THE JOURNAL OF PHILOSOPHY

VOLUME LXXII, NO. 5, MARCH 13, 1975

GRUE *

THIS paper is concerned with an aspect of the problem of describing or specifying those inductive practices we take to be rational.

At the level of description, there is no doubt that one common inductive practice we take to be rational is to project common properties from samples to populations, to argue from certain *F*s being *G* to certain other *F*s being *G*. There are many ways we can try to spell out this practice in semi-formal terms: by saying '*Fa* & *Ga*' confirms '$\forall x[Fx \supset Gx]$', or 'All examined *A*s are *B*' supports 'All unexamined *A*s are *B*', or 'Fa_1 & $\cdots$ & Fa_n' gives a good reason for 'Fa_{n+1}', and so on. The precise way chosen will not particularly concern us, and I will simply refer to the kind of inductive argument pattern reflected in the various formalizations as the *straight rule* (SR). The discussion will be restricted to the simplest case where everything in a sample, not merely a percentage, has the property we are concerned with.

To say that the SR is one common inductive argument pattern we all acknowledge as rational, is not to say that it is the most fundamental inductive argument pattern, or the most important in science, or the pattern that must be justified if induction is to be justified; it is simply to say what is undeniable—that we all use it on occasion and take it as rational to do so. This paper is not concerned with how important or fundamental the SR is—for example, vis-à-vis hypothetico-deduction—it is concerned with the *description* of those applications of the SR which we regard as rational.

Since Nelson Goodman's 1946 paper[1] and the development of it in *Fact, Fiction, and Forecast*,[2] it has been very widely supposed that

* This paper has benefited considerably from discussions with colleagues, particularly with Robert Pargetter.

1 "A Query on Confirmation," this JOURNAL, XLIII, 14 (July 4, 1946): 383–385.

2 Cambridge, Mass.: Harvard, 1955; Indianapolis: Bobbs-Merrill, 1965; ch. 3.

the rough description of the SR given above—as certain *F*s being *G* supporting certain other *F*s being *G*—requires the insertion of a substantial proviso to the effect that the properties or predicates (or, in an alternative terminology, the hypotheses) involved be projectible.[3] The notion is that, though there are certain values of '*F*' and '*G*' for which it is manifestly true that the SR applies, there are other values for which it is manifestly false that the SR applies.

This gives rise to a new problem (Goodman's new riddle) in inductive logic—that of demarcating the projectible predicates from the nonprojectible. The extensional aspect of this problem has not been so controversial as the intensional. There has been reasonable agreement about which predicates go into which class: 'green', 'blue', 'round', etc. into the projectible; 'grue', 'bleen', 'sampled', into the nonprojectible. But there has been enormous controversy over the *rationale* for this division; over what makes, for example, 'grue' nonprojectible and 'green' projectible. It has, to say the least, proved difficult to give a plausible, *nonarbitrary* account of the projectible/nonprojectible distinction other than the circular, useless one that a predicate is projectible just if the SR applies with respect to it.

I believe we can resolve the apparently interminable conflict over what it is about nonprojectible predicates that makes them so, by challenging its very foundation. I will argue in this paper 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.

The almost universal view that we need a distinction between projectible and nonprojectible predicates and hypotheses has had, I believe, 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 will consider these matters in turn.

I. THE THREE WAYS OF DEFINING 'GRUE'

In this section I will consider the three common kinds of ways of defining 'grue' by considering typical instances of each. I will argue that the first two ways do not pose even a prima facie problem for the SR, leaving us with the third way to consider in later sections.

[3] I will talk primarily in terms of the projectibility or otherwise of properties, predicates, or open sentences (the differences among these three not being relevant to the arguments that follow), rather than hypotheses; that is, I will follow Goodman's usage in "A Query" rather than in *FFF*.

A typical example of the first way is:

D_1. *x* is grue iff *x* is green before *T* and blue thereafter.

where *T* is a chosen time in the future.[4]

On D_1, 'grue' is atemporal—an object is grue or not once and for all, it cannot be grue at one time and not grue at another—and in this respect differs from 'green'.

There seems no case for regarding 'grue' as nonprojectible if it is defined in this way. An emerald is $grue_1$[5] just if it is green up to *T* and blue thereafter, and if we discovered that all the emeralds so far examined had this property, then, other things being equal, we would probably accept that all emeralds, both examined and unexamined, have this property of being green to a certain time and then turning blue; or, at least, would regard this hypothesis as supported.

We would in this case be regarding emeralds as like tomatoes and oranges, one of those things which change color dramatically during their life cycles. No doubt we would seek an explanation for the fact that the change in emeralds occurs at a fixed time, *T*; but there would in principle be no impossibility about finding a satisfactory explanation. For example, we might discover that emeralds contain a radioactive element the radiation of which makes them green instead of blue, and that the level of this radiation is due to drop below a crucial figure at *T*.

[A puzzling feature of the discussions of the new riddle of induction by those who employ a D_1-type definition is that they take it as not in dispute that all emeralds observed to date are grue, as well as green. For example, Stephen Barker simply asserts as if it were an evident truth that "all the numerous emeralds that we have observed have been grue" (*op. cit.*, p. 189)—but what is an evident truth is that these emeralds were green at the time of observation; what we all believe is that they are always green; and what none of us believe is that they are grue, for none of us believe they will change to blue in the year 2000 (Barker's choice for *T*).[6]]

[4] D_1-type definitions appear in the discussions of projectibility by: H. Kyburg, *Probability and Inductive Logic* (Toronto: Macmillan, 1970); I. Hacking, *The Logic of Statistical Inference* (New York: Cambridge, 1965); S. Barker, *Induction and Hypothesis* (Ithaca, N.Y.: Cornell, 1957).

[5] When it is not clear from the context, '$grue_i$' is used for: 'grue' defined according to D_i.

[6] I have kept my discussion of D_1 brief, since similar points have been well made by S. Blackburn, "Goodman's Paradox," *American Philosophical Quarterly*, Monograph no. 3, 1969; and M. Kelley, "Predicates and Projectibility," *Canadian Journal of Philosophy*, I, 2 (December 1971): 189–206.

A typical example of the second way[7] of defining 'grue' is:

D_2. x is grue at t iff (x is green at t & $t < T$)
or (x is blue at t & $t \geq T$).

'Grue' on this definition is like 'green' in being temporal: an object may be grue_2 at one time and not at another.

It sometimes seems to be thought that D_2 really amounts to $D_{2.1}$: 'x is grue' means 'x is green' before T and 'x is blue' after T, which is an explicit case of ambiguity; and, consequently, that it raises no problem for the SR.[8] When we read the SR as licensing the projection of a common predicate, it is understood that the predicate has the same meaning throughout.

The two definitions are not, however, equivalent. The appearance of equivalence arises from a failure to be explicit about time in $D_{2.1}$, and if we write in a time variable to give: 'x is grue at t'' means 'x is green at t'' before T, and 'x is blue at t'' after T, the disparity becomes obvious. Consider a time t_1 before T, and whether a green emerald is grue at t_1. According to D_2, the answer is an unequivocal yes; but, according to $D_{2.1}$, the answer depends on the time at which the question is being asked. If the question is asked before T, the answer is yes; because before T, 'x is $\text{grue}_{2.1}$ at t'' means 'x is green at t'', and the emerald is green at t_1: if asked after T, the answer is no; because after T, 'x is $\text{grue}_{2.1}$ at t'' means 'x is blue at t'', and the emerald is not blue at t_1. In short, D_2 and $D_{2.1}$ are not equivalent because the time at which we consider the question of an object's grueness is relevant on $D_{2.1}$ and not relevant on D_2—on D_2, the time at which the object is green or blue is relevant, but not the time at which we consider the matter.

There is, I believe, no getting away from the fact that D_2 is a perfectly proper, intelligible definition. Nevertheless, D_2 does not give rise to a paradox or "new riddle" when conjoined with the SR, and so does not give grounds for supposing that there are non-projectible predicates of which 'grue_2' is the best-known example.

The contrary view has arisen from confusion over whether we are considering the SR in conjunction with 'grue_2' as applied to objects that endure through time, that is, four-dimensional objects, or as

[7] This kind of definition appears in W. Salmon, "On Vindicating Induction," in H. Kyburg and E. Nagel, eds., *Induction: Some Current Issues* (Middletown, Conn.: Wesleyan, 1963); P. Achinstein and S. Barker, "On the New Riddle of Induction," *Philosophical Review*, LXIX, 4 (October 1960): 511–522; B. Skyrms, *Choice and Chance* (Belmont, Calif : Dickenson, 1966). The tendency (e.g., by Barker and Kyburg) to slip between D_1 and D_2 may be due to the fact that 'x is $\text{grue}_1 \equiv \forall t$($x$ is grue_2 at t)' is true.

[8] See, e.g., Kelley, *op. cit.*, §III.

applied to three-dimensional objects *at* times, that is, time-slices of the four-dimensional objects.

If we are considering the SR as applied to enduring objects like tables and emeralds, if we take the members of the samples and populations we discuss to endure through time, then we must read a temporal factor into the predicates with which the SR is concerned. Enduring objects aren't red, or green, or square, *simpliciter*: they are red at t_1, green at t_2, and so on. A tomato isn't both red and green; it is green early in its life history and red later.

From this it follows that when we read the SR (applied to enduring objects) as licensing the projection of common predicates from samples to populations, we must incorporate a temporal factor into these predicates. What we project must be understood as at a time; not just being green but being green at t. Only when this is overlooked does the appearance of paradox arise from applying SR with D_2, because the apparently paradoxical result only comes about with projections across T. To illustrate with the usual emerald case, suppose we have a sample of emeralds that are green at t_1, where t_1 is before T, then they will also be grue$_2$ at t_1; and, hence, the SR will equally lead to 'All emeralds are green at t_1' and 'All emeralds are grue$_2$ at t_1'. And these two universals are in no way incompatible. Whereas for time t_2 after T, it is impossible that a sample of emeralds be both green at t_2 and grue$_2$ at t_2, and so we cannot be led by the SR to hold together the incompatible universals: 'All emeralds are green at t_2' and 'All emeralds are grue$_2$ at t_2'.

It is only if we slide illegitimately from t_1 to t_2 that an appearance of paradox arises. 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 at t_2', do we obtain an apparent paradox.

It may be objected that my insistence on the distinction between the predicates 'x is grue (green) at t_1' and 'x is grue (green) at t_2', forces an unwelcome restriction on the role of the SR: sometimes we use the SR to argue from certain examined emeralds being green *now* to others being green *now*; sometimes from certain emeralds being green at one time, now, say, to certain others being green at a different time, in the future, say; and it may be thought that the second kind of use—when we go from the present to the future—requires ignoring the distinction between being F at t_1 and being F at t_2.

But this is to overlook the application of the SR to time-slices of objects as distinct from enduring objects. When we argue from the greenness of present emeralds to the greenness of future emeralds, we do best to view this as an application of the SR to temporal parts of emeralds, and so as an application involving, not being green at t true of an enduring emerald, but rather being green *simpliciter* true of the temporal part at t of an emerald. When we wish to explicate our intuitive feeling that emeralds being green now supports their being green in the future by reference to the SR, by reference to the projection of common properties, we ought not fudge the clear distinction between being green now and being green in the future; rather we should regard the projected property, being green, as a tenseless characteristic of present emerald temporal parts which is being projected to future temporal parts in accord with the SR.

(A question that might well be asked now is what happens to D_2 if we recast it as a predicate on temporal parts instead of enduring objects. What happens, as can easily be seen, is that D_2 becomes like D_3, below, in all respects essential to whether there is a 'grue' paradox; and hence does not call for separate treatment.)

Although D_1 and D_2 figure prominently in the 'grue' literature, they are not the kind of predicate with which Goodman launched it.[9] Goodman's predicates are of the kind, '$(x$ is green & $\emptyset x) \vee (x$ is blue & $\sim\emptyset x)$', where '$\emptyset x$' is chosen so that its extension includes all the sampled (observed, examined, etc.) emeralds, that is, the emeralds from which we are imagined to be projecting, and so that the extension of '$\sim\emptyset x$' includes the other emeralds, those to which we are projecting. A simple way of doing this is to introduce a temporal factor into '$\emptyset x$', which is Goodman's usual but not invariable practice; in particular, the following definition is close to that he uses in *Fact, Fiction, and Forecast*:

D_3. x is grue at t iff (x is examined by T and x is green at t) or (x is not examined by T and x is blue at t).

As indicated by the 'at t' in D_3, this definition is for enduring objects. To avoid tedious repetition of the 'at t', we will conduct our discussion in the editorial present. Likewise, we will commonly drop the 'at T' by taking T to be a moment in the near future such that 'examined by T' just amounts to 'examined (to date)', and 'not examined by T' amounts to 'unexamined (to date)'. Both these procedures are implicitly adopted by Goodman, so that being grue_3 can be simply characterized as being green and examined, or being

[9] As Goodman points out in *Problems and Projects* (Indianapolis: Bobbs-Merrill, 1972); see p. 359.

blue and unexamined. The paradox D_3 appears to lead to, as we will see, is not essentially time-linked. It is not essential that we consider the sampled emeralds at one time, the remaining at another, to get an apparently paradoxical result; so that, with D_3, by contrast with D_2, there is no objection to making things simpler by fudging a bit with respect to time.

D_3—the correct definition in the sense that it gives rise to more trouble than D_1 and D_2, as well as being Goodman's—will be the only definition we will be concerned with in the following sections, and when I refer to 'grue' and the alleged associated paradox or new riddle, I will mean 'grue' as defined in D_3.

II

Just what is the 'grue' paradox supposed to be; just what objectionable result is obtainable? In outline, the picture is clear enough. The idea is that, by suitable choice of predicates, the SR can be deployed to reach two incompatible conclusions starting from the same evidence. In particular, it is argued that a certain fact about emeralds when expressed in terms of 'green' leads to one projection about other emeralds when we apply the SR, and the same fact expressed in terms of 'grue' leads to another, incompatible projection when we apply the SR.

Though the picture is clear enough in outline, it starts to get murky as soon as we try to fill in the details. If, to fix our discussion, we consider a series of emeralds, $a_1, \ldots, a_n, a_{n+1}$, such that $a_1, \ldots, a_n$ are known to be green and examined, while a_{n+1} is known to be unexamined and is the emerald whose color we are concerned to predict; precisely how does the SR lead to incompatible projections about a_{n+1} from equivalent evidential bases?

Well, if we use 'Grx' for 'x is green', 'Ex' for 'x is examined', 'Bx' for 'x is blue', and 'Gux' for 'x is grue' = '$(Grx \,\&\, Ex) \vee (Bx \,\&\, {\sim}Ex)$', we are given

(1) $$Gra_1 \,\&\, \cdots \,\&\, Gra_n$$

and

(2) $$Gua_1 \,\&\, \cdots \,\&\, Gua_n$$

But, first, (1) and (2) are not equivalent (neither entails the other), so there is no objection to the SR leading to different predictions ('Gra_{n+1}' and 'Gua_{n+1}', respectively) regarding a_{n+1}; second, the predictions are not inconsistent (neither entails the denial of the other); and, finally, neither (1) nor (2) embodies our total evidence.[10]

[10] As, in effect, R. Carnap points out in "On the Application of Inductive Logic," *Philosophy and Phenomenological Research*, VIII, 1 (September 1947): 133–147; see §3.

Our total evidence (or near enough for present purposes) is, rather expressed by

(3) $Gra_1 \ \& \ Ea_1 \ \& \cdots \& \ Gra_n \ \& \ Ea_n$

which is, of course, equivalent to

(4) $Gua_1 \ \& \ Ea_1 \ \& \cdots \& \ Gua_n \ \& \ Ea_n$

But what (3) and (4) support by the SR is

(5) $Gra_{n+1} \ \& \ Ea_{n+1}$

and

(6) $Gua_{n+1} \ \& \ Ea_{n+1}$

respectively; which, far from being incompatible, are equivalent.

Perhaps it will be argued that (5) entails (as it does)

(7) $\sim Ea_{n+1} \supset Gra_{n+1}$

and that (6) entails (as it does)

(8) $\sim Ea_{n+1} \supset Gua_{n+1}$

which is equivalent to

(9) $\sim Ea_{n+1} \supset Ba_{n+1}$

And that (7) expresses the prediction that, if a_{n+1} is not examined, it is green, whereas (9) expresses the incompatible prediction that, if a_{n+1} is not examined, it is blue. So we have derived opposite, incompatible predictions from equivalent bases, (3) and (4).[11]

But this is like arguing that our observations of black ravens support white ravens being black, as follows: our observations support Joey, an as yet unobserved raven, being black. But 'Joey is black' entails 'Joey is white $\supset$ Joey is black' so that our observations support the prediction that if Joey is white, then he is black.

The fallacy here is obvious. We do have support for 'Joey is white $\supset$ Joey is black', but only because we have support for the falsity of the antecedent. Likewise, we do have support on the basis of (3) and (4) for (7) and (9), but only because we have support for the falsity of their antecedents. It may be replied here that we do not have support for the falsity of their antecedents, that is, for a_{n+1} being examined, because being examined or '*Ex*' is not projectible. But this is to *assume* that there are nonprojectible properties, in the course of an argument designed to show that there are; moreover, we will be

[11] I take this to be essentially the argument in H. Leblanc, "That Positive Instances Are No Help," this JOURNAL, LX, 16 (Aug. 1, 1963): 452–462.

giving reason later for allowing that being examined is projectible (in sec. IV).

III. THE COUNTERFACTUAL CONDITION

So far I have not used the fact that we are given that a_{n+1} is not examined. And in Goodman's view our knowledge that there are unexamined emeralds is essential to deriving a paradox. He says, for instance,

> If the hypothesis that all emeralds are green is also projected [i.e., in addition to 'All emeralds are grue'], then the two projections disagree for unexamined emeralds. In saying these projections thus conflict, we are indeed *assuming that there is some unexamined emerald* to which only one of the two consequent-predicates applies, but it is upon just this assumption that the problem arises at all (*FFF*, 2nd ed., p. 94; my emphasis).

But just how can we use the fact that a_{n+1} is not examined? It sometimes seems to be thought that it is proper to add in this additional information in a more or less mechanical fashion, somewhat 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.

There is no question that we have here genuinely incompatible predictions about the color of a_{n+1}, for we have categoricals, not material implications. But we also have a pattern of argument that is quite certainly fallacious.

The pattern is: If a proposition, p, which we know to be true, supports a conjunction, $q \& r$, one conjunct, r, of which we know independently to be false; we have, over-all, support for $q \& \sim r$, and so, for anything it entails.

Once this pattern is explicitly set out, I doubt if anyone would assent to it; for it leads easily to an inconsistency, as follows: p supports $(q \& r)$ if and only if p supports $[(q \& r) \vee (\sim q \& \sim r)] \& r$, for the latter is truth-functionally equivalent to $(q \& r)$. Hence, by the argument pattern just displayed, when I know r to be false on independent grounds, I have, over-all, support equally for $(q \& \sim r)$ and for $[(q \& r) \vee (\sim q \& \sim r)] \& \sim r$, which are truth-functionally inconsistent [the latter is equivalent to $(\sim q \& \sim r)$].

It is equally clear from actual examples that this argument pattern is fallacious. Suppose a reliable friend tells me that Hyperion won the cup by five lengths, then I have support for the conjunction, 'Hyperion won the cup and Hyperion won by five lengths.' Further suppose I have quite decisive, independent evidence that the winning margin in the cup was only three lengths, but that this evidence is neutral as to who won by that margin. Do I have, over-all, evidence for Hyperion winning, though not by five lengths? If the argument pattern in question were valid, the answer would be an invariable yes; but in fact the answer is that it all depends on the circumstances. In some it will be most rational for me to take the error as to winning margin as indicating that my normally reliable friend is having one of his few off days and so is not to be trusted concerning the identity of the winner either; in other circumstances it will be most rational for me to take it that my friend regarded the winning margin as a relatively unimportant detail compared to the identity of the winner, and was his usual reliable self concerning the latter.

What we have here is, of course, just an aspect of the universally acknowledged fact that inductive support is defeasible; and it is strange how often this defeasibility is overlooked in the context of discussions of 'grue'. For example, it is common to find it suggested that by means of a 'grue'-type maneuver it is easy to show that an unrestricted SR leads to the unacceptable consequence that *any* n objects support some $(n + 1)$st object being G, for *any* 'G',[12] as follows: For any $(n + 1)$ objects, there will be an 'Fx' such that it is true of the first n, but not the $(n + 1)$st. But if 'Fx' is true of the first n, so is '$Fx \vee Gx$', for any 'Gx'; therefore, runs the argument, the (unrestricted) SR supports the $(n + 1)$st being F or G. But it is given as not being F; so it is concluded that we are led to the absurdity that we have support for the $(n + 1)$st object being G, for any 'G'. Now if something like: 'If p supports q, then $(p \;\&\; r)$ supports $(q \;\&\; r)$' were valid, all would be well with this argument; *but we all know that nothing like this is valid*, and so, that the information that the $(n + 1)$st object is not F cannot be incorporated in so simple a fashion.

We must, therefore, proceed *very* carefully when attempting to incorporate the additional information that a_{n+1} is unexamined, and, in particular, we must, I think, see the matter in context.

The general context is this: we have a sample, $a_1, \ldots, a_n$, each of which has a property, being examined, in addition to the particular

[12] For just one, typical example, see Skyrms, *op. cit.*, pp. 61, 62.

properties we are interested in (being green and being grue) and which is given as not being possessed by a_{n+1}. This kind of situation arises virtually whenever we use the straight rule. When we use the SR to project common properties from a sample to members of the population from which the sample comes, there are nearly always features common to every member of the sample which we know are not features of all (or any) members of the population outside the sample. Some of these common sample features are normally disregarded as being unimportant, indeed trivial, like being sampled, being one of a_1, a_2, ..., and being examined (before . . .); while others clearly cannot be disregarded, as, for instance, in the following cases: Every diamond I have observed has glinted in the light. Does this support the contention that the next diamond I observe will glint in the light? Clearly, yes. But suppose we add a detail to the story, namely, that the next diamond that I observe is unpolished. Now all the diamonds I have observed so far have been polished, and, moreover, I know that they glint *because* they have been polished—that is, if the diamonds had not been polished, then they would not have glinted. It is clear that once we add this detail, it is no longer reasonable for me to regard it as likely that the next diamond I observe will glint in the light. The fact that all the diamonds I have observed glint in the light supports the next diamond I observe will glint; but the fact that all the polished diamonds I have observed glint when taken in conjunction with my knowledge that they would not have glinted if unpolished, does *not* support an unpolished one glinting.

A similar example is afforded by lobsters. Every lobster I have observed has been red. This supports that the next lobster I observe will be red, and no doubt it will be. But every lobster I have observed has been cooked, and I know that it is the cooking that makes them red—that is, that the lobsters I have observed would not have been red if they had not been cooked. Hence I do not regard myself as having good evidence that the next uncooked lobster I observe will be red.

We have here two cases where certain *F*s being *G* supports, by the SR, other *F*s being *G*, but certain *F*s which are *H* being *G* does not support other *F*s which are not *H* being *G*; in each case the reason being that it is known that the *F*s that form the evidence class would not have been *G* if they had not been *H*. The condition: that certain *F*s which are *H* being *G* does not support other *F*s which are not *H* being *G* if it is known that the *F*s in the evidence class would not have been *G* if they had not been *H*, will be referred to as *the counter-*

factual condition. I cannot think of any way of proving it, as opposed to illustrating it as I just have, but also I cannot think that anyone will seriously deny it. (For ease of reading, I have expressed the condition in a conditional form. Strictly, it should be expressed as that the *conjunction* of certain *F*s which are *H* being *G* *with* these *F*s being such that if they had not been *H*, they would not have been *G*, does not support other non-*H* *F*s being *G*.)

We are now in a position to discuss the incorporation of the additional information that a_{n+1} is unexamined. When we argue from examined emeralds $a_1, \ldots, a_n$ being green to the unexamined a_{n+1} being green, we are arguing from certain *F*s which are *H* being *G* to an *F* which is not *H* being *G*, in the special case got by replacing '*F*' by 'emerald', '*G*' by 'green', and '*H*' by 'examined'. Hence the counterfactual condition is that the emeralds $a_1, \ldots, a_n$ would still have been green even if they had not been examined; and, in the world as we know it, this condition is satisfied. The emeralds we have examined are green not because they have been examined but because of their chemical composition and crystalline structure, and so, like most objects in our world, they would have had the color they do have whether or not they had been examined.

Precisely the opposite is the case with 'grue'. We know that an emerald that is grue and examined would *not* have been grue if it had not been examined; for if it is grue and examined, it is green and examined, and, as noted already, if it had not been examined would still have been green; but then it would have been green and unexamined, and so, not grue. In other words, a green, examined emerald would have been a green, unexamined emerald if it had not been examined, and so $a_1, \ldots, a_n$ would not have been grue if they had not been examined. Therefore, to use the SR to yield the prediction that a_{n+1} is grue (and unexamined) is to violate the counterfactual condition.

In sum, the position is this. 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. Moreover, not only don't we get an inconsistency, we *cannot* get one, because it cannot be the case both that if X had not been H, it would not have been G, and if X had not been H, it would have been G—at least, on standard views about the logic of counterfactuals.

Our discussion of the SR has been couched in terms of constants, 'a_1', . . ., 'a_{n+1}', taken to designate emeralds. It is common to discuss the SR in terms of universals. The counterfactual condition shows, I think, that it can be misleading to characterize the SR as 'All examined As are B' supports 'All unexamined As are B'.

There are cases where it is absurd to take 'All examined As are B' as supporting 'All unexamined As are B'. Some properties of the elementary particles of physics are known to be affected by examination of the particles (hence the indeterminacy principle). It would be absurd to argue, for such a property, that, since all examined particles have it, so do all unexamined particles; just because we know that if the particles in question had not been examined, they would not have had the property.

Moreover, we do not have to turn to recondite entities like submicroscopic particles for examples of properties such that something would not have them if they had not been examined. Examined emeralds have a property of just this kind, namely, being grue. Take an emerald that is green and examined, and so, grue. If it had not been examined, it would still have been green, because examining emeralds (and indeed examining most things) doesn't alter their color; therefore, if the emerald had not been examined, it would have been green and unexamined, and so, not grue. Hence, it is a mistake to argue from 'All examined emeralds are grue' to 'All unexamined emeralds are grue', not because 'grue' is intrinsically nonprojectible, but simply because the counterfactual condition is violated.

Parallel remarks apply to functor expressions of the SR such as: 'All examined As are B' supports 'The first unexamined A is a B'. We get an apparently simple and decisive development of the 'grue' paradox by noting that: 'All examined emeralds are green' and 'All examined emeralds are grue' are equivalent, and that: 'The first unexamined emerald is green' and 'The first unexamined emerald is grue' are inconsistent.[13] But, evidently, it is reasonable to use this kind of version of the SR only when being B is appropriately independent of being examined, and this is not the case when being B is being grue.

[13] As in W. V. Quine, "Natural Kinds," in *Ontological Relativity* (New York: Columbia, 1969).

It is, perhaps, unfortunate that being examined (observed, sampled, etc.) appears so frequently in statements of the SR. The SR is intended as an essentially *relational* principle of inductive support concerning whether p supports q, quite independently of whether p is known. Examining, observing, sampling, and so on, are how we—human beings—come to know that certain As are B; but our knowing this is separate from these As being B supporting certain other As being B. *What* we come to know does the supporting (if any), not our coming to know it.

Though it is a fact about our world that the emeralds we have examined would still have been green if they had not been examined, it might not have been a fact. We might have lived in a world in which they would not have been green if they had not been examined. For example, we might have lived in a world in which all examined emeralds were green and in which investigation of the crystalline structure of these emeralds reveals that they are naturally blue; this structure being affected by the light necessarily involved in examining them in such a way that emeralds turn green instantaneously on being examined.[14] In this world, all emeralds we have direct observational evidence concerning are green and examined and grue. What ought we believe about those not examined? Obviously, that they are blue, and, hence, that all emeralds are grue. Our counterfactual condition explains this. In this world, examined emeralds are both green and grue, as in our world, but, as not in our world, if they had not been examined, they would have been grue, not green.

IV. THE PROJECTIBILITY OF BEING SAMPLED

Our counterfactual condition also bears on the question of the projectibility of such properties as being sampled, being examined, and being one of $a_1, \ldots, a_n$. Richard Jeffrey holds that—whereas it may just be doubted that 'grue' is nonprojectible—it is beyond doubt that such properties as these are nonprojectible.[15]

Why is he so certain? No doubt the kind of case he has in mind is where I am drawing marbles from a barrel and noting that each marble is red. Normally we suppose this to support that the remaining marbles are red. But, equally, each marble drawn will have the property of being sampled, and we do not normally regard the proposition that the remaining marbles are sampled as being supported by this.

[14] And we could bring in the time factor by, for instance, supposing the method of examining changes at T.

[15] R. C. Jeffrey, "Goodman's Query," this JOURNAL *of Philosophy*, LXIII, 11 (May 26, 1966): 281–288; see p. 288. He actually has 'bleen' for 'grue' in the relevant passage.

But it would be too hasty to infer from this point that being sampled is not projectible. Suppose my reason for thinking that all the marbles drawn out have been sampled is that they each have Jones's finger prints on them, and so must have been sampled (in the past, by Jones). Then it is clear that I will be entitled to increase my degree of belief that the remaining marbles have been sampled.

What is the explanation for the dramatic change whereby it is evidently absurd to increase one's expectation that the remaining marbles are sampled in the first case, and evidently not absurd in the second? I think it would be a mistake to explain this change in terms of the projectible/nonprojectible distinction by saying that *being sampled by me now* is nonprojectible, whereas *being sampled by Jones in the past* is projectible. For suppose that in the first case I am Jones and that after drawing the red marbles I go out for a cup of coffee; on my return I am confronted by a group of marbles all of which have the property of being sampled by Jones in the past. Do I now increase my expectation that the remaining marbles have this, allegedly projectible, property? Quite obviously no. The projectible/nonprojectible property distinction cannot explain the divergence in our inductive behavior in the two cases—and, surely, this is just the kind of case that the distinction, if it is worth making, ought to help us with.

What does explain the divergence is our counterfactual condition. In the first case, we have certain marbles, all of which are sampled and all of which have just been drawn from the barrel, and are concerned with whether we have support for certain other marbles, not drawn from the barrel, being sampled. We do not, because we know how it is that the sampled marbles came to be sampled, namely, by being drawn out. Hence, if the marbles had not been drawn out, they would not have been sampled; and our counterfactual condition is violated. On the other hand, in the second case (where I discover that the marbles have been sampled by observing Jones's fingerprints on them), the marbles drawn out would still have been sampled (by Jones, in the past) even if they had not been drawn out by me. The counterfactual condition is not violated, and we, therefore, have in the second case support for the marbles not drawn out being sampled.

There is nothing intrinsically nonprojectible about being sampled. In some cases, it is perfectly reasonable to project it, and, in those cases where it is not, the explanation does not have to do with the nature of the property or the meaning of the corresponding predicate, that is, does not relate to a feature of being sampled that calls for a label such as 'nonprojectible', but is rather that the counterfactual condition is violated. Exactly similar remarks apply to being examined and to being one of $a_1, \ldots, a_n$. I will look briefly at the latter.

Despite the frequency and confidence with which it is said that properties of the being one of $a_1, \ldots$ kind are not projectible, it is easy to describe the counter cases. Suppose I am a policeman investigating a series of cat burglaries, and I discover that in each case the person responsible is one of Tom, Dick, and Harry; then I will be entitled to regard 'The person responsible for the next cat burglary will be Tom, Dick, or Harry' as supported. Again, if I am drawing marbles from a barrel and find each one stamped with a name, and the name is always one of 'a_1', 'a_2', ... 'a_n', I will have increasing support for the next marble being one of $a_1, \ldots, a_n$. (Of course, after I have drawn out all of $a_1, \ldots, a_n$ marbles *and if* I am drawing without replacement, I won't expect the next marble to be one of $a_1, \ldots, a_n$; but this is because I am acquainted with the necessary truth that n things cannot be identical with $n + 1$ things, and shows, not nonprojectibility, but the role of additional negative evidence).[16]

By way of contrast, if I don't find the names already stamped on the marbles, but *give* the names to the marbles as they are drawn out, I won't expect marbles not drawn out to be identical with one of $a_1, \ldots, a_n$; because the counterfactual condition is violated. I know that the marbles drawn out would not have the names they do if they had not been drawn out.

Whenever we apply the SR, we know, as it were, too much. I am drawing marbles from the ubiquitous barrel, and, in consequence, the drawn marbles are in my hand, recently exposed to light, and observed. These are all things I know about the marbles which I would not dream of projecting to the marbles remaining in the barrel: not because these properties are intrinsically nonprojectible—there are obviously many cases where we would project, for instance, being recently exposed to light—but because I know how the drawn marbles came to be recently exposed to light (to single this property out for discussion), namely, as a result of being sampled. Therefore, if they had not been drawn out, they would not have been recently exposed to light, and so, the argument from the drawn marbles being recently exposed to light to the undrawn marbles being so exposed, violates the counterfactual condition.

I expect that two objections will be generated by the prominent role of the counterfactual condition in the above discussion. The first is the general objection that counterfactuals raise some of the most

[16] Cf. Kelley, *op. cit.*, p. 196.

difficult problems in philosophy. This is true, but the fact remains that we do, on occasion, know with certainty that certain counterfactuals are true, despite the difficulties in analyzing just what it is that we know on such occasions and how we know it. Perhaps one day we will have a good theory of counterfactuals, or a way of eliminating the need for them; until then we must put up with them.

The second objection is the more particular one that, by appealing to the counterfactuals that I appeal to, I am introducing a kind of circularity. Take, for example, my reason for saying that the SR favors unexamined emeralds being green rather than grue: that the emeralds we have in fact examined would have been green, not grue, if they had not been examined. There is no disputing the fact that we do know this—it is as certain as any knowledge of the form: if *a* had not been X, it would have been Y, is—but it might be objected that we know this only because we know unexamined emeralds are green. Hence, on pain of circularity, we cannot appeal to this fact to explain why the SR leads to the prediction that unexamined emeralds are green.

However, our knowledge that the examined emeralds would still have been green if they had not been examined is knowledge about the *examined* emeralds, not about the unexamined ones. It is knowledge we might have had even if there were no unexamined emeralds to be green or not green. If it turned out that there were very many fewer emeralds than we at first thought, and that in fact every emerald has been discovered and examined, this would not alter the fact that if the examined emeralds had not been examined they would have been green. Moreover, this fact is quite consistent with the unexamined emeralds turning out to be, to our great surprise, red; the result, say, of the emeralds so far examined all coming from regions in the world where certain minerals that make things green abound, and those not so far examined coming from regions containing minerals that make things red. This surprising discovery would not undermine our belief that the *so-far examined* emeralds would have been green even if not examined.

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.

The point is more obvious in the marble-barrel case. I may know that each of the three red marbles that I drew out of the barrel would still have been red even if it had not been drawn out, without knowing the color of the remaining marbles—indeed, I commonly will know the former without knowing the latter. To know the former is to know something about the lack of connection between the color of an object and whether or not it is examined in the given case, whatever that color may be; and is not dependent on knowledge of the particular color of a particular object or objects, be they drawn out or not. Similar remarks apply in the converse case where I discover that the red marbles are painted with a special paint that turns red immediately on contact with a human hand (due, say, to the warmth). In this case, the marbles would not have been red if not drawn out (by hand), and we would not increase our expectation that the remaining marbles are red for just this reason. It is quite obvious that I may know the relevant facts about the paint without knowing or having any idea of the color of the remaining marbles. There is, thus, no circularity. (Likewise, Goodman's appeal to the entrenchment of predicates isn't circular, though it also involves appeal to inductively gained knowledge. The objection to entrenchment is rather that it is excessively anthropocentric.)

V. SUMMARY

The over-all position is this. The SR: certain *F*s being *G* supporting other *F*s being *G*, does not lead to incompatible predictions when combined with 'grue' and like predicates.

When we apply the SR in practice, we commonly argue on the modified pattern: certain *F*s which are *H* being *G* supports *F*s which are not *H* being *G* ('*H*' often being 'examined', 'sampled', etc.). When we argue on this modified pattern, we take it that it is not the case that the *F*s which are *H* would not have been *G* if they had not been *H*. This guarantees that we can never be led from the same evidence to opposite predictions concerning whether the non-*H* *F*s are *G*. For, though *F*s are *H* and *G* if and only if they are *H* and G^*, where 'G^*x' = '$(Hx \,\&\, Gx) \vee ({\sim}Hx \,\&\, {\sim}Gx)$', we cannot be led both to the non-*H* *F*s being *G* and to their being G^*, and so—as a non-*H* *F* is G^* just if ${\sim}G$—to opposite predictions. This is because we know from the logic of counterfactuals that it cannot both be the case that the *F*s which are *H* and *G* would have been *G* if they had not been *H*, and that they would have been G^* if they had not been *H*; for this amounts to '$p \,\square\!\!\rightarrow q$' and '$p \,\square\!\!\rightarrow {\sim}q$' being true together, since a non-*H* is G^* only if ${\sim}G$.

To arrive at counterfactuals of the required form, we must, of course, draw on our knowledge of the world. Just which knowledge is as controversial as counterfactuals—that is, very. But it is clear that 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. And, of course, it is not controversial that *applying* the straight rule in a particular case requires reference, *inter alia*, to knowledge gained inductively from *other* applications of the SR. Even knowing that certain *F*s are *G* requires trusting one's senses, memory, the reliability of reference books, and so on. This may well raise fundamental problems at the level of justification, in the context of the "old problem of induction," but this has not been our concern here. Our concern here at the level of description has been to urge that the SR can be specified without invoking a partition of predicates, properties, or hypotheses into the projectible and the nonprojectible.

FRANK JACKSON

La Trobe University

SAVING LIFE AND TAKING LIFE

THE purpose of this paper is to examine the distinction between "negative" and "positive" duties. Special attention will be given to certain criticisms raised against this distinction by Michael Tooley.

A PARADIGM CASE

If someone threatened to steal $1000 from a person if he did not take a gun and shoot a stranger between the eyes, it would be very wrong for him to kill the stranger to save his $1000. But if someone asked from that person $1000 to save a stranger, it would seem that his obligation to grant this request would not be as great as his obligation to refuse the first demand—even if he had good reason for believing that without his $1000 the stranger would certainly die. Refraining from the action of killing is a kind of "inaction" which it seems appropriate to call a "negative" duty. Saving is a kind of "action" which it seems appropriate to call a "positive" duty.[1] In this particular example, it seems plausible to say that a person has a greater obligation to refrain from killing someone than to save someone, even though the effort required of him ($1000) and his motivation toward the stranger be assumed identical in both cases. None of this is meant as exact analysis, but rather as an initial indication of what seems to be a plausible view.

[1] Philippa Foot defends the view that we are more obligated to meet negative duties of not injuring people than to meet positive duties of helping them. See "The Problem of Abortion and the Doctrine of the Double Effect," *The Oxford Review*, v (1967): 5–15.