



Okasha, S. (2007). What does Goodman's 'Grue' Problem Really Show?: re-reading of Frank Jackson, 'Grue', *Journal of Philosophy* 5 (1975), pp.113-131. *Philosophical Papers*, 36(3), 483-502. 10.1080/05568640709485211

Link to published version (if available):
[10.1080/05568640709485211](https://doi.org/10.1080/05568640709485211)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/pure/about/ebr-terms.html>

Take down policy

Explore Bristol Research is a digital archive and the intention is that deposited content should not be removed. However, if you believe that this version of the work breaches copyright law please contact open-access@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline of the nature of the complaint

On receipt of your message the Open Access Team will immediately investigate your claim, make an initial judgement of the validity of the claim and, where appropriate, withdraw the item in question from public view.

What Does Goodman's 'Grue' Problem Really Show?

Samir Okasha

Re-reading of Frank Jackson, 'Grue', *Journal of Philosophy* 5 (1975), pp.113-131.*

Introduction

Nelson Goodman devised his famous 'grue' example in order to show that the simple 'straight rule' of induction leads to inconsistency unless restricted to the so-called 'projectible' predicates. In his searching critique of Goodman's discussion, Frank Jackson (1975) argued that on careful examination, Goodman's alleged paradox dissolves—the naïve straight rule does not lead to inconsistency, and does not need to be restricted to predicates of any particular type. Jackson's article seems to have been overlooked by much of the subsequent literature on 'grue'; indicative of this is that virtually all philosophers today believe that Goodman *did* show what, according to Jackson, he did not show.¹

The structure of this paper is as follows. Section 1 briefly rehearses Goodman's main arguments regarding 'grue', to set the stage for Jackson's critique. Sections 2 to 5 expound Jackson's argument, commenting on a number of points. Section 6 discusses the consequences of Jackson's position, and offers some suggestions about

* *Editor's Note*: 'Re-Readings' is a regular feature in *Philosophical Papers*. Authors are invited to write on a past article, book, or book chapter that they deem, for whatever reason, to deserve renewed attention. Authors are encouraged, where appropriate, to discuss the work's reception by and influence upon the philosophical community.

¹ Though an important recent exception is P. Godfrey-Smith, 'Goodman's Problem and Scientific Methodology', *Journal of Philosophy* 2003, pp. 573-590; see the discussion in Section 6 below.

why Goodman's discussion of 'grue' has so frequently been thought to contain a deeper philosophical moral than in fact it does.

1. The Straight Rule of Induction

In *Fact, Fiction and Forecast*, Nelson Goodman argued that the project of trying to *justify* inductive reasoning—the 'old problem of induction'—was fundamentally misguided, and should be replaced by the project of trying to *describe* inductive reasoning, i.e., of specifying which inductive inferences we regard as rational and which not. The natural starting point for this more modest descriptive project, Goodman argued, is the so-called 'straight rule' of induction, i.e., arguing from 'some Fs are G' to 'all Fs are G' or 'the next F will be G'. (Following Jackson, I use 'SR' as an abbreviation for 'straight rule'.) Since Hume, this argument pattern has often been regarded as *the* basic schema for inductive inference, covering many of the actual inferences we make in everyday life. However, Goodman argued that the straight rule, in its unqualified version above, provides a very imperfect description of real-life inductive practice, and in fact leads to paradox. He called this the 'new riddle of induction'.

The paradox arises, Goodman claims, because from equivalent evidential bases, the straight rule can lead us to incompatible predictions. More precisely, from the logically equivalent statements 'all observed emeralds are green' and 'all observed emeralds are grue'—where 'grue' is defined as 'green and observed or blue and unobserved'—the straight-rule leads to the predictions 'the next emerald to be observed will be green', and 'the next emerald to be observed will be grue', which are incompatible. Unrestricted application of the straight rule therefore leads to inconsistency.

Goodman's own solution to this supposed paradox was to restrict the class of predicates to which the straight-rule can be applied. Predicates such as 'green' are *projectible*, he argued, while ones such as 'grue' are unprojectible. The problem of describing our inductive reasoning—at least in so far as it follows the 'straight rule' pattern—

thus becomes the problem of distinguishing the projectible from the unprojectible predicates. As Jackson notes, the extensional aspect of this problem has usually been regarded as straightforward: most philosophers, following Goodman, have assumed that we know which predicates belong in the projectible class and which not. However, there has been widespread disagreement over the *intension* of this class, i.e., over what *makes* a predicate projectible. Goodman's own solution—that 'entrenchment' is what makes a predicate projectible—has met with widespread criticism, and alternative theories of projectibility abound.² On two points there is near universal consensus, however: firstly that there *is* a projectible/non-projectible distinction, and secondly, that particular instances of the straight-rule inference schema are only rationally acceptable if 'F' and 'G' are replaced by projectible predicates.

Jackson launches a forthright attack on this consensus. He writes: 'I will argue [that] there is no "new riddle of induction", by arguing that all (consistent) predicates are projectible and that there is no paradox resulting from "grue" and like predicates' (114). So in short, Jackson argues that the philosophical mainstream has made a bad mistake in its appraisal of Goodman's ideas. He traces this mistake to three sources: 'one, a tendency to conflate three different ways of defining "grue"; two, a lack of precision about just how, in detail, the "grue" paradox or new riddle of induction is supposed to arise; and three, a failure to note a counterfactual condition that governs the vast majority of our applications of the SR'. I turn to each of these three points in turn.

2. Jackson on the Definition of 'Grue'

Jackson identifies three different ways of defining 'grue', each of which is found in the literature. The first two do not even pose a *prima facie* problem for the straight rule, he argues, and can be dealt with swiftly;

² See for example the papers in D. Stalker (ed.) *Grue*, Chicago: Open Court, 1994.

the third does provide a *prima facie* problem, but it dissolves on closer examination.

An example of the first type of definition is:

D₁: x is grue iff x is green before T and blue thereafter

where T is a chosen time in the future. As Jackson notes, on this definition 'grue' is an atemporal predicate, unlike green: an object cannot be grue at one time and fail to be grue at another time.

Is 'grue' as defined by D₁, or 'grue₁' for short, an unprojectible predicate? Jackson argues, surely correctly, that the answer is 'no'. To be grue₁, an object must change from being green to blue at time T. If it turned out that all examined objects of a certain type, e.g., emeralds, did indeed change colour at time T, we would most likely conclude that there was some reason for this, and thus that the emeralds we have not examined also underwent a colour change at time T. Emeralds would then simply be objects that undergo a colour change during their life-cycle, like tomatoes, the only difference being that they all change colour at one particular point in time. So although grue₁ denotes a somewhat odd property, it seems perfectly projectible.

An example of the second type of definition is:

D₂: x is grue at t iff (x is green at t & t < T) or (x is blue at t & t ≥ T)

As Jackson notes, grue as defined by D₂—grue₂ for short—is like green in being a temporal predicate—an object may be grue₂ at one time and not at another.

Does the predicate grue₂, when used with the straight rule of induction, lead to a paradox? According to Jackson, the answer is 'no'. The widespread view to the contrary, he argues, arises from a confusion about whether 'grue₂' is supposed to apply to enduring objects or to time-slices of enduring objects.

Suppose we are doing straight-rule induction in relation to enduring objects. We must then temporally index the predicates that we are projecting. For enduring objects don't have the property of being green

simpliciter, but rather being green at a particular time. Only if this point is overlooked, Jackson argues, does the appearance of paradox arise when we use grue_2 in conjunction with the SR.

To see this, suppose the emeralds in our sample are all green at t_1 , where t_1 is before T; by definition, they are also grue_2 at t_1 . Applying the SR, we get ‘all emeralds are green at t_1 ’ and ‘all emeralds are grue_2 at t_1 ’; but as Jackson points out, these universals are perfectly compatible. On the other hand, for any time t_2 later than T, it is impossible that the sampled emeralds be green at t_2 and also grue_2 at t_2 ; so there is no way that the SR will lead us to the incompatible universals ‘all emeralds are green at t_2 ’ and ‘all emeralds are grue_2 at t_2 ’. Jackson writes: ‘only if we start from the fact that the sampled emeralds are both green at t_1 and grue at t_1 , and then, by conflating being green (grue) at t_1 with being green (grue) at t_2 , wrongly take the SR to provide support equally for the incompatible “all emeralds are green at t_2 ” and “all emeralds are grue_2 at t_2 ”, do we obtain an apparent paradox’ (117).

Jackson next notes that although D_1 and D_2 figure prominently in the ‘grue’ literature, Goodman’s own definition of grue was different. Goodman was concerned with predicates of the form ‘(x is green and Φx) or (x is blue and $\neg\Phi x$)’, where ‘ Φx ’ is so defined that its extension includes all the sampled, or observed, emeralds, while the extension of $\neg\Phi x$ includes all the other emeralds, i.e., those which we haven’t sampled. A convenient way of achieving this is to introduce a temporal factor into Φx , e.g., by defining it as ‘x is examined before the present’, but this is inessential; the essential point is just that Φx must be true of all and only the objects in the sample from which we are projecting. Thus one simple definition of grue, identical in all relevant respects to Goodman’s own, is:

D_3 : ‘x is grue at t iff x is green and examined, or blue and unexamined’

where ‘examined’ means ‘examined to date’, and ‘unexamined’ means ‘unexamined to date’. As is clear from the presence of ‘at t’ in D_3 , the

definition applies to enduring objects, not temporal parts of objects; but as Jackson notes, it is possible to conduct the discussion in the editorial present (as Goodman does), thus eliminating the need for the ‘at t ’. Grue as defined by definition D_3 raises more problems than those raised by either D_1 or D_2 , Jackson argues, and is the definition on which the literature should have focused. In what follows, ‘grue’ will be taken to mean ‘grue as defined by D_3 ’.

3. How Does the Supposed Paradox Arise?

Recall what the ‘grue’ paradox is supposed to be. Goodman claimed that the straight-rule of induction could lead to *incompatible* conclusions, from the very same evidence, depending on how the evidence is described. Specifically, if the evidence is described using the predicate ‘green’ we reach one conclusion; if the same evidence is re-described using ‘grue’, we reach another, incompatible conclusion.

Jackson urges, correctly, that we must play close attention to how *exactly* Goodman reaches this result. Jackson considers a series of emeralds $a_1 \dots a_n, a_{n+1}$, such that $a_1 \dots a_n$ are known to be green and examined; a_{n+1} is known to be unexamined, and is the emerald whose colour we are trying to predict. Following Jackson, we let ‘Grx’ stand for ‘x is green’, ‘Ex’ for ‘x is examined’, ‘Bx’ for ‘x is blue’ and ‘Gux’ for ‘x is grue’. So by definition, ‘Gux’ = ‘(Grx & Ex) \vee (Bx & \neg Ex)’. Since each of $a_1 \dots a_n$ are known to be green, we have:

$$(1) \quad \text{Gra}_1 \& \dots \& \text{Gra}_n \quad (\text{‘emeralds } a_1 \dots a_n \text{ are green’})$$

We also have:

$$(2) \quad \text{Gua}_1 \& \dots \& \text{Gua}_n \quad (\text{‘emeralds } a_1 \dots a_n \text{ are grue’})$$

If we apply the straight rule to (1), we get the prediction Gra_{n+1} (‘emerald a_{n+1} is green’); while if we apply the straight rule to (2), we get the prediction Gua_{n+1} (‘emerald a_{n+1} is grue’). However, *this* can be hardly be Goodman’s point, for two reasons. Firstly, the two predictions Gra_{n+1} and Gua_{n+1} are not inconsistent, for neither entails the denial of

the other; and secondly, the two evidence statements (1) and (2) are not logically equivalent. So far, therefore, we do *not* have a case where the same evidence statement, described in two ways, licenses two incompatible predictions via the SR, as Goodman claimed. Rather, we have two *different* evidence statements licensing two different predictions, which are in any case not incompatible.

Jackson also points out, following Carnap, that neither (1) nor (2) embodies our total evidence; specifically, they do not incorporate the fact that the emeralds $a_1 \dots a_n$ have been examined. So our total evidence is:

- (3) $Gra_1 \& Ea_1 \& \dots \& Gra_n \& Ea_n$ ('emeralds $a_1 \dots a_n$ are green and examined')

From the definition of 'Gux', this is equivalent to:

- (4) $Gua_1 \& Ea_1 \& \dots \& Gua_n \& Ea_n$ ('emeralds $a_1 \dots a_n$ are grue and examined')

So statements (3) and (4) are logically equivalent descriptions of our total evidence, one in terms of 'green', the other in terms of 'grue'. Direct application of the SR to (3) and (4) leads respectively to:

- (5) Gra_{n+1} and Ea_{n+1} ('emerald a_{n+1} is green and examined')

and

- (6) Gua_{n+1} and Ea_{n+1} ('emerald a_{n+1} is grue and examined')

But as Jackson points out, (5) and (6) are not incompatible predictions; on the contrary, they are actually logically equivalent. So there is still no hint of paradox.

The natural objection to Jackson's argument to this point is that he hasn't taken account of an important fact, namely that emerald a_{n+1} is *not* examined, i.e., the fact that $\neg Ea_{n+1}$. This piece of information is surely crucial to Goodman's puzzle; for it is only when we take into account that a_{n+1} is given as unexamined, that we can infer that, if a_{n+1} is grue then it

must be blue. In short, Goodman's idea seems to be that we get paradox in the following way. We use the SR to infer that a_{n+1} is green, and that it is grue; we then invoke the fact that a_{n+1} is unexamined—a fact which, when combined with ' a_{n+1} is grue', implies ' a_{n+1} is blue', which contradicts ' a_{n+1} is green'. Jackson is aware that this is how Goodman wants to proceed; however, he insists that the logic of this argument must be spelled out precisely.

Jackson suggests that Goodman is in effect reasoning as follows:

Our evidence supports a_{n+1} is green and examined. We know independently that a_{n+1} is not examined, hence our over-all evidence supports a_{n+1} is green and not examined. Equally, as far as the SR goes our evidence supports a_{n+1} is grue and examined, and so, via the same line of argument, we arrive at our over-all evidence supporting a_{n+1} is grue and not examined, which entails that a_{n+1} is not green (121).

However, Jackson then points out the underlying pattern of argument here is clearly fallacious. The pattern is: 'a proposition p , which we know to be true, supports a conjunction, q and r , one conjunct, r , of which we know independently to be false; we have, overall, support for q and $\neg r$, and so, for anything it entails.' Valid instances of this schema can no doubt be found, but it is not generally valid; for from the fact that p supports a conjunction one of whose conjuncts, r , we know to be false, we might well decide that, overall, we do *not* have good reason to believe q and r , despite the support provided for that conjunction by p . There can be no principle that dictates that, in such a situation, we *do* have reason to believe q and $\neg r$, for this ignores the familiar fact that inductive support is defeasible. The moral Jackson draws is that 'we must ... proceed *very* carefully when attempting to incorporate the additional information that a_{n+1} is not examined' (122).

4. The Counterfactual Condition

Jackson has now set the stage for the crucial move in his argument. The epistemic situation we are in is as follows. We have a sample of emeralds, $a_1 \dots a_n$, each of which has a certain property, namely greenness, that we

are interested in projecting to the emerald a_{n+1} . However, each of $a_1 \dots a_n$ also has *another* property, namely being examined, that we know is not possessed by a_{n+1} . So the type of straight-rule inference we are dealing with does not simply go from certain Fs being G to certain other Fs being G—which is perhaps the simplest version of the SR. Rather, the inference goes from certain Fs, which are also H, being G, to certain other Fs, which are not H, being G. (In Goodman's example, 'F' stands for emeralds, 'H' for examined, and 'G' for green.)

Jackson notes that most real applications of the straight-rule have this more complicated structure—for in real-life cases, there are almost always features common to all members of our sample, which we know not to be possessed by members of the population outside the sample. (Trivially, 'being in the sample' is one such property.) So the question arises, under what situations are inferences of this form reasonable? When can we legitimately infer from certain Fs being both G and H, to certain other Fs, which are known not to be H, being G? Jackson proposes the following necessary condition on the legitimacy of this inference. If we know that the Fs in the evidence class would not have been G if they had not been H, then we cannot reasonably infer that the Fs which are not-H will also be G. He calls this the 'counterfactual condition', and claims that it governs all uses of the straight rule.

Jackson's illustrates the counterfactual condition with a simple example. All of the lobsters he has observed have been red; but he does not take this to support the conclusion that the lobsters in the sea, which he hasn't observed, are also red. Why not? Because the lobsters in his evidence sample—the observed lobsters—have all been cooked, and he knows that cooking makes lobsters red. Had the lobsters in the evidence sample not been cooked, they would not have been red; so the evidence does not support the lobsters in the sea, which are uncooked, being red. The inference thus fails because the counterfactual condition is violated. There is a property H possessed by all the Fs in the evidence sample, which is such that, if they hadn't possessed it, they wouldn't have been G; so the evidence does not support the conclusion that other Fs, which do

not possess H, are also G. This explains why we do not accept the inference in question.

Armed with the counterfactual condition, Jackson can return to the issue of how to incorporate the additional information that a_{n+1} is unexamined. Consider firstly the inference from emeralds $a_1 \dots a_n$ being green and examined, to the unexamined emerald a_{n+1} being green. In this case, the counterfactual condition is satisfied. Had the emeralds $a_1 \dots a_n$ not been examined, they would still have been green—given that examining an emerald does not generally affect its colour. So while the $a_1 \dots a_n$ do share a feature—being examined—that a_{n+1} is known to lack, we know that the property we are projecting—being green—does not depend on this feature. So the straight-rule can legitimately be used to infer that emerald a_{n+1} is green.

By contrast, if we are projecting ‘grue’ then the counterfactual condition is violated. Each of the $a_1 \dots a_n$ possesses the property of being grue and examined. But we know that, had these emeralds not been examined, they would not have been grue, given that we live in a world where examining emeralds doesn’t alter their colour. Had the $a_1 \dots a_n$ not been examined, they would have been green and unexamined, hence not-grue. So if we use the straight rule to infer that a_{n+1} is grue, we violate the counterfactual condition. We ignore the fact that the objects in our sample possess a property—being examined—which is not possessed by other objects in the population; and which is such that, if the objects in our sample had not had it, they would not have had the property—grue—that we wish to project.

What exactly does this show? The main moral that Jackson draws is simply that the projectible/unprojectible distinction is unnecessary, at least in so far as we are trying to characterize rational inductive inference. He writes; ‘it is a mistake to argue from “all examined emeralds are grue” to “all unexamined emeralds are grue”, not because “grue” is intrinsically non-projectible, but simply because the counterfactual condition is violated’ (125). So there is no need to restrict the SR to the so-called projectible predicates; the rule can be applied to

any predicate so long as the counterfactual condition is satisfied, and can never be applied where that condition is not satisfied.

In support of this diagnosis, Jackson points out that there are many cases where we wouldn't infer from 'all examined As are B' to 'all unexamined As are B', precisely because we know that the act of examining the As is what causes them to be B, even though 'A' and 'B' are quite ordinary predicates. The lobster example is one such; another is afforded by quantum physics, which teaches us that certain (perfectly real) properties of elementary particles are affected by being examined. For such a property, we obviously would not take the fact that all examined particles have the property to support the unexamined particles having the property too. So there is clearly much to be said for Jackson's thesis that where the counterfactual condition is violated, we refrain from apply the straight-rule, even when dealing with perfectly ordinary, non-gerrymandered, predicates.

If Jackson is right that it is violation of the counterfactual condition, not the inherent unprojectibility of 'grue', that leads us to refrain from using grue in straight-rule induction, then it should presumably be possible to describe a hypothetical case where we *are* prepared to project grue, because the counterfactual condition *is* satisfied. Jackson describes just such a case. Suppose that we lived in a world where all examined emeralds are green, but investigation of their crystalline structure shows that they are naturally blue; however, the structure is affected by light, in such a way that the act of examining the emeralds makes them go blue. In this world, as in the actual world, all examined emeralds are grue. But in this world we *would* infer that unexamined emeralds are grue; for we know that, if the emeralds in our sample had *not* been examined, they would still have been grue—for then they would have been unexamined and blue. So the counterfactual condition is satisfied, which explains why we make the inference. (Conversely, in the imagined world, we would not project 'green' to the unexamined emeralds, for we know that had the emeralds in our sample not been examined, they would not have been green.)

If Jackson's arguments are correct, as they appear to be, it follows that in order to characterise the class of straight-rule inferences that we intuitively take to be rational, it is unnecessary to invoke a projectible/unprojectible distinction; the counterfactual condition does the job instead. In fact, a stronger conclusion follows: we *must not* appeal to the projectible/unprojectible distinction, on pain of mischaracterising the class of rational SR inferences. For as Jackson's examples show, there are cases where we *do* use the SR even with predicates that would traditionally be thought non-projectible; and there are cases where we don't apply the SR even with perfectly kosher predicates. In both types of case, it is the counterfactual condition that explains whether or not we regard the application of the SR as rational, not the status of the predicates.

Might someone object to Jackson that the counterfactual 'if the emeralds in our sample had not been examined, they would still have been green', is *not* something that we know to be true; or at least, that this knowledge cannot legitimately be appealed to in the context of Goodman's riddle? In order to assess this matter, it is important to remember that Goodman's project, and Jackson's, is to *describe* the class of inductive inferences that we take to be rational. The question of what *justifies* us in making those inferences is quite different—and was regarded by Goodman, rightly or wrongly, as a pseudo-question. (Recall Goodman's distinction between the 'old' and 'new' riddles of induction.) It was in the context of his descriptive project that Goodman raised the 'grue' problem; it was meant to show that the SR provides a very inadequate description of our inductive practices. So there is nothing *per se* wrong with Jackson's appealing to inductively-acquired knowledge—such as that examining gemstones does not generally affect their colour—as there would be if he were concerned with questions of justification.

A related point is that although Jackson states the counterfactual condition in terms of knowledge, it could in fact be stated in terms of *belief* and still perform the same role in his argument. To see this, imagine a

person who believes that the possible world described above, in which emeralds in their natural state are blue but the act of examining them makes them go green, *is* the actual world. This belief is obviously wildly false; but given that the person in question holds it, it would obviously not be rational for them to infer that unexamined emeralds are green, while it would be rational for them to infer that unexamined emeralds are grue. (This illustrates the familiar point that which inductive inferences it is rational to make, from a given data set, depends on one's background beliefs; but whether those beliefs are themselves rational is a separate issue.) Therefore, in order to characterise those applications of the SR that we regard as rational, what matters is that the agent *believe* that the counterfactual condition is satisfied; whether, in addition, she knows it to be satisfied does not matter.

5. The Threat of Inconsistency Defused

Goodman himself drew two conclusions from the 'grue' example: firstly, that we only regard applications of the SR as rational if it is restricted to projectible predicates; secondly, unless it is so restricted, the SR leads to inconsistency. Jackson's arguments, summarized above, successfully refute the first of these conclusions; what about the second?

The threat of inconsistency arises, to recall, because the statements 'all examined emeralds are green' and 'all examined emeralds are grue' are logically equivalent; and the SR, applied to each, leads us to 'unexamined emeralds are green' and 'unexamined emeralds are grue', which are incompatible. Now Jackson's counterfactual condition implies that the 'grue' inference is impermissible—given that we know the falsity of 'if the examined emeralds had not been examined, they would still have been grue'—which means that inconsistency is averted *in this case*. But what about other cases? Can a general argument be made that, so long as the counterfactual condition is satisfied, the SR will not lead us to make inconsistent predictions via a grue-type manoeuvre?

Jackson provides just such an argument, by exploiting a simple feature of the logic of counterfactuals. Consider again the modified SR

pattern: ‘some Fs are H and G, therefore other Fs, which are not H, are also G’. (Again, let ‘F’ be emerald, ‘H’ be examined, and ‘G’ be green, to get the standard example.) Now consider the predicate G^*x , defined as: $G^*x = (Hx \text{ and } Gx) \text{ or } (\neg Hx \text{ and } \neg Gx)$. Clearly, an object is H and G if and only if it is H and G^* , so (H and G) and (H and G^*) are logically equivalent renderings of our data. The threat of inconsistency arises because the straight-rule appears to license the prediction that the non-H Fs are G and that they are G^* - which conflict, since a non-H is G^* if and only if it is $\neg G$.

However, the counterfactual condition immediately blocks the inconsistency. For the statements ‘if the Fs that are H and G had not been H, they would have been G’ and ‘if the Fs that are H and G^* had not been H, they would have been G^* ’ cannot both be true. For this would amount to the joint truth of ‘ $p \Box \rightarrow q$ ’ and ‘ $p \Box \rightarrow \neg q$ ’, since a non-H is G^* if and only if it is $\neg G$. And on standard views about the logic of counterfactuals, the statements $p \Box \rightarrow q$ and $p \Box \rightarrow \neg q$ cannot both be true. (To illustrate, it cannot be true *both* that if the emeralds in our sample had not been examined they would have been green *and* that if they had not been examined they would have been grue.) So if we only apply the SR when the counterfactual condition is satisfied, grue-type predicates cannot lead us to inconsistency.

Of course, imposing the counterfactual condition does nothing to show that the SR cannot lead us to incompatible predictions starting from *different* data; indeed it clearly can. Goodman’s claim, however, was that from the *same* data, i.e., from two logically equivalent data statements, the SR leads to incompatible predictions unless restricted to ‘projectible’ predicates. It is this claim that Jackson is contesting.

6. Consequences

If we believe, as I do, that Jackson’s diagnosis of the grue problem is correct, what follows? The first and most important consequence is that the standard philosophical reactions to Goodman’s problem must be mistaken. Almost without exception, philosophers have accepted

Goodman's own diagnosis of the grue case, namely that it shows that the SR leads to inconsistency unless restricted to the so-called projectible predicates. (This is as true today as it was in 1975, when Jackson's article appeared.) Certainly, many philosophers have disagreed with Goodman over how the projectible/unprojectible distinction should be drawn—in particular, over whether 'entrenchment' provides a good account of projectibility—but that such a distinction is necessary in the first place is generally regarded as conclusively established by the grue example. If Jackson's arguments are correct, this consensus is wholly misplaced.

To my mind this is a welcome conclusion, for the very idea of a projectible/unprojectible distinction is in some ways quite odd, especially if it is assumed that we know which predicates are which. (This assumption is standard; as Jackson notes, the disagreements in the literature have concerned the intension, not the extension, of the class of projectible predicates.) Projectible predicates are supposed to be the ones suited for use in straight-rule induction, i.e., that can be rationally projected from samples to populations. The assumption that we know which predicates are projectible and which not, therefore, amounts to the assumption that we know the respects in which nature is uniform. However, the latter is not something that can be known *a priori*—our beliefs about it change from time to time. We used to think that the predicate 'obeys the laws of classical mechanics' held true of every object in the universe; now we know better. So the idea of a fixed partition of predicates into the projectible and unprojectible, constructed on *a priori* grounds, seems deeply unempiricist.

I suspect that the wide acceptance of the projectible/projectible distinction stems from the fact that similar distinctions have been thought necessary for other philosophical purposes. For example, in his well-known paper 'New Work for a Theory of Universals', David Lewis argues that a distinction between natural and unnatural properties, however hard to draw satisfactorily, is essential for understanding a host of philosophical topics, such as supervenience, materialism, mental content, causation and law. Natural properties, Lewis tells us, are ones

whose sharing makes for objective resemblance between objects, and which are relevant to their causal powers; indeed, he includes 'grue' in his list of unnatural properties. Lewis's view enjoys widespread support among contemporary philosophers; this contributes, I suspect, to belief in the projectible/unprojectible distinction.

However, the kinship between Goodman's notion of projectibility and Lewis's notion of naturalness, however strong, should not blind us to the fact that they use quite different arguments to motivate their respective distinctions. Therefore, accepting Jackson's arguments against Goodman's projectible/unprojectible distinction does *not* commit us to rejecting Lewis's natural/unnatural distinction. Jackson's arguments show only that the distinction is not necessary in order to characterise rational straight-rule induction, as Goodman had claimed. (Additionally, it is important to remember that Goodman's claim is much stronger than Lewis's. Goodman claims that the projectible/unprojectible distinction is needed in order to avert *inconsistency* when we do straight-rule induction. Lewis claims only that the natural/unnatural distinction is needed in order to facilitate the philosophical analysis of certain key notions.)

The supposed link between Goodman's projectibility and Lewis' naturalness is an instance of a more general trend in the literature, namely to assimilate the grue problem to other philosophical problems with which it is superficially similar; this in turn has detracted attention from the details of Goodman's actual argument, and reinforced the opinion that he was making a deep philosophical point. One egregious example of this assimilation is the idea, propounded in innumerable philosophy of science courses, that the 'grue' example is simply a dramatic way of raising the curve-fitting problem. The curve-fitting problem is the problem of how to choose between alternative hypotheses about the functional relation between two continuous variables, x and y , from finite data. As is familiar, an infinity of different curves will fit any finite number of data points, and yet make conflicting predictions concerning unmeasured values of x . In practice, scientists tend to favour simple over more complex curves where possible—but the justification

for this preference (and the precise meaning of ‘simple’) are points of ongoing philosophical controversy.

To see why it is wrong to equate the grue problem with the curve-fitting problem, consider a simple example. We have two data points: (4, 32) and (2, 16), and are trying to choose between the conflicting hypotheses $H_1: y = 8x$, and $H_2: y = x^2 + 2x + 8$, both of which fit our data, but make conflicting predictions about the value of y for any other values of x . According to a widely held view, this is analogous to trying to choose between ‘unexamined emeralds are green’ and ‘unexamined emeralds are grue’, in Goodman’s example. To probe this analogy, let us explicitly re-cast the curve-fitting problem in the format of a straight-rule inference. Consider a ‘rational’ scientist who opts for the simpler linear hypothesis H_1 . His inference may be represented as: ‘all examined cases satisfy the relation $y = 8x$, therefore all unexamined cases satisfy the relation $y = 8x$ ’. Of course, the inference of a ‘deviant’ scientist who opts for the more complex H_2 can be similarly represented: ‘all examined cases satisfy the relation $y = x^2 + 2x + 8$, therefore all unexamined cases satisfy the relation $y = x^2 + 2x + 8$ ’. And since the hypotheses $y = 8x$ and $y = x^2 + 2x + 8$ are logically incompatible, this *seems* similar to Goodman’s point: from the same data, the straight rule generates incompatible predictions. So opting for the simple hypothesis is analogous to projecting green, while opting for the complex hypothesis is analogous to projecting grue. The challenge of saying why we prefer the linear to the quadratic hypothesis is thus analogous to the challenge of saying why we project green but not grue.

However, a moment’s reflection shows this to be mistaken. For in the curve-fitting case, the premises of the two inferences, i.e., ‘all examined data points satisfy $y = 8x$ ’ and ‘all examined data points satisfy $y = x^2 + 2x + 8$ ’ are *not* logically equivalent. So the fact that they lead to incompatible predictions, via the straight-rule, is unsurprising. Crucially, Goodman’s claim was that from *logically equivalent* data statements, we get incompatible predictions via the straight rule. (Recall that the propositions ‘all examined emeralds are green’ and ‘all examined

emeralds are grue' are logically equivalent.) The fact that we can get incompatible predictions when we start with different data statements is hardly surprising. So it is quite wrong to assimilate the curve-fitting problem to the grue problem.

This point can be seen from another angle. When the curve-fitting problem is re-cast as a straight-rule inference problem, then of necessity, a logical weakening of the data occurs. In the example above, our total evidence is actually the proposition, 'when $x = 4$, $y = 32$, and when $x = 2$, $y = 16$ '. This proposition entails, but is not entailed by, both 'all examined data points satisfy $y = 8x$ ' and 'all examined data points satisfy $y = x^2 + 2x + 8$ '. Seen in this light, the curve-fitting problem is the problem of which of the two logical weakenings is the most reasonable to accept, i.e., which pattern should we read into our data, for the purposes of performing a SR induction? This is a real problem, but it is not the same as the problem that Goodman purported to have discovered with 'grue'. For in the curve-fitting case, we do not even have a *prima facie* example of two logically equivalent data statements leading to incompatible predictions via the straight rule, as we do in the grue case if Jackson's counterfactual condition is ignored.

The near-universal belief that Goodman proved the necessity of restricting the straight-rule to 'projectible' predicates has given rise to certain other associated beliefs, which have also entered philosophical orthodoxy. For example, Glymour and Spirtes express a widely held view when they say that the grue example shows the impossibility of a 'syntactic' treatment of inductive logic, of the sort that Carnap hankered after—for it shows that the logical form of an inductive inference does not determine its 'inductive validity' or reasonableness.³ This view goes hand-in-hand with a belief in the projectible/unprojectible distinction. If we think that the inference from 'all examined emeralds are green' to 'all unexamined emeralds are green' is good because 'green' is

3 C. Glymour and P. Spirtes 'Real Grue', in D. Stalker (ed.) *Grue*, Chicago: Open Court, 1994.

projectible, but the corresponding inference with *grue* is bad because '*grue*' is unprojectible, it clearly follows that logical form is a poor guide to inductive validity. However, if Jackson is right that the difference between the two inferences is simply that the counterfactual condition is satisfied in one case but not the other, then this conclusion does not follow. Obviously, this does not mean that Carnap's conception of inductive logic is salvaged—for it faces many other real objections—but it does mean that one standard reason for rejecting it can be discounted.

In a recent discussion of '*grue*'—one of the few that takes full account of Jackson's points—Peter Godfrey-Smith makes an interesting observation concerning the way that many philosophers have reacted to Goodman's problem.⁴ Godfrey-Smith notes that the various attempts in the literature to say what the difference is between projecting green and *grue* share a common assumption, namely 'that the two inductions share all the features that might be addressed by a general discussion of good methods of projection *within* science' (2003 p. 573). That is, philosophers assume that the *grue* example raises a problem that is in a sense prior to, or more fundamental than, the standard problems concerning inferences from samples to populations that are dealt with in statistics textbooks. Thus for example, John Pollock writes: 'it seems likely that ... [a] strong projectibility constraint should be imposed upon all familiar patterns of statistical inference. The need for such a constraint seems to have been ignored in statistics' (quoted in Godfrey-Smith 2003 p. 574n).

Godfrey-Smith argues, building on Jackson's work, that this assumption is incorrect. The problem raised by the *grue* example is not categorically unlike the problems discussed in books on statistical inference and data analysis, but rather is an instance of the much-discussed problem of statistical bias. In inferring from sample to population, statistics books tell us to try to ensure that sample is *random*, i.e., each object in the population has an equal chance of getting into the sample. This helps ensure that the objects in our sample will be

4 P. Godfrey-Smith, 'Goodman's Problem and Scientific Methodology'.

representative of the population as a whole. Where random sampling is impossible, it is still advisable to control for known confounding factors, i.e., to ensure that the objects in our sample do not differ, in some respect we know about, from the objects in the population as a whole. Thus for example, in trying to determine the safety of a drug for general use in the population, we would not test the drug solely on men—for then our sample would differ, in a respect that might be important, from the population as a whole.

In the grue case, we know that the objects in our sample all have a property—being examined—that is not shared by objects in the larger population, and that this property affects whether the objects are grue—for we know that if the examined emeralds had not been examined, they would not have been grue. So we know that our sample is likely to be unrepresentative of the population as a whole, in the very respect that we are interested in projecting. Familiar maxims of statistical inference therefore counsel against projecting grue to the whole population, Godfrey-Smith argues. If this argument is correct, as I think it is, it follows that the grue example does not contain a moral that students of statistical inference have ignored, as many philosophers appear to believe, but rather contains a familiar moral in unfamiliar garb.

Goodman's discussion of grue has achieved the status of a 'modern classic' in philosophy; as a result, it is usually taken for granted that Goodman was making a deep philosophical point, and the actual details of his argument tend to be forgotten. Undergraduate around the world are taught about the 'grue paradox'—as if it were a genuine paradox on a par with, for example, Russell's paradox—and are taught that the notion of a 'projectible predicate' is needed to avoid the paradox. But Jackson's probing analysis shows, entirely convincingly in my view, that this is just wrong. The straight-rule does *not* lead to inconsistency, so long as the counterfactual condition is satisfied; so there is no paradox, and no need for a projectible/unprojectible distinction at all.