Explanation as a Guide to Induction

Roger White

Philosophers' Imprint
<www.philosophersimprint.org/005002/>
Volume 5, No. 2
April 2005

In what circumstances should we draw an inductive inference? Gilbert Harman (1965) answered: When our conclusion provides the best explanation of our evidence. Despite the wide influence of Harman's suggestion, little work has been done in applying it specifically to induction in the narrow sense of enumerative induction, the projection of observed regularities to unobserved instances.[1] I think there is plenty more to be said, and my purpose here is to illustrate and defend the idea that explanatory considerations can serve as a powerful guide to enumerative inductive reasoning.

The slogan 'inference to the best explanation' might suggest something stronger than what I am defending here, not only because the best explanation we can come up with may be a lousy one. For I wish to leave open the question of what role, if any, non-explanatory considerations should play in epistemic evaluation. My aim is simply to show that explanatory considerations have an important role to play in the evaluation of inductive hypotheses. This thesis is not so weak as to be uncontroversial, but most of the interest here lies in the fruitfulness of its application.

The three main sections of the paper are progressively specific. Section 1 clarifies the notion of explanatory value and discusses a crucial explanatory virtue. Section 2 concerns general issues about explanation and induction, especially getting clear on what are the relevant explananda and explanantia. Section 3 applies the explanatory strategy to a number of specific problems and principles of induction.

[1] Two important exceptions are Lipton (2004) and Peacocke (2004). Others such as Armstrong (1983), Bonjour (1998), and Foster (1982, 2004) have suggested that explanatory considerations might be crucial to answering the traditional problem of induction, without getting into the details of how we should reason inductively.

Roger White is Assistant Professor of Philosophy at New York University.

1. Explanatory Value

In this section I want to begin by clarifying what is involved in taking explanatory value as a guide to inductive inference. I will then spell out the main explanatory virtue that I appeal to in later sections, and discuss why it is virtuous.

1.1 Explanation and Guidance

Perhaps sometimes when we speak of a hypothesis being the best explanation of some fact, we simply mean that it is plausibly the correct explanation, all things considered. As Lipton (2004) emphasizes, if this were the only dimension along which we can evaluate an explanation, it would be trivial to take explanatory value as a guide to inference. Of course we should take plausibility as such a guide. If explanatory considerations are to guide our reasoning, there must be a dimension of explanatory value that can be assessed without simply asking whether the explanation seems correct.

I would suggest that the relevant notion of explanatory value has to do with the degree of *satisfaction* that an explanation should deliver, if assumed to be correct. In asking a why-question we are seeking to satisfy a peculiar kind of curiosity; we are seeking *understanding* and trying to *make sense* of things. The kind of satisfaction that a good explanation can deliver does not simply consist in our taking ourselves to have discovered the truth. For even if known to be correct, an explanation can remain deeply unsatisfying. One semester I received three term papers which were almost identical. It is possible that none of the students copied from another or from a common source, but all three just happened to write the same paper. But such an explanation is unsatisfying. Our dissatisfaction here is not simply a matter of finding the story *implausible*. An oracle might reveal to us that this is in fact what happened. He might give us a detailed psychological explanation of how each student independently happened to form the same thoughts and expressed them in the same words. He might even trace these independent causal paths back, showing us that this state of affairs was an inevitable outcome of the initial state of the universe. Since the oracle is infallible, we would have to admit that his explanation is not only correct but the deepest and most thorough explanation that can be given. But we would remain utterly mystified. Compare this with our learning that they each copied from a Googled web page. This kind of explanation is satisfying in that it makes sense of what happened.

The way then to assess the value of an explanation is to ask: Suppose we were to learn for certain that this explanation is correct. How satisfying should it be in our quest for understanding and making sense of things? On this conception of explanatory value, to answer the question 'How good an explanation does this hypothesis provide?' is not simply to judge its *plausibility* as an explanation. For we are factoring out the question of its truth by considering how satisfying it should be, *if* we were to know that it is true. The non-trivial thesis I am exploring here is that we are to take the degree to which an explanation should satisfy us were we to know that it is true, as a guide to whether it is true.

1.2 Explanatory Urgency and Stability

Perhaps in many cases we can judge explanatory value without the aid of explicit guidelines. But it can be illuminating to consider what are some of the factors that make for a good explanation. To outline the main explanatory criterion that I will appeal to, we first need the notions of explanatory *urgency* and *stability*.

1.2.1 Urgency

What it takes to provide a satisfying explanation of some fact depends in part on how urgently it requires an explanation. Some facts stand out as puzzling and calling for explanation. Others are mundane and we are more content to ignore them. This distinction can be illustrated with any number of examples. A monkey's typing "I want a banana" calls for an explanation; her typing "l;kawje oiflkm wesdf" does not. For any two students' papers, there is some possibly very complex function mapping the characters of one onto the other. For most of these functions, the fact that two term papers stand in this relation does not call for an explanation; that they stand in the *identity* relation does. Most haphazard arrangements of pebbles strewn on the ground do not call for an explanation; their arrangement in thousands of identical rings as they are found in parts of Alaska does.

While I will not offer a thorough account of what it takes for something to call for an explanation—I will be relying on your intuitive grasp of the distinction—the following necessary condition will be helpful.

> **Salience Condition**: If *F* is a member of a homogeneous partition of *G*, and *G* was bound to obtain, then *F* does not call for an explanation.

A *homogeneous* partition of *G* is a class of states of affairs (pair-wise inconsistent, and whose disjunction is equivalent to *G*) such that no member of the class stands out as *any more* in need of explanation than any other would, had it obtained.

The principle underlies many of our judgments that a fact does not call for an explanation. We often gloss our reasoning in such cases along the lines of "Well, something had to happen, and it might as well have been this as anything else." Of course I am astonished when I win the lottery. But there is nothing deeply puzzling about it. To ask with great urgency "Why me?" would be misguided. After all, someone was bound to win; if it hadn't been me then it would have been someone else. There is nothing about *my* good fortune at having won which stands out as needing an explanation, any more than anyone else's would had they won instead. When the monkey types "lakj lwkerfo lkw," this sequence is a member of a large class of irregular, meaningless strings. That one or another of these should be typed is all but certain, as they make up the vast majority of all possible strings. This class of possible outcomes of the monkey's typing make up a homogeneous partition of the fact that an irregular, meaningless string was typed—no such sequence stands out any more than the others—and hence no such sequence calls for an explanation.

We can briefly note why this principle holds. A fact calls for an explanation when it challenges our assumptions about the circumstances that brought it about. When a monkey types an English sentence, or students turn in identical papers, or billions of pebbles are found arranged in simple geometric patterns, we might not know exactly what to conclude, but we have reason to question our initial assumptions about the circumstances that brought this about (e.g. we should wonder whether the monkey was typing randomly and whether the students worked independently). Now suppose that *F* is a member of a homogeneous partition of *G*, which was bound to obtain. It cannot be that for *every* member of this partition, had it obtained, we would have reason to question our assumptions. For we know that one of these states of affairs is bound to obtain since *G* is bound to obtain. So then we shouldn't need to learn specifically that *F* obtains in order for our initial assumptions to be challenged. But this can only show that *F* does not challenge our assumptions, and hence does not call for an explanation.

To illustrate briefly: To whatever extent I am tempted to think that my having won the lottery calls for an explanation, I should have to admit that the situation is parallel for each of the participants in the lottery. I'm no more significant than anyone else. Any reason to suppose that *my* winning calls for an explanation would also have to be a reason to think that *Smith's* winning would call for an explanation, had he won, and similarly for Jones, had he won, and so on. But someone had to win, and it cannot be that no matter who won, we would have reason to doubt our basic assumptions about the lottery mechanism, such as that it was fair. This goes to show that no matter who won, his having won does not call for an explanation.

1.2.2 Stability

Some explanations display a certain stability or robustness characterized as follows.

> **Stability**: An explanation of a fact *F* is *stable* to the extent that according to this explanation, *F* couldn't easily have failed to obtain.

A certain radium atom decayed between 12:00 pm and 12:01 pm today. Why then? According to the standard answer it was a matter of pure chance. The most thorough explanation cites quantum mechanical laws which assign this outcome a low probability. Even given all of the causally relevant factors at play, the atom could just as easily have decayed much earlier or later. This is one kind of unstable explanation which involves indeterminism. But even thoroughly deterministic explanations can be unstable in the relevant sense. A coin is tossed repeatedly, landing in the following sequence: HHTHTHHHTTHTTTHTHHTT. It could very easily have landed differently, as each toss is extremely sensitive to numerous parameters. Small variations in height, speed, coin texture, or landing-surface conditions on any toss would have made a difference to how the coin landed. Now, perhaps the dynamics of coin-tossing are roughly or even thoroughly deterministic. If so, then in principle we could give an explanation citing very specific antecedent conditions given which the coin was guaranteed to land heads. Such an explanation would still be very unstable in our sense. For as these conditions which determine how the coin lands could themselves very easily have differed, the coin could very easily have landed differently. Indeed it would be extremely *difficult* to recreate these conditions in order to obtain the same sequence again.

Compare the explanation of the coin's falling toward the ground. Here we might appeal to the mass of the coin and the earth, and gravitational law. This gives a very stable explanation as these factors all but determine that the coin will fall, and they could not easily have been otherwise. The closest possible worlds in which the coin doesn't fall are ones which have either very different physical laws, or in which the earth was much smaller (in either case, it is doubtful that the coin would even exist). Even if it is possible for the coin not to have fallen when dropped, this is not a possibility that could easily obtain.

1.3 The Stability Criterion

Explanatory urgency is a feature of facts to be explained. Explanatory stability is a feature of potential explanations. The two notions come together in the following principle

> **Stability Criterion**: Stability is a virtue of an explanation to the extent that its explanandum calls for an explanation.

There is nothing striking about our atom's decaying within a certain minute interval—it was going to decay sometime, and it might just as well be then as any other time. So it is not especially unsatisfying that we can only give an unstable explanation of this occurrence. The same goes for the coin's landing HHTHTHHHTTHTTTHTHHTT. This sequence does not stand out as urgently requiring an explanation. So it is not troubling that according to our best explanation, the conditions which led to this outcome were extremely precarious. It would be a different matter if the coin were to land heads a thousand times in a row. This sort of outcome cries out for explanation, and instability in a given explanation is correspondingly unsatisfying. It would be quite extraordinary to discover that it was just an ordinary symmetrical coin tossed in the usual way, but that if my hand had shifted just a millimeter on any toss it would not have landed heads. An explanation that makes far more sense of the phenomena suggests that there is some kind of mechanism biasing the coin toward heads (perhaps both the coin and landing surface are magnetized). On this explanatory story, the long run of heads is no longer just a fluke. The tosses couldn't easily have turned out otherwise, as in a very wide range of initial conditions—e.g., velocity of spin, distance from landing surface, wind resistance, etc.—the coin is bound to land heads.

1.4 The Value of Explanatory Stability: Expectation and Regress

When we survey a range of possible why-questions, we find that the most urgent ones are most satisfyingly answered by stable explanations.[2] Rather than multiply examples, I leave this to the reader as an exercise. But I think it can be illuminating to consider what is so valuable about stability. There are two elements to the answer, as there are two ways in which an explanation can be unstable: (a) the explanandum can remain improbable even given the explanans, or (b) the explanans could very easily have failed to obtain.

1.4.1 Expectation

One way for an explanation to be unstable is for the explanandum to remain unlikely given the explanans. At least part of what makes something an urgent candidate for explanation is that it violates our expectations. We are taken aback when the coin keeps landing heads *in part* because this is not what we thought would happen. In those puzzling cases where we are urged to ask why it is that *P*, it initially strikes us that alternative possibilities were equally open. An unstable explanation given which we still shouldn't expect that *P* will leave us still wondering why it is that *P*.

That expectedness of the explanandum can be an explanatory virtue is reflected in the fact that influential philosophers such as Hempel (1965) have insisted that it is simply a necessary condition on something's counting as an explanation at all. Most contemporary philosophers do not follow Hempel this far. On a wide variety of theories of explanation (e.g. Salmon [1984], Lewis [1986], Railton [1978]), we can explain why our radium atom decayed when it did by citing the quantum mechanical laws that confer a very low chance on this occurrence. We needn't fuss over whether this counts as a genuine explanation. What matters for our purposes is that such low-chance explanations provide us with little of what we are looking for in an explana-

[2] Some cases must be handled with care. Janitor Joe is suddenly seen driving a Mercedes. This is surprising and appears to call for an explanation. It is entirely satisfying to learn that he won the lottery. Yet this explanation is unstable in that he could very easily have failed to win. I think what we should say about this case is that any apparent explanatory urgency here is misleading. We are surprised to see Joe in a Mercedes as he usually has to take the subway. But someone was bound to win the lottery, and it is no more puzzling that it should be Joe than anyone else. Nor is it puzzling that he should splurge in this way if he should win.

tion, at least when the need for explanation seems particularly urgent.

To illustrate, suppose I were to levitate off my chair and float across the room and out the window. I should very much like to learn why this was happening. Now, contemporary physics suggests that this is quite possible, although the chance of its happening is extremely low. Randomly moving air molecules may all happen to align and move upwards with enough momentum to lift me. It may be that the correct and most thorough explanation of why this happened appeals to nondeterministic laws which assign it an extremely low probability. As I drift across Manhattan, were I somehow to learn that such an explanation is correct, including every detail of the physical laws, I would hardly respond, "Oh, okay, so that's why I'm floating—for a while there I was puzzled." Indeed I might be even more puzzled by this explanation than if I learnt that it was because a witch put a spell on me. The latter explanation would be hard to swallow, since I don't believe in witches. But at least if there *were* witches we might expect this sort of thing to happen. It cannot possibly alleviate my bafflement to be told, "These are the relevant physical laws and conditions, and one possibility that they allow for is the one you're in. It is far more probable for this not to occur. Nevertheless, it did."[3]

1.4.2 Explanatory Regress

As we have noted, an unstable explanation needn't assign a low chance to its explanandum. An explanation may be deterministic, citing conditions and laws which guarantee that its explanandum will obtain, and yet still be extremely unstable, as these determining conditions might be very precarious. Suppose that the laws of physics turn out to be

[3] See Strevens (2000) for a defense of the explanatory value of high probabilities in less fanciful cases.

thoroughly deterministic. My levitation would then be the inevitable result of some extremely precise initial conditions. Various particles happened to be arranged in just the right positions, with just the right velocities, that they would move and collide and ultimately pick me up and toss me out the window. A full and precise explanation would render my behavior highly probable, even certain. Yet it couldn't be any more satisfying than the low-chance explanation considered above. Why not?

The problem is that in answering one explanatory question we have just raised an equally urgent one. As my two-year-old daughter has discovered, any answer to a why-question can be the subject of a further why-question, whose answer can raise a further why-question, and so on *ad infinitum*.[4] All explanations must end somewhere, but some stopping points are more satisfying than others. Suppose as I float across to Brooklyn I am able to calculate the precise positions and momenta of all the surrounding air molecules two seconds before my launch. From these and the deterministic laws I am able to deduce my precise trajectory. All this could do is raise the question of why the molecules were in *this* very state to begin with. After all, they could very easily have been in a vast array of possible states, almost none of which lead to anything like this strange occurrence. And of course it will not help to trace the causes back two seconds earlier, or five minutes, or a billion years. Any explanation that terminates in some extremely precarious conditions simply passes the buck by raising the question of why these precise conditions held. That particles were once in a state that lawfully leads to my levitation is no less ur-

[4] I've discovered that in special cases the two-year-old explanatory regress has limits. A recent exchange went as follows: "We have to go home soon." "Why?" "Because I'm really tired." "Why?" "Because I'm sick and the medicine I took me makes me drowsy." "Why?" "Because it has a dormative virtue." "…Oh."

gently in need of explanation than the levitation itself.[5]

One final example illustrates how unexpectedness and explanatory buck-passing are unsatisfying features of unstable explanations. You enter a room and look around at the walls and ceiling through a pinhole in a piece of cardboard, seeing nothing but purple. Why have you seen only purple? It might be that the room is mostly white, with a thin squiggly purple line sprayed along the walls and ceiling by a vandal. If this were so, your observation would be a highly improbable coincidence. An explanation which stopped there would be hopelessly unsatisfying, since given this story, your observation would go against all expectation.

But we could always strengthen this explanation to make our explanandum highly probable given the explanans. A richer explanation might describe the exact shape and location of the purple line with respect to you, the angle of rotation of your head, and give a geometrical proof that in these circumstances you were bound to see only purple, even though the rest of the room is white. It might go further and describe the series of muscular contractions that you and the spray-painter underwent which caused this. Here the explanans might guarantee the explanandum, but the explanation is equally unstable. The slightest variations in muscular movement would have caused you to see the white wall. Learning that this explanation was correct could only force the question of why the vandal's and your muscular activity were so perfectly coordinated that your line of vision happened to exactly match the contours of the irregular purple line. The weaker explanation was unsatisfying because the explanandum was not even to be expected given the explanans. The strengthened explanation is unsatisfying because it passes the buck. Of course the far more stable and satisfying explanation is that the room is painted mostly purple. It is only on this story that your having seen only purple was not highly precarious.

[5] Perhaps the universe has no beginning, and every prior state of the universe is lawfully determined by an earlier one, each of which ultimately determines that today I will be lifted up and away and set down gently in a deck-chair on Coney Island. That the universe was ever in such a state remains deeply puzzling, even though our explanation of any particular state needn't ultimately terminate in an appeal to some earlier state.

2. Explanations and Induction: General Considerations

The explanatory criterion outlined above has broad application, but I will restrict my attention now to enumerative induction. The main purpose of this section is to get clear on what facts are to be explained in inductive reasoning, and how in general they might best be explained.

2.1 Explaining Observations: Instances and Absences

The basic idea is that our support for the generalization that all *F*s are *G* depends on how well it can explain our data, when we have observed many *G F*s and no non-*G F*s. But here we need to note a typically overlooked but crucial ambiguity. There are at least two explanatory tasks we might have in mind here:

E1: explaining, concerning the *F*s we have observed, why they are *G*
E2: explaining why all observed *F*s are *G*

The explanatory tasks are very different, as they involve different explananda:

F1: the fact that *these F*s are *G*
F2: the fact that all observed *F*s are *G*

These are quite different facts, since although *these* may actually be the observed *F*s, they need not have been, as we needn't have observed the *F*s that we actually did. F1 is a claim just about certain *F*s, namely that they are *G*; F2 is a claim about *us*, namely that we haven't set eyes on a non-*G* *F*. In E1 we are explaining *instances* of the generalization that all *F*s are *G*; in E2 we are explaining the *absence* of observations of counter-instances.

Explaining why *these F*s are *G* can be a very different matter from explaining why the observed *F*s are *G*, even though these are the observed *F*s. The reason why all observed stars lie within a trillion light years of the earth is that light would not have had time to reach us from any further stars. But our relative position as observers does nothing to explain the fact that these very stars are located where they are. It is often claimed that a *lawful* generalization explains its instances: e.g., the law that gases expand when heated (at constant pressure) explains why the gas in this balloon expanded when it was heated. But even if this is so, there is no guarantee that this generalization explains why all observed portions of gas have expanded upon heating. Perhaps a demon has a firm policy of shielding us from observing any non-expanding heated gases, if there were any. But he could easily change the laws to make gases contract when heated. In this case it is not the physical law but the demon's policy of shielding our observations that explains why all observed portions of gas expand when heated (the physical law may still explain why *this* portion of gas expanded).

This distinction is not typically drawn in the literature, which often leads to confusion. For example, in the following passage, Armstrong (1983) argues that the mere fact that all *F*s are *G*s *cannot* explain the fact that all observed *F*s are *G*s.

> That all *F*s are *G*s is a state of affairs which is in part *constituted* by the fact that all observed *F*s are *G*s. 'All *F*s are *G*s' can even be rewritten as 'All observed *F*s are *G*s and all unobserved *F*s are *G*s'. As a result, trying to explain why all observed *F*s are *G*s by postulating that all *F*s are *G*s is a case of trying to explain something by appealing to a state of affairs part of which is the thing to be explained. But a fact cannot be used to explain itself. And that all *unobserved F*s are *G*s can hardly explain why all observed *F*s are *G*s. (p. 40)

To see why Armstrong is wrong here, note that we can rewrite "All observed *F*s are *G*s" as "No non-*G F*s have been observed," and "All *F*s are *G*s" as "There are no non-*G F*s." Now our explanatory question and answer become: (Q) Why have no non-*G F*s been observed? (A) Because there are none. This appears to be a perfectly good explanation. The reason no one has set eyes on a flying pig is that there aren't any. A deeper explanation may go further and say why there are no flying pigs, but this does not invalidate the explanation we have given.

What about Armstrong's complaint that the fact that all unobserved *F*s are *G*s can hardly explain why all the observed ones are? The plausibility of this argument seems to depend on a conflation of the two explananda we distinguished above. If we are asking of *these F*s—the ones we happen to have observed—why they are *G*, then it does seem irrelevant to note that the other *F*s which we have not seen are *G* also. A number of philosophers have denied this, insisting that we can explain certain *F*'s being *G* by appealing to all *F*'s being *G*.[6] With Armstrong, I find this implausible. If we collect all the ravens we have observed in a cage and ask, concerning these ravens, why they are black, it hardly helps to be told that all other ravens are black too. Presumably *these* ravens would have been black no matter what color the other ravens were or if other ravens even existed. Perhaps the ravens in the cage are all the ravens there are. In this case the claim that all the other ravens are black

[6] According to Harman (1973), "That all emeralds are green does not cause a particular emerald to be green: but it can explain why that emerald is green." See also Lewis (1994) and Loewer (1996).

may be vacuously correct, but could hardly explain anything. An adequate explanation will tell us, for each raven, why it is black, and this should appeal to some feature of each raven by virtue of which it is black. Such an explanation will be adequate to explain why these ravens are black even if all the others are pink or green. Furthermore, if the other ravens are all black too, we should expect our explanation of why these ones are black to carry over to explain why the others are too.[7] And the fact that the other ravens are black could hardly play this role.

So I think Armstrong is right to be suspicious of an explanation of certain *F*'s being *G* which merely appeals to the fact that certain other *F*s are *G*. However, the fact we were originally trying to explain was not that any particular *F*s are *G* but that *all observed* F*s are* G. The color of the unobserved *F*s *is* relevant to explaining this fact. The crucial point here is that we need not have observed these very *F*s. Had our investigations gone slightly differently we would have observed some *F*s other than these ones. If the unobserved *F*s are not *G*, then had we observed not these *F*s but others, then it would not be the case that all observed *F*s are *G*. If the unobserved *F*s are *G*, then it is to be expected that we only observe *F*s which are *G*, regardless of which *F*s we observe. If they are not, it remains a mystery that we should happen to have avoided observing these non-*G F*s. If what we are seeking to explain is our failure to make certain kinds of observations, then the properties of certain unobserved things is crucial, for it may make our explanandum more *stable*, i.e., less dependent on arbitrary conditions which could very easily have failed to obtain.[8]

[7] This needn't always turn out to be the case. Perhaps these ravens are black because of their genes, but the rest are albinos that got caught in an oil spill. Still, the fact that the other ravens are black is doing no explanatory work here.

[8] Dretske (1977) presents an argument similar to Armstrong's. He suggests that what explains my having seen only red marbles from an urn is not merely that all marbles in the urn are red but "the fact (law) that you cannot draw nonred marbles from an urn containing only red marbles" (p. 256). That you cannot draw non-red marbles if there are none strikes me as a trivial, necessary truth; it is hard to see what crucial explanatory work it is doing.

In the literature on enumerative induction stemming from Hempel (1945), there has been an unhealthy emphasis on the *instances* of inductive generalizations. Our knowledge of such instances need not provide much evidence for a generalization at all. It is often easy enough to find positive instances of a false generalization. If I empty a bag of change on the floor, there are bound to be plenty of instances of the generalization 'All coins in the pile are heads-up'. Suppose I am given a collection of coins with heads facing up and told that these are among the coins in the pile. I learn that this coin from the pile is heads-up and this one is too, and so on. How much help is this in assessing the claim that all coins from the pile are heads-up? Very little. What I have learnt is that *some* of the coins landed heads. This hardly suggests that all of them did. The further fact that *these* coins are among the heads adds nothing significant to my evidence. Note that this has nothing to do with the fact that the all-heads hypothesis is unlikely, to begin with. Perhaps I know that there are bags of double-headed coins about, and the emptied bag might be one of those. Being handed some heads-up coins gives me very little evidence that they are from a double-headed bag. This is because the double-headed–bag hypothesis provides, at best, only a marginally better explanation of some of the coins having landed heads.[9]

It is a familiar point that the evidential strength of our

[9] The explanatory difference may be enhanced if I know what proportion of the pile I have been handed. If I was given considerably more than half of the coins, the all-heads hypothesis becomes more plausible. But as we will see shortly, even a small proportion heads from the pile can strongly confirm that they are all heads if we know they were selected randomly.

data depends on whether and how our sampling procedure was *biased*. If I know that my friend who passed me the coins deliberately picked only heads, then my evidence that they are all heads may be very weak, or nil. If I knew he had closed his eyes and picked the very same coins at random, then my evidence will be much stronger. This has struck some authors as paradoxical.[10] How can there be this difference, if in either case I am handed the same bunch of coins, and hence learn that each of these coins show heads? The answer is that the aspect of our evidence that is most relevant here is not that certain coins show heads, but that all the coins that I have observed show heads. For if the sampling procedure was biased toward coins showing heads, I can fully explain why I have seen nothing but heads from the pile, without any appeal to the remaining coins at all. The random-sampling assumption rules out this hypothesis which is potentially in competition with the all-heads hypothesis.

2.2 Laws and Explanations of Explanatory Generalizations

Many philosophers have thought that there is an important connection between laws of nature and enumerative induction. Armstrong (1983) agrees that induction involves an inference to an explanation, but argues that a Humean regularity account of laws is not up to the job and hence leads to inductive skepticism. According to Armstrong, a law "serves first to explain why all observed *F*s are *G*s, and, second, to entail that any unobserved *F*s there are will be *G*s" (p. 55). So our pattern of inference runs *observed instances*→ *law*→ *unobserved instances*. But since on a regularity account a law just consists in the fact that all *F*s are *G*s, our inference is just *observed instances*→ *observed instances + unobserved instances*→*unobserved instances*, which reduces to *observed instances*→ *unobserved instances*. And without an explanatory mediator, such an inference is unjustified.

The major problem with this argument has been diagnosed above as involving a conflation of the fact consisting of instances of the generalization, and the fact that no counter-instances have been observed. But I think that there is an important point lurking here: explanatory appeal to a generalization can often be more satisfying if we have some idea of how the generalization itself might be explained. And consequently the strength of an inductive inference depends on our prospects for explaining the explanatory generalization. Indeed, in many cases where we are sure that there can be no explanation of the generalization, our observations give us little or no evidence for the generalization. If I toss a coin many times and observe nothing but heads, I have some reason to think that it has always landed heads and always will. But the strength of this evidence depends on our answer to the question 'Why does it keep on landing heads?' Suppose we are quite certain that the coin is evenly weighted and tossed in the ordinary way, and so if it does land heads every time this is just a fluke. In this case we have no reason to expect any other tosses to land heads. We have reason to think that all tosses of the coin have landed heads only to the extent that we can suppose that, say, the coin is double-headed, or something of this sort that could explain the generalization. This fits well with the explanatory criteria we have discussed. To be told that we have observed only heads tosses because the coin always lands heads can only prompt the question 'Why does it always land heads?' That it should just happen to land heads every time by a fluke is every bit as baffling as if it lands heads about half the time but by some fluke we see only the heads landings. In either scenario, the explanation of our observation is extremely unstable. Had we tossed the coin on other

[10] See Stuart (1962) and Howson and Urbach (1993).

occasions in even slightly different circumstances it would not be the case that all observed tosses landed heads.

The claim of Armstrong and other anti-Humeans (e.g., Dretske 1977) is that its being a *law* that all *F*s are *G*, can explain why all *F*s are *G*. This fits nicely with my account, since one thing that distinguishes genuine laws from accidental generalizations is their *stability*. It is accidental that all the coins in my pocket are copper, but it is a law that copper conducts electricity. As it happens, all the coins in my pocket are copper, but this could easily not have been so had I ordered a different sandwich for lunch, or grabbed some quarters for the parking meter, or the like. But all these copper coins would conduct electricity regardless of particular circumstances. If there is a possible situation in which they do not, it is one that would not obtain, except in very peculiar circumstances.

Nevertheless, I think that the role of laws in induction has been vastly overrated. Armstrong's argument suggests that we can infer that all *F*s are *G*s, having observed *G F*s only if there is a law that all *F*s are *G*, which can mediate the inference. This has been endorsed by many philosophers (e.g., Goodman [1950], Sheffler [1963]) not all of whom share his anti-Humeanism. Goodman goes so far as to suggest that confirmability by accumulation of instances is the defining feature of laws as opposed to accidental generalizations. This is clearly a mistake, as Goodman's own examples can illustrate. Suppose Jane has a deep pocket full of a hundred coins. Having pulled out ninety-nine copper coins, we should be fairly confident that the remaining one is copper also.

But while it is not a law that all coins in her pocket are copper, our inference to this generalization should be driven by considerations of what might explain the generalization. In this case a basic explanation is that she chose only copper coins to fill her pocket. This would give us some understanding of why all the coins in her pocket are copper. The explanation, and hence the inference, is strengthened to the extent that we can explain why she might do this, for instance because only pennies are copper and they are not very useful and she is at the bank where they can be deposited. The overall lesson is that the strength of an inductive inference to a generalization can depend on our prospects for explaining the generalization. But this higher-order explanation may or may not directly appeal to a law.

However, just as laws aren't always needed for induction, we shouldn't overestimate the necessity of higher-order explanations. In a discussion very much in the spirit of mine, Peacocke (2004) claims that one is entitled to infer that all *F*s are *G* from observations of *G F*s *only if* our data also make it plausible that there is some condition *C* that explains why all *F*s are *G*. I think this is too strong. Suppose it's raining cats and dogs. About equal numbers of cats and dogs are falling randomly from the sky, many of them landing in the trash bins on your block.[11] A reliable source tells you that while most of the bins may have collected about as many cats as dogs, at least one of them happens to contain only dogs. Having no idea which one it is, you start emptying one of the bins. Each of the many animals you retrieve is a dog. This can give you rather strong inductive evidence that all the animals in the bin are dogs. The reasons for this are the same as usual: there being no cats in the bin explains well why you haven't seen any. This inference may be appropriate even if you are quite certain that there is no condition *C* explaining why all the animals in the bin you are emptying are dogs. For the cats and dogs fell randomly, and your choice of bin to unload was entirely accidental. Perhaps it is

[11] The cats and dogs are stuffed toys that fell from the cargo hold of an airplane. No animals were harmed in the making of this thought-experiment.

to some extent unsatisfying that the inductive generalization remains unexplained. But given the size of the downpour, it is not so surprising that some trash bin should collect dogs but not cats, even if this occurrence is rare. What gives the generalization a boost of credibility in this case is an explanatory relation in the opposite direction. The bin's containing only dogs can explain why you received the report that you did.

2.3 Instantial and Selectional Explanations

Most philosophers who emphasize the importance of laws to induction focus on a law's ability to explain its instances. To take a standard case, its being a law that gases expand when heated at constant pressure explains why the portion of gas in this balloon expanded, given that it was heated at constant pressure. Now, I am not sure how great an explanation this is, as it offers me only the barest understanding of why the gas expanded. I gain a fuller understanding by learning of the molecular structure of gases, how heating involves an increase in kinetic energy, and so forth. But let us grant that its being a law that all *F*s are *G* is at least a minimal explanation not only of why all *F*s are *G*, but also for each *F*, why it is *G*.

Not all explanations of a generalization which support inductive inferences extend to an explanation of its instances. Picking objects out of a barrel, I find them all to be green. I find a grasshopper, an emerald, a Boston Celtics T-shirt, a cucumber, a plastic St. Patrick's Day hat, a painted wooden frog, a photograph of a field in Vermont… . After seeing dozens of green things, I should expect to see more. Now if all contents of the barrel are green, this is not because it is a *law* that they are; perhaps it is because whoever filled the barrel selected only green things. But this policy of selecting only green things does nothing to explain why the grasshopper is green, or why the T-shirt is green, or why anything is. There *is* no unified explanation concerning the barrel contents, why *they* are green—the explanation is very different in each case.[12] What is explained, rather, is for various non-green things, why they are not in the barrel. So we have two possible kinds of explanation for why all *F*s are *G*, which we might call the *instantial* and the *selectional*. In an instantial explanation, we explain why all *F*s are *G* in a way that can explain for each *F*, why it is *G*. The explanation in each case is that it is an *F*, and it is a law that all *F*s are *G* (perhaps with a further elaboration of why the law holds; we often appeal to a feature *H*, common to the *F*s, which causes them to be *G*). In a selectional explanation we explain why all *F*s are *G* in a way that can explain for a non-*G* thing, why it is not an *F*. The explanation in each case cites the fact that it is not *G*, together with the conditions that prevent anything non-*G* from being an *F*.[13]

Selectional explanations have been neglected in discussions of induction. But such explanations are not at all uncommon and are not restricted to contrived cases where an agent chooses to select things of a certain sort. We might find all objects caught in the bushes in a flowing stream to be of roughly the same size. Perhaps this is because smaller objects can pass between branches, while larger objects, being heavier, have enough momentum to push through. This explanatory hypothesis can support an inference to further

[12] Inspired by Nozick's (1974) discussion of hidden-hand explanations, Lipton (2004) makes a similar point with the example of a club that admits only redheads. The admission policy does not explain why Arthur, a club member, has red hair. In a similar vein, Sober (1984) argues that natural selection can explain the frequency of a trait within a population but not why any particular individual possesses it.

[13] Lipton (2004) argues that the crucial difference between the hypotheses that all emeralds are green and that they are all grue is that the latter "does not explain why an object is grue, because 'grue' does not pick out the sort of feature that can be an effect" (p. 101). Given the arguments in this section, this cannot be the right diagnosis of the grue problem.

objects in the bush being the same size, while it fails to explain why any object is of that size. Darwinian explanations involve a combination of instantial and selectional explanations. Very roughly, we explain why all current *F* organisms have trait *G*, by noting that the *F*s share certain genes and citing the law that all organisms with these genes have *G*. We explain their sharing these genes in turn by appeal to the ancestry of the *F*s. But at some point we ask why none of a certain group of organisms lacking the *G*-gene had descendents which are current *F*s. Our answer is that their lack of *G* prevented them from surviving and reproducing.

3. Explanation and Induction: Specific Applications

It is time to focus on a variety of specific problems and principles of inductive reasoning. In each of the cases discussed in this section, a competent reasoner should find it fairly obvious which inferences one should draw or refrain from drawing. My strategy will be to show that explanatory considerations endorse lines of reasoning that we can independently recognize as correct.[14]

3.1 Deviant Inferences

A great many inductive inferences fit a simple pattern: having observed a number of *F*s which are *G* and no non-*G* *F*s, we infer that all *F*s are *G*, or at least that the next *F* we come across will be G. It is a mistake however, to suppose that criteria for good inductive inference can be captured by such a schema. For there are plenty of cogent lines of reasoning that deviate from this general pattern, and plenty of fallacious ones which fit it well. I will consider three such deviant inferences, and discuss how explanatory considerations make sense of them.

First, there are situations in which discovering many *F*s which are *G* should lead us to expect that the next *F* will *not* be *G*, while also confirming that all *F*s are *G*.[15] If we have a good reason to believe that someone at the party is not a philosopher, then as we meet lots of people and find that they are all philosophers, our confidence that the next person we meet will be a philosopher should *decrease*. For we are narrowing down the pool in which the non-philosopher might be. But at the same time, our meeting only philosophers suggests that perhaps everyone at the party is a philosopher.

Second, there can be cases where the accumulation of instances (and no counter-instances) of the claim that all *F*s are *G* disconfirms, not just the next observed *F*'s being *G*, but the generalization itself. I'm looking for *F*s in a certain region, say, weeds in the garden. As I find lots of weeds all over the place, there will still be a scattered region R in which I have not looked and hence not found weeds. The evidence I gain by finding lots of weeds outside of region *R* does not confirm the hypothesis that all weeds lie outside of *R*—it disconfirms it. That there are so many weeds surrounding region *R* makes it very plausible that there are some in *R* also.

Third, it may even be that our finding many *F*s which are *G*, and no *F*s which are not *G*, *disconfirms* the generalization that all *F*s are *G*, while *confirming* that the next observed *F* *will* be *G*. I dispatch my researchers to collect weeds and report where they came from. They come back with weeds from all but one corner of the garden. This may suggest that there are no weeds in that corner; i.e., it may support the hypothesis that all weeds lie outside the corner. But given different background assumptions, it may not. There is a

[14] Many of the cases I discuss here have received attention from other approaches—most prominently, Bayesianism. There are subtle questions concerning the relation between Bayesianism and "explanationism". In some cases my story dovetails neatly with a Bayesian approach. In others the relationship is more complicated. I can't address these matters here. My strategy is just to present the explanationist approach as it is to be judged on its own merits.

[15] Here and throughout I use 'confirms' in the sense of "provides some evidential support". If E confirms H, then upon learning E one's confidence in H should rise by some degree.

prickly bush in that corner, but I am not sure how much of a deterrent it will be to looking there. The lack of weeds brought back from the corner will confirm that the prickly bush did deter them, and hence confirms that the next weed I see will not be from that corner, even though the many weeds found surrounding the corner suggests that there are weeds in the corner also (i.e., it disconfirms the hypothesis that all weeds lie outside the corner).

Each of these cases can be neatly accounted for by explanatory considerations. The hypothesis that everyone at the party is a philosopher might well explain why I have only seen philosophers so far. So I have gained at least a little support for this generalization. If it were not for my prior conviction that not everyone there is a philosopher, the support for this generalization would carry over to its consequence that I will meet another philosopher next. The inference to the next observed instance must be mediated by the explanatory generalization, for my meeting another philosopher next cannot help explain why I've met only philosophers so far. But in the case as described, the support for the generalization is minimized to the extent that I have strong initial reasons to doubt it. Any support that might otherwise have accrued to the next observed instance is offset by the fact that I am more likely to meet a non-philosopher if there are fewer people left among which to find him.

In the first case of the weeds, we have the fact that all observed weeds are in the scattered region $\overline{R}$ (the complement of R). In explaining this fact we have no need to appeal to the hypothesis that all weeds are in $\overline{R}$. For we already know the correct explanation which covers it. The reason we have seen weeds only within $\overline{R}$ is that that is the only place we have looked so far. Now, this explanation, we should note, is unstable. For we could easily have chosen to look in a region other than $\overline{R}$, so if there are weeds beyond $\overline{R}$, then it could very easily not have been that all observed weeds are in $\overline{R}$. But by the Stability Criterion this should not concern us, as the fact that we have only observed $\overline{R}$-located weeds was not in urgent need of an explanation in the first place. No matter where we looked, it would be the case that all observed weeds were within this observed region. And the fact that all observed weeds are in $\overline{R}$ no more stands out in need of explanation than that they have been observed only in some other region. We had to look somewhere, and it is no more surprising that it was in $\overline{R}$ than anywhere else.

So the fact that all observed weeds are in $\overline{R}$ does not support the generalization that all weeds are in $\overline{R}$. But there is another aspect of our data that could do with an explanation: the fact that we have observed many weeds. A nice explanation of this is that there are weeds scattered throughout the garden. The suggestion that all the weeds are just within region $\overline{R}$ provides a very poor explanation of this fact. If this were the case, it would be surprising that we have seen so many weeds, since we could easily have looked in a different region and seen fewer or none. In addition to the fact that we have *observed* so many weeds, there is the fact that there *are* so many weeds. A good explanation of this is that the weeds tend to spread easily and the garden has been neglected. This should also lead us to expect weeds outside of region $\overline{R}$. If there are so many weeds within $\overline{R}$ but none elsewhere, we would have a hard time explaining why they haven't spread. Perhaps the weeds outside of $\overline{R}$ have just recently been removed. But if a gardener were to pull up so many weeds, why didn't he do it more systematically than leaving a widely scattered weedy region? And even if he did, isn't it surprising that the region he left untouched is the very one in which we looked? These lines of explanation are less plausible as they don't terminate in a satisfying way.

The situation is different in the third case when my researchers bring me many weeds found all over the garden except one corner. One part of a good explanation of their having found so many weeds in such a wide region of the garden is that they searched very broadly and thoroughly. This suggests that they looked in all corners, and so a good explanation of their having found none in one corner is that there are none in that corner. But our knowledge of the prickly bush undercuts this explanation. For now we have a competing explanation: the researchers found no weeds in that corner because they didn't look there. This explanation becomes viable only because we have a nice explanation of it in turn. Simply to suggest that they failed to look in this corner would not be satisfying by itself. If they looked everywhere else, why didn't they look there? The deterrence of the prickly bush provides the satisfying explanation, as it makes sense that they should want to avoid getting prickled. As in the case of region $\bar{R}$, the fact that there are so many weeds surrounding this corner confirms that there are some in the corner, too (i.e., it disconfirms the hypothesis that all weeds lie outside that corner). The hypothesis that the researchers are avoiding this corner does not undermine the evidence that there are weeds in there, while it does suggest that the next observed weed will not be from the corner.

3.2 The Raven Paradox

The familiar puzzle of the ravens arises as follows (Hempel [1945]). It would seem that the generalization that all ravens are black is confirmed by observations of its instances, namely, black ravens. But 'All ravens are black' is equivalent to 'All non-black things are non-ravens,' and instances of the latter are non-black non-ravens such as blue jeans and green markers. Surely if E confirms *H*, then E confirms anything known to be equivalent to *H*. So this seems to allow us to do indoor ornithology: amassing examples of non-black non-ravens around the office we are gathering evidence that all ravens are black.[16]

The key to untangling the puzzle of the ravens is to focus on the fact that all observed ravens are black, rather than the instances of black ravens or non-black non-ravens. While the instances of a generalization and its contrapositive are entirely distinct, the fact that all observed non-black things are non-ravens just is the fact that all observed ravens are black. Whether we are out in the field searching for ravens that all turn out to be black, or hunting around in the office for non-black things which turn out not to be ravens, we learn that we have observed no non-black ravens. One explanation for this is that there are none out there to be seen, that is, that all ravens are black. But how *necessary* this explanation is depends on how we have gone about making our observations. If we have gone outside to find ravens, then our failure to find non-black ones does seem best explained by there not being any to find. For if there are non-black ravens we should expect to have come across one by now. That we have not would just raise a further explanatory question which would be difficult to answer: why have we happened to miss all the non-black ravens? It is unsatisfying to be left with no answer to this question. It won't do simply to answer that something or other is preventing us from seeing the non-black ravens. We need to consider some more concrete suggestions.

Perhaps I have some cognitive quirk that is triggered by the presence of a non-black raven causing me to hallucinate that it isn't there, or that it is black. But now even if we are

[16] My response to the puzzle is in the same spirit as Horwich's (1983) Bayesian response, although I hope to have been clearer about the nature of our evidence. One advantage of my account may be that it makes intuitive sense without knowledge of probability theory. Lipton (2004) proposes an explanationist solution of a very different kind.

to entertain this odd hypothesis, it just raises a further explanatory question: why do I have this strange quirk? It is hard to imagine a satisfying evolutionary explanation. Such an explanation would have to involve some arbitrary conditions that somehow favored organisms having such a quirk, plus the occurrence of the necessary mutations. But it would be a curious fact that conditions happened to obtain which would result in people having such a quirk. We should still be compelled to ask why this is the case. No explanatory satisfaction seems to lie down this path.

I am suggesting that all observed ravens are black because all of them are black. Doesn't this just raise the further question of why all ravens are black? Yes, but this line of questions seems to terminate in a more satisfying way. It seems plausible that we could explain for just one black raven, why it is black. Presumably it has something to do with the raven's genes, and how they are expressed its development. Once we have this individual explanation, we have what we need to explain why they are all black, namely, that they all share certain genes. Why do they all share these genes? Because they got them from a common source, as they are all related. This line of questions does not seem to go on forever, with each answer just as puzzling as the previous one. So if we have gone out looking for ravens, the hypothesis that all ravens are black seems to be the best explanation for all observed ravens being black.

Not so if we have just sat in our armchairs and catalogued all the non-black things in the office. In this case it is no surprise that all observed ravens are black (or equivalently, all observed non-black things are non-ravens) since we shouldn't expect to come across any ravens, let alone any non-black ones, regardless of what proportion of all the ravens are black. In this case the hypothesis that all ravens are black is explanatorily redundant. We should not expect what we have observed to be any different regardless of whether the hypothesis that all ravens are black is true or not. Even if we stepped out and went about looking at things in general, noting whether they are non-black ravens, we should not be surprised at failing to see a non-black raven, even if there are some. For whatever proportion of the raven population is non-black, the non-black ravens will make up a small proportion of things in general, since ravens of any sort are not that common.

To make our failure to find a non-black raven matter the most to our general hypothesis, we need to try our hardest to find the non-black ravens, hence ruling out or minimizing the plausibility of alternative explanations of our finding only black ravens. Strictly speaking, however, just trying but failing to find non-black ravens needn't be sufficient for inductive support. Suppose our raven hunt takes us to breeding and feeding grounds that we expect to be teeming with ravens. But oddly enough, we don't find a single one. Now we have failed to find a non-black raven, even though we should expect to, given our method of inquiry, if not all ravens are black. Should this strongly suggest that all ravens are black?

Obviously not. And this might suggest that the *instances* of a generalization are still very relevant to its confirmation, as without them we have no confirmation at all. Let's first see what has gone wrong in this case. True, all observed ravens are black, that is, we've failed to find non-black ravens. But there is a stronger fact calling for explanation here: we've failed to find *any* ravens. Whatever might account for this—perhaps the local population has been wiped out by a regional disaster—will perfectly well explain why we have specifically found no non-black ravens. However we manage to find no ravens, we will find no non-black ones regardless of whether all ravens are black. So the hypothesis that all ravens are black will in this case be explanatorily redundant.

Typically, the strongest inductive support for a generalization is in cases where we do observe some of its instances. For this restricts the kind of alternative explanations available. We can't say that the local raven population has been wiped out, since we have seen some ravens. We need a more restricted hypothesis, but restrictions tend to introduce instability and thus raise further explanatory queries. Perhaps only the non-black ravens were destroyed in a local bush fire. But why only the non-black ones? If any ravens managed to escape, they could just as easily have been pink or green ones as black ones. We might appeal instead to some selective disadvantage among the non-black raven population. Perhaps there were fewer of them, they couldn't breed with black ones, and competed for food, eventually getting wiped out. But if this is a good explanation, then we should expect it to apply not just locally but everywhere, thus only confirming that there are *no* non-black ravens. If we have seen black ravens but no non-black ones, it is hard to find a rival to the explanation that all ravens are black.

It is a mistake, however, to suppose that in the case where we find no ravens at all, we have necessarily gained no support for the hypothesis that all ravens are black. Initially there may be room for the hypothesis that there is a local population of green ravens in the region we are about to investigate. Environmental conditions may be markedly different there, allowing for the possibility of a satisfying explanation of there being some green ravens in this region only. When we go there and find no ravens, we can eliminate this possibility. So while the hypothesis that all ravens are black does not explain our failure to find the local green ravens, it is indirectly supported by the elimination of an alternative hypothesis.

3.3 The New Riddle of Induction

All observed emeralds are green. What color are the rest? It won't do to rely on a schema like 'All observed *F*s are *G*, therefore all *F*s are *G*.' Goodman showed that we can define 'grue' as being either observed and green or unobserved and blue.[17] Then if we put "grue" in for "*G*" we get the conclusion that the unobserved emeralds are blue. The challenge is to show how our preference for the all-green over the all-grue hypothesis is not arbitrary.

Let's see how the all-green and all-grue hypotheses stand up as explanations. That all the emeralds are green would nicely explain why all the observed ones are green. Any other suggestion regarding the color of the remaining emeralds would offer a poorer explanation, for it would render it surprising that we haven't encountered a non-green one. But the observed emeralds are also grue, since they are all observed and green. It might then seem—by reasoning analogous to that in the green case—that the hypothesis that they are all grue provides a more satisfying explanation of our observing only grue emeralds than one that says that the remaining emeralds are not grue. So far, the cases for projecting greenness or grueness appear symmetrical.

But let's be clear on what these "gruesome" claims amount to. The fact that all observed emeralds are *grue* is just the fact that all observed emeralds are *green* (if something is green and observed then it is grue; if something is grue and observed, it is green, since it is not blue and unobserved). And the hypothesis that all emeralds are grue says that all the observed ones are green while the rest are blue. So the explanation 'All observed emeralds are grue, because all emeralds are grue' really amounts to 'All observed emer-

[17] Goodman's original definition was in terms of having been observed *before time t*. For simplicity I am dropping this by letting *t* be the present.

alds are green, because all observed emeralds are green and the unobserved ones are blue.' This doesn't seem like much of an explanation at all. We have just repeated the fact to be explained and tacked on a claim about the remaining emeralds' being a different color. The first part offers no explanation, and the second just makes it worse, by raising the question of why we have failed to see these blue emeralds.

Our concern here should be to see whether this line of reasoning can be mirrored by one of equal force in terms of 'grue'. That would leave us open once again to the charge of arbitrariness in our choice of properties to project. The Martians, we are told, find it natural to inductively reason in terms of the schmolors grue and bleen (x is bleen iff x is blue and observed or green and unobserved). They respond as follows.

> These Earthlings want to explain, as they would put it, the fact that all observed emeralds are green, by suggesting that they are all green. But let's be clear on what these "greensome" hypotheses amount to. The fact that all observed emeralds are green is just the fact that all observed emeralds are grue. And the hypothesis that all emeralds are green says that all the observed ones are grue while the rest are bleen. So the explanation 'All observed emeralds are green, because all emeralds are green' really amounts to 'All observed emeralds are grue, because all observed emeralds are grue, and the rest are bleen.' This doesn't seem like much of an explanation at all. We have just repeated the fact to be explained, and tacked on a claim about the remaining emeralds being a different schmolor. The first part offers no explanation, and the second just makes it worse, by raising the question of why we have failed to see these bleen emeralds.

Is it arbitrary to insist on taking the greensome line of reasoning rather that the gruesome one? A natural response is that there is something fishy about the way that, unlike the color terms 'green' and 'blue', schmolor terms like 'grue' and 'bleen' are defined in terms of observation. Goodman was quick to point out that taking 'grue' and 'bleen' as primitives, 'green' and 'blue' must be defined in terms of observation. The key difference, I would suggest, is rather that unlike a thing's color, its schmolor is *counterfactually dependent on observation*.[18] Whereas an unobserved green thing would still have been green had we observed it, an unobserved *grue* thing would have been *bleen* had we observed it. For if something is grue and unobserved then it is blue, and would have been blue had we observed it, in which case it would have been observed and blue and hence bleen. Similarly, an unobserved bleen thing would have been grue had we observed it. For if something is bleen and unobserved then it is green and would have been green had we observed it, in which case it would have been green and observed and hence grue.[19] This helps us to see why, somewhat surprisingly, our datum

[18] Godfrey-Smith (2003), following Jackson (1975), appeals to this dependence in response to Goodman's puzzle. He suggests that something similar to standard principles of statistical inference can help us rule out various grue-like inferences. To the extent that he is right in the application of these principles, I think that my explanatory account can be developed to show why this application is appropriate. But I will not develop this here.

[19] I am helping myself to counterfactuals that Goodman himself would have found just as problematic as the projectability of 'green' over 'grue'. So my discussion does not solve all of the problems that worried Goodman. But I assume that few would want to deny these counterfactuals, and it is not a trivial matter to show how they relate to correct inductive inference.

D: All observed emeralds are grue.

is better explained by

E1: All observed emeralds are grue and the unobserved ones are bleen.

than it is by

E2: All emeralds are grue.

Our choice of which emeralds to examine was quite arbitrary; had our investigations gone slightly differently, we would have observed some of those emeralds which are actually unobserved. On the assumption of E1, had we observed some of the unobserved emeralds, D would still obtain. For the unobserved bleen emeralds would have been grue, had we observed them. But on the assumption of E2, had we observed some of the unobserved emeralds, D would not obtain. For the unobserved grue emeralds would not have been grue but bleen if we had observed them. E1 provides a better explanation for D, for it makes D independent of a fact which could easily have failed to obtain, namely that we observed those very emeralds that we happened to observe. But of course E1 is just the claim that all emeralds are green, which is just the conclusion that we should expect.

If we suppose that the unobserved emeralds are not green, then any explanation which renders the explanandum likely will have to have a clause noting that we have only observed the green ones. The claim that all emeralds are grue does this. It says that not all the emeralds are green, but we have only observed the green ones. But this fact would itself then seem to call for an explanation. On the all-grue hypothesis we have two properties, greenness and having been observed, which are co-instantiated by the same subclass of emeralds. This is a striking fact which seems to call for an explanation. We should not expect apparently causally independent properties to be correlated in this way in a large number of instances. But it is very hard to see what kind of explanation could be given, which does not itself call for explanation in much the same way.

The Martians might complain that on the all-green hypothesis, the properties of grueness and having been observed are surprisingly correlated. For on this hypothesis, there are both grue and bleen emeralds but we happen to have seen all and only the green ones. But there should be no surprise about this at all. On the hypothesis that all emeralds are green, the correlation between grueness and having been observed does not call for some unusual causal explanation. We see why the correlation holds by noting the causal *independence* of observation and greenness. Since observation does not affect greenness, if all emeralds are green, then they would all have been green regardless of which ones had been observed. But then it follows that all and only the grue ones would have been observed, regardless of which ones had been observed. For no matter which of these green emeralds were observed, the observed green ones would be grue and the unobserved green ones would not be grue. So there is no real mystery about the correlation between grueness and observation on the assumption that all emeralds are green. We see once again the explanatory asymmetry between the all-green and all-grue hypotheses, which results in the asymmetry of rational inductive inference.

3.4 Grue and Gruer

I will not defend the claim that explanatory considerations solve all the puzzles associated with Goodman's riddle in its various manifestations. But Goodmanesque variations provide nice illustrations of the power of explanatory reasoning. I will discuss a well know variation adapted from Davidson (1966) and then one of my own.

3.4.1 Emeroses

By definition, something is an *emerose* if it is either an observed emerald or an unobserved rose. All observed emeroses are green. To conclude that all emeroses are green is to conclude that all the roses that we haven't seen are green. But surely, observations of emeralds teach us nothing of the color of roses.

First note that our datum that all observed emeroses are green is equivalent to the proposition that all observed emeralds are green. (Since only unobserved roses are emeroses, the observed emeroses are just the observed emeralds). And as we have seen, a good explanation of this fact is that all emeralds are green. But why isn't the hypothesis that all emeroses are green an equally good explanation of our data? The Martians may insist that the reason why we have failed to see non-green emeroses is simply that there are none to see.

The hypothesis that all emeroses are green just says that the observed emeralds are green and the unobserved roses are green. This has two explanatory weaknesses. First, it tells us nothing about the color of unobserved emeralds. These, of course, are not emeroses. But they would have been emeroses if we had observed them since they still would have been emeralds, and observed emeralds are emeroses. The color of these unobserved emeralds is relevant to explaining our failure to observe non-green emeroses. For if there were non-green emeralds we could easily have observed some, in which case we would have observed non-green emeroses. Only the hypothesis that all emeralds are green can explain our data in this respect. Second, the color of unobserved emeroses is irrelevant to explaining our failure to observe non-green ones. Unobserved emeroses are roses, and would still be roses were they observed, in which case they would be not be emeroses. So no matter what color the unobserved emeroses are, they could make no difference to our observations of the color of emeroses. For were we to observe an emerose which is actually unobserved, it would not be an emerose, and hence its color would be irrelevant to whether all observed emeroses were green. As the hypothesis that unobserved emeroses are green plays no explanatory role, we have no reason to infer it.

*3.4.2 Grue**

A crucial part of my solution to the original problem lay in noting that unlike greenness, grueness is counterfactually dependent on observation. We can get around this solution by defining 'grue', not in terms of observation, but in terms of some property that is contingently coextensive with having been observed. For example, let's say that something is grue* if it is either green and in *R*, or blue and located elsewhere, where *R* is the region occupied by the observed emeralds. The same problem arises, since all observed emeralds are grue*, and if they are all grue* then the unobserved ones are blue. But unlike an object's schmolor, its schmolor* is unaffected by observation. An unobserved grue* emerald would still be grue* if it were observed, for it would have the same color and location.

Note also that the fact that all observed emeralds are green is not the same as the fact that all observed emeralds are grue*, so we have two possible explananda. One advantage of the all-green over the all-grue hypothesis is that the former provides a much better explanation of our having observed only green emeralds. For if all emeralds are grue*, then had we observed different emeralds we would not have seen only green ones. But by similar reasoning, the all-grue* hypothesis might appear to provide a better explanation of our having observed only grue* emeralds. For if all emeralds are green, then had we observed different emeralds we would not have seen only grue* ones. There is a symmetry here that cannot be broken by appeal to observation-dependence.

We can break the symmetry, first, by comparing our prospects for explaining these two explanatory generalizations. Supposing all emeralds are green, what might explain this? Plausibly, the emeralds share a certain micro-physical composition and structure which explains for each emerald why it is green. Furthermore, if they do share this composition and structure, it is arguably a necessary truth that they do, for that is simply what it is to be an emerald (even though this cannot be determined *a priori*). So even if we do not have all the physical details, we can sketch an explanation which terminates in a satisfying way. The explanation is very stable. If emeralds share these micro-physical properties we can expect them to be green regardless of when they are observed, where they are located, their shape, size, age, and so forth.

Suppose instead that all emeralds are grue*, i.e., the ones in scattered region *R* are green, but the rest are blue. Any explanation that relied on this fact would appear to be very unstable. If magma had cooled in different crevasses, or miners had dumped dirt in different piles, or women wearing emerald necklaces had walked in different directions, then not all emeralds would be grue*. It is hard to imagine a stable explanation of the all-grue* hypothesis, as there are just too many ways in which it could easily have been false. So the all-green hypothesis provides a more stable explanation of the fact that we have observed only green emeralds, than the all-grue* hypothesis does of the fact that we have observed only grue* emeralds.

It might be argued that the fact that we have observed only grue* emeralds is not in urgent need of an explanation in the first place. This fact does not entail that we have seen any green emeralds or any blue ones. It is just that if we have seen green ones, then they were in *R*, and if we have seen blue ones, then they were somewhere else. Even if we were to see both green and blue emeralds, it is no more remarkable that they should be demarcated by region *R* than by any other region. But if our having observed only grue* emeralds does not particularly call for an explanation, then an unstable explanation such as that all emeralds are grue* should not be so unsatisfying.

This point is correct, but it does not help the case for the all-grue* hypothesis. For if our having observed only grue* emeralds does not require a stable explanation, we can just as well explain it by the all-green hypothesis, noting that *R* is the only region that we happen to have looked in so far. For if all emeralds are green and we have looked only in *R*, then all the observed emeralds are green and in *R* and hence are either green and in *R* or blue and elsewhere. This explanation may be unstable, since if we had observed different green emeralds we would not have observed only grue* ones. But in this respect it is no worse off than the hypothesis that all emeralds are grue*. Since the all-green hypothesis does give us a nice explanation of why all observed emeralds are green, it is the better explainer, all things considered.

3.5 Jackson's Condition and Selectional Explanations

Frank Jackson's (1975) influential response to Goodman's problem also appeals to the observation-dependence of grueness, by subsuming it under a general condition on enumerative induction.

> **Jackson's Condition:** The evidence that certain *F*s which are *H* are also *G* does not support the conclusion that other *F*s which are not *H* are *G*, if it is known that the *F*s in the evidence class would not have been *G* if they had not been *H*.

To illustrate: All the lobsters I've seen have been red. But I know that these lobsters were all cooked, and that had they not been cooked, they would not be red. So by Jackson's Condition my evidence does not support the conclusion that all other lobsters, including the uncooked ones, are red. Applying this to the grue inference, the emeralds that I've observed have been grue, but they have also been observed, and had they not been, they would not be grue. So our evidence does not support the conclusion that unobserved emeralds are grue.

Jackson's Condition is highly intuitive, and examples supporting it can be multiplied. Even if correct, it would be nice to have an account of why it is correct. However, I will argue that it is false and try to provide a deeper understanding of why something like it is correct. For any *F* and *G*, let *H* be the property of being one of *these F*s and being *G*, where 'these *F*s' refers *rigidly* to those *F*s we have observed. Only the observed *F*s are *H*. And if these *F*s had not been *H*, then they would not have been *G*. For none of these *F*s could have failed to have been one of these *F*s, and hence if one of these *F*s were not both one of these *F*s and *G*, then it must not have been *G* (either by being non-*G* or not existing). So Jackson's condition rules out *any* inference from a certain a collection of observed *F*s' being *G* to others' being *G*.

While this objection is decisive, it may also seem shallow, as there seems to be something right about Jackson's condition. I think we take a step in the right direction if we replace Jackson's counterfactual clause and put it in terms of explanation. We might say that an inductive inference is undermined if we know of some *H* such that only the observed *F*s are *H*, and their being *G* is *explained* by their being *H*. This would get around my objection. For the lobster's being red is explained by its having been cooked, and an observed grue emerald is grue in part because we chose to observe it. But it is never the case that what explains a certain *F*'s being *G* is that it is one of these *F*s and it is *G*.

However, this is still too strong and does not get to the heart of the matter, as the following case illustrates. All the girls I have seen in a certain high school social clique have blonde hair. I haven't met Jane but have inductive evidence that she is blonde, too, as she hangs with the same group. The strength of this inference is not diminished much upon learning that unlike the others, Jane is not a natural blonde but dyes her hair some color. A reasonable guess is that she dyes it blonde like her friends. But this inference fails both Jackson's and our modified condition. For Jane lacks those genes which explain, for each girl in my evidence class, why she is blonde.

To understand what is going on here we need to draw on the distinction between instantial and selectional explanations. In Jackson's lobster case, we may be tempted to explain our having seen only red lobsters by supposing that all lobsters share some feature in virtue of which they are red. This explanation is refuted by the discovery that the lobsters we have seen are red because they have been cooked, and that not all lobsters are cooked. As the inference supporting explanation is undermined, so is the inference to the unobserved lobsters' being red. In the case of the blonde clique, the explanation for my having only met blondes might be that you have to be blonde to hang out with this group. This is a selectional explanation which does nothing to explain, for any particular girl in the clique, why she is blonde. So this explanation, and hence the inference that it supports, is not threatened by our knowledge of the differences in what explains the hair color of Jane and the other girls.

In some cases we may have a competition between instantial and selectional explanations. Visiting a foreign country we find many similar distinctively shaped boxes on the streets, all of them red. We reasonably infer that all similarly shaped boxes in the vicinity are red. Why would they all be red? Perhaps they are trash cans and their color is just a byproduct of the most efficient process of manufacturing them. Or perhaps they are mailboxes, and their color is chosen to make them identifiable. Which explanation is most plausible, all things considered, should affect how the strength of our inductive inference is diminished upon learning that it fails to meet a Jackson-style condition. Suppose we learn that there is a new neighborhood development where the boxes are made of a new material and by a different, more cost-effective process. These new boxes lack the property shared by those in our evidence class which explains their redness. If we suppose the boxes are trash cans, this new information significantly undermines our inference that the new ones are red. For that inference depended on the explanatory hypothesis that all the boxes were colored by the same process. If they are mailboxes, the inference is not undermined nearly as much. For we should expect that by no matter what process boxes are made, only red ones will be put on street corners to be identified as mailboxes. This selectional explanation is unaffected by the discovery that there is no common explanation for the color of each box.

Similar kinds considerations may apply in biological cases. At the risk of over-simplification I will sketch the idea. All organisms that we observe sharing collection of traits F also have trait G, so we expect that other Fs will be G. It is discovered that there are Fs which are distant enough that they could not be descendents of the first ancestor of the observed Fs to possess G. Should we expect these less closely related Fs to have G? That might depend in part on the role that G plays in enhancing the fitness of those organisms that possess it. If F organisms that lack G would be considerably less capable of reproduction, this might well explain the scarcity of non-G Fs in our observations. This in turn would suggest that if there are Fs of different ancestry, the trait G may have developed by an independent evolutionary path. Such an inference depends, of course, on numerous other factors. But we can see that the strength of such an argument is diminished if the difference in fitness between G and non-G Fs does not explain well the absence of non-G Fs. If we were relying on an instantial explanation of all Fs' being G, our inference will be weakened upon learning of the different ancestry of some of the Fs.

3.6 Arbitrary Disjunctions

The literature on inductive paradoxes contains various examples with disjunctions of predicates in the antecedent or consequent of a generalization. While these are in the same vein as Goodman's, they don't involve the dubious devices of defining predicates in terms of observation, time, or location.

3.6.1 Disjunctive Antecedents

The following case is adapted from Skyrms (1966). Green emeralds are instances of the generalization that everything that is either an emerald or a frog is green. So collecting green emeralds should confirm this generalization, which entails that all frogs are green. But surely looking at emeralds tells us nothing about frogs.

As usual, we should focus not on the generalization's instances but rather the fact that everything we've observed which is either an emerald or a frog has been green. Now our having seen only green emeralds confirms that all emeralds are green, and this together with the fact that we haven't been looking at frogs explains well why we haven't seen a non-green emerald or frog. Given these facts, the general color of frogs would make no difference to our observations and so is explanatorily redundant. There is no paradox in supposing that our emerald observations lend support to the disjunctive generalization. For the latter is equivalent to the conjunction 'All emeralds are green and all frogs are green.' Our data support this conjunction by removing some of our doubt in the first conjunct only. But the degree of this support is limited, no matter how many green emeralds we observe. For these observations can do nothing to remove our doubts about the second conjunct. The mistake is to suppose that the support we obtain for the disjunctive generalization must carry over to the consequence that all frogs are green. It needn't do so since the color of frogs plays no role in explaining our observations.[20]

3.6.2 Disjunctive Consequents

The following case is discussed by Jackson & Pargetter (1980).

> Suppose every member of a certain club who I have met is in the Social Register, then each member I have met is

[20] In his discussion, Skyrms had been assuming Hempel's (1945) Special Consequence Condition ("If an observation report confirms a hypothesis H, then it also confirms every consequence of H"), which Hempel took to be a condition of adequacy on a theory of confirmation. The example is better seen as a counterexample to Hempel's condition.

> either in the Social Register or has been to Pisa. But it would be absurd to infer that Smith, a member I have not met, who I know not to be in the Social Register, either is in the Social Register or has been to Pisa. Because that would be allowing [enumerative induction] to lend support to Smith's having visited Pisa on obviously irrelevant data. (p. 418)

There is clearly something fishy about the appeal to a disjunctive hypothesis in this inference. But the problem cannot lie solely with its disjunctiveness. For compare this inference with one which is not quite so absurd. I have met many members who are not in the Social Register and many who have not been to Pisa, but everyone that I have met is either in the Social Register or has been to Pisa. In this case it seems that I do have at least *some* reason to suppose that Smith, who is not in the Social Register, has been to Pisa. For as odd as it might be, it is not out of the question that there is a reason this disjunctive generalization should hold. Perhaps the club's eccentric founder decreed that members must be in the Social Register unless they have been to Pisa. This would explain my observations thus far, while entailing that Smith has been to Pisa.[21] In the original case, however, we seem to have gained precisely *no* evidence that Smith has been to Pisa. The problem has not merely to do with the disjuctive nature of the hypothesis but the *arbitrariness* of the disjunction. We have no data concerning whether anyone has been to Pisa or not. We could just as easily have added a disjunct to conclude that Smith rides a Kawasaki or has a dog named 'Rufus.'

Now, we know that each member that we have met is either in the Social Register or has visited Pisa only because

[21] Oddly enough, Scheffler's (1963) original case, from which Jackson & Pargetter's is modified, is of this form, yet he appears to think that we have no evidence that Smith has been to Pisa. Perhaps this is because it would take a large number of instances to make this degree of confirmation more than negligible. Imagine there are ten thousand members in the club, and I have met every one but Smith. Surely I have at least some reason to think that Smith also fits the pattern.

we know more specifically that they are all in the Register. And this stronger fact requires an explanation. We can think of a number of candidate explanations. Perhaps this is a very exclusive club, and it is a condition of membership that one be in the Social Register. Or perhaps those not in the Register have been avoiding me, as they feel awkward given my superior social status. Either hypothesis could explain why I have only met members who are in the Register and hence are either in the Register or have been to Pisa. But neither hypothesis lends any support to the claim that those not in the Register have been to Pisa. Note that the weaker hypothesis that everyone in the club is either in the Register or has been to Pisa provides a poor explanation of why I have only met members who are in the Register. For given this I might still have easily met someone not in the Register. But more importantly, it provides *no better* explanation than the hypothesis that each member is either in the Register or has *not* been to Pisa. While both hypotheses may be confirmed to some small degree, they cancel each other out with respect to whether Smith, or anyone else has been to Pisa.

Notice how a small modification to the story makes the inference more reasonable. There has been a rumor going around that membership requires that one either be in the Social Register or to have been to Pisa. We don't believe this odd rumor, but the fact that it is going around gives it a little more credibility than the hypothesis that club membership is restricted to those who are either in the Register or have *not* been to Pisa. This slight imbalance is enough to allow our having met only members of the Register to lend a little support to Smith's having visited Pisa. For this hypothesis now does additional explanatory work. In addition to providing a weak but not hopeless explanation of my having met only members in the Register, it explains the existence of the rumor.[22]

3.7 The Significance of Variety

It is a standard cannon of inductive reasoning that the strength of our evidence increases with the variety of the instances in our data. Some philosophers such as Strawson (1952) have taken this to be a straightforward conceptual truth, just a part of what we mean by such expressions as "strong evidence" and the like. It would be nice to give a deeper account of the significance of variety to inductive support.

While I have argued that Jackson's Condition is too strong, the kinds of cases that it handles illustrate well how lack of variety can weaken an inductive inference. Even if we did not know that the lobsters we have seen are red because they were cooked, the fact that they were all cooked should make us cautious in inferring that all lobsters are red. Our evidence that they are all red would be a lot stronger if we had sampled both cooked and uncooked lobsters and found them all to be red. Van Inwagen (2002) recounts the following story. In 1936 *The Literary Digest* predicted on the basis of a telephone poll that Alf Landon would win the election. Yet Roosevelt won by a landslide. Why did the pollsters miss all the Roosevelt supporters? Because the poor, who benefited most from Roosevelt's New Deal, could not afford telephones.[23] In each case the observed *F*s (lob-

[22] This last modification of the case shows the inadequacy of Jackson & Pargetter's proposed solution. They appeal to a modification of Jackson's Condition: We can infer from all *F*s which are *H* being *G* to all *F*s which are not *H* being *G* only if it is reasonable to believe that if the *F G*s which are *H* had not been, they would still have been *F G*s. Let us further stipulate that according to the rumor, membership requires those not in the Register to have *already* visited Pisa, before learning of the membership requirements. In this case it is not reasonable to believe that the members in the Social Register would have been either in the Social Register or have been to Pisa if they had not been in the Register.

[23] Obviously in the actual case some of those polled voted for Roosevelt, but for simplicity I will pretend that none of Roosevelt's supporters could afford a telephone. Most of what I say here and elsewhere in the paper extends naturally to inferences from $n\%$ of observed *F*s being *G* to (about) $n\%$ of all *F*s being *G*.

sters/voters) share a property *H* (being cooked/being wealthy) which might help explain why those *F*s are *G* (red/Republicans), and in addition we can explain why we have only observed those *F*s which are *H* (we only see lobsters when we eat them, and they must be cooked to be edible/the poll was conducted by telephone, and only the wealthy could afford telephones). Put these together and you have an explanation of why all observed *F*s are *G*. This explanation can compete with the hypothesis that all *F*s are *G*, but only to the extent that it is likely that there are some non-*H* *F*s.

This last clause concerning the likely existence of non-*H* *F*s is crucial. According to Sainsbury (1995), our data confirm that all *F*s are *G* only if "the data do not say that there is, or even that *there is quite likely to be*, a property, *H*, such that the examined *F*s are *G* only in virtue of being *H*" (p. 88). This is far too strong and rules out most ordinary inductive inferences. All observed ravens are black, and it is likely that they also share some genetic property *H* in virtue of which they are black. But far from diminishing our evidence that all ravens are black, this fact strengthens it. For we have no particular reason to doubt that all ravens have genetic feature *H*, and our having observed only ravens with *H* suggests that they all share *H*, which in turn supports the all-black hypothesis. If there were no common property that explains the blackness of the observed ravens, we would have a hard time explaining why all ravens are black. And as we saw in Section 2.2, this generalization gives a less satisfying explanation, and hence is less well supported, if it cannot itself be explained. The reason that our having observed only cooked lobsters was so damaging to the inference that all lobsters are red is that we know that there are plenty of uncooked lobsters.

Any set of data will be varied in some respects and simi-

lar in others. Which respects matter for the strength of our evidence? It is those that are best fit to play a role in a competing explanation for all observed *F*s' being *G*. More precisely, for any property *H* possessed by some but not all of the *F*s, and such that we could potentially explain why all *H* *F*s are *G*, and why we have observed only *H* *F*s; our evidence that all *F*s are *G* will be stronger if we have observed both *H* and non-*H* *F*s and found them to be *G*. Note that it is not only those properties which potentially explain the *G*ness of the observed *F*s which we should want to vary in our data. A lobster's edibility does not explain its redness, but we are wise to examine both edible and inedible lobsters for their color. In this case edibility and redness are correlated as they share a common explanation in having been cooked. In short, the more varied our data, the more they will tend to cut across whatever categories might help explain away the regularities in our observations.

Now, of course no matter how varied our data may be, we can always find some disjunction that fits our data but not all of the unobserved instances of our inductive generalization. But such gerrymandered hypotheses do not serve well in explanations. Suppose the election poll had surveyed the same number of voters, but the group was varied in terms of wealth, race, religion, education, and so on, finding them all voting for Landon. In this case the inference that all (or most) will vote for Landon is a lot stronger. For to find a feature shared by the observed voters but not all voters, we would have to resort to something like 'being either a six-foot-tall redhead living in a blue house at the bottom of a hill, or a plumber with a recently sprained ankle, or…' . As this disjunction is made up of properties that tend to be explanatorily unrelated, an explanation in which it figures will tend to be weak. It is hard to imagine how we could explain why everyone meeting this disjunctive condition voted for Landon, unless we appeal to some *other* feature that they all share but is shared also by those we have not surveyed. Furthermore, it could very easily not have happened that we surveyed only voters with this property. Indeed, if we had not surveyed the very set of voters that we did, the set that we surveyed would probably not meet this disjunctive condition. So this hypothesis provides us with a very unstable explanation of our having surveyed only Landon supporters and hence does not compete well with the hypothesis that all (or most) are voting for Landon.

3.8 Random Sampling

In sampling *F*s to evaluate whether they are all *G*, it is usually a good policy to try to sample them randomly—for instance, by selecting them out of a shaken urn. Why? Part of the answer is that it is one effective way to obtain a *varied* sample. We are not always in a position to ensure that we sample both *H* and non-*H* *F*s, for various properties *H* that might help explain away the regularities in our observations. By sampling them randomly we maximize our chances of obtaining a varied sample. This is because we are maximizing the causal independence between our sampling and the features of the *F*s. If I sample *F*s in some *systematic* way—say, by knocking on consecutive doors down the street—then I will obtain a sample of *F*s that have their general location in common. And these may well share other features such as race and income bracket that bear explanatory relations to location. By this method, the location of the *F*s is guiding me in my sampling. If instead I put the names of all the *F*s in a hat, pick some out, and examine those, I am unlikely to latch onto certain properties and obtain a homogeneous sample.

But there is more to it than this, for however varied our sample may be, the knowledge that it was selected randomly may enhance our evidence. To see why, it will help to be a little clearer on what it is for some *F*s to be sampled randomly. Sampling methods that are random in the relevant sense include picking *F*s out of a shaken urn, putting their names on the wall and throwing a dart, rolling dice to get numbers corresponding to *F*s, and so on. To be random in this sense cannot entail being indeterministic, as we would take these sampling methods to be random in the sense relevant to statistical inference even if determinism turned out to be true. What these sampling methods have in common is that whatever objects are selected, we could just as easily have selected any of the others. Of course, if determinism is true, then *given* some very specific prior conditions, the objects that were selected were bound to be selected. But it is of the nature of processes such as dice rolling or selecting from a shaken urn that the conditions which determine the outcome could themselves very easily not have obtained. The upshot of this is that if *F*s are sampled randomly, then we cannot give a *stable* explanation of our having selected only *G* *F*s, without assuming that all (or at least most) of the *F*s are *G*. Suppose I randomly select only green balls from a shaken urn. While insisting that the remaining balls are not green, I may be able to give a *kind* of explanation for my having selected only green balls. It would be one that appealed to the very precise initial arrangement of the balls in the urn and their color distribution, the exact manner in which it was shaken, and so on. But such an explanation would be terribly unsatisfying due to its instability. The situation is very different if our sampling was not random but *biased*. If green balls tended to be lighter than others and hence tended to find their way to the top of the urn after repeated shaking, we could give a more stable explanation of our having selected only green balls without supposing that the rest are green. Like variety in our data, random sampling has the advantage of diminishing the force of explanations of the fact that all observed *F*s are *G*, which are rivals to the hypothesis that all *F*s are *G*.

4. Conclusion

In surveying a variety of inductive inferences and principles of inference, I've been arguing that explanatory considerations can sort out the intuitively good inferences from the bad and support what we take to be the right principles. We could apply these results in a few ways. Following Lipton (2004), we could appeal to these data in the purely descriptive task of identifying the rules that we tacitly follow in inductive reasoning. That the inductive hypotheses we accept score well by explanatory criteria might itself be best explained by our tacitly following explanatory principles.

A different question concerns which considerations *ought* to guide us in inductive reasoning. We might seek to answer this by the method of Goodman (1955) and Rawls (1999), in which particular judgments and general principles are mutually adjusted to achieve reflective equilibrium. That the principle of preferring stable explanations fits well with our considered judgments will support the further application of this criterion to cases that are less obvious.

Lastly, there is a certain philosophical satisfaction in being able to subsume an array of inductive judgments and rules of thumb under a central guiding principle. We have the resources for a distinctly philosophical explanation of epistemic value in terms of explanatory value. Perhaps part of what makes good inductive inferences good is that they involve certain explanatory virtues. And the reason why certain inductive principles are correct is that they promote the aim of explaining well.[24]

[24] Thanks to Adam Elga, Hartry Field, Gil Harman, Matt Kotzen, Peter Lipton, Chris Peacocke, Jim Pryor, Ted Sider, Judy Thomson, Achille Varzi, and two *Philosopher's Imprint* referees for conversations, criticism, and encouragement. Some of this material was presented at my NYU Graduate Epistemology Seminar, the NYU Undergraduate Philosophy Club, and the Corridor group. I'm grateful for the feedback I received on these occasions.

References

Armstrong, D. M. 1983: *What is a Law of Nature?* Cambridge: Cambridge University Press.

Bonjour, L. 1998: *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Davidson, D. 1966: 'Emeroses By Other Names.' *Journal of Philosophy*, 63: 778-780.

Dretske, F. 1977: 'Laws of Nature.' *Philosophy of Science*, 44: 248-68.

Foster, J. 1982: 'Induction, Explanation, and Natural Necessity.' *Proceedings of the Aristotelian Society* 83: 87-101.

______, 2004: *The Divine Lawmaker*. Oxford: Oxford University Press.

Godfrey-Smith, P. 2003: 'Goodman's Problem and Scientific Methodology.' *Journal of Philosophy*, 100: 573-90.

Goodman, N. 1955: *Fact, Fiction, and Forecast*. Cambridge Mass.: Harvard University Press.

Harman, G. 1965: 'The Inference to the Best Explanation.' *Philosophical Review*, 74: 88-95.

______ 1973: *Thought*. Princeton: Princeton University Press.

Hempel, C. G. 1945: 'Studies in the Logic of Confirmation.' *Mind*, 54: 1-26, 97-121.

______ 1965: 'Aspects of Scientific Explanation,' in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Horwich, P. 1983: *Probability and Evidence*. Cambridge: Cambridge University Press.

Howson, C., & Urbach, P. 1993: *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court.

Jackson, F. 1975: 'Grue.' *Journal of Philosophy*, 72: 113-131.

Jackson, F., & Pargetter, R. 1980: 'Confirmation and the Nomological.' *Canadian Journal of Philosophy*, 10: 415-438.

Lewis, D. K. 1986: 'Causal Explanation', in *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.

______ 1994: 'Humean Supervenience debugged.' *Mind*, 103: 473-90.

Lipton, P. 2004: *Inference to the Best Explanation,* 2nd edition. London: Routledge.

Loewer, B. 1996: 'Humean Supervenience.' *Philosophical Topics,* 24: 101-28.

Peacocke, C. 2004: *The Realm of Reason*. Oxford: Oxford University Press.

Railton, P. 1978: 'A Deductive-Nomological Model of Probabilistic Explanation.' *Philosophy of Science*, 45: 206-26.

Rawls, J. 1999. *A Theory of Justice,* revised edition. Cambridge, Mass.: Harvard University Press.

Sainsbury, M. 1995: *Paradoxes*. Cambridge: Cambridge University Press.

Salmon, W. C. 1984: *Scientific Explanation and the Causal Structure of the World*. Princeton, N. J.: Princeton University Press.

Scheffler, I. 1963: *The Anatomy of Inquiry*. New York: Alfred Knopf.

Skyrms, B. 1966: 'Nomological Necessity and the Paradoxes of Confirmation.' *Philosophy of Science* 33: 230-249.

Sober, E. 1983: *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Chicago: University of Chicago Press.

Strawson, P. F. 1952: *Introduction to Logical Theory*. London: Methuen & Co.

Strevens, M. 2000: 'Do Large Probabilities Explain Better?' *Philosophy of Science* 67: 366-90.

Stuart, A. 1962: *Basic Ideas in Scientific Sampling*. London: Charles Griffen.

Van Inwagen, P. 2002: *Metaphysics*, 2nd edition. Colorado: Westview Press.