ROBUSTNESS AND THE NEW RIDDLE REVIVED

Adina L. Roskies

Abstract

The problem of induction is perennially important in epistemology and the philosophy of science. In response to Goodman’s ‘New Riddle of Induction’, Frank Jackson made a compelling case for there being no new riddle, by arguing that there are no non-projectible properties. Although Jackson’s denial of nonprojectible properties is correct, I argue here that he is mistaken in thinking that he thereby shows that there is no new riddle of induction, and demonstrate that his solution to the grue paradox fails to rule out the possibility of equally justified contradictory inductions. More importantly, in illuminating where Jackson’s argument fails, the paper casts a new light on the problem of induction, locating the problem not in the nature of the next (unexamined) x, but in the counterfactual robustness of properties of already examined x’s.

The traditional problem of induction challenges us to justify our belief that regularities observed in the past can be expected to hold in the future; all attempts at justification appear to be circular, in that they appeal to some version of the principle to be justified.¹ In ‘The New Riddle of Induction’, Nelson Goodman claimed that the traditional problem has a solution, and that Hume himself pointed the way to it.² Goodman raised a new riddle, however, that poses a problem equally as vexing for our inductive practices as the original problem of induction: how to distinguish predicates that support inductive inferences from those that do not. In the absence of a principled basis for making distinctions between so-called projectible and nonprojectible predicates, the worry is that induction licenses contradictory inferences with equivalent justificatory basis. This problem is famously illustrated by the grue paradox, in which our evidence to date

equally supports the inductions that ‘all emeralds are green’ and ‘all emeralds are grue,’ where grue is the property of being green and examined before time T or blue and examined after time T. Our evidence therefore provides us equal reason to predict that an emerald examined after T will be green and that it will be blue.

In his 1975 paper ‘Grue,’ Frank Jackson argues that Goodman was mistaken – there is no new riddle of induction, because all predicates are projectible.\(^3\) That is, all predicates license application of what Jackson calls the Straight Rule (SR) – the straightforward projection of past regularities to future cases. Thus, from the evidence that certain \(F\)s are \(G\), application of the SR allows us to conclude that certain other \(F\)s are \(G\). Grue-type paradoxes are avoided by taking note of what Jackson calls the Counterfactual Condition (CC): that in cases in which we know that whether or not the property about which we are making an induction obtains depends on the obtaining of some other property, about which we lack information, we are barred from application of the SR. This, according to Jackson, prevents the generation of conflicting grue-type inductive inferences. If Jackson is correct about the new riddle, and Goodman is correct about the traditional problem, there is no deep-seated problem with induction.

But there is a problem with induction. I argue here that Jackson fails to dissolve the grue problem, not because there are non-projectible predicates (I agree with Jackson here), but because his proposed solution points toward yet another face of the new riddle. This involves discriminating between robust and nonrobust counterfactuals, and justifying our beliefs about when they hold. This problem is not a problem with induction per se, but because induction relies upon the robustness of counterfactuals, if we are unable to justify our beliefs in which counterfactuals are robust, our inductive practices will be subject to grue-type paradoxes.

Jackson’s argument

Jackson’s main point in ‘Grue’ is that all predicates, green and grue alike, are projectible. He further explains that the grue scenario only seems paradoxical because of muddy thinking.

Jackson diagnoses our error in the following way: it comes from unclarity in the definition of grue and in understanding just how the grue paradox is supposed to arise, and from failing to take into account the CC, which, when satisfied, blocks the application of the SR. Consistent interpretation of grue and heeding the CC dissolves the paradox.

Jackson demonstrates that on two possible interpretations of ‘grue’ that have found their way into the literature, no paradox arises, for on these interpretations grue is projectible: no grue-type paradoxes arise when the SR is applied to grue under these interpretations. The only interpretation under which a paradox is generated is a third, the reading that Goodman originally intended. That is:

**Grue**: $x$ is grue at $t$ if (1) $x$ is examined by $T$ and $x$ is green at $t$  
or (2) $x$ is not examined by $T$ and $x$ is blue at $t$.

Here $T$ is some time in the near future (which determines the extension of the set of examined objects) and $t$ indexes the time of examination. The paradox arises when observations before $T$ are used as an inductive base for a prediction after $T$: Suppose there is a class of $x$’s examined before $T$ that are found to be green. Application of the SR leads us to expect the next $x$ to be green, regardless of when it is examined. So we predict that an $x$ examined after $T$ will be green. By the definition of grue$_3$, above, the class of $x$’s examined by $T$ are also grue. Application of the SR on the basis of examined grue $x$’s will lead us to expect future $x$’s to be grue, regardless of when they are examined. Thus, we predict that an $x$ not examined until after $T$ will be grue, which, according to grue$_3$, means it will be blue. In short, the properties of observed $x$’s are consistent with their being both green and grue. The SR applied on the basis of this evidence to make an induction after $T$ yields conflicting predictions: in one case that the next $x$ will be green, and in the other that it will be blue.

The root of the problem is that in projection from the cases of, say, examined emeralds to unexamined ones, we have failed to take into account all our evidence. For along with the color of the examined emeralds, our observations also yield the fact that those emeralds were examined. Thus, Jackson points out, what we are

---

really licensed to infer from our past observations that all examined emeralds are (green and examined) or (grue and examined), is that the next unexamined emerald will be, respectively, (green and examined) or (grue and examined). Since by hypothesis the emerald in question is unexamined, we mistakenly allow that information to simply override the information provided by our inductive base, namely, that all the emeralds we have evidence about to date have been examined, while retaining information we have as to their color. We thus, erroneously, conclude that the next emerald will be green and unexamined (therefore green), or grue and unexamined (therefore blue). Jackson lays out clearly the fallacious argument patterns at work. When the logic is made plain, we can see where the problems originate that lead to the supposed paradox.

Jackson cautions that we must be very careful in incorporating the additional information that the next emerald is unexamined. In particular, he argues, this information must be incorporated bearing in mind the following principle: the SR cannot be applied if what he calls the Counterfactual Condition (CC) is satisfied. As Jackson puts it ‘certain Fs which are H and G do not support other Fs which are not H being G if it is known that the Fs in the evidence base would not have been G if they had not been H.’ The CC, then, is the italicized clause, and we are prevented from using the SR when that clause is satisfied. So, if, for example, all lobsters examined to date are red and are also boiled, and it is known that they are red because they are boiled, we are barred from inferring that the next lobster seen will be red, in the absence of knowing that it was boiled.

According to Jackson, the new riddle is thus dissolved: All predicates are projectible, and grue-type paradoxes cannot arise as long as we take proper note of the CC.\(^5\)

---

5 Jackson, ‘Grue’, p. 123, my emphasis.
6 Not everyone agrees that Jackson has dissolved the grue paradox. For instance, some argue that Jackson’s CC does not rule out all grueish instances, because there are cases in which we do not know that the Fs in the evidence base would not have been G if they had not been H, and these admit of contradictory inductions. For instance, Charles Chihara, in his 1981 paper ‘Quine and the counterfactual paradoxes,’ (Midwest Studies in Philosophy, pp. 425–452) argues that his CC is not sufficiently restrictive. He points out that Jackson’s CC fails to rule out all cases which can lead to grue-type paradoxes: there are cases in which it is not known that that the Fs in the evidence base would not have been G if they had not been H, so the CC cannot be invoked, yet nonetheless the predicates in question are intuitively not of the projectible kind, and, like grue, permit the induction of contradictory inferences.
The new riddle revived

I think that Jackson’s defense of the SR falls prey to a still deeper worry. Let us think of the CC this way: Its satisfaction requires knowledge of the counterfactual robustness of $G$ with respect to $H$, and the SR is not licensed if it is known that $G$ness is not counterfactually robust with respect to $H$. Consider again how Jackson claims to resolve the grue paradox. He writes:

If we use the SR with the evidence that $a_1, \ldots, a_n$ are green and examined, and grue and examined, ignoring the fact that $a_{n+1}$ is unexamined, we get support for ‘$a_{n+1}$ is green and examined’ and for ‘$a_{n+1}$ is grue and examined’; which, far from being inconsistent, are equivalent. If we bring in the fact that $a_{n+1}$ is unexamined, we no longer are dealing with a case of certain $F$s being $G$ supporting other $F$s being $G$, but of certain $F$s which are $H$ being $G$ supporting certain other $F$s which are not $H$ being $G$; and, hence, must take note of the counterfactual condition. But if we take note of this condition, we do not get an inconsistency because – although $a_1, \ldots, a_n$ would still have been green if they had not been examined – they would not have been grue if they had not been examined.”

My worry is this: Jackson smuggles into the application of his Counterfactual Condition knowledge that we are not entitled to, namely, that greenness of the observed emeralds is counterfactually robust whereas grueness is not. In particular, Jackson asserts that we know that the CC is violated (so the SR can be applied) in the case of green emeralds, for if the examined emeralds had not been examined, they would still have been green, whereas the CC is satisfied (and the SR blocked) in the case of grue emeralds, for if the examined emeralds had not been examined, they would not have been grue, since they would have been green and unexamined. But here is the rub: How is it we are able to evaluate the CC with regard to the observed emeralds without supposing knowledge of their counterfactual properties, which in turn depends upon their actual properties, in particular that they are green and not grue? Jackson appeals to background information about the emeralds that have already been examined, namely that ‘they are

---

7 Jackson, ‘Grue’, pp. 124–125, my emphasis.
green not because they have been examined but because of their chemical composition and crystalline structure.\textsuperscript{8} While that is certainly the reasoning we do use when thinking about the grue paradox, we must bear in mind the paradox is one of justification, and we must ask whether we are entitled to that knowledge, or what it would take to be so entitled.

We would, for instance, be entitled to that knowledge that the observed emeralds are green and not grue if we knew, for instance, that all emeralds are green and not grue. But that is precisely the sort of claim that induction is supposed to establish, not one it is to rely upon.

Jackson recognizes this, for he considers a circularity objection at the end of ‘Grue’. It differs from my own, and exploring it will help clarify my objection. Jackson acknowledges the potential objection that the account he gives for emeralds being green and not grue is circular, because his claim that emeralds in the observation base would have been green even if they had not been examined is one that is justified because we know that unexamined emeralds are green. He counters, however, that we have this knowledge not because we have knowledge about unexamined emeralds, but that we have it because we know something about examined emeralds: namely that they would have been green even if they had not been examined. Our knowledge about the counterfactual properties of examined emeralds is thus independent of the nature of unexamined emeralds, and circularity is avoided. Jackson writes:

It follows that our knowledge that the examined emeralds would be green even if not examined does not tacitly rest on our knowledge that unexamined emeralds are green. It is knowledge we might have had even if unexamined emeralds were not green or, indeed, were nonexistent, and so, is knowledge we may appeal to without circularity in describing our application of the straight rule in a way that makes clear why we have support for unexamined emeralds being green rather than grue.\textsuperscript{9}

The form of circularity that Jackson considers is this: In an attempt to project to the nature of as yet unexamined emeralds,

\textsuperscript{8} Jackson, ‘Grue’, p. 124.

\textsuperscript{9} Jackson, ‘Grue’, p. 130.
an appeal to the nature of those very emeralds would be circular. However, since his appeal is to the nature of already examined emeralds, not the unexamined ones, he claims immunity to charges of this type of circularity: 'the knowledge required is not the knowledge at issue in the particular application of the SR in question, and so it is not circular to appeal to it.' However, the form of circularity with which I am charging him is different. My claim is that appeal to the counterfactual knowledge of the nature of the already-examined emeralds is what leads to circularity: in the context of worries about induction, the problems I raised for application of the CC cannot be addressed without circularity by appeal to knowledge of what the color of examined emeralds would have been had they not been examined. The background knowledge Jackson appeals to is precisely the knowledge at issue, for it is not the knowledge required in applying the SR, but rather the knowledge required in applying the CC that is the problem.

Can the CC be evaluated purely on the basis of knowledge about observed emeralds? It appears not. Not on the basis of their observation alone can Jackson claim knowledge of their counterfactual properties, namely that 'they are green not because they have been examined but because of their chemical composition and crystalline structure.' This claim appeals to knowledge about emeralds as a class (in particular, to knowledge of their microstructure and its color-consequences). And, if we begin with the knowledge that all emeralds have a microstructure $m$ that makes them green, the grue problem never arises, for not only do we have the information that emerald $a_{n+1}$ is unexamined, but also that emerald $a_{n+1}$ has microstructure $m$ and is therefore green and not grue. And since the grue problem doesn’t arise, we don’t need the CC to solve it.

What is available to us through observation of the emeralds in our inductive base is their color and perhaps their microstructure. That alone entails nothing relevant to solving the paradox, for what we know is merely that all emeralds in the inductive base have microstructure $m$ and look green. In order to assess the counterfactual to infer that the CC was not satisfied, we would need to know other things. Namely, we would need to know (1)
how structure determines color, and (2) that objects with that structure are green and not grue. Among other things, requiring (1) implies that before we had methods for investigating the microstructure of emeralds we would not have been justified in taking the next emerald to be green and not grue, which seems absurd. Surely the problem of induction cannot be considered to be dissolved if we can only be justified in making inductive generalizations about predicates for which we possess a microstructural or otherwise reductive account (which would allow our prediction of the color of the next emerald to rest on deduction, rather than induction). Indeed, we even now have no such account for color. And secondly, in the absence of a complete reductive account, it is our prior observations of objects with microstructure $m$ which provide the basis for our belief in the statement ‘objects with such-and-such a chemical and crystalline structure are green.’ But all our evidence to date is equally consistent with the hypothesis that objects with microstructure $m$ are grue and not green. And if all objects with microstructure $m$ are grue, we can deduce that emeralds are too.

If we admit that we are not entitled to the knowledge that grounds the counterfactual claim that the already observed emeralds would have been green if not examined and not grue, then we are unable to resolve the paradox. To illustrate, suppose we gave equal credence to the hypotheses that the already examined emeralds are green, and that they are grue. How then to evaluate the CC? If the examined emeralds are green, then they would have been green even if not examined, and the CC is violated. So we would be permitted to invoke the SR and predict that the next emerald will be green. But if they are grue, then they would have been grue if unexamined, and thus blue. Again, the CC would be violated, so we could invoke the SR and predict that the next emerald will be grue.

What we take to be the counterfactual properties of objects in our evidence base depends upon what we take their actual properties to be, and that is what the new riddle shows to be problematic.

To continue along these lines, suppose we antecedently believed that the observed emeralds were grue instead of green. Such a belief is consistent with all our evidence: to date, all emeralds examined would look just as they always have – they look green, and thus grue. Would the CC prohibit us from inferring that the next emerald would also be grue? Indeed no, the CC would not be satisfied, for, since all predicates are projectible, we would conclude that the very emeralds which we have examined...
would have been grue if they had not been examined – that is, they would have been blue. Indeed, it is both logically possible and consistent with all our evidence that all unexamined emeralds are blue. What is more, in cases in which we antecedently believed that observed emeralds were grue, the CC would prevent us from applying the SR in the case of greenness, preventing the inference that the next emerald we examine will be green. So although Jackson’s CC succeeds in preventing grue-type paradoxes from arising, we see that it provides equal license for mutually incompatible predictions, unless we appeal to other knowledge. Which of these predictions is allowed depends solely on our prior beliefs (or, according to Jackson, our knowledge), but we see now that these prior beliefs themselves lack independent justification: the hypotheses that all examined emeralds are green and that all examined emeralds are grue are both equally supported by our previous observations, and no observational evidence gives us insight into their counterfactual properties with regard to observation. If we are mistaken at the outset in our beliefs about how the world is, our inductive practices will consistently lead us astray.

Given that the new riddle of induction seems to reduce to a problem about counterfactual robustness, the best strategy for resolving the puzzle seems to be to provide arguments for why some predicates are counterfactually robust in the relevant respects while others are not. Jackson claims that we know the observed emeralds would still have been green if they had not been examined, whereas they would not have been grue if they had not been examined, but I have questioned the legitimacy of this knowledge. Imagine what would be the case if we lived in a grueish world – if we adopted Jackson’s solution we would be consistently wrong about our actual judgments that emeralds are green and not grue, and the proper inductive move would be to project grueness, not greenness, of emeralds.13 The grueish world looks just like our world, so how do we know that we don’t live in a grueish world?14

13 Bayesians may argue that this problem is not a problem for them, because Bayesian inferences ultimately converge to the same answers regardless of the distributions of priors one has. While this may be so, appeal to Bayesian principles is not likely to resolve this issue as long as T remains in the future, because Bayesian updating will strengthen one’s credence in one’s antecedent beliefs about greenness or grueness, regardless of the nature of emeralds. The Bayesian answer will hold only if one updates on observations for long enough, where long enough already entails a non-Bayesian solution to the grue problem.

14 This is not to say that I take this problem to be a metaphysical problem about the nature of the world. This is an epistemological problem, but one way in which to illustrate
As Jackson notes, the problem of counterfactuals is a thorny one. Although observation can often provide insight into how to evaluate counterfactuals, observation itself cannot ground counterfactual knowledge. Logically speaking, neither the claim that emeralds are green or grue should command greater credence. Now one might be tempted to think that we are entitled to opt for the green over the grue possibility for various straightforward reasons. For instance, we might opt for greenness because green is a well-behaved or projectible predicate, while grue is not. This would be a mistake. First, Jackson himself cannot avail himself of this option because the need for it arises in a context where he is arguing that there is no such thing as a nonprojectible predicate. Secondly, I am convinced by Jackson’s arguments that all predicates are projectible, in that all license, in the absence of countervailing reasons, application of the SR. One might instead think that it may be sufficient to ground our belief that emeralds are green on the basis of our antecedent beliefs about emeralds being green, but, as noted above, that puts us on a road to skepticism.

It is tempting to argue that we know that green is counterfactually robust because it is observation-independent, whereas grueish colors are observation-dependent. After all, only grueish colors predict a shift in color relative to T. Unfortunately, this move won’t work – it is just to overlook the symmetry of normal and grueish colors noted by Goodman in his original chapter: green and blue can be defined in terms of grue and bleen, and given this linguistic basis, green and blue, and not grue and bleen, are the observation-dependent predicates, and grue and bleen would then be the counterfactually robust ones. One cannot on this basis dismiss the problem as merely one of linguistic variants, however, for they license different predictions about the nature of the objective world: our choice of language should not make it the case, in one instance, that the next emerald we see will look green, and in the other, blue. Moreover, this move seems to ignore that the notion of counterfactual robustness needed for application of the CC is not something that applies to a predicate across the board: in some cases color seems to be counterfactually robust, in

lack of justification is to show that we could make the very same set of observations in a grueish world, and our reliance on Jackson’s CC would consistently direct us to make the wrong inductions.
The relevant notion of counterfactual robustness is robustness with respect to some other property or properties, so no argument about the observation-dependence or response-dependence of a property full-stop appears to be a promising avenue of criticism.  

Thus, the problem with Jackson’s solution to the grue paradox is that application of his counterfactual condition requires appeal to knowledge that application of the condition is supposed to justify. This is a pernicious form of circularity, in the absence of independent arguments to shore up our intuitions that greenness with respect to observation is counterfactually robust whereas grueness is not. So the question arises again: on what basis are we entitled to think that emeralds would have been green and not grue had they not been observed? For had the observed emeralds not been observed, if they would have been grue they would not have been green, and conversely. This is not a question about induction, in other words, a question about what the next emerald will be like – it is a question of counterfactual robustness. In what cases are we licensed to conclude that we know that the $F$ in the evidence base would not have been $G$ if they had not been $H$? Applicability of the SR to novel instances may provide information about counterfactual robustness, but since we are only justified in using the SR if we antecedently know the robustness of the predicate, we find ourselves in a tight circle.

I, as much as Jackson, am convinced that in using induction in everyday life we do, and do effectively, appeal to background knowledge in deciding when applying the SR, and our reasoning often if not always tracks reasoning articulated by the CC. However, we must recall that the problem of induction is a problem of justification. In the context of questions about justification of principles of induction, we must press further than we tend to do in everyday contexts, and question whether the background knowledge is knowledge that we are entitled to use. The claims which Jackson employs to ground the counterfactual are themselves subject to some version of the grue problem, and, in the case of color, to the very same version of the problem that appeal to the CC is supposed to solve. So in applying the CC, we implicitly make use of knowledge we are not entitled to. Although Jackson avoids the criticism of tight circularity because he does

\[ \text{I thank an anonymous reviewer for bringing this to my attention.} \]
not appeal directly to a general statement about the color of emeralds in order to assess the CC, he is open to charges of a looser circularity when he appeals to specific statements about the nature of already examined emeralds. In doing so he appeals indirectly to knowledge about unobserved emeralds: either to general claims based on implicit knowledge that all emeralds, not just the observed ones, are green, or, more worrying still, to other general statements equally vulnerable to the grue paradox as our original statement about the color of emeralds. Either way, assessment of the CC involves a circular appeal to the knowledge it is meant to yield.

There are, no doubt, other avenues to consider; for instance, it may be that a noncircular account of justification of counterfactual knowledge of the type required for application of the the CC can be given by relying upon background knowledge justified by the coherence of entire scientific theories, rather than the truth about particular claims about the nature of things. My object here was to demonstrate that Jackson’s solution to the new riddle of induction is at best incomplete, and at worst ultimately fails. However, in doing so it points at a deeper issue, one for which we as yet have no satisfactorily articulated response. Problems of induction are problems of justification. Goodman’s new riddle was about understanding the rationale for sorting predicates into two classes: the projectible and the nonprojectible. Jackson argued convincingly that predicates do not fall into these two classes, so there is no rationale to supply. However, to apply Jackson’s counterfactual condition, and thereby prevent ourselves from faulty inductions, we already need to have resolved a riddle, though not the classical form of Goodman’s riddle. Rather, it is the epistemological problem that arises in justifying, in a given context, which of the infinitely many predicates consistent with our evidence are the counterfactually robust ones, and thus the appropriate ones to use as an inductive base. If green is counterfactually robust, then grue is not, and vice versa. But which one is the correct one to apply? If our inductive rules of inference equally sanction the generalizations that emeralds are grue and that they are green, how can we justify our supposed knowledge that they are green and not grue? Without some answer to this

16 See, for instance, Peter Godfrey-Smith’s article, ‘Goodman’s problem and scientific methodology,’ The Journal of Philosophy, 100 (2003), 573–590.

© 2008 The Author
Journal compilation © 2008 Blackwell Publishing Ltd
riddle about justification, use of Jackson’s counterfactual condition is subject to the same kinds of epistemological worries as are our intuitions about the lawlikeness of ‘green’ and the failure of such with ‘grue’. If so, grue-type paradoxes can still arise.17

Department of Philosophy
Dartmouth College
Hanover, NH 03755
adina.roskies@dartmouth.edu

17 I am indebted to David Braddon-Mitchell, Peter Godfrey-Smith, Frank Jackson, Walter Sinnott-Armstrong and Roy Sorensen, as well as to two anonymous reviewers, for their insightful comments. Kim Sterelny enabled me to spend time at the ANU with a grant from the Australian Research Council.