PAUL THAGARD AND RICHARD E. NISBETT

VARIABILITY AND CONFIRMATION *

> Why is a single instance, in some cases, suffi-
> cient for a complete induction, while in others
> myriads of concurring instances, without a
> single exception known or presumed, go
> such a very little way towards establishing
> an universal proposition? [1]

Hume left us with two problems of induction. The *strong* problem concerns
the global justification of induction: how do we know that the future will be
like the past? This problem seems to be insoluble, since no *a priori* justifica-
tions are forthcoming, and any *a posteriori* justification would be circular.
We can do much more with the *weak* problem, which is to find a set of rules
and principles which describe and justify, at least locally, our inductive
inferential behavior. As Goodman pointed out, we can develop a set of induc-
tive principles by paying attention to accepted inductive practice, while at
the same time adjusting our inferential practice in the light of new principles.[2]
The procedure of mutual adjustment of principles and practice is more com-
plex than Goodman indicated,[3] but Goodman's general point, that the des-
cription and justification of inductive principles go hand in hand, provides
a basis for pursuing Hume's weak problem.

   Yet even the weak problem is very difficult. There has been limited
progress in specifying the principles that describe and justify inductive
practice. We believe that this is partly because many philosophers concerned
with induction have assumed that inductive principles would be like deductive
ones in being formulatable in terms of the syntactic structure of the premises
and conclusions of inductive inferences. Whereas deductive principles can be
based solely on the logical form of the relevant sentences, inductive rules
must, we shall argue, make essential reference to the *content* of the premises
and conclusion of the inference. The validity of inductive inferences depends
in part on the nature of the objects and events about which one is reasoning.

Inductive principles therefore are content relative in a way in which deductive ones are not.

We shall illustrate the contention that inductive rules are content relative by considering the classic problem that concerned Hume — the confirmation of generalizations by their instances. Much recent work on inductive reasoning has focused on giving a characterization of what it is for a generalization 'All F are G' to be confirmed by an instance, 'Fa and Ga'.[4] As Mill noticed, the degree to which an instance confirms a generalization is highly variable. Some instances provide nearly decisive confirmation, whereas others hardly increase our confidence in a generalization at all. We shall argue that the degree to which an instance confirms a generalization depends primarily on very rich background knowledge about the *kinds* of entities and properties the generalization concerns. Specifically, our confidence in inferring a generalization 'All F are G' depends on background knowledge about how *variable* F's tend to be with respect to G's. This knowledge is based on F being a kind of K_1, G being a kind of K_2, and on the variability of things of kind K_1 with respect to things of kind K_2.

Section I contains experimental, thought-experimental and anecdotal evidence for the claim that this sort of background knowledge does in fact play a rôle in our inductions. Section II describes more rigorously how variability of kinds plays a rôle in confirmation. In III we briefly consider some of the philosophical problems concerning natural kinds, causality, and variety of instances which our account of confirmation raises. Section IV contrasts our account with more familiar models of inductive reasoning. The concluding Section V contains brief reflections on confirmation and on the methodology of using empirical work on inferential behavior as a guide to normative inductive logic.

I

Consider the following thought experiment. Imagine you are exploring a newly discovered island. You encounter three instances of a new species of bird, called the shreeble, and all three observed shreebles are blue. What is your degree of confidence that *all* shreebles are blue? Compare this with your reaction to the discovery of three instances of a new metal floridium, all of which when heated burn with a blue flame. Are you more or less confident of the generalization 'All floridium burns with a blue flame' than you were of

the generalization 'All shreebles are blue'? Now consider a third case. All three observed shreebles use baobab leaves as nesting material, but how confident do you feel about the generalization 'All shreebles use baobab leaves as nesting material'?

Most people feel more confident about the floridium generalization than about either of the shreeble generalizations, and more confident about shreebles being blue than about their using baobab leaves as nesting material. Nisbett *et al* found that undergraduate subjects share such intuitions with professional philosophers and psychologists, and in fact often spontaneously articulate a version of the justification that we prefer.[5]

We know that metals tend to be quite constant with respect to such physical properties as flame color and electric conductivity, whereas we know that birds are more variable with respect to their color: different sexes of the same species can be different colors, and in some kinds such as parrots there is wide variation in color. Background knowledge also suggests that birds are even more variable with respect to what sort of materials they use for nests. This is why three shreebles' nests are so much less persuasive than three floridium flames.

It seems clear that people may differ in their background knowledge in such a way as to affect markedly the inductive generalizations they make from a given instance. This point has been illustrated in an experiment by Quattrone and Jones.[6] They found that subjects were more prone to make generalizations about members of an out-group than about members of an in-group. For example, Rurgers students were more likely to generalize about the behavior of Princeton students on the basis of the behavior of one Princeton student than were Princeton students, while Princeton students were more likely to generalize about Rutgers students after observing one Rutgers student than were Rutgers students. Quattrone and Jones conjecture that the reason for the discrepancy is the assumption by the in-group that members of the out-group are less variable in their properties than members of the in-group.

One implication of the Quattrone and Jones study is that the difference in people's beliefs about variability may be a function of the degree of their experience with events of the kind in question. For kinds of events that are characterized by high variability, the novice and the expert may thus make quite different generalizations. Nisbett *et al.* (op. cit.) found that people with a good deal of athletic experience were less willing to assume that a

superb performance at a football try-out was indicative of generally superb abilities than were people who had less athletic experience. Similarly, people with acting experience were less willing to assume that a superb performance at an audition was indicative of generally superb acting abilities than were people who had no acting experience. Apparently the experienced subjects were more aware that a single instance of football playing, or of acting, may ˜ail to reflect the individual's general level of ability.

The history of science provides examples of cases where a very limited number of instances proved to be decisive. We believe that these are well understood in terms of background knowledge about variability. For example, Einstein's general theory of relativity implied that light could be deflected by an intense gravitational field. Eddington's celebrated experiment during the 1919 eclipse set out to test the generalization that light is so deflected by considering whether starlight is bent by the sun. Only a few photographs sufficed to convince most scientists that light was indeed bent by a gravitational field.[7] Background knowledge tells us not to expect much variability in the behavior of light, so a few instances suffice. We do not crucially need to consider other eclipses, other light sources, other gravitational fields, or other mountain tops. Light is not the kind of thing for which these factors would matter. Compare the situation in social science, where variability is so great and we are so lacking in a coherent account of inductively reliable kinds of people and behavior, that we must constantly be wary of spurious correlations with no causal significance.

The above examples indicate that background knowledge about variability in kinds is important for confirmation theory. The importance has been largely unnoticed in recent philosophical work, which has followed Hempel in trying to give a general account of what it is for an instance to confirm a generalization, or Carnap in trying the build a formal model for quantitative confirmation.[8] Previous, less formal inductive logicians including Mill, Keynes, Russell, and Harrod noticed the relevance of background knowledge about variability and kinds but did not develop the insight.[9] We need a more rigorous and general account of how background knowledge about variability affects the degree to which a generalization can be confirmed by its instances.

II

That 'All floridium burns with a blue flame' is highly confirmed by a few in-

stances is due to the fact that metals are highly invariant with respect to physical properties such as combustion. That 'All shreebles use baobab leaves as nesting material' is poorly confirmed by a few instances is due to the fact that birds are more variable with respect to the kinds of nesting materials they use. To make this precise, we need a formalism which enables us to compare the degree of variability in different kinds of things. Then we can make the claim that degree of confirmation is a function of degree of invariance.[10]

Let H be the hypothesis '$(x)(Fx \supset Gx)$', and let E be the evidence 'Fa and Ga'. Suppose F's are a kind of K_1 and G's are a kind of K_2; for example, shreebles are a kind of bird, and blue is a kind of color. Then if $C(H, E)$ is the degree to which H is confirmed by E, and $I(K_1, K_2)$ is the invariance of K_1 with respect to K_2, our basic claim can be represented:

$$C(H, E) = f(I(K_1, K_2)).$$

The claim needs to be fleshed out in two ways, by further characterizing the function f, and by indicating how the metric $I(K_1, K_2)$ is to be calculated. Confirmation is a function of more than just variability of kinds: full determination of $C(H, E)$ may require reference to other sorts of background information in addition to what we discuss, and of course degree of confirmation is partly a function of number of instances. But we maintain that how rapidly degree of confirmation increases with number of instances depends *primarily* on variability of kinds. Degree of confirmation is presumably a monotonic, negatively accelerated function of number of instances, but the rate of acceleration depends primarily on considerations of variability. With high $I(K_1, K_2)$, degree of confirmation increases very rapidly with only a few instances; but if invariance is low, degree of confirmation will increase only slowly as instances mount up.

How can we measure $I(K_1, K_2)$? In the metal case, estimation of $I(K_1, K_2)$ is based on our background knowledge that metals tend to be invariant with respect to physical properties like combustion and conductivity. This knowledge in turn is based on experience with many kinds of metals. We know that whatever particular physical property a few instances of the metal have, the other instances of that kind of metal will probably have it too. If one instance of aluminum conducts electricity, probably they all do. If one instance fails to take on a magnetic charge, probably they all do, given standard conditions. With birds, however, we can think of kinds which

CONFIRMATION

High I (K_1, K_2)

Moderate I (K_1, K_2)
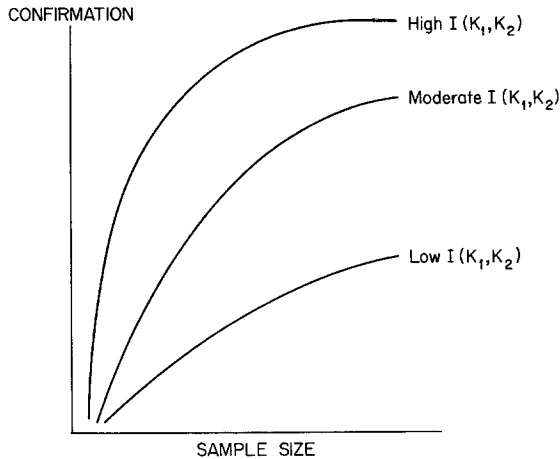
Low I (K_1, K_2)

SAMPLE SIZE

Fig. 1. Confirmation as a function of sample size and $I(K_1, K_2)$.

assume different colors and use different sorts of nesting material. There is a greater degree of association between kinds of metal and combustion properties than there is between kinds of birds and either colors or kinds of nesting material.

Various measures of association between classifications have been discussed by statisticians. Different measures are appropriate for different circumstances, but one particularly useful measure, discussed by Goodman and Kruskal,[11] is based on optimal prediction: degree of association of two classifications into kinds is a function of how well knowing that something is of a kind from the first classification will enable you to predict its kind in the second classification, compared to how well you would predict without using the classifications. If kinds in the second classification are associated with high probability with specific kinds from the first classification, then knowing the classification will greatly improve prediction. For the formal details, see Goodman. and Kruskal. It can easily be verified that their measure of association between classification can be adapted to provide a measure of degree of invariance between kinds, $I(K_1, K_2)$. In ordinary life, we do not have as part of our background knowledge a measure of invariance as precise as that calculated by Goodman and Kruskal, but their measure could be viewed as an ideal approximation to assessments of invariance such as people make in the metal and bird cases.

In assessing the degree of confirmation of $(x)(Fx \supset Gx)$, it is of course essential to consider the *relevant* kinds, K_1 and K_2. Normatively, selecting the relevant kind is analogous to the problem of selecting the appropriate reference class for a single object or event. To get an estimate of the probability that Fred will have a heart attack, we need to place him in the class such that subdivision is statistically irrelevant.[12] Similarly, we need to select a reference class for the class of shreebles; this is the broadest kind such that considering subkinds to which shreebles belong does not give a different estimate of variability with respect to color. The more one knows about a subject, the more relevant do subkinds become. Perhaps we should consider shreebles as a kind of waterfowl, or as a kind of Australian waterfowl, if these subdivisions give us a different judgment of variability from merely considering shreebles as a kind of bird. Placing shreebles in the most homogeneous class can make an enormous difference in the inferrability of generalizations. If our reference class is vertebrates, we have too much variability to be able to generalize usefully about color. We are reduced essentially to simple enumeration. At the other extreme, knowledge of the appropriate reference class can mean that a single case is "sufficient for a complete induction". Thus normally our inference making will be enormously more furthered by finding a narrower, more homogeneous reference class than by collecting more instances of shreebles.[13]

III

Philosophers accustomed to the syntactic austerity of Hempelian confirmation theory may find the above use of the notion of *kind* ontologically profligate. What are kinds, and how can we legitimately use such a vague notion in inductive logic? In responding to these questions, it should be noted at the outset that we already know from Goodman's grue problem and other paradoxes that a purely syntactic account of confirmation is impossible. Some background knowledge must be brought in, if only to select predicates which are *projectible* in Goodman's sense. The above discussion indicates that we also need knowledge about kinds, and we shall now try to show that the notion of kind possesses some psychological and philosophical respectability.

Much recent work in cognitive science concerns how human knowledge is organized with respect to kinds and categories. In cognitive psychology, the

organization of objects into categories is considered to be critical to deter-mining how humans store and process information.[14] Similarly, a standard technique in artificial intelligence is the organization of concepts into A-KIND-OF hierarchical networks.[15] Thus it is entirely plausible that people do naturally employ the organization in terms of kinds which our account of confirmation presupposes.

The notion of a kind is also philosophically important. There is more to saying that a robin is a kind of bird than just saying that all robins are birds. Simply to write (x) $(Rx \supset Bx)$ would not bring out the information needed to run the model of confirmation described in Section II. As Quine points out in an important essay on natural kinds, kinds can be seen as sets, but not all sets are kinds. Quine claims that humans have an "innate similarity sense" which provides us with the initial ability to organize the world into kinds. However, natural selection has also endowed man with the ability to transcend this initial organization:

> He has risen above it by developing modified systems of kinds, hence modified similarity standards for scientific purposes. By the trial-and-error process of theorizing he has regrouped things into new kinds which prove to lend themselves to many inductions better than the old.[16]

Systems of theoretical kinds can be built up through inductive experience. Our model of confirmation makes use of such systems. Through past expe-rience, we know a great deal about different kinds of metals and kinds of birds, and such knowledge cannot be neglected when analyzing degree of confirmation.

Quine describes important connections between the notion of a kind and more philosophically familiar notions of subjunctive conditionals: we can say that if x were a robin, it would be a bird. Kinds also are relevant to understanding singular causal statements: "To say that one event caused another is to say that the two events are of *kinds* between which there is invariable succession".[17] Mere sets will not suffice, since the two events might fall into any number of trivial sets which accidently overlap.

Talk of causality raises an important question for the account of confirm-ation we have given. We have described degree of confirmation in terms of background knowledge concerning kinds and variability. To what extent is this background knowledge *causal*? Certainly, we presume there are causal explanations of why metals behave as they do and why birds have the colors they do. If we were actually in possession of good causal explanations for

these phenomena, our confidence in the degree to which we generalize from a given number of instances might increase. But we do not need this information in order to make inductive inferences. Although it is assumed that there is some causal underpinning to the classification into kinds, we do not need to know the full causal story in order to use the information about kinds and variability to assess degree of confirmation.

Much remains to be said about the philosophical and psychological bases for considering organization into kinds as a fundamental part of our understanding of the world, but we shall pursue that task no further here. The point of our discussion has been to show that the importance of the notion of a kind goes far beyond its rôle in our account of degree of confirmation.

IV

In this section we shall relate our discussion to several more general accounts of inductive reasoning — theory choice construed as inference to the best explanation, Hempel's qualitative confirmation theory, and Bayesian probabilistic confirmation theory.

Thagard has used case studies from the history of science to develop an account of scientific theory choice as selection of the best of competing explanations of the evidence.[18] This account departed from standard hypothetico-deductive accounts of theory confirmation by using a non-syntactic, highly contextual measure of explanatory power, taken from William Whewell.[19] Briefly, one theory is more *consilient* than another if it explains more classes of facts. The notion of a class of facts is highly pragmatic, depending on the organization of scientific knowledge at a given time. For example, refraction and reflection — the basic laws of each — constitute two classes of facts which optical theories must explain. Consilience is most relevant to assessing the explanatory power of *theories*, but it also has an application to generalizations. A generalization $(x) (Fx \supset Gx)$ is consilient if there is *variety* among the objects a such that Fa in conjunction with the generalization explains Ga. It is not the number of instances which matters so much as their variety.

That conclusion can be understood in terms of the account of confirmation given above. How do we assess "variety among instances"? Such an assessment requires background knowledge about the *kinds* of things involved. To test, for example, Snell's law of refraction, we would measure

different substances at different temperatures, and so on, but not worry about such matters as the particular city where the results were achieved: the light and its media are not the kinds of things which are variable with respect to geographical location. When background knowledge tells us that F is of a kind K_1 which is variable with respect to K_2, of which G is a kind, then we know we need to find a variety of instances before we can infer that (x) $(Fx \supset Gx)$. Invariance based on background knowledge licenses not only restrictions to fewer instances, but also less concern about variety of instances. It is therefore background knowledge about the variability of kinds which enables us to assess the consilience of a generalization and infer it as the best explanation.

Variability in kinds is a consideration which goes beyond Hempel's syntactic account of confirmation,[20] but is of some help in understanding a well-known paradox in that account. It is a consequence of Hempel's conditions on confirmation that, because 'All F are G' is logically equivalent to 'All non-G are non-F', the first generalization is confirmed by anything which is non-G and non-F. Thus we get the paradoxical result that an instance of a white shoe confirms 'All ravens are black'. One popular resolution of the paradox is to grant that 'All ravens are black' is confirmed by a white shoe, but to point out that this confirmation is much *less* than that gained by an instance of a black raven.[21] The reason for the difference in degree of confirmation is that there are far more non-black things than ravens. Our model suggests that there is more to the difference in degree of confirmation than just the different numbers of ravens and non-black things. Our background knowledge tells us that ravens are kinds of birds, and black is a kind of color, and that birds are fairly invariant with respect to color. However, we have no analogous background knowledge about non-black things and non-ravens. 'Non-black' and 'non-raven' are not kinds of anything. With those properties, we are relegated to doing the kind of induction by simple enumeration which requires us to gather very many instances before we can have any confidence that we have more than an accidental correlation. In contrast, 'raven' and 'black' fit into our knowledge system in such a way that we can use information about variability of kinds to establish a high degree of confirmation on the basis of relatively few instances. This, in addition to size of the relevant classes, allows us to judge that 'All ravens are black' is much better confirmed by a black raven than 'All non-black things are non-ravens' is confirmed by a white shoe.

In probabilistic confirmation theory, $C(H, E)$ is generally identified with $P(H \mid E)–P(H)$.[22] The more a piece of evidence raises the probability of a hypothesis, the more it is said to confirm the hypothesis. $P(H \mid E)$ is generally calculated by means of Bayes' Theorem, a simple form of which is:

$$P(H \mid E) = P(H) \times P(E \mid H) / P(E).$$

There are numerous problems with Bayesian, probabilistic confirmation theory which we shall not attempt to review here.[23] But a discussion of our claims concerning the importance of variability to confirmation would not be complete without relating it to the Bayesian tradition.

At the top level, our variability model fits very well with probabilistic confirmation theory. Three instances of blue-burning floridium increase the probability of the relevant generalization more than three instances of blue shreebles increase the probability of 'All shreebles are blue', in some sense of probability. Hence degree of confirmation is greater in the former case. We might expect also that the background knowledge about what kinds of things are involved could play a direct rôle through Bayes' theorem, but the simple Bayesian model does not seem rich enough to bring out the relevant reasoning.

Three instances of blue-burning floridium give us a high probability for 'All floridium burns with a blue flame', but why? This is not due to high prior probability $P(H)$, because until we get the first instance of floridium, $P(H)$ would seem to be low, if we could sensibly give it a value at all. Moreover, high $P(H)$ cannot be the main contributor to high $P(H \mid E)$, since then $P(H \mid E)–P(H)$ and hence $C(H, E)$ would not be high. In addition, there seems to be no difference between the relevant values of $P(E \mid H)$ and $P(E)$ in the floridium and shreeble cases. Hence the Bayesian approach, narrowly construed, does not seem adequate to account for the variability effect.

A Bayesian could say that use of background information about variability is really deductive, and therefore need not be taken into account in a model of inductive reasoning. We might for example have the principle that any combustion property of a metal will be completely invariant. Letting 'MF' stands for 'F is a kind of metal' and 'CG' stand for 'G is a kind of combustion property' we could write the generalization:

$$(x)(y)(F)(G) \, [(MF \, \& \, CG \, \& \, Fx \, \& \, Gx \, \& \, Fy) \supset Gy] \, .$$

This implies that once you have one instance of something which is both F and G, it follows deductively that the next F will be G. More generally, we can assume the presence in our background knowledge of information of the form:

$$(x)(y)(F)(G) \ \{P[Gy \mid (K_1 \ F \ \& \ K_2 \ G \ \& \ Fx \ \& \ Gx \ \& \ Fy)] \ = n\}$$

Using this information, we could deductively get from a single instance '$Fa \ \& \ Ga$' a high value for $P(Gb \mid Fb)$ and hence for $P((x)(Fx \supset Gx))$ and $P(H \mid E)$. Thus, it could be argued, the difference between the floridium and shreeble cases is that the former case gives a much higher value for n. Undoubtedly, some such construction could be made, but it does not go far in illuminating the logic of variability and confirmation. Perhaps the metal case is deductive, but the two shreeble cases, one involving color and the other nesting behavior, are clearly more complex. One instance would not generate a stable value for $P(Gx)$: getting extra cases *does* matter when confirming the generalizations that all shreebles are blue and that all shreebles use baobab leaves for nesting materials, and the important result to hang on to is that the extra case matters in different amounts to the two generalizations. Hence, we need the general relation $C(H, E) = f(I(K_1, K_2))$. So the existence of deductive relations is not enough to enable the Bayesian to take variability into account.

Mary Hesse has made a Bayesian proposal for accounting for the importance of variety of instances, which, we saw earlier in this section, is correlative to the problem of variability. It is indeed a problem for the Bayesian to explain why variety of evidence is inductively important. Hesse proposes that the concern for variety of instances is a corollary of eliminative induction.[24] Suppose $S = \{H_1, H_2, ... H_n\}$ is a set of finite, exhaustive, and mutually exclusive hypotheses. Then $P(H_1 \mid E) + P(H_2 \mid E) + ... P(H_n \mid E) = 1$, and any disconfirmation of any proper subset of S by new evidence increases the sum of the probabilities of the remaining subset. Variety of instances is then seen to be desirable, since greater variety will presumably lead to disconfirmation of more of the H_i, leaving the remaining ones better confirmed.

In the shreeble case, this might work as follows. Thinking deterministically, our alternative hypotheses, prior to any evidence, might be:

H_1 All shreebles are blue.

H_2 All shreebles are green.

...

H_n All shreebles are brown.

Then if $P(H_1 \vee H_2 \vee ... H_n) = 1$, the observation of even one blue shreeble will eliminate all alternative hypotheses and establish that $P(H_1 \mid E) = 1$. Reference to variability and kinds would be unnecessary.

However, we have to consider the possibility that shreebles will vary in color. This possibility requires us to add:

H_{n+1} Shreebles come in various colors.

Then the observation of a blue shreeble will still rule out $H_2 ... H_n$, but we would have only the result: $P(H_1 \vee H_{n+1} \mid E) = 1$. If $P(H_{n+1})$, the prior probability that shreebles vary in color, is very high, then the probability that shreebles are blue will *not* be greatly increased by the evidence. Conversely, if shreebles are highly invariant with respect to color, then the probability of H_1 is greatly elevated. In order to get an objective estimate of $P(H_{n+1})$ we need a method for assessing invariance along the lines suggested above in Section II.

We do not doubt that our concerns about the relevance of variability to confirmation could be incorporated into some more elaborate Bayesian account of inductive inference. Our point has not been that there is any incompatibility between Bayesian inference to generalizations as we construe it, but rather that any confirmation theory including Bayesian will have to take into account the effects of variability on degree of confirmation.

<div align="center">V</div>

Our account of confirmation illustrates our contention that inductive principles are content relative. Assessment of degree of confirmation must go beyond syntactic matters to consider the kinds of things under investigation. Further illustration of this principle could be found in the logic of theory choice, where the criteria for selecting the theory which provides the best explanation of the evidence are content relative, and in the logic of statistical analysis of data, where appropriateness of a particular statistical test in an empirical context depends on background assumptions about the nature of the relevant population.

In discussing the validity of deductive reasoning, we can abstract from the content of the sentences in question, since deductive validity is a function only of logical form. Even here, however, abstraction from content is unrealistic at the practical level. Since no finite reasoner has the time or resour-

ces to infer all the consequences of his or her beliefs, what actually gets inferred when must be a function of more than just what legitimately *may* be inferred by logical rules. The content of what we believe and the nature of our general concerns will determine what inferences are selected from the infinite number we could make.

In inductive reasoning, even which inferences we *may* make is in part a function of the content of our beliefs. Content can play a rôle in different ways. It affects what rules it will be appropriate to use. The rules for accepting theories, for adopting generalizations as well confirmed, and for accepting or rejecting statistical hypotheses, apply in different contexts. One needs to know what one is talking about in order to select the appropriate rules for the context. In addition, as our discussion of degree of confirmation showed, content affects how much evidence is needed before an inference becomes legitimate. Hence the validity of our inductive inferences depends on beliefs about the nature of the events in question.[25]

Since Carnap, most work on inductive logic has been highly analytical and mathematical, emphasizing syntactic constructions or Bayesian formalisms. In company with Goldman's approach to epistemology,[26] we recommend a more empirical approach to inductive logic. The inductive logician can take as data experimental results such as we described in Section I, or case studies from the history of science, or even the more informal anecdotal evidence that is usually all that logicians have considered. The empirical approach obviates distractions which arise from artificial constructs such as 'grue'.[27] Inductive logic is primarily concerned with how people actually do reason. Of course, there is no simple move from how people do reason to how people *should* reason, which is the concern of the normative discipline of inductive logic. But empirical results and reflection on them can contribute greatly to the development of normative models.[28] Empirical studies help us to identify pervasive features of human reasoning, and sound normative models must either account for those features, or explain why those features are not desirable parts of a normative model. We hope to have shown in this paper that any satisfactory normative model of confirmation must take into account the rôle of background knowledge concerning kinds and variability.

University of Michigan at Dearborn

University of Michigan at Ann Arbor

NOTES

* For helpful comments and suggestions, we are grateful to Jonathan Adler, Susan Carey, Lisbeth Fried, Daniel Hausman, John Holland, Keith Holyoak, David Krantz, Lance Rips, Glenn Shafer, Stephen Stich, and Stephen White.

[1] J. S. Mill: 1970, A System of Logic, Eighth edition (Longman London), Book III, Ch. III, Section 3.

[2] N. Goodman: 1965, Fact, Fiction, and Forecast, 2nd edition (Bobbs-Merrill, Indianapolis), p. 63f.

[3] See S. Stich and R. E. Nisbett: 1980, 'Justification and the psychology of human reasoning', Philosophy of Science 47, pp. 188–202; and P. Thagard: 1982, 'From the descriptive to the normative in psychology and logic', Philosophy of Science 49, pp. 24–42.

[4] See, e.g., C. Hempel: 1965, Aspects of Scientific Explanation (Free Press, New York); and R. Swinburne: 1973, An Introduction to Confirmation Theory (Methuen, London).

[5] R. E. Nisbett, D. H. Krantz, C. Jepson, and Z. Kunda; 'Intuitive models and statistical heuristics', (forthcoming).

[6] G. Quattrone and E. Jones: 1980, 'The perception of variability within in-groups and out-groups: implications for the law of large numbers', Journal of Personality and Social Psychology 38, pp. 141–152.

[7] R. Clark: 1972, Einstein, The Life and Times (Avon, New York), p. 290.

[8] Hempel, op. cit.; R. Carnap: 1950, Logical Foundations of Probability University of Chicago Press, Chicago).

[9] Mill, op. cit.; J. M. Keynes: 1921, A Treatise on Probability (Macmillan, London); B. Russsel: 1948, Human Knowledge: Its Scope and Limits (Allen and Unwin, London); R. Harrod: 1956, Foundations of Inductive Logic (Macmillan, London).

[10] We use the term 'invariance' rather than 'invariability' for reasons of euphony. Our notion of variability here is not the same as the statistical notion of variance of a sample or population, which is calculated by reference to a mean of some random variable. In our qualitative cases, we have no mean value. Statistical variance does however have an effect essentially the same as our variability: if we know that the variance in a population is low, then only a very small sample suffices to give us an accurate estimation of the mean.

[11] L. A. Goodman, and W. H. Kruskal: 1954, 'Measures of association for cross classifications', Journal of the American Statistical Association 49 pp. 732–764, 740ff. Reprinted in Goodman and Kruskal: 1979, Measures of Association for Cross Classifications (Springer Verlag, New York).

[12] See Wesley C. Salmon: 1967, The Foundations of Scientific Inference (University of Pittsburgh Press, Pittsburgh), pp. 91ff. The sort of inference discussed by Salmon also will often require consideration of kinds. If we have complete statistical information, the appropriate reference class can be determined mathematically. But in the more common situations where information is limited, we will select a reference class for Fred based on what kinds of things we know him to be.

[13] It should be noted that there are ways of getting an estimate of variability other than by considering kinds. For example, we might have heard on the radio that Ford is economizing by making all its new subcompacts the same color. Then, deductively, we can infer from one observation of a blue Ford subcompact that all are blue.

[14] For surveys see: A. Glass, K. Holyoak, and J. Santa: 1979, Cognition (Addison Wesley, Reading, Mass.), Ch. 10; and E. Rosch and B. B. Lloyd (eds.): 1978, Cognition and Categorization (Erlbaum, Hillsdale, N.J.).

[15] See, for example P. Winston: 1977, Artificial Intelligence: (Addison Wesley, Reading Mass.), Ch. 7; and S. Fahlman: 1979, NETL: A System for Representing and Using Real World Knowledge (MIT Press, Cambridge).

[16] W. Quine: 1968, Ontological Relativity and Other Essays (Columbia University Press, New York), p. 128.
[17] *Ibid.*, p. 133.
[18] P. Thagard: 1978, 'The best explanation: criteria for theory choice', Journal of Philosophy, 75, pp. 76–92.
[19] W. Whewell: 1967, The Philosophy of the Inductive Sciences, reprint of second edition (Johnson Reprint Corporation New York).
[20] Hempel, *op. cit.*
[21] See for example, J. L. Mackie: 1963, 'The paradox of confirmation', British Journal for the Philosophy of Science 13, pp. 265–277.
[22] Carnap, *op. cit.*; M. Hesse: 1974, The Structure of Scientific Inference (University of California Press, Berkeley).
[23] C. Glymour: 1980, Theory and Evidence (Princeton University Press, Princeton), Ch. III.
[24] M. Hesse, *op. cit.*, pp. 137ff.
[25] For further discussion, see R. E. Nisbett, D. H. Krantz, C. Jepson, and G. T. Fong: 1982, 'Improving inductive inference', in D. Kahneman, P. Slovic, and A. Tversky (eds.): Judgment Under Uncertainty: Heuristics and Biases (Cambridge University Press, New York), pp. 445–459.
[26] A. Goldman: 1978, 'Epistemics: the regulative theory of cognition', Journal of Philosophy, 75, pp. 509–523.
[27] Goodman's (*op. cit.*) problem about grue does not arise in our account of confirmation. That all emeralds are green is highly confirmed by a few green emeralds, since we know that gems are highly invariant with respect to color. We have no similar information for assessing the degree of confirmation of 'All emeralds are grue', since grue is not a kind of anything represented in our background knowledge. Without background knowledge about the invariance of gems with respect to properties like grue, we cannot expect high confirmation of 'All emeralds are grue' by a few grue emeralds.
[28] See the articles cited in Note 3, and M. Friedman: 1979, 'Truth and confirmation', Journal of Philosophy 76, pp. 361–382. The relevant psychological discussions include: W. Edwards: 1968, 'Conservatism in human information processing', in B. Kleinmuntz (ed.): Formal Representation of Human Judgment (Wiley, New York); P. Slovic and S. Lichtenstein: 1971, 'Comparison of Bayesian and regression approaches to the study of information processing in judgment', Organizational Behavior and Human Performance 6 pp. 649–744; A. Tversky and D. Kahneman: 1974, 'Judgment under uncertainty: heuristics and biases', Science 185 pp. 1124–1131; R. E. Nisbett and L. Ross: 1980, Human Inference: Strategies and Shortcomings of Social Judgment (Prentice-Hall, Englewood Cliffs, N.J.).