-maker. We can easily imagine, for example, that the fake sheep is at the edge of the field while the real sheep grazes on the adjoining hillside: the two are close enough to satisfy the "company" stipulation, yet strictly speaking the real sheep is not in the field. This means that the real sheep's being *in the field* (rather than on the neighboring hillside) is a purely accidental spatial fact – one that plays no crucial role in the explanation of Harold's belief. In that case, Condition (1) is not satisfied: Harold's belief that there is a sheep *in the field* is not saliently explained by the fact that there is a sheep in the field. ME correctly rules that Harold does not know there is a sheep in the field.[11]

Alternatively, one may read "keeping company" more strictly, as requiring genuine spatial co-location: wherever the fake sheep is placed – in a field, on a hill, or anywhere else – the real sheep must be within that very same strictly demarcated region. On this strong reading, the co-presence of the real and fake sheep in the field is no longer a mere coincidence but is guaranteed by the standing arrangement described in the case. Under this setup, ME does appear to count Harold's belief as knowledge. But this, we think, is not an implausible result. At the very least, it is much less clear that Harold lacks knowledge in this scenario. The driving force behind the Gettier intuition is knowledge-undermining luck – that the subject's belief is true merely by luck or accident. In the traditional, coincidental version of the sheep case, the real sheep's presence behind the decoy is a sheer fluke, which drives our intuitive verdict that Harold lacks knowledge. By contrast, under the strict co-location setup, the real sheep's presence in the same area is no longer a coincidence: the fake sheep is placed there because a real sheep is there, and this relation is maintained across the relevant nearby circumstances. Once this feature is built into the case, the classic Gettier intuition is considerably weakened.

Indeed, under this stipulation, the relation between the fake sheep and the real one resembles the relation in the following case. A person looks at what he takes to be a window and has a visual experience as of a tree outside, thereby coming to believe that there is a tree outside. Unbeknownst to him, however, he is looking at an ultra-high-definition screen that displays, perfectly and in real time, the actual tree standing directly behind it. In such a case, it does not seem unreasonable to say that he knows there is a tree outside.[12] Harold appears to be in a relevantly similar position. On the strict reading,

[11]See also Bogardus and Perrin (2025) for a similar difference-maker-based treatment of a Gettier case that also involves a deviant causal chain.

[12]This intuition is not uncontroversial. Cases of this general form have a substantial history in the epistemological literature. As Boyce and Moon (2023) note, a related case was originally raised by Lehrer and Paxson (1969: 234) against Unger's non-accidentality account of knowledge and was later adapted as an objection to Nozick's (1981: 190) tracking theory. In a recent version, Boyce and Moon present the case of

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

the fake sheep is present in the relevant region because a real sheep is there, and this dependence is not a one-off coincidence but a stable feature of the setup. Thus, on the strict reading, the intuition that Harold lacks knowledge seems substantially weakened. The case is no longer a straightforward counterexample to ME.

6. Comparison with virtue epistemology: testimonial knowledge and Gettier cases

In contemporary epistemology, another attractive approach to knowledge is virtue epistemology. According to it, knowledge is a kind of cognitive achievement – a true belief whose correctness is credited to the agent's intellectual virtues or competences. However, virtue epistemology faces a dilemma when it comes to the social dimension of knowledge – a dilemma that highlights the distinctive advantage of ME.

The dilemma faced by virtue epistemology, as Lackey (2007) has pointed out, stems from the inherent tension in its "credit" condition. On the one hand, to effectively eliminate Gettier-style epistemic luck, virtue epistemologists need a strong credit condition, requiring that the acquisition of a true belief be primarily credited to the agent's own intellectual virtues. In paradigmatic Gettier cases such as the blood type case, the correctness of the belief is primarily a product of luck rather than the agent's own intellectual virtues. So a strong credit condition correctly excludes it from knowledge.

On the other hand, when we consider testimonial knowledge, the situation is reversed. Consider the following case:

> **Asking for Directions in Chicago:**
>
> Morris has just arrived at the Chicago train station and wants to go to the Sears Tower. He looks around and asks a passerby who appears to be a local resident for directions. This person happens to be a resident who is very familiar with Chicago and gives Morris accurate directions. Morris unhesitatingly believes the person and forms a corresponding true belief. (Lackey 2007)

Intuitively, Morris acquires knowledge regarding the Sears Tower's location through this exchange. Yet under a strong credit condition, the correctness of Morris's belief is primarily credited to the local resident's geographical knowledge and honest character, not to Morris's own intellectual virtues – he merely happens to ask a passing local. If the correctness of a belief must be primarily credited to the agent's own intellectual virtues, Morris's knowledge is in jeopardy. To accommodate such testimonial knowledge, virtue epistemologists are typically forced to weaken their core condition, adopting a weak

Micha, who appears to see a vase on a pedestal but is in fact seeing a holographic projection generated from a real vase hidden inside the pedestal. The real vase explains why the corresponding holographic image appears, and Micha forms his belief on the basis of that image. Boyce and Moon nevertheless maintain that Micha clearly does not know that there is a vase in front of him, and present the case as a challenge to explanationism. Bogardus and Perrin (2025) offer a more qualified diagnosis. They agree that Micha does not know the specifically locational proposition that there is a vase in front of him, since the vase's being located in front of Micha is not a difference-maker in the explanation of his belief. But they argue that, if Micha instead forms the more general belief that there exists a vase, it is much less clear that he lacks knowledge; indeed, they are inclined to say that he knows.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

credit condition according to which it suffices for knowledge that the correctness of the belief is only *partially* credited to the agent's own intellectual virtues.

But an overly lenient credit condition would wrongly classify Gettier cases as knowledge. In the blood type case, John does exhibit some basic intellectual virtues (following instructions for drawing blood, reading the report, reasonably trusting hospital equipment). The weak condition thus wrongly attributes knowledge to John.

Virtue epistemology thus faces an uncomfortable trade-off: strengthening the credit condition to handle Gettier cases undermines its explanation of testimonial knowledge, and weakening the condition to accommodate testimonial knowledge reopens the door to Gettier cases.

Although we do not think this difficulty is insurmountable for the virtue epistemologist,[13] the fact that ME does not face it at all speaks in its favor. ME's theoretical framework naturally circumvents this dilemma. Virtue epistemology focuses on the question of agential attribution: to what extent can the correctness of the belief be credited to the agent's own virtues? ME, however, focuses on a question of objective connection: is the belief explained by the relevant worldly fact? This objective explanatory chain can – and often does – span across multiple agents.

In the Chicago case, ME delivers the following diagnosis. Condition (1) is satisfied because a complete explanatory chain runs from the fact regarding the actual location of the Sears Tower, through the local resident's genuine knowledge and sincere testimony, to Morris's belief. This is a chain that originates in a worldly fact and passes through a reliable intermediary before reaching its endpoint in Morris's belief. Condition (2) is also satisfied: in the vast majority of nearby worlds in which Morris forms his belief in the same way – say, by sizing up the situation and asking someone who appears to be a local adult – the person he asks is reliable, and the explanatory connection between his belief and the actual location of the Sears Tower is preserved.[14] ME therefore correctly classifies Morris as knowing where the Sears Tower is.

At the same time, ME remains immune to knowledge-undermining luck involved in Gettier cases. As mentioned earlier, in the blood type case, although John exercises his ordinary cognitive abilities in reading and trusting the result, the fact that makes his belief true (i.e., that his blood type is indeed B) and the source of his belief (i.e., the report provided by the malfunctioning machine) are explanatorily disconnected. Condition (1) therefore fails, and John does not know. In sum, ME is flexible enough to accommodate the social dimension of knowledge, yet rigorous enough to rule out luck involved in Gettier cases.

7. Conclusion

The Gettier problem involves two structurally distinct forms of knowledge-undermining luck. Modalism and explanationism exhibit complementary strengths and limitations in addressing these two forms of luck. Modalism grounds knowledge in counterfactual stability. It thus rules out environmental luck, but struggles with certain cases of intervening luck. Explanationism, by contrast, grounds knowledge in an explanatory connection between belief and fact in the actual world. It therefore rules out intervening

[13]See, for example, Carter and Kallestrup (2026), Greco (2010, 2012, 2020), Pritchard (2012), Sosa (2007) for virtue-theoretic responses to this challenge.

[14]If the case involves environmental luck so that he could easily have asked an unreliable speaker, then the intuition that he knows is also undercut.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

luck, but does not adequately handle environmental luck. These complementary limitations suggest that neither modal stability nor actual-world explanation, taken on its own, adequately captures all forms of knowledge-undermining luck. Modal Explanationism combines the two. It requires that a belief be explained by the fact that makes it true, and that this explanatory connection remain stable across a range of nearby possible worlds. We have argued that this integrated account diagnoses both forms of luck, handles the challenge from deviant causal chains, and avoids the credit dilemma afflicting virtue epistemology.

Acknowledgements. We would like to thank an anonymous reviewer and an associate editor of this journal for their helpful comments and suggestions on earlier versions of this paper.

Funding statement. Haicheng Zhao's research was supported by the National Social Science Fund of China (grant number: 25CZX037).

References

Beddor B. and Pavese C. (2020). 'Modal Virtue Epistemology.' *Philosophy and Phenomenological Research* **101**(1), 61–79.

Bogardus T. and Perrin W. (2022). 'Knowledge is Believing Something Because It's True.' *Episteme* **19**(2), 178–96.

Bogardus T. and Perrin W. (2025). 'A Defense of Explanationism against Recent Objections.' *Episteme* **22**(1), 35–46.

Boyce K. and Moon A. (2023). 'An Explanationist Defense of Proper Functionalism.' In L.R.G. Oliveira (ed), *Externalism about Knowledge*. New York, NY: Oxford University Press.

Carter A. and Kallestrup J. (2026). 'Virtuous Deferral.' *Noûs*, e70041.

Carter A. and Pritchard D. (2015). 'Knowledge-How and Epistemic Luck.' *Noûs* **49**(3), 440–53.

Chisholm R. (1966). *Theory of Knowledge* (1st edn). Englewood Cliffs, NJ: Prentice Hall.

Clarke-Doane J. (2012). 'Morality and Mathematics: The Evolutionary Challenge.' *Ethics* **122**(2), 313–40.

Dutant J. (2016). 'How to Be an Infallibilist.' *Philosophical Issues* **26**(1), 148–71.

Faraci D. (2019). 'Groundwork for an Explanationist Account of Epistemic Coincidence.' *Philosophers' Imprint* **19**(4), 1–26.

Gettier E. (1963). 'Is Justified True Belief Knowledge?' *Analysis* **23**(6), 121–3.

Goldman A. (1976). 'Discrimination and Perceptual Knowledge.' *The Journal of Philosophy* **73**(20), 771–91.

Goldman A. (1982). 'An Explanatory Analysis of Knowledge.' *The Journal of Philosophy* **79**(11), 718–9.

Greco J. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. New York: Cambridge University Press.

Greco J. (2012). 'A (Different) Virtue Epistemology.' *Philosophy and Phenomenological Research* **85**(1), 1–26.

Greco J. (2020). *The Transmission of Knowledge*. New York: Cambridge University Press.

Jenkins C. (2006). 'Knowledge and Explanation.' *Canadian Journal of Philosophy* **36**(2), 137–64.

Lackey J. (2007). 'Why We Don't Deserve Credit for Everything We Know.' *Synthese* **158**(3), 345–61.

Lehrer K. and Paxson T. (1969). 'Knowledge: Undefeated Justified True Belief.' *The Journal of Philosophy* **66**(8), 225–37.

Miracchi L. (2015). 'Competence to Know.' *Philosophical Studies* **172**(1), 29–56.

Mortini D. (2024). 'The Explanationist and the Modalist.' *Episteme* **21**(2), 371–86.

Nozick R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Pritchard D. (2012). 'Anti-Luck Virtue Epistemology.' *The Journal of Philosophy* **109**(3), 247–79.

Pritchard D. (2016). 'Epistemic Risk.' *The Journal of Philosophy* **113**(11), 550–71.

Sainsbury M. (1997). 'Easy Possibilities.' *Philosophy and Phenomenological Research* **57**(4), 907–19.

Smith M. (2010). 'What Else Justification Could Be.' *Noûs* **44**(1), 10–31.

Smith M. (2016). *Between Probability and Certainty: What Justifies Belief*. Oxford: Oxford University Press.

Sosa E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford: Oxford University Press.

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press

Strevens M. (2011). *Depth*. Cambridge, MA: Harvard University Press.
Valaris M. (2022). 'Normality, Safety and Knowledge.' *Philosophy and Phenomenological Research* **106**(2), 394–401.
Zagzebski L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge, England: Cambridge University Press.
Zhao H. (2024). 'Why Better Safe than Sensitive.' *Philosophy and Phenomenological Research* **109**(3), 838–55.
Zhao H. (2025). 'Solving the Lottery Problem: From Modal Accounts to Explanationism.' *Canadian Journal of Philosophy* **55**(4), 387–404.
Zhao H. (2026). 'Against Knowledge as Justified True Belief.' *Erkenntnis* **91**(1), 463–71.

Ying Zhang is a PhD student in the Department of Philosophy at Xiamen University. Her research interests lie primarily in epistemology. Email: 10420250158062@stu.xmu.edu.cn

Haicheng Zhao is a faculty member in the Department of Philosophy at Xiamen University. His main research interests cover epistemology and normative ethics. Email: haichengzhao90@gmail.com

Cite this article: Zhang Y. and H. Zhao (2026). "A Modal Explanationist Solution to the Gettier Problem." *Episteme* 1–14. https://doi.org/10.1017/epi.2026.10165

https://doi.org/10.1017/epi.2026.10165 Published online by Cambridge University Press