Inductive Rules, Background Knowledge, and Skepticism

Daniel Steel

Department of Philosophy

503 S. Kedzie Hall

Michigan State University

East Lansing, MI 48823-1032

Email: steel@msu.edu

Abstract

This essay defends the view that inductive reasoning involves following inductive rules against objections that inductive rules are undesirable because they ignore background knowledge and unnecessary because Bayesianism is not an inductive rule. I propose that inductive rules be understood as sets of functions from data to hypotheses that are intended as solutions to inductive problems. According to this proposal, background knowledge is important in the application of inductive rules and Bayesianism qualifies as an inductive rule. Finally, I consider a Bayesian formulation of inductive skepticism suggested by Lange. I argue that while there is no good Bayesian reason for judging this inductive skeptic irrational, the approach I advocate indicates a straightforward reason not to be an inductive skeptic.

1. Do People really use Inductive Rules?

According to an understanding of the topic going at least as far back as David Hume, inductive reasoning is a matter of following inductive rules and one major task of epistemology is to articulate those rules and explain how they are justified. Yet that traditional picture of inductive reasoning has not gone unchallenged (cf. van Fraassen 1989, 2000; Okasha 2001). For example, Samir Okasha (2001, 315-24) argues that people do not use inductive rules at all and proposes that a proper appreciation of this point leads to an answer to Hume's inductive skepticism. In Okasha's slogan: "No Rules of Induction, No Humean Argument" (2001, 315). Okasha argues that inductive rules are neither desirable nor necessary: undesirable because they ignore background information (2001, 318-9) and unnecessary because people might be Bayesians and Bayesian updating is not an inductive rule (2001, 316). The claim that Bayesian updating is not an inductive rule is also made by Bas van Fraassen (1989, 151).

In this essay, I defend the traditional view that inductive rules are central to inductive reasoning. I begin by proposing a general account of what an inductive rule is. According to the proposal I suggest, inductive rules are sets of functions from samples of data to subsets of a partition of hypotheses. Inductive rules are considered in relation to an inductive problem, and rules are selected for a problem on the basis of whether they achieve a success condition, such as convergence to truth in the long run or being likely to get in the neighborhood of the right answer in the short term. In section 2, I explain how the principle of induction can be identified with an inductive rule that is necessary and sufficient for convergence to truth in simple cases of enumerative induction. I also explain how statistical significance tests qualify as inductive rules according to the account I propose.

Given this proposal, I describe two ways in which inductive rules allow a role for background knowledge in their application. First, since inductive rules are normally designed for particular types of problems, background knowledge is often relevant in deciding whether a given case is one in which the rule may by applied. Secondly, applying an inductive rule requires selecting one function from the set that the rule encompasses, and background knowledge is often relevant to deciding which function to choose. In addition, I show that Bayesian inference qualifies as an inductive rule according to my account as well as by a criterion proposed by Okasha himself. I also show that Bayesianism entails the principle of induction as formulated in section 2. Finally, I consider a formulation of inductive skepticism suggested by Marc Lange (2002) as an objection to Okasha's answer to Hume's problem of induction. I explain that the inductive skeptic as suggested by Lange is best construed as a person who refuses to commit to any one of the probability distributions consistent with the data. I argue that, although Bayesianism provides no basis for judging such a person irrational, there is a straightforward reason for not being an inductive skeptic that follows from the position I advocate. Thus, I do not agree that inductive skepticism is an inevitable consequence of an epistemology that emphasizes inductive rules.

2. What are Inductive Rules?

In general, rules are prescriptions about how to solve problems of various sorts. For instance, traffic rules are mostly designed to solve coordination problems that arise when two or more motorists use the same road at the same time. Likewise, rules of distributive justice are typically designed to solve social problems that arise when the interests of two or more individuals conflict with regard to the possession of material goods. Given this general perspective on rules, inductive rules are naturally regarded as prescriptions about how to solve various types of inductive problems. Thus, it is necessary to say what an inductive problem is before attempting to explicate the notion of an inductive rule.

I suggest that an inductive problem consists of four basic components.

1. A set of possible sequences of observations or data streams. I label this set $\mathcal{D}$.
2. A body of background knowledge, which I label $\mathcal{K}$.
3. A partition of hypotheses, which I label $\mathcal{H}$.
4. A success condition, which I label $\mathcal{S}$.

It will be helpful to say a bit more about each of these components. The sequences of observations or data streams are presumed to be potentially infinite, so that there is no limit in principle to how much data one can accumulate. The set of possible data streams $\mathcal{D}$ indicates the form that the data takes, for instance, records of the age, body mass index, and blood pressure of a list of individual people. Or to take a philosopher's example, a sequence of reports of emeralds indicating the color of each. The background knowledge $\mathcal{K}$ can exclude some data streams from consideration while indicating that others are live possibilities. In the example of the emeralds, $\mathcal{K}$ might say that every emerald is either green or blue, so that that one would only need to consider the subset of $\mathcal{D}$ containing permutations of those two colors. In addition, $\mathcal{K}$ might place a ranking or weighting on those data streams deemed possible, for instance, by judging some more probable than others. In the emerald example, $\mathcal{K}$ might indicate that the most probable data stream is the one in which all emeralds are green. Since $\mathcal{H}$ is a partition, its members are mutually exclusive and collectively exhaustive given $\mathcal{K}$. Hypotheses might be predictions about some finite set of data, generalizations about the data stream as a whole, or claims about the causes that generate the data. Finally, the success condition is whatever (presumably, epistemic) aim that one hopes to achieve. In an inductive problem, the most desirable epistemic aim—namely, being guaranteed to decisively indicate the true hypothesis—is not attainable by any method. Hence, if an inductive problem is to have any solution at all, the success condition must be something less ambitious. Examples of epistemic aims that might be attainable in an inductive problem include converging to the truth in the long run and having a reasonably high probability of getting in the neighborhood of the truth in the short term. Naturally, people may disagree about what the success conditions should be, and central disputes among rival theories of inductive inference often turn on this very issue (cf. Steel 2005).[1]

I define an *inductive rule* as a non-empty set of functions from initial segments of members of $\mathcal{D}$ to subsets of $\mathcal{H}$. I will call a *single* function from initial segments of members of $\mathcal{D}$ to subsets of $\mathcal{H}$ a *strict inductive rule*. Thus, a strict inductive rule specifies which hypothesis (or disjunction of hypotheses) to infer given a set of data, while an inductive rule is a set of strict inductive rules. A solution to an inductive

[1] My conception of an inductive problem is largely borrowed from the literature on formal learning theory (cf. Juhl 1993; Kelly 1996; Schulte 1999a, 1999b) with one modification. In particular, I keep $\mathcal{D}$ and $\mathcal{K}$ distinct because I want to allow that $\mathcal{K}$ may do more than merely rule out some possible data streams. In my approach, $\mathcal{K}$ may also impose a ranking or weighting on $\mathcal{D}$, such as a probability distribution. This allows my account of inductive rules to accommodate methods that make assumptions about probabilities (cf. Harman and Kulkarni 2007).

problem consists of specifying a set of strict inductive rules each of which achieves the success condition $\mathcal{S}$ given $\mathcal{D}$, $\mathcal{K}$, and $\mathcal{H}$. A few simple examples to illustrate these concepts will be helpful.

Consider a case that I call *simple enumeration*. Suppose that one is examining a series of objects and observes in each case whether or not the object has the property *A*. Then $\mathcal{D}$ would contain all possible infinite sequences composed of *A*'s and ¬*A*'s. Let $\mathcal{H}$ be {∀x*A*x; ¬∀x*A*x}, where *x* ranges over the objects in the sequence of observations. In other words, given this hypothesis set, the challenge is to distinguish the sequence that consists of an endless string of *A*'s from the other sequences. The only assumption about $\mathcal{K}$ is that it does not indicate whether the universal generalization is true or false, and allows that there might be only one counter instance to the universal generalization. Finally, let the success condition $\mathcal{S}$ be logical reliability. A strict inductive rule is said to be *logically reliable* if with respect to $\mathcal{D}$, $\mathcal{K}$, and $\mathcal{H}$ if, for every data stream in $\mathcal{D}$ consistent with $\mathcal{K}$, the rule eventually permanently indicates the hypothesis in $\mathcal{H}$ that is true of that data stream. In other words, a rule is logically reliable in a given situation just in case it is assured of eventually settling on the truth, whatever that truth may be.[2] A set of strict rules will be said to be logically reliable with respect to an inductive problem just in case every strict rule in the set is.

It is easy to show that the following inductive rule solves the inductive problem of simple enumeration.

[2] See Kelly (1996, chapters 1 and 2) for further discussion of the concept of logical reliability.

> Principle of Induction: If a sufficient number of objects have been observed and all are *A*, then conjecture that $\forall x Ax$; if an object that is not *A* has been observed, conclude that $\neg\forall x Ax$.

I label this inductive rule the "principle of induction" because of its similarity to standard formulations of that proposition, which usually go something like: If a large number of *A*'s have been observed and all have been *B*, then it is probable that all further *A*'s are *B*.[3] My formulation of the principle of induction is differs only in being a prescription about what to infer given the data rather than a descriptive statement about probabilities. However, my formulation is not unique in this regard (cf. Ladyman 2002, 28-29). The principle of induction as formulated here is a set of strict inductive rules that differ regarding the number of positive instances that suffice for the universal generalization to be conjectured. However, for each strict rule, there is some number *n* of positive instances that is sufficient. The strict rules encompassed in the principle of induction may also disagree about what to do before *n* observations have been made. For example, a rule might recommend suspending judgment until *n*, or it might recommend that one begin by conjecturing that the universal generalization is false.

It is easy to show that the principle of induction is a solution to simple enumeration. Suppose that the universal generalization is true. Then the observations consist of an infinite sequence of *A*'s. In this case, any strict inductive rule that is consistent with the principle of induction will eventually conjecture the universal generalization and never change its conjecture after that. On the other hand, suppose that the universal generalization is false. In that case, an object that is not *A* will eventually

[3] For example, see Russell (1961, 153) and Chalmers (1999, 47).

be observed, and any strict inductive rule consistent with the principle of induction will conclude that the universal generalization is false. Hence, following the principle of induction is sufficient for logical reliability in cases of simple enumeration.

The principle of induction is also necessary for logical reliability with respect to simple enumeration in the sense that any strict rule *not* included in the set demarcated by the principle is *not* logically reliable in this inductive problem. To see this, consider a strict inductive rule that is not consistent with the principle of induction. One (or both) of the following must be true of this rule: (1) there is no number of positive instances sufficient to make the rule conjecture the universal generalization; (2) the rule does not conclude that the universal generalization is false when a counter instance is observed. Case (1) corresponds to an inductive skeptic who refuses to conjecture that the universal generalization is true no matter how much evidence accumulates in its favor. The skeptic is not logically reliable, since when the universal generalization is true, he never settles on the true hypothesis. Case (2) corresponds to what might be called stereotypical reasoning, since it involves ignoring counter instances to a universal generalization. Clearly, rules falling into case (2) are not logically reliable, since they do not reject the universal generalization even after an exception to it has been encountered.

The principle of induction, therefore, marks out the set of all and only those strict inductive rules that are logically reliable in cases of simple enumeration. It is important to recognize, however, that this result pertains to a particular type of inductive problem—simple enumeration—and does not demonstrate that the principle of induction is a universal prescription underlying inductive inferences of all kinds.[4] For example, an

[4] Thus, I am inclined to agree with Okasha (2005) that there is no single inductive principle assumed by all inductive inferences.

essential feature of simple enumeration is that our background knowledge is consistent with both the universal generalization and its negation, in other words, that one does not know at the start of inquiry whether or not the universal generalization is true. That is a fairly minimal requirement, since if it did not obtain, one would not have an inductive problem at all. Nevertheless, several authors object to the principle of induction on the grounds that it is not a good idea to follow it when our background knowledge entails that the universal generalization is false.[5] For example, Okasha (2001, 309) observes that, although all of the Costa Ricans he has ever met are philosophers, he would not infer that all Costa Ricans are philosophers. Yet since it is part of our background knowledge that there is no nation populated solely by philosophers, this case is crucially distinct from simple enumeration. In this example, one knows the answer from the start, and there is no inductive problem to solve.[6]

A related case is Colin Howson's (2000, 30) claim that Nelson Goodman's (1954) famous new riddle of induction shows that principle of induction is inconsistent. Goodman's riddle consists of providing some reason to think that all emeralds are green rather than grue, where an object is grue if it is observed before some future date *t* and green or not observed before *t* and blue. According to Howson, the principle of induction would lead one to predict that future emeralds will be both green and blue. But this is not

[5] A different type of objection to the principle of induction confuses it with stereotypical reasoning. For example, van Fraassen asserts that there is "exactly one shining ubiquitous example of induction and abduction: racism, ethnic prejudice, and stereotyping" (2000, 264). But promulgators of such stereotypes systematically ignore counter examples to their generalizations, and hence are not following the principle of induction. Indeed, if one reasons inductively, one would be driven to the conclusion that personality traits vary enormously within ethnic and racial groups, and consequently that simpleminded racial or ethnic stereotypes are bound to be false.

[6] Russell (1961, 155) makes a similar objection to the principle of induction: from observing that a child has grown taller each year, we would not infer that her height will continue to increase forever. Again this is not a case of simple enumeration, since our background knowledge includes the information that all humans are mortal and hence will not continue growing forever.

a correct application of the principle of induction. The principle requires that for any universal generalization we deem possible, there must be a finite number of positive instances that would suffice for us to conjecture its truth. But this does not entail that the sufficient number of positive instances is the same for each universal generalization. Thus, the principle of induction does not entail that there must be some *n* such that, after that after *n* observations, one conjectures that emeralds are both green and grue. Given the account proposed here, Goodman's riddle is a distinct inductive problem from simple enumeration because it has a different hypothesis set, and the principle of induction does not indicate what one should do when the hypothesis set includes more than one universal generalization. As Howson points out, it would be absurd to conjecture all universal generalizations that are consistent with the data. An inductive rule for Goodman's riddle, therefore, should include some device to ensure that no more than one universal generalization is conjectured at once. It is easy to design such an inductive rule: generate a list of the universal generalizations, and conjecture the first generalization on the list that is consistent with the data (and if none are, conclude that they are all false).[7] This inductive rule is logically reliable in Goodman's riddle.[8] It is also consistent with the principle of induction: for every universal generalization on the list, there is a sufficient number of positive instances that will result in it being conjectured. Of course, this rule does not tell us what to conjecture first, green or grue. Thus, Goodman's new riddle of induction is one way of making the point that convergence in the long run is consistent

[7] Indeed, Howson's Bayesian answer to Goodman's riddle (2000, 73-4) consists of the observation that a Bayesian updater with a countably additive prior will follow this rule.

[8] Suppose that one of the universal generalizations is true. Then the rule will eventually be driven to conjecturing that universal generalization and will never change its conjecture after that. On the other hand, if none of the universal generalizations are true, then decisive evidence for this will eventually be encountered for this (for example, a green emerald after a blue one or an emerald that is neither green nor blue).

with predicting almost anything in the short term. If one wants an account of what short term inferences to draw, one will have to consider success conditions in addition to logical reliability. Oliver Schulte (1999b) makes some interesting suggestions about what these further success conditions might be in the case of Goodman's riddle.

Let us consider one more example of an inductive rule, namely, statistical significance tests in which one decides whether the data justify rejecting a null hypothesis, for example, that a treatment has no effect. As before, an inductive rule must be associated with an inductive problem that it is intended to solve. Consider the textbook example of deciding whether a flipped coin is biased for heads or tails. In this case, the null hypothesis would be that the coin is not biased, that is, that the probability of heads is .5. In a significance test, the number of trials is typically set in advance, say, 100 flips of the coin. Hence, in this case, $\mathcal{D}$ is the set of possible permutations of 100 coin tosses. The background knowledge $\mathcal{K}$ would typically include information relevant for constructing a probability distribution over $\mathcal{D}$ given the assumption that the null hypothesis is true. The most commonly made assumptions of this kind are that the outcomes are independent and identically distributed. Independent means that the probability of heads on one flip is unaffected by the outcomes of other flips, while identically distributed means that the probability of heads is the same on every flip. Given these conditions, the probability of getting any number of heads in 100 flips on the assumption that the null hypothesis is true can be derived from the binomial distribution. The hypothesis set in this example simply consists of the null hypothesis and its negation. Finally, the success condition is that the probability of rejecting the null hypothesis when it is true be suitably small (say, .05), while the probability accepting the null hypothesis when it is false be as small as possible given this first requirement. Significance tests are inductive rules that recommend that the null hypothesis is rejected if the actual experimental result falls within a “critical region” of this distribution (for instance, at least two standard deviations from the mean, which in a normal distribution corresponds to the .05 significance level). Given the full details of a case, a significance test is a strict inductive rule. For example, in the above example, the significance test might tell us to reject the null hypothesis if the proportion of heads is less than one third or greater than two thirds. In contrast, the general recipe for devising significance tests is a set of strict rules that differ in accordance with such details as the sample size and the desired significance level.

In sum, there are two key features of inductive rules highlighted by the account proposed here. First, inductive rules are sets of functions from data to hypotheses, where the membership of the set is determined by a general recipe or principle, as illustrated by significance tests and the principle of induction. I call each individual function a strict inductive rule, and the strict rules encompassed by an inductive rule often differ by the value assigned to some parameter. For example, strict inductive rules consistent with the principle of induction can differ with regard to the number of positive instances that suffice for the universal generalization to be conjectured. Secondly, inductive rules are advanced as solutions to inductive problems, where an inductive problem is defined by a set of possible sequences of data, a body of background knowledge, a partition of hypotheses, and a success condition. Since a rule might be a solution to one problem but not another, evaluating an inductive rule requires that one keep clearly in mind what inductive problems the rule is intended to solve.

3. Okasha's Case against Inductive Rules

Okasha gives an initial characterization of inductive rules that, on its face, appears compatible with the account provided in the foregoing section.

> An inductive or ampliative inference rule, if such a thing existed, would be a rule for forming new beliefs on the basis of evidence, where evidence does not entail the belief. (2001, 315)

The qualification that the evidence does not entail the belief is another way of putting the point that inductive rules are intended as solutions to inductive problems. In an inductive problem, the evidence might never logically entail the true hypothesis. But the phrase "if such a thing existed" in the above quotation suggests that Okasha doubts whether there are any inductive rules. In fact, Okasha makes it quite clear that he is skeptical that people use inductive rules precisely because he doubts whether any exist.

> The basic case against the idea that we use inductive rules is straightforward: no one has yet come close to saying what these rules actually are. (2001, 317)

Okasha's motivation for calling the existence and use of inductive rules into question is the hope that Hume's inductive skepticism will thereby be avoided.

> Hume's inductive skepticism would be quite right, if we did use rules of inductive inference. But if we do not, then Hume's skeptical argument will not go through. (2001, 320)

According to Okasha, therefore, Hume's problem of induction is a consequence of the mistaken belief that inductive rules play a role in learning from experience. I do not agree that Hume's inductive skepticism is an unavoidable consequence of using inductive rules, but that is not the issue I wish to address now. Let us begin by considering the claim that people, and scientists in particular, do not use inductive rules.

As significance tests illustrate, it is easy to find examples of inductive rules that are widely used in science. Moreover, significance tests are inductive rules not only according to the proposal described in section 2 but also according to Okasha's characterization quoted at the head of this section. A significance test is a rule for forming a new belief on the basis of evidence that does not guarantee the truth of that belief. A significance test recommends that one reject the null hypothesis if the observed result falls in the critical region, yet the null hypothesis might be true even when this happens. For example, given the standard .05 significance level, when the null hypothesis is true it will be rejected approximately once in every twenty trials. Nor are significance tests the only example of a statistical inference rule that is pervasive in science. Other examples include regression analysis and maximum likelihood estimators. Given examples of widely used inductive rules such as significance tests, what basis could there be for claiming that there are no inductive rules or that inductive rules are not used in science?

In fact, Okasha's case against inductive rules does not include a survey of scientific practice or a discussion of empirical studies of human reasoning. Instead, he argues on conceptual grounds that inductive rules are undesirable and unnecessary. These charges would certainly be very significant if true, and in what follows I propose to focus on Okasha's arguments for these claims. His chief argument for the undesirability of inductive rules is that using them would force one to ignore relevant background information.

> One powerful consideration against 'inductive rules' concerns the role that background beliefs play in mediating our responses to experience. The same experience affects our expectations of the future in different ways depending on which background beliefs we hold. Bayesian models accommodate this point easily, of course, in the form of prior probability assignments. But the point is highly problematic for the 'inductive rule' idea. For any inductive rule that we allegedly follow, it is easy to imagine background beliefs which would lead us to disregard it. (2001, 318)

According to this objection, inductive rules are mechanical imperatives whose application entirely ignores relevant background information. Okasha considers the response that inductive rules are intended to apply only to those cases in which one is entirely ignorant, but rejects this argument on the grounds that, in real life, relevant background information is almost always present (2001, 318-9). Okasha's argument that inductive rules are unnecessary is also suggested in the above quotation: Bayesian inference is not an inductive rule, and people could be Bayesians. I will examine that argument in section 5. For now, let us consider the argument that inductive rules ignore background knowledge.

4. Rules and Background Knowledge

In this section, I critically examine the argument that it would be a bad thing to follow inductive rules because they force one to ignore relevant background knowledge. I argue to the contrary that rules in general and inductive rules in particular typically allow a role for background knowledge in at least two ways. First, since rules are usually designed for specific types of circumstances, background knowledge is often involved in deciding whether a rule may be legitimately applied in a given case. Secondly, rules often contain vague terms or adjustable parameters that must be given interpretations or numerical values on the basis of domain specific information, and background knowledge often plays a role in this process. Moreover, both of these features of inductive rules are highlighted by the account provided in section 2.

One of the main themes of the account of inductive rules presented in section 2 is that an inductive rule is an intended solution to a particular type of inductive problem. Hence, to decide whether one may legitimately apply an inductive rule in a particular instance, one needs to know what inductive problem one is faced with. Moreover, since a body of background knowledge is one of the components of an inductive problem, one cannot decide whether a rule should be applied without knowing something about what assumptions are justified in the case hand. This point is illustrated by significance tests as well as by other inductive rules commonly found in statistics texts. For instance, such rules typically rely on assumptions about the appropriate probability model, for example, that the outcomes are independent and identically distributed. Consider one additional example among the many that could be chosen. The simplest type of linear regression, known as ordinary least squares, is asymptotically reliable only when the dependent variable is a linear function of the independent variable and an error term, whose distribution is normal and homoscadastic. (Homoscadastic means that the variance of the distribution of the error term is constant, that is, it does not change with the value of the independent variable.) If your background knowledge indicates that this probability model is not a decent approximation of the case at hand, then you shouldn't use ordinary least squares.

This point is also relevant to Okasha's argument that behaving in accordance with an inductive rule in some circumstances but not others cannot count as following the rule, since we follow a rule only if "we accept *all* substitution-instances" (2001, 318; italics in original). The expression "substitution-instances" is ambiguous in this context because it is not clear whether it includes some reference to the type of inductive problem that the rule is intended to solve. If it does not, one might argue that ordinary least squares is not a rule that scientists follow because it is not used in every case in which one wishes to predict one real valued variable by means of another. Likewise, one might argue that no one follows the principle of induction because it is not used when background knowledge unequivocally indicates that the universal generalization is false. But such objections are mistaken, since an inductive rule is normally recommended as a solution to a specific type of inductive problem, not for every circumstance in which it could possibly be applied. If one keeps that point in mind, then it is easy to explain how following an inductive rule is consistent with accepting it in some "substitution-instances" but not others.

Background information can also be relevant for how a rule applies in cases in which it is relevant. Rules often contain parameters whose values must be specified on the basis of information about the case at hand, something that is especially true of rules that purport to be very general in scope. For example, consider the principle of utility, according to which one should always choose the act that will produce greatest total sum of happiness for all those concerned. One can draw informative conclusions about what should be done in a particular case from this rule only given some idea of how to measure happiness and knowledge of how the various possible actions will impact happiness so interpreted. Such judgments will typically rely upon background knowledge. More narrowly restricted rules can also contain vague or open-ended terms that must be interpreted in light of background knowledge. For instance, consider the rule that a police officer should not use excessive force when apprehending a suspect. What "excessive force" means in a given case is a matter of judgment that depends on such things as the nature of the threat posed by the suspect to the police officer and by-standers. Given that it is often impossible to anticipate all of the circumstances in which a rule might be applied, vague or open-ended terms that must be interpreted in light of background knowledge are an unavoidable feature of rules.

The account of inductive rules proposed in section 2 aims to capture this simple observation. Since inductive rules are sets of functions from data to hypotheses, judgments about how to apply an inductive rule in a given case involve a decision about which of the functions to select. And background knowledge usually plays a role in that decision. For example, as formulated above, the principle of induction does not specify how many positive instances should suffice for the universal generalization to be conjectured, and such judgments may be influenced by background knowledge. For example, if the universal generalization is improbable in light of background knowledge, one might require a very large number of positive instances. The same point goes for significance tests: they are a general recipe whose specific implementation in a given case often depends on relevant background knowledge.

As ordinarily understood, rules allow a role for background knowledge in their application. That point goes equally for familiar examples of inductive rules and is accommodated by the proposal in section 2. There is, therefore, very good reason to reject the claim that following inductive rules requires ignoring background knowledge.

5. Bayesian Inference and Inductive Rules

Okasha also argues that inductive rules are unnecessary because people may be Bayesians and Bayesian updating is not an inductive rule (2001, 315-7). Okasha's position on this point is inspired by a similar viewpoint advocated by van Fraassen (1989, chapter 7; 2000). Van Fraassen argues that from the orthodox Bayesian perspective "IBE [inference to the best explanation], and indeed the whole species of ampliative rules, is incoherent" (1989, 151). His argument for this claim presumes that Bayesian updating itself is not an inductive rule and then points out difficulties that ensue from attempting to graft inductive rules onto it. In this section, I argue to the contrary that Bayesian updating *is* an inductive rule. I explain how this conclusion follows both from the account of inductive rules advanced here and from a criterion for identifying inductive rules suggested by Okasha. I also show that Bayesian updating is committed to the principle of induction as formulated in section 2.

Bayesian inference is based upon the following theorem of the probability calculus (known as Bayes' theorem):

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}\text{, where } P(E) > 0\,.$$

The notation "$P(H \mid E)$" is read as the probability of H given E, and is known as the posterior probability. Thus, Bayes' theorem says, among other things, that the more H

raises the probability of *E*—that is, the more $P(E \mid H)$ exceeds $P(E)$—the greater the probability of the hypothesis given the evidence. Likewise, the greater the prior probability of the hypothesis, $P(H)$, the greater the probability of the hypothesis given the evidence. In the Bayesian approach, probabilities are interpreted as degrees of confidence or belief that an agent has in the statements in question. Let $P_{old}(\bullet)$ be the agent's probability function before learning the evidence *E*, and let $P_{new}(\bullet)$ be the agent's probability function after learning *E*. Then the rule of strict conditionalization says that that $P_{new}(H)$ should equal $P_{old}(H \mid E)$. Using Bayes' theorem in tandem with the rule of strict conditionalization is sometimes referred to as Bayesian inference, updating, or conditionalization.

Okasha argues that Bayes' theorem conjoined with strict conditionalization is not an inductive rule:

> Updating by Bayesian conditionalization does not count as an inductive inference in the strict sense, of course. For a rule of inductive inference is supposed to tell you what beliefs you should have, given your data, and the rule of conditionalization does not do that. If you apply the rule of conditionalization to your data, the state of opinion you end up in depends on the state you were in previously; whereas if you apply an inductive rule to your data, the state of opinion you end up in depends on the instructions contained in the rule. So there is the following crucial difference between the rule of conditionalization and the inductive rules whose existence I am disputing: an inductive rule is only going to be truth-conducive if the world is arranged in a particular way, but the same is not true of the rule of conditionalization. (2001, 316)

There are two arguments in this passage for the claim that Bayesian inference is not an inductive rule, which I will examine separately.

First, Okasha argues that Bayesian inference is not an inductive rule in the strict sense, because which inferences one draws are affected by one's choice of prior probability distribution and Bayesianism does not tell you which prior distribution you must have. Okasha's notion of an inductive rule in the strict sense corresponds closely to my notion of a strict inductive rule defined in section 2. Both are functions from data to subsets of a partition of hypotheses. The question of whether Bayesian inference is a strict rule depends on exactly what one means by "Bayesian inference." If by that term one refers to Bayesian updating with a *specific probability distribution*, then Bayesian inference is indeed a strict inductive rule. Given numerical values for all of the probabilities in Bayes' theorem, Bayesian updating indicates which hypotheses to infer based on the evidence and with what degree of certitude. On the other hand, by "Bayesian inference" one might intend only the following general prescription: for any inductive inference problem, devise a prior probability distribution, and then proceed by Bayesian updating. In this case, Bayesian updating is a recipe that demarcates a set of strict inductive rules. Thus, Bayesian inference in this more general sense is a clear example of an inductive rule according the account advanced here: it defines a set of functions from data to hypotheses, where the particular function chosen from this set in a given application is guided by pertinent background knowledge. As explained in the foregoing section, this account has the advantage of accommodating the fact that inductive rules, like rules of other kinds, often contain vague terms or free parameters that must be specified on a case-by-case basis.

Okasha's second argument in the above passage rests on a criterion for judging whether something qualifies as an inductive rule. An inductive rule must be truth-conducive if the world is one way but not if the world is otherwise. Okasha then asserts that Bayesian inference does not satisfy this criterion. But in fact the opposite is the case. For example, suppose that *H* is a true universal generalization, but that the agent mistakenly believes with certainty that *H* is false, that is, that $P(H) = 0$. In this situation, Bayesian updating might never raise the probability of *H* no matter how many positive instances are observed. This is a simple case in which whether Bayesian updating is truth-conducive depends on what the world is actually like. A Bayesian updater who assigns a prior probability of zero to a universal generalization must hope that it is false in the real world, since otherwise she may be perpetually mistaken. There is also a more interesting sense in which the actual state of the world matters for whether Bayesian inference is truth-conducive. For simplicity, suppose that there are only three possible hypotheses: H_1, H_2, and H_3. Then we can rewrite Bayes' theorem as follows.

$$P(H_1 \mid E) = \frac{P(E \mid H_1)P(H_1)}{P(E \mid H_1)P(H_1) + P(E \mid H_2)P(H_2) + P(E \mid H_3)P(H_3)}$$

The $P(E \mid H_i)$'s are known as *likelihoods*. Applications of Bayes' theorem require devising a probability model that indicates particular values for these likelihoods. If that probability model is correct, then it is possible to demonstrate that Bayesian updating is truth-conducive (cf. Goldman 1999, 115-23). However, if one's probability model is seriously mistaken, then all bets are off. And of course, whether the probability model is approximately correct or significantly wrong depends on the actual state of the world. This is essentially the same point made above about inductive rules that one typically finds in statistics textbooks. They depend on assumptions about a probability model, and if those assumptions are wrong the rule may be very unreliable. Thus, whether Bayesian updating is truth conducive can depend on what the world is actually like. So, Bayesian inference *is* an inductive rule according to Okasha's criterion.

Moreover, it is straightforward to show that a Bayesian updater must accept the principle of induction as formulated in section 2 for cases of simple enumeration. Let us say that a Bayesian conjectures H given the evidence E if $P(H \mid E) > .5$ and that the agent concludes that H is false given the evidence E if $P(H \mid E) = 0$. Now consider a case of simple enumeration. Presented with a series of objects, we observe in each case whether or not the property A is present. Then $\mathcal{D}$ consists of all possible countably infinite sequences of A's and $\neg A$'s. The hypothesis set $\mathcal{H}$ is $\{\forall xAx; \neg\forall xAx\}$, where x ranges over the observations listed in the sequence of data. Moreover, simple enumeration requires that the background knowledge $\mathcal{K}$ be compatible with $\forall xAx$ and its negation, which in Bayesian terms would be represented by having $P(\forall xAx)$ be strictly between 1 and 0. Given this set up, it is easy to see that a Bayesian agent will follow the principle of induction as formulated in the section 2.

Recall that the principle of induction requires two things: that there be an n such that the universal generalization is conjectured if n positive instances and no negative instances have be observed, and that it be concluded that the universal generalization is false if a negative instance is observed. Let A_i indicate that the ith object was observed to have the property A. Obviously, for any i, $\neg A_i$ entails that $\forall xAx$ is false.
Therefore, $P(\forall xAx \mid \neg A_i) = 0$, which means that a Bayesian updater will conclude that is $\forall xAx$ false upon observing a single counter instance. Moreover, since $P(\forall xAx) > 0$, there

is a sufficient number of positive instances that will make the Bayesian updater conjecture that $\forall xAx$ is true. To see this, write out Bayes' theorem as follows.

$$P(\forall xAx \mid A_1 \,\&\, ... \,\&\, A_n) = \frac{P(\forall xAx)P(A_1 \,\&\, ... \,\&\, A_n \mid \forall xAx)}{P(A_1 \,\&\, ... \,\&\, A_n)}$$

Note that since $\forall xAx$ entails A_1 & … & A_n, the term $P(A_1 \,\&\, \ldots \,\&\, A_n \mid \forall xAx) = 1$ and drops out of the equation.

$$P(\forall xAx \mid A_1 \,\&\, ... \,\&\, A_n) = \frac{P(\forall xAx)}{P(A_1 \,\&\, ... \,\&\, A_n)}$$

Moreover, since $\forall xAx$ is equivalent to $(A_1 \,\&\, A_2 \,\&\, A_3 \ldots)$, $P(\forall xAx)$ is equal to $\lim_{n\to\infty} P(A_1 \,\&\, ... \,\&\, A_n)$. Hence, so long as $P(\forall xAx) > 0$, $P(\forall xAx \mid A_1 \,\&\, ... \,\&\, A_n)$ converges to 1 as $n \to \infty$. Therefore, in cases of simple enumeration, there is an n such that $P(\forall xAx \mid A_1 \,\&\, ... \,\&\, A_n) > .5$, which means that there is finite number of positive instances that is sufficient to make the Bayesian updater conjecture that $\forall xAx$ is true. Consequently, in cases of simple enumeration, a Bayesian updater must follow the principle of induction as formulated in section 2.[9]

There are multiple reasons, then, for concluding that Bayesian inference is an inductive rule. It qualifies as such according to the account proposed here and by the criterion suggested by Okasha. In addition, in cases of simple enumeration, Bayesian inference is committed to the principle of induction, which is a standard example of an inductive rule. The upshot of this is that arguments in favor of Bayesian inference should proceed in the same way as arguments for any other proposed inductive rule: specify

[9] The same observation is made by Howson (2000, 71-2).

desirable epistemic ends and show that Bayesian inference attains those ends in various types of inductive problems.

6. Lange's Inductive Skeptic

But for the sake of argument, let us suppose that Bayesian inference is not an inductive rule and consider what consequences might ensue regarding Hume's problem of induction. Okasha argues that it follows that there is no need to be concerned about justifying Hume's principle of induction. A person has no choice but to begin with beliefs that can be represented by a probability distribution, and the best she can do is modify those beliefs in a rational manner as new evidence comes in (2001, 321-3). According to this proposal, inferences about the future based on past experience do not depend on "inductive rules" but merely upon substantive prior beliefs about the actual state of the world. This picture is also advocated by van Fraassen (1989, chapter 7; 2000), although Okasha (2001; 2005) explores its relevance to Hume's problem in greater depth. Lange (2002) criticizes Okasha's answer to Hume's problem on the grounds that our prior knowledge is sometimes insufficient to indicate which inductive inferences ought to be drawn. Lange provides several examples from the history of science to illustrate this point, and suggests that an inductive skeptic can be understood in Bayesian terms as a person whose prior probability distribution does not license any inductive generalizations. In this section, I examine Lange's inductive skeptic and its relevance to the account of inductive rules proposed here.

Lange proposes that an inductively neutral prior probability distribution can be constructed in the following way.

> If the sceptic is asked to recommend a prior probability, he should suggest a distribution that makes no probability assignment at all to any claim about the world which concerns logically contingent matters of fact. (By this, I do not mean the assignment of *zero* subjective probability to such a claim. That would be to assign it a probability, namely, zero. Nor do I mean assigning it a *vague* probability value. I mean making no assignment at all to any such claim, unless we have directly observed it to hold.) (2002, 228; italics in original)

But the notion of a probability distribution that fails to assign any numerical value at all to some measurable subsets of the outcome space is puzzling, since it directly contradicts the usual definition of a probability measure (cf. Billingsley 1995, 19-23). Lange suggests that he means to employ an alternative conception of probability that does not require assigning a number to every member of a σ-field[10] of subsets of the outcome space (2004, 202-3). However, this maneuver comes at a cost, since there is no reason to suppose that any of the familiar theorems of probability hold true of this alternative, weakened conception. I think a much simpler approach is to leave the concept of a probability as it is, while interpreting Lange's skeptic as a person whose state of belief is represented by *a set of probability distributions*.

In the discussion of Bayesian inference and the principle of induction in the previous section, I presumed that a Bayesian updater must adopt a particular probability distribution. And of course that is the usual Bayesian way of doing things. However, there are situations in which one is not merely uncertain but also ignorant in the sense of not even having an idea of the precise probabilities to associate with various possibilities.

[10] A σ-field is a set of subsets of the outcome space that is closed under the formation of complements and countable unions.

In such cases, it is natural to represent the agent's state of belief by a set of probability distributions that are consistent with her judgment, say, that one possibility is more probable than the others. An inductive skeptic can be thought of as an extreme case of this sort of situation. In Bayesian terms, an inductive skeptic is a person whose state of belief is represented by the *set of all probability distributions* that are compatible with the data.

Consider how a skeptic thus defined behaves the case of simple enumeration described in the previous section. The outcome space consists of the set of all infinite sequences of A's and $\neg A$'s, and let A_i indicate that the ith object observed has the property A. At the beginning of inquiry, the inductive skeptic's state of belief is represented by the set **P** of all probability distributions over this outcome space that give $\forall xAx$ a probability strictly between 1 and 0. As evidence comes in, each of these probability distributions is updated using the rule of strict conditionalization. For example, if the data so far are $(A_1 \& ... \& A_n)$, then for each probability distribution P in **P**, the updated probability of $\forall xAx$ is equal to $P(\forall xAx \mid A_1 \& ... \& A_n)$. The inductive skeptic conjectures $\forall xAx$ if and only if the updated probability of $\forall xAx$ is greater than .5 for every probability distribution in **P**. Given this set up, it is easy to show that the inductive skeptic will never conjecture that the universal generalization is true, no matter how many positive instances he observes. That is because, for any n, it is possible to find probability distribution in which $P(\forall xAx) > 0$ but $P(\forall xAx \mid A_1 \& ... \& A_n) < .5$. Here is one simple recipe for constructing such a probability distribution: $P(A_1 \& ... \& A_n) = 1$,

$P(A_{n+1}) = .01$, and for any $m > 1$, $P(A_{n+m}) = 1$. In this case, $P(\forall xAx) = P(A_{n+1}) = .01$. And since $P(A_1 \& ... \& A_n) = 1$, $P(\forall xAx \mid A_1 \& ... \& A_n) = P(\forall xAx)$.

Formulating inductive skepticism in this way highlights the crucial weakness in Okasha's Bayesian answer to inductive skepticism. Okasha presumes that it is necessary to be committed to a single probability distribution, and then challenges the skeptic to say which distribution should be chosen (2001, 323; 2003, 422). But the skeptic would deny the presupposition of this challenge. He would claim that he is under no obligation to commit to a particular probability distribution and that, as a skeptic, he can hardly be expected to tell others what to believe! The real issue here is not whether the skeptic can tell us what prior probability distribution to adopt, but whether there is some reason not to be an inductive skeptic. Is there some reason why it would be a bad thing to conjecture only those hypotheses that receive a probability of at least .5 in every probability distribution consistent with the data? We can consider this question from two perspectives. First, we can ask if there is a specifically *Bayesian* reason why it is bad to be an inductive skeptic. I argue that the answer to this question is *no*. Secondly, we can whether there is any *other* reason why it is bad to be an inductive skeptic. I argue that the answer to this question is *yes*: an inductive skeptic is not logically reliable in cases of simple enumeration. To the extent that logical reliability is a desirable epistemic end, there is good reason not to be an inductive skeptic.

There are two primary types of arguments given by Bayesians for the claim that a rational person should be *coherent*, that is, should have degrees of belief consistent with the axioms of probability. An example of incoherence is a person with a degree of belief of .6 that it will rain tomorrow and a degree of belief of .5 that will not rain. One basis for claiming that incoherence is irrational is the Dutch book argument, which attempts to show that an incoherent person would be willing to accept a series of bets in which he is sure to suffer a net loss.[11] But it is clear that the Dutch book argument cannot impugn the rationality of the inductive skeptic, since the inductive skeptic does have degrees of belief that violate the probability axioms. Far from it: the beliefs of an inductive skeptic are consistent with infinitely many probability distributions. In short, the Dutch book may provide some support for the claim that, *if you have degrees of belief*, then your degrees of belief should obey the rules of probability. However, nothing in the Dutch book argument provides any support for the claim that rationality requires having precise degrees of belief in every possibility under consideration, which is what is at issue when it comes to inductive skepticism.[12]

Representation theorems are the other basis commonly given for the claim that coherence is a necessary condition for rationality. Representation theorems state that if a person has preferences that satisfy certain seemingly reasonable requirements, then that person can be represented as acting to maximize expected utility according to a utility and probability function. However, to get the result that the utility-probability function pair is unique, it is necessary to assume a condition known as connectedness: for every two options, the person either prefers one to the other or is indifferent between them. As Patrick Maher (1993, 19-21) points out, there is little plausibility in the claim that connectedness is a necessary condition for rationality. For example, difficult, real-world decisions typically involve tradeoffs between incommensurable values and hence defy simple preference rankings. Thus, Maher's version of the representation theorem does

[11] See Howson and Urbach (1993, chapter 5) and Earman (1992, 38-40).
[12] Lange also makes this point (2004, 203).

not assume connectedness, and consequently only requires that there be at least one utility-probability function pair consistent with the person's preferences. There is no reason why an inductive skeptic cannot satisfy this condition.

Although there do not appear to be any specifically Bayesian reasons for judging an inductive skeptic irrational, there is a straightforward reason not to be an inductive skeptic that follows from the discussion in section 2. Recall that an inductive rule is logically reliable in an inductive problem if it assured of eventually settling on the true hypothesis, whatever the truth may be. The reason not to be an inductive skeptic, then, is that an inductive skeptic fails to be logically reliable in cases for which logically reliable methods are easily devised. An inductive skeptic may not be irrational, but he is not very good at finding the truth either. Thus, if epistemology aims to find methods that lead to the truth, then there is an epistemic reason not to be an inductive skeptic, whether or not the inductive skeptic is irrational. The thesis that rationality may not be all there is to epistemology conflicts with van Fraassen's "epistemological voluntarism" (1989, chapter 7). According to epistemological voluntarism, all beliefs and changes in belief are permitted so long as they do not violate some canon of rationality, and Bayesianism is the only account of rationality pertinent to inductive reasoning that van Fraassen describes. Hence, it is difficult to see what argument against inductive skepticism could be made on the basis of epistemological voluntarism. For one who expects an adequate epistemology to provide some reason not to be an inductive skeptic, this is a serious shortcoming.

7. Conclusion

This essay has defended the traditional notion that, when it comes to inductive reasoning, epistemology is a matter of articulating inductive rules and providing reasons why those rules should be followed in various sorts of circumstances. Two primary objections to this conception of inductive inference were considered: inductive rules are undesirable because they ignore background knowledge and unnecessary because Bayesian updating is not an inductive rule. In defense of the traditional view, I proposed an account of inductive rules according to which an inductive rule is a set of functions from data to hypotheses intended to solve an inductive problem. Given this proposal, I argued that inductive rules do allow a role for background knowledge, and moreover that Bayesian updating qualifies as an inductive rule. Finally, I explained how the inductive rules approach provides a straightforward reason not to be an inductive skeptic, despite the fact that there appears to be no Bayesian reason to judge an inductive skeptic irrational.

References

Billingsley, P. (1995), *Probability and Measure*. New York, NY: John Wiley & Sons.

Chalmers, A. F. (1999), *What is this thing called Science?* 3rd edition. Indianapolis, IA: Hackett Publishing.

Earman, J. (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Goldman, A. (1999), *Knowledge in a Social World*. New York, NY: Oxford University Press.

Goodman, N. (1954), *Fact, Fiction, and Forecast*. Cambridge, MA, Harvard University Press.

Harman, G. and S. Kulkarni (2007), *Reliable Reasoning: Induction and Statistical Learning Theory*. Cambridge, MA: MIT Press.

Howson, C. (2000), Hume's Problem: Induction and the Justification of Belief.

Howson, C. and P. Urbach (1993), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Juhl, C. (1993), "Bayesianism and Reliable Scientific Inquiry", *Philosophy of Science* 60: 302-19.

Kelly, K. (1996), *The Logical of Reliable Inquiry*. Oxford, UK: Oxford University Press.

Ladyman, J. (2002), *Understanding Philosophy of Science*. New York, NY: Routledge.

Lange, M. (2002), "Okasha on Inductive Skepticism", *Philosophical Quarterly* 52: 226-32.

_____ (2004), "Would 'Direct Realism' Resolve the Classical Problem of Induction?", *Noûs* 38: 197-232.

Maher, P. (1993), *Betting on Theories*. New York, NY: Cambridge University Press.

Okasha, S. (2001), "What did Hume really show about Induction?", *Philosophical Quarterly* 51: 307-27.

_____ (2003), "Probabilistic Induction and Hume's Problem: A Reply to Lange", *Philosophical Quarterly* 53: 419-424.

_____ (2005), "Does Hume's Argument Against Induction Rest on a Quantifier-Shift Fallacy?", *Proceedings of the Aristotelian Society* 105: 237-255.

Russsell, B. (1961), *The Basic Writings of Bertrand Russell, 1903-1959*. Edited by R. Egner and L. Denonn. New York, NY: Simon and Schuster.

Schulte, O. (1999a), "The Logic of Reliable and Efficient Inquiry", *Journal of Philosophical Logic* 28: 399-438.

_____ (1999b), "Means-Ends Epistemology", *British Journal for the Philosophy of Science* 50: 1-31.

Steel, D. (2005), "The Facts of the Matter: A Discussion of Norton's Material Theory of Induction", *Philosophy of Science* 72: 188-197.

Van Fraassen, B. (1989), *Laws and Symmetry*. New York, NY: Oxford University Press.

_____ (2000), "The False Hopes of Traditional Epistemology", *Philosophy and Phenomenological Research* 60: 253-80.