dialectica

Vol 75 : 3
September 2021

Gruesome Counterfactuals

Jared Warren

doi:10.48106/dial.v75.i3.01

Jared Warren. 2021. "Gruesome Counterfactuals."
Dialectica 75(3): 1–1. doi:10.48106/dial.v75.i3.01.

Gruesome Counterfactuals

Jared Warren

One of the most popular answers to the grue puzzle appeals to counterfactual dependence. An observed green emerald is grue. But while it still would have been green even if unobserved, if unobserved it would not have been grue. Because of this, or something like it, we can project "green" but not "grue" to the unobserved emeralds. Counterfactual theories have been offered by Frank Jackson and Peter Godfrey-Smith, among others. But there is a worry that all counterfactual approaches to grue fail for the same reason—counterfactual symmetry. The grue theorist can endorse symmetrical counterfactuals: an observed green emerald would have been grue but not green if unobserved. It then seems that counterfactual responses to grue beg the question. Here I argue that there are two ways to understand this challenge and that they both fail, but for different reasons. I close by drawing some general lessons about philosophical fair play regarding the twentieth century's many broadly semantic, broadly skeptical challenges—grue, quus, gavagai, and the like.

1 Counterfactuals and Grue

All of the emeralds we've observed have been green. They've also been grue—either green and observed or blue and unobserved. Yet we take our observations to support or confirm the hypothesis that all emeralds are green, but not the hypothesis that all emeralds are grue. We inductively project the "green" predicate, but not the "grue" predicate. The new riddle of induction is roughly the challenge of explaining and vindicating these inductive policies.[1]

One popular strategy for answering the riddle appeals to *counterfactual* differences between green and grue. For a given observed green emerald, *e*, the following counterfactuals seem true: Observed(e) Green(e)

(1) ¬ Observed(\$e)

1 The riddle derives from Goodman (1946, 1955).

(2)

]{#one} ¬Observed(e) □→ Green(e)

(II) ¬Observed(e) □→ ¬Grue(e)

If *e* hadn't been observed, it still would have been green. But if it hadn't been observed, it wouldn't have been grue. On this strategy, counterfactuals like these are key to why projecting "green" is warranted but projecting "grue" is not.

The counterfactual strategy comes in different forms. Not all of them appeal to (1) and (2).[2] And not all that appeal to (1) or (2) do so in the exact same way. For some it is our *knowledge* of such counterfactuals that is key.[3] For others it is enough that we *believe* such counterfactuals.[4] For still others, the *truth* of (1) and (2) is what matters.[5] These differences are important, and will be relevant at several points below, but I won't belabor them. It has been claimed that *every* counterfactual approach to grue fails for the same reason—counterfactual symmetry.[6] I will respond to the challenge here, but obviously, even if my response is successful, counterfactual responses to grue might fail for other reasons.

2 Counterfactual Symmetry

The grue theorist can respond to counterfactual approaches by rejecting (1) and (2) and instead accepting:

(III) ¬Observed(e) □→ ¬Green(e)
(IV) ¬Observed(e) □→ Grue(e)

Given that the antecedent is possible—◇¬Observed(e)—(1) and (III) are incompatible, as are (2) and (IV). The challenge is roughly that if we can vindicate our practices by appealing to (1) and (2), the grue theorist can vindicate gruesome practices by instead appealing to (III) and (IV). The exact nature of the appeal will, of course, depend on the details of the counterfactual

2 See Freitag (2015, 2016) for a counterfactual approach that doesn't appeal to (1) or (2).
3 See Jackson (1975).
4 See Schramm (2014) and Okasha (2007). In Jackson and Pargetter (1980) justified belief is appealed to.
5 See Godfrey-Smith (2003, 2011).
6 See Roskies (2008) and Dorst (2016, 2018); Schwartz (2005) makes some related points.

account on offer. Taken generally, the claim is that the gruesome practice is symmetrical to our non-gruesome practice, in the epistemically relevant way.

Some counterfactual theorists will be fine with this. Alfred Schramm merely attempts to show that at most one out of "all emeralds are green" and "all emeralds are grue" is confirmed by *our* evidence.[7] And Samir Okasha's position makes warrant for an inductive inference relative to our beliefs.[8] If you believe (1) and (2), you are warranted in concluding that all emeralds are green. If you instead believe (III) and (IV), you are warranted in concluding that all emeralds are grue. These versions of the counterfactual strategy aim for a somewhat limited conclusion. They want to show that projecting "green" but not "grue" is warranted *for us*, given *our beliefs*.

This is compatible with gruesome practices being warranted relative to alternative beliefs. Admitting this is not paradoxical, nor need it collapse into epistemic relativism. Nearly everyone accepts this kind of distinction. Different background beliefs warrant different conclusions. The kind of epistemic warrant being used here is broadly *internalist*; it is based on factors internal and accessible to the agent, such as their beliefs. Call this kind of internalist epistemic warrant *justification*. There is also an externalist kind of epistemic warrant, call it *entitlement*. Entitlement can depend on factors completely outside of and inaccessible to an agent. In the case at hand, perhaps the mere truth of (1) and (2) means that we are entitled to "green" projections, but not "grue" projections.

So both justification and entitlement to "green" projections can be secured using counterfactual strategies. This is significant; most attempted replies to grue don't even get this far. But while it certainly shouldn't be dismissed, we may hope for a bit more. On the internalist front, we might hope for non-question-begging arguments that beliefs in (1) and (2) are justified, while beliefs (2) and (IV) are not. On the externalist front, we might hope for non-question-begging arguments that (1) and (2) are true, while (2) and (IV) are not. These and other hopes seem to hinge on our being able to successfully defend a counterfactual theory based on (1) and (2) against a gruesome counterfactual theory based on (2) and (IV).

Can counterfactual attempts to solve grue successfully rebut the challenge of counterfactual symmetry? This is our question. To answer it, we'll need to

7 In Schramm's (2014) terminology, people who believe (III) and (IV) while disbelieving (1) and (2) have *different evidence* than people—like us—who have the opposite beliefs.

8 See Okasha's (2007) commentary on Jackson (1975).

doi: 10.48106/dial.v75.i3.01

distinguish between two different ways of pushing the symmetry challenge. I think that both ways fail, but they fail for different reasons.

3 The Epistemological Version

Goodman originally defined "grue" as being either green and observed before t, or blue and unobserved before t. Above, I followed the common practice of simply assuming that t is *now*. Either way, the definition of "grue" involves a temporal element. A natural first thought is that this temporal element is what defeats the projection of "grue". In response to this, Goodman introduced a parallel term, "bleen", meaning observed and blue or unobserved and green. He then noted that "green" and "blue" are definable using "grue" and "bleen" (Goodman 1955). In a language that starts with "grue" and "bleen", it is "green" and not "grue" that has an explicit temporal element in its definition. This shows that whether a term's definition includes a temporal element—and even whether a term is defined at all—is language-relative.

This point is much cleaner than the counterfactual symmetry point. All Goodman needed to show was that a definition existed for "green" that included an explicit temporal element. The correctness of Goodman's definition is common ground in the debate. Some have doubted the possibility of a grue/bleen mother tongue, and others have pointed to lingering temporal asymmetries between "grue" and "green", but nobody challenges the accuracy of Goodman's definition of "green" in terms of "grue" and "bleen". The definitional facts are agreed upon by all parties.

Not so for the counterfactual facts. Counterfactual claims are about the world and its features. This match *would* light, if struck. This sugar cube *would* dissolve, if placed in water. That star *would* collapse into a black hole, if it were twice as massive. Those who disagree with us about the world and its features may well disagree with us about which properties are independent of our observational procedures. So unlike grue theorists who adopt Goodman's symmetrical definitions, grue theorists who endorse symmetrical counterfactuals disagree with us about *matters of fact.*

The question is whether there is truly symmetry here and, if so, what kind? One idea is that there is *epistemological* symmetry. Adina Roskies (2008) has pushed something like this form of the counterfactual symmetry point, calling it the problem of "counterfactual robustness". Her direct target is Jackson's theory, which required knowledge of counterfactuals like (1) and (2).

Roskies claims that any route to this knowledge is question-begging against proponents of (III) and (IV).[9]

This version of the counterfactual symmetry challenge isn't merely the original grue puzzle in another guise, at least on the most natural understanding. It is instead akin to a *skeptical* challenge based on grue. It asks us how we know that the world isn't radically different than we think it is. How do we know that color features don't depend on observations? The challenge is now to vindicate our overall picture of physical reality, not just to defend some of our local inductive policies.

Of course, from the earliest presentations, Goodman tied the grue puzzle to issues of laws and counterfactuals. So you might question my claim that the epistemological symmetry challenge substantively differs from the original grue puzzle. I agree that in its more expansive formulations, Goodman's puzzle concerns not only induction proper but also more general explanatory reasoning, including abduction or "inference to the best explanation." This point was perhaps first clearly made in the literature by John Moreland in an excellent but little-noticed discussion published in 1976:

> What is misleading in Goodman's formulation of the Riddle is that it mixes questions of induction with questions of abduction. It is not just a question of which [hypothesis] to project. We have seen that in the appropriate circumstances either might be projected. We wish normally to reject [the gruesome hypothesis] out of hand (regardless of the evidence) because in most situations [the gruesome hypothesis] would not be accepted as an explanatory hypothesis; and this is a question of abduction, not induction. [...] it does seem important to distinguish between the question of whether or not [the gruesome hypothesis] is to be projected in a given situation and the question of whether or not [the gruesome hypothesis] would ever be formulated as an explanatory hypothesis and, thus, made a candidate for inductive confirmation. [Moreland (1976), 376.][10]

In the case of Jackson's counterfactual theory, a division like this falls out quite naturally. We start with the question of which predicates we can project,

9 A related argument is in Schwartz (2005), targeting Godfrey-Smith (2003).

10 I tracked down and read Moreland's paper after first reading a detailed synopsis of it in Stalker's incredible annotated bibliography of the first fifty years of grue literature, found in Stalker (1994).

doi: 10.48106/dial.v75.i3.01

but answering that question involves an appeal to background knowledge of counterfactuals. This background knowledge is what is questioned by Roskies. There are at least two questions here, and they are not the same. Proponents of counterfactual theories are independently committed to distinguishing them.

In fact, *everyone* must distinguish between questions of induction and questions of abduction, not just counterfactual theorists.[11] So this isn't an *ad hoc* move of desperation in the face of refutation. Nor does distinguishing these questions mean that a unified inductive logic, covering all reasonable non-deductive reasoning, is impossible.[12]

With this distinction noted, Jackson can plausibly explain how it is that we know (2) (which is what, on his account, blocks the projection of "grue"). The overall answer is likely a very long story.[13] In short: an extended process of observation, induction, deduction, and—most crucially—abduction led to our overall theory of the natural world. This overall scientific story entails (2), so knowing this, our overall theoretical knowledge transfers from our background theory to (2). Unless *everything* is *always* up for grabs, it is perfectly legitimate to appeal to our fundamental beliefs about the natural world when evaluating some particular inductive inference involving a newly introduced predicate. No question is begged in the process. Toward the end of her discussion, Roskies herself indicates openness to this type of reply to her challenge. She says her goal was only to show that a Jackson-style counterfactual account required supplementation.

I don't disagree completely, but we should put the point somewhat differently. We should say that Jackson's account of projection is fine as it stands, but add that it appeals to background knowledge that must itself be explained, in the long run. That explanation will involve not the original anti-grue reasoning, but instead general explanatory reasoning about the world, so there is

11 In addition to the introduction of the "IBE" terminology, Harman (1965) argued that enumerative induction should be understood using inference to the best explanation. This is either unacceptable or compatible with the point I'm making here, depending on exactly how the claim is understood.

12 Fraassen (1989) has argued that IBE contradicts conditionalization, and so cannot be integrated into a standard Bayesian framework. But his argument is based on an implausible way of combining Bayesianism with explanatory reasoning. For a better strategy for integration, see Huemer (2009) and Weisberg (2009).

13 I believe Jackson was always aware of this. Douglas Stalker told me that Jackson once told him that a full development and defense of his (Jackson's) approach to grue would take a very long book, not just a paper.

no circularity. If general explanatory considerations tell against the overall grue position, including the alternative counterfactuals, then Jackson has an answer to the symmetry challenge.[14]

It is worth noting that in requiring *knowledge* of counterfactuals like (2), Jackson's account is extremely demanding.[15] Every other counterfactual theory of projection requires much less of us. This is important. I already mentioned that variant theories like Okasha's and Schramm's require only *belief* in the relevant background counterfactuals, not *knowledge*. And other counterfactual theories, like Godfrey-Smith's, require only the *truth* of the relevant counterfactuals. I highlight these points to stress that, by considering the symmetry challenge as aimed at Jackson's original counterfactual approach, we have been considering it in its strongest form. Other counterfactual theories should do *at least* as well at answering the challenge.

Whatever form the counterfactual theory takes, there is no epistemological symmetry between us and the grue theorists with respect to these counterfactuals. If all parties understand counterfactuals as we do, then there are good reasons for preferring (1) and (2) over (III) and (IV). These reasons are general and theoretical and explanatory, but they aren't question-begging. Of course, this assumes that the challenge is posed using *our* understanding of counterfactuals. There is another way of pressing the counterfactual symmetry challenge. This more radical approach has recently been pursued by Christopher Dorst in critical discussions of the theories of both Alfred Schramm and Wolfgang Freitag.[16] Here I'll be discussing the general merits of the challenge, not its justice as an objection to any *particular* counterfactual theory.

4 The Similarity Version

Consider how we semantically evaluate counterfactuals like (1), (2), (III), and (IV). Obviously, we used and asserted and evaluated counterfactuals long

14 In addition to Moreland's (1976) discussion, my response here also dovetails with Godfrey-Smith's (2011) response to related objections, which he attributes to Laura Schroeter and Ira Schall. Though in some ways his discussion makes the challenge he is addressing sound more like the alternative similarity version discussed below.

15 Jackson later altered his account in several ways—see Jackson and Pargetter (1980).

16 See Dorst (2016, 2018), Schramm (2014), and Freitag (2015, 2016). Schramm's approach is similar to that of Jackson (1975) in many ways, but different in others. But despite how Dorst interprets him, Freitag doesn't appeal to knowledge of or belief in either (1) or (2)—see Freitag (2019) for further clarifications about this.

doi: 10.48106/dial.v75.i3.01

before anyone came up with an explicit semantic theory for counterfactuals. Still, a semantics is useful for codifying the truth conditions our practices assign to counterfactuals. The usual counterfactual semantics derives from Stalnaker and Lewis and uses a similarity metric over the space of possible worlds.[17] Here's a simplified version of this kind of semantics:

> A counterfactual $\ulcorner \phi \mathrel{\square\!\!\rightarrow} \psi \urcorner$ is true at a world w just in case in all of the most similar worlds where ϕ is true, ψ is also true.[18]

What exactly *similarity* comes to here has been much discussed.[19] There is broad agreement over cases, but the precise analysis is tricky. Sometimes "similarity" is claimed as subjective—including by Goodman himself.[20] Yet if subjectivity about similarity is combined with a similarity-semantics for counterfactuals, and then fed into our scientific and inductive practices as the counterfactual strategy requires, absurdities result.[21] This will be illustrated below.

Let's first assume that the relevant notion of similarity, though context-sensitive, is not completely subjective. Given what *we* mean by "most similar" in this semantic clause, the only way for (III) and (IV) to be true while (1) and (2) are false is for the world to be wildly different in the manner discussed in the previous section. Yet there is another option. Grue theorists could appeal to radically different "similarity" judgments, and then use those judgments in their counterfactual semantics without otherwise disagreeing with us about the world.

This involves saying that a world in which an observed green emerald *e*—this very one—is unobserved and blue, is *more similar to our world* than is a world where *e* is unobserved but green. This is bizarre to us, given what we mean by "similar." Perhaps there are possible worlds where *e*, this very thing, is blue and not green. But given everything we know and believe about physics, chemistry, optics, and more, such a world must be very dissimilar to

17 See Stalnaker (1968) and Lewis (1973). A relatively popular minority alternative analyzes counterfactuals using strict conditionals ($\ulcorner \phi \dashv \psi \urcorner$) and context-sensitivity.

18 Stalnaker and Lewis make different framework assumptions that lead to differences in their respective counterfactual logics. These differences won't be of concern here. Lewis's approach doesn't require a sphere of similarity containing *only* antecedent worlds, but I have simplified.

19 See Fine (1975), Lewis (1979), and Bennett (2003) for important contributions.

20 Not by proponents of the similarity semantics for counterfactuals—see Lewis (1983a). For Goodman's treatment, see his (1972, ch. IX).

21 Something like this bullet is bitten by Ullian (1961).

our world. The imagined grue theorist denies this. They agree with us about which worlds are possible. And they also agree with us about the facts in this world, but they disagree about how similar certain worlds are to this world.

Something like this reply might be implicit in some of Goodman's later discussions of gruesome matters. More recently it has been explicitly pursued by Dorst in reply to recent counterfactual approaches:

> We are thus examining the same world in both cases, so only one of the two counterfactuals can possibly be true. [...] But which one is true? That will evidently depend on the similarity metric we impose on the space of possible worlds. On our traditional understanding of 'similarity,' the closest (most similar) world where the emeralds in our evidence class were not observed before 2020 will be one in which they are green and not grue. Surely, however, a 'grue'-speaker would have exactly the opposite conception of 'similarity.' After all, he thinks grue things all "look alike," so it is only natural that his conception of similarity would reflect that. [...] So if we appeal to counterfactuals to justify the 'green' induction over the 'grue' induction, the 'grue'- speaker will have a precisely symmetric justification open to him. (Dorst 2016, 153) [22]

This understanding of the counterfactual symmetry challenge differs from the epistemological understanding discussed above. In some ways it is a more radical and troubling challenge.

There is no accounting for taste, and maybe there is no accounting for weird similarity judgments either. Yet meaning is determined by use. It's plausible that anyone who clearheadedly used the term "similar" so differently would no longer mean what we mean by the term. If they then used their alternative notion of similarity in giving a counterfactual semantics, this difference in meaning will also infect terms like "would" and "counterfactual." But the real issue is not about semantic theory. The real issue is *use*—the use that the formal semantic theory was meant to codify. Drastic changes in use lead to changes in meaning. If these grue theorists use counterfactuals in a way that aligns with their "similarity" judgments and not ours, then *they no longer mean what we mean by counterfactuals*.

22 A similar passage also occurs in Dorst (2018, 181).

doi: 10.48106/dial.v75.i3.01

If this "change of meaning" charge is true, it provides a response to the similarity version of the counterfactual symmetry challenge. The response is that, in adopting this version of the challenge, grue theorists have changed their language significantly. They have changed it so much, in fact, that *they no longer disagree with us*. Our dispute has devolved into a merely verbal dispute, with no direct disagreement.

In order to see this, it is important to understand that the kind of linguistic change involved here is *not* the simple change of moving to a language in which there are primitives for "grue" and "bleen" but all else remains the same. In that type of language, "similar" still has the same meaning it has in our language. So those grue speakers will agree with us about (1), (2), (III), and (IV) or rather, about their *translations* into the grue language. That change did not amount to a difference in worldview, only a difference in language. This shows that the radical counterfactual similarity charge is not backed up by the possibility of grue/bleen languages of the kind discussed by Goodman.[23] Instead a much more radical linguistic change is required, one that systematically alters the truth conditions of counterfactuals.

Some may quibble. Has meaning *really* been changed, they will ask. Anyone who thinks meaning is closely tied to use will say *yes*. And since almost everyone thinks that meaning is closely tied to use, almost everyone will say *yes*. Even Quine, the arch-critic of analyticity, argued that drastic meaning changes undermine simple homophonic translations (Quine 1970).[24] So I don't think my claims about meaning change beg any significant questions about analyticity or the like.

We could argue for meaning change here *theoretically*, by appealing to widely accepted theoretical principles of interpretation or translation—charity, humanity, rationality, and so on.[25] But the central point is probably best illustrated more simply, by reflecting on simple applications of our actual practices of translation and interpretation. Imagine that you encounter someone who clear-headedly makes "similarity" judgments that align with those of our imagined grue theorist. Even after all of the facts are in, they continue to disagree with you. They say that grue things are "more similar" to each other than green things, even more similar with respect to "color," and that grue things, but not green things, "look alike." After you convinced yourself that these divergences are not caused by some perceptual deficiency or a mistake

23 Dorst (2016, 2018) sometimes seems to deny this. See also Schwartz (2005).
24 See Warren (2018) for an updated version of the argument.
25 For such principles, see Grandy (1973), Lewis (1974), Hirsch (2011), and Warren (2016).

about the factual situation, you would conclude that your interlocutor simply spoke a different language than you did. They simply do not mean what you mean by "similar" or the like.

This meaning change diagnosis is the best and most appealing way to understand the apparent disagreement here. To some extent though, it can be left to one side. The crucial point concerns the differences in practical language use. Even those who think there is a difference of opinion, not a difference of meaning, must admit that the gruesome practices differ wildly from our own. Let me provide a concrete illustration of this by considering what happens *after* time t, where t is the time to which the definition of "grue" is indexed. Let's update Goodman's original definition with a predicate for "observed before 2020":

$$\text{Grue}_{2020}(\alpha) \leftrightarrow (\text{Green}(\alpha) \wedge \text{Observed}_{2020}(\alpha)) \vee (\text{Blue}(\alpha) \wedge \neg \text{Observed}_{2020}(\alpha))$$

Something is grue2020 just in case it is either green and observed before 2020 or blue and not observed before 2020. Since it is now past 2020, we can observe previously unobserved emeralds without them being observed2020. What happens when we do is instructive. On January 1st, 2020, the grue defender is committed to the following for previously observed emerald, *e*:

(III*) $\neg \text{Observed}_{2020}(e) \mathrel{\square\!\!\rightarrow} \neg \text{Green}(e)$
(IV*) $\neg \text{Observed}_{2020}(e) \mathrel{\square\!\!\rightarrow} \text{Grue}_{2020}(e)$

Now let us observe a previously unobserved emerald, m.

1. $\neg \text{Observed}_{2020}(m) \wedge \text{Emerald}(m)$ (assumption)
2. $\forall x(\text{Emerald}(x) \rightarrow \text{Grue}_{2020}(x))$ (inductive projection made by the grue defenders, backed up by (III*) and (IV*))
3. $\text{Grue}_{2020}(m)$ (1,2)
4. $\text{Blue}(m)$ (1,3 and the definition of "Grue2020")

1.	$\neg \text{Observed}_{2020}(m) \wedge \text{Emerald}(m)$	(assumption)
2.	$\forall x(\text{Emerald}(x) \rightarrow \text{Grue}_{2020}(x))$	(inductive projection made by the grue defe
3.	$\text{Grue}_{2020}(m)$	(1,2)
4.	$\text{Blue}(m)$	(1,3 and the definition of "Grue2020")

doi: 10.48106/dial.v75.i3.01

(1) $\neg$Observed2020(m) $\wedge$ Emerald(m)
(2) $\forall x$(Emerald(x) $\rightarrow$ Grue2020(x))
(3) Grue2020(m)
(4) Blue(m)

In other words, these grue theorists can prove to themselves that m is blue, and then they look to the world and see that it's green. Saying that m is green does not beg any questions here. This is because the similarity-based symmetry challenge differs from the epistemological challenge. The radical grue theory under consideration is supposed to agree with us about all of the physical facts, including facts about the color of emeralds. They were supposed to differ from us only over similarity claims.

Related arguments have been used elsewhere in the massive grue literature, for different purposes.[26] The purpose here is to illustrate that when counterfactuals connect to induction, as proponents of the counterfactual strategy believe that they do, our practice of evaluating counterfactuals is not isolated. It instead feeds into a cluster of related physical notions, including *nomological modality*, *laws*, *dispositions*, and *causes*.[27] So if you change what you count as "relevantly similar," you change a great deal indeed.[28] Someone who changes what is meant by counterfactual terms would be ill-advised to fit their alternative "counterfactual" notions into the same conceptual space as our notions. Doing so leads to radically different ways of reasoning about and interacting with the same natural world.

Claiming that similarity itself is entirely subjective doesn't change this. If you say that, and then use similarity-relations to analyze the counterfactuals

26 From the comprehensive annotated bibliography in Stalker (1994), I learned that something like this reasoning has been used by Bayesians like Cohen (1989) to assign the grue hypothesis a low prior probability. Although there Cohen seemingly used a definition of "grue" more in line with Barker and Achinstein's (1960)—see Jackson (1975) for criticism of this. Also, Cohen's reasoning requires an additional step about grue-like predicates that goes beyond anything in my argument so far. That step concerns future grue predicates. Here we don't need that, since if the grue defender doubles down, with grue2021, grue2022, and so on, the same situation recurs. This is secured by the assumption that the radical grue defender agrees with us about all particular, physical facts.

27 See Putnam (1990) and Maudlin (2007) for related views. An aside: my view of alethic modality is (at least) tripartite. Logical and conceptual modality is a projection of our conventions, while physical modality is fully factual and objective. Finally, metaphysical modality is a mixed case—see Warren (2022).

28 Hesse (1969) made some related points. Like her, I don't think the strangeness of the grue theorist's conceptual scheme is itself a response to grue. Recall though that the dialectic here is that we are answering an objection to counterfactual approaches to grue.

which underwrite justified inductive inferences, you descend into a subjectivist nightmare.[29] Use any alternative counterfactual practice and you will likely find yourself with many false beliefs and many frustrated expectations. You could get lucky, but I wouldn't bet on it. Neither would you. Induction is not a pointless game we play for our amusement, it is instead a crucial part of how we reason about and master the physical world that surrounds and includes us.

So on neither reading does the counterfactual symmetry charge lead to genuine and troubling symmetry between our position and the grue theory. If the symmetry challenge is posed using *our* counterfactual notions, then we have non-question-begging *epistemic* reasons for favoring our counterfactuals over theirs. And if it is posed using some *alternative* notion of counterfactual similarity, then we have non-question-begging *practical* reasons for favoring our counterfactual practice over theirs. Either way, we have non-question-begging reasons for favoring our practices over the gruesome practices. Counterfactual approaches to grue might fail for other reasons, but the counterfactual symmetry charge doesn't stick. And despite its superficial appeal, the argumentative strategy it exemplifies is quite risky. I will close by explaining this.

5 Philosophical Fair Play

Twentieth-century philosophy was replete with overtly *semantic*, broadly *skeptical* challenges. These challenges attacked some of our most cherished doctrines using clever semantic tricks, principally clever redefinitions of crucial terms. The targets differed, as did the particular semantic tactics employed. Yet a general similarity between these challenges is easily recognized, provided it isn't overstated. Quine's translation argument, Putnam's model-theoretic argument, and Kripkenstein's skeptical paradox all fit into this model.[30] So too, does Goodman's grue puzzle.

Seen from this perspective, the overall dialectic surrounding the counterfactual symmetry challenge is quite familiar. A challenge has been posed by a semantic skeptic. One of our treasured assumptions is under threat. We rush in gallantly to offer a defense. Alas, the semantic skeptic uses a version of the original re-definition move yet again. This time on the very defense we have

29 This is arguably the exact path that Goodman followed to reach the radical irrealist position of his (1978). Even Goodman's most committed followers, for instance, Scheffler (2001), were unable to follow this path all the way to the end.

30 See Quine (1960), Putnam (1980), and Kripke (1982).

doi: 10.48106/dial.v75.i3.01

offered. The defense itself is seen by the skeptic as "just more theory" to be reinterpreted, just more grist for the skeptical mill.[31]

More often than not though, this move is not quite fair. When a constraint is used to screen off some skeptical reinterpretation, reinterpreting the *statement of the constraint* misses the mark. If we respond to Kripkenstein by claiming we mean addition and not quaddition by "plus" because we execute the addition algorithm in response to "plus" queries, talk of "quaddition algorithms" misses the point. The constraint concerns what we *do*, not what we *say about what we do*. Likewise with Quine's challenge, and Putnam's. Likewise too, with Goodman's grue challenge.

Almost the same exact dialectic pops up again and again, all across the philosophical landscape, so the point is worth belaboring. Skeptical reinterpretation is *risky*. Great care must be taken whenever the move is attempted. In the present context, we have seen that blithe appeals to gruesome counterfactuals come with baggage. The counterfactual symmetry claim has hidden costs. Either a commitment to absurd factual claims or an unnoticed change of topic. In contexts like this, we must always take care to tease out all ramifications of the skeptic's maneuvering. The semantic skeptic's tricks are ever so easy to apply, but they can very quickly take us into uncharted waters, where monsters lurk. In these waters, merely ersatz symmetry is often mistaken for the real thing.*

Jared Warren
0000-0003-4028-7969
Stanford University
jaredwar@stanford.edu

References

BARKER, Stephen F., and Peter ACHINSTEIN. 1960. "On the New Riddle of Induction." *The Philosophical Review* 69(4): 511–522, doi:10.2307/2183485.

BENACERRAF, Paul, and Hilary PUTNAM, eds. 1964. *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. Second edition: Benacerraf and Putnam (1983).

31 Compare Lewis's (1984) discussion of the dialectic surrounding Putnam's model-theoretic argument.

* Thanks to Rosa Cao, Christopher Dorst, Eli Hirsch, Alfred Schramm, Douglas Stalker, and Daniel Waxman.

—, eds. 1983. *Philosophy of Mathematics: Selected Readings*. 2nd ed. Cambridge: Cambridge University Press. First edition: Benacerraf and Putnam (1964).

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press, doi:10.1093/0199258872.001.0001.

Cohen, Laurence Jonathan. 1989. *An Introduction to the Philosophy of Induction and Probability*. Oxford: Oxford University Press.

Dorst, Christopher. 2016. "Evidence, Significance, and Counterfactuals: Schramm on the New Riddle of Induction [on Schramm (2014)]." *Erkenntnis* 81(1): 143–154, doi:10.1007/s10670-015-9733-2.

—. 2018. "Bet Accepted. A Reply to Freitag (2015)." *The Philosophical Quarterly* 68(270): 175–183, doi:10.1093/pq/pqx016.

Fine, Kit. 1975. "Critical Notice of Lewis (1973)." *Mind* 84(335): 451–458, doi:10.1093/mind/lxxxiv.1.451. Reprinted in Fine (2005, 357–365).

—. 2005. *Modality and Tense. Philosophical Papers*. Oxford: Oxford University Press, doi:10.1093/0199278709.001.0001.

Fraassen, Bas C. van. 1989. *Laws and Symmetry*. Oxford: Oxford University Press, doi:10.1093/0198248601.001.0001.

Freitag, Wolfgang. 2015. "I Bet You'll Solve Goodman's Riddle." *The Philosophical Quarterly* 65(259): 254–267, doi:10.1093/pq/pqu093.

—. 2016. "The Disjunctive Riddle and the Grue-Paradox." *Dialectica* 70(2): 185–200, doi:10.1111/1746-8361.12136.

—. 2019. "Why Doxastic Dependence Defeats Grue: A Response to Dorst's Reply [to Dorst (2016)]." *The Philosophical Quarterly* 69(274): 156–165, doi:10.1093/pq/pqy029.

Godfrey-Smith, Peter. 2003. "Goodman's Problem and Scientific Methodology." *The Journal of Philosophy* 100(11): 573–588, doi:10.2307/3655745.

—. 2011. "Induction, Samples, and Kinds." In *Carving Nature at Its Joints. Natural Kinds in Metaphysics and Science*, edited by Joseph Keim Campbell, Michael O'Rourke, and Matthew H. Slater, pp. 33–52. Topics in Contemporary Philosophy n. 7. Cambridge, Massachusetts: The MIT Press, doi:10.7551/mitpress/9780262015936.001.0001.

Goodman, Nelson. 1946. "A Query of Confirmation." *The Journal of Philosophy* 43(14): 383–385, doi:10.2307/2020332. Reprinted in Goodman (1972, 363–366).

—. 1955. *Fact, Fiction and Forecast*. Cambridge, Massachusetts: Harvard University Press.

—. 1972. *Problems and Projects*. Indianapolis, Indiana: Bobbs-Merrill Company Inc.

—. 1978. *Ways of Worldmaking*. Indianapolis, Indiana: Hackett Publishing Co.

Grandy, Richard E. 1973. "Reference, Meaning, and Belief." *The Journal of Philosophy* 70(14): 439–452, doi:10.2307/2025108.

Hahn, Lewis Edwin, ed. 1992. *The Philosophy of A.J. Ayer*. The Library of Living Philosophers n. 21. LaSalle, Illinois: Open Court Publishing Co.

HARMAN, Gilbert H. 1965. "The Inference to the Best Explanation." *The Philosophical Review* 74(1): 88–95, doi:10.2307/2183532.

HESSE, Mary B. 1969. "Ramifications of 'Grue'." *The British Journal for the Philosophy of Science* 20(1): 13–25, doi:10.1093/bjps/20.1.13.

HIRSCH, Eli. 2011. *Quantifier Variance and Realism: Essays in Metaontology*. Oxford: Oxford University Press.

HUEMER, Michael. 2009. "Explanationist Aid for the Theory of Inductive Logic." *The British Journal for the Philosophy of Science* 60(2): 345–375, doi:10.1093/bjps/axp008.

JACKSON, Frank. 1975. "Grue." *The Journal of Philosophy* 72(5): 113–131, doi:10.2307/2024749.

JACKSON, Frank, and Robert PARGETTER. 1980. "Confirmation and the Nomological." *Canadian Journal of Philosophy* 10(3): 415–428, doi:10.1080/00455091.1980.10715734.

KRIPKE, Saul A. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, Massachusetts: Harvard University Press.

LEWIS, David. 1973. *Counterfactuals*. Cambridge, Massachusetts: Harvard University Press. Cited after republication as Lewis (2001).

—. 1974. "Radical Interpretation." *Synthese* 27(3): 331–344, doi:10.1007/bf00484599. Reprinted, with a postscript (Lewis 1983c), in Lewis (1983b, 108–119).

—. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13(4): 455–476, doi:10.2307/2215339. Reprinted, with a postscript (Lewis 1986b), in Lewis (1986a, 32–51).

—. 1983a. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61(4): 343–377, doi:10.1080/00048408312341131. Reprinted in Lewis (1999, 8–55).

—. 1983b. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

—. 1983c. "Postscript to Lewis (1974)." In *Philosophical Papers, Volume 1*, pp. 119–121. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

—. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62(3): 221–236, doi:10.1080/00048408412340013. Reprinted in Lewis (1999, 56–77).

—. 1986a. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

—. 1986b. "Postscript to Lewis (1979)." In *Philosophical Papers, Volume 2*, pp. 52–66. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

—. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511625343.

—. 2001. *Counterfactuals*. Oxford: Basil Blackwell Publishers. Republication of Lewis (1973).

MAUDLIN, Tim. 2007. *The Metaphysics Within Physics*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199218219.001.0001.

Moreland, John. 1976. "On Projecting Grue." *Philosophy of Science* 43(3): 363–377, doi:10.1086/288693.

Okasha, Samir. 2007. "What Does Goodman's 'Grue' Problem Really Show? ." *Philosophical Papers* 36(3): 483–502, doi:10.1080/05568640709485211.

Putnam, Hilary. 1980. "Models and Reality." *The Journal of Symbolic Logic* 45(3): 464–482, doi:10.2307/2273415. Reprinted in Benacerraf and Putnam (1964) and Putnam (1983, 1–25).

—. 1983. *Realism and Reason. Philosophical Papers, Volume 3*. Cambridge: Cambridge University Press.

—. 1990. "Is It Necessary That Water Is H_2O?" In *Realism with a Human Face*, pp. 54–79. Cambridge: Cambridge University Press. Reprinted in Hahn (1992, 429–454).

Quine, Willard van Orman. 1960. *Word and Object*. Cambridge, Massachusetts: The MIT Press. New edition: Quine (2013).

—. 1970. *Philosophy of Logic*. Cambridge: Cambridge University Press. Second edition: Quine (1986).

—. 1986. *Philosophy of Logic*. 2nd ed. Cambridge, Massachusetts: Harvard University Press. First edition: Quine (1970).

—. 2013. *Word and Object*. Cambridge, Massachusetts: The MIT Press. First edition: Quine (1960).

Roskies, Adina L. 2008. "Robustness and the New Riddle Revived ." *Ratio* 21(2): 218–230, doi:10.1111/j.1467-9329.2008.00396.x.

Scheffler, Israel. 2001. "My Quarrels with Nelson Goodman." *Philosophy and Phenomenological Research* 62(3): 665–677, doi:10.1111/j.1933-1592.2001.tb00084.x.

Schramm, Alfred. 2014. "Evidence, Hypothesis, and Grue." *Erkenntnis* 79(3): 571–591, doi:10.1007/s10670-013-9524-6.

Schwartz, Robert. 2005. "A Note on Goodman's Problem." *The Journal of Philosophy* 102(7): 375–379, doi:10.5840/jphil2005102715.

Stalker, Douglas F., ed. 1994. *Grue!* LaSalle, Illinois: Open Court Publishing Co.

Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Studies in Logical Theory*, edited by Nicholas Rescher, pp. 98–112. American Philosophical Quarterly Monograph Series n. 2. Oxford: Basil Blackwell Publishers.

Ullian, Joseph S. 1961. "More on 'Grue' and Grue." *The Philosophical Review* 70(3): 386–389, doi:10.2307/2183381.

Warren, Jared. 2016. "Internal and External Questions Revisited." *The Journal of Philosophy* 113(4): 177–209, doi:10.5840/jphil2016113411.

—. 2018. "Change of Logic, Change of Meaning ." *Philosophy and Phenomenological Research* 96(2): 421–442, doi:10.1111/phpr.12312.

—. 2022. "Inferentialism, Conventionalism, and *A Posteriori* Necessity." *The Journal of Philosophy* 119(10): 517–541, doi:10.5840/jphil20221191034.

Weisberg, Jonathan. 2009. "Locating IBE in the Bayesian Framework ." *Synthese* 167(1): 125–143, doi:10.1007/s11229-008-9305-y.