On Novel Confirmation

James A. Kahn; Steven E. Landsburg; Alan C. Stockman

The British Journal for the Philosophy of Science, Volume 43, Issue 4 (Dec., 1992), 503-516.

Stable URL:
http://links.jstor.org/sici?sici=0007-0882%28199212%2943%3A4%3C503%3AONC%3E2.0.CO%3B2-D

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The British Journal for the Philosophy of Science is published by Oxford University Press. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/oup.html.

The British Journal for the Philosophy of Science
©1992 The British Society for the Philosophy of Science

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

http://www.jstor.org/
Tue Aug 12 11:43:35 2003

Brit. J. Phil. Sci. 43 (1992) 503–516 *Printed in Great Britain*

On Novel Confirmation

JAMES A. KAHN, STEVEN E. LANDSBURG AND
ALAN C. STOCKMAN

ABSTRACT

Evidence that confirms a scientific hypothesis is said to be 'novel' if it is not discovered until after the hypothesis is constructed. The philosophical issues surrounding novel confirmation have been well summarized by Campbell and Vinci [1983]. They write that

> philosophers of science generally agree that when observational evidence supports a theory, the confirmation is much stronger when the evidence is 'novel' ... There are, nevertheless, reasons to be skeptical of this tradition ... The notion of novel confirmation is beset with a theoretical puzzle about how the degree of confirmation can change without any change in the evidence, hypothesis, or auxiliary assumptions ... There have not yet appeared any obviously satisfactory solutions to these problems

Much of the literature on novel confirmation relies on the Bayesian analysis of conditional probabilities. Let H represent a hypothesis, E an event that confirms the hypothesis, and B some relevant background information. Denote by $Pr(x \mid y)$ the conditional probability of x given y. There are various plausible measures of the degree of support that E lends to H. Among these are:

(1) Degree of support $= Pr(H \mid E)$
(2) Degree of support $= Pr(H \mid E \,\&\, B) - Pr(H \mid B)$
(3) Degree of support $= Pr(H \mid E \,\&\, B) - Pr(H \mid B - \{E\})$
(4) Degree of support $= Pr(H \mid B \,\&$ knowledge that $H \rightarrow E) - Pt(H \mid B)$

In the work cited, Campbell and Vinci offer a somewhat more involved Bayesian interpretation.

Formula (1) is discussed by Gardner [1981] who points out that under this formulation there can be no role for novelty. 'The function $Pr(H \mid E)$ contains no third slot in which to insert a temporal relation between the invention of H and the inventor's learning E. Obviously, then, this relation could not possibly affect E's support of H.' Formula (2), on the other hand, suggests a role for novelty. Bayes's Theorem allows us to rewrite the formula as

$$\text{(2}'\text{)} \qquad \text{Degree of support} = Pr(H \mid B) \times \left[\frac{Pr(E \mid H \,\&\, B)}{Pr(E \mid B)} - 1\right].$$

We can use $1/Pr(E \mid B)$ as a measure of the novelty of E. Then (2′) shows that the degree of support increases with novelty of E. In the paper already cited, Campbell and Vinci discuss shortcomings of this analysis.

Formula (3) expresses an alternative offered by Howson [1984]. In that formula, $Pr(H \mid B - \{E\})$ represents the probability of H assuming (counterfactually) that only $B - \{E\}$ is known. This allows for the possibility of non-novel facts generating support for hypotheses.

Niiniluoto [1984] argues for a variant along the lines of (4), in which we account for the possibility that the theorist was unaware that his hypothesis entails E.

One problem in deciding among these approaches is that the choice of a definition for the degree of support appears arbitrary. What kind of argument could justify the choice of one definition over another?

It is our position that there can be no basis for addressing this question in the absence of an explicit model of the process by which hypotheses are generated. Only in the presence of such a model can the various conditional probabilities be given meaningful interpretations. We provide such models in Sections 1, 2 and 4. The simple model of Section 1, incorporating strong assumptions, yields the conclusion that novelty is irrelevant. When these assumptions are relaxed in the later sections, novelty becomes relevant for a variety of reasons.

It is at least potentially the case that scientists have more information about their own abilities than is publicly available, and this information might influence their decisions about whether even to attempt novel prediction. If this is so, then it should be incorporated into the model of hypothesis generation. This requires an explicit discussion of how scientists respond to incentives and how the incentives themselves evolve, which in turn takes us into the realm of economic theory. We have addressed these issues in another paper, written for an audience of economists. The results of this research are summarized in Section 3.

I A SIMPLE MODEL

In this section we present the simplest model of hypothesis generation that we have been able to devise. It incorporates three fundamental assumptions, each of which will be explicated further as we proceed. First, we assume that the abilities of scientists are public knowledge. Second, we assume (in the terminology of economics) that scientists' research strategies are determined *exogenously*, which in this context means that scientists do not select their strategies in response to any existing incentive structure. This could come about, for example, if scientists have no control over their own research strategies, being either theorists or empiricists by birth or by exogenous training. It could also come about if their research strategies are dictated for them arbitrarily by a research director. Third, we assume that all observations and experimental results are always interpreted correctly (that is, if scientists believe that they observe a precession in the orbit of Mercury, then the orbit of Mercury actually precesses). These assumptions will be relaxed in Sections 2, 3 and 4 respectively.

Our main conclusion in this section will be that in the presence of these three assumptions, the novelty of evidence is irrelevant. Our main conclusion in succeeding sections will be that when any of these assumptions is relaxed, novelty becomes important, for a variety of reasons.

We consider a single scientist who works on a project that involves generating one new hypothesis H and making one new observation O. We interpret the notion of the scientist's making an observation quite broadly. Making an observation can consist of looking through a telescope, performing an experiment and noting the results, or being told by another scientist what *he* saw when looking through a telescope or performing an experiment.

The scientist can follow either of two research strategies. He can 'theorize first', in which case he first generates the hypothesis and then makes the observation, reporting his hypothesis only if the observation agrees with it. Alternatively, he can 'look first', in which case he first makes the observation and then generates a hypothesis that is compatible with it. We can ignore the results of the failed theorizer, and want to compare the believability of a hypothesis generated by a successful theorizer with that of the same hypothesis when it is generated by a looker. That is, we want to compare the two expressions

$$Pr(H \mid T \,\&\, E)$$

and

$$Pr(H \mid L \,\&\, E)$$

where T is the statement 'the scientist theorizes first', L is the statement 'the scientist looks first', and E is the statement 'the observation O agrees with the hypothesis H'.

Of course, in the second expression, E is redundant in view of the fact that it is implied by L.

No interpretation can be given to these conditional probabilities in the absence of a model. Our next task is to provide that model.

We divide hypotheses into four types: type A hypotheses are true, and hence compatible with all previously existing observations as well as with O. Type B hypotheses are false, but nevertheless compatible with all previously existing hypotheses as well as with O. (Presumably their falsehood could be revealed by some future observation.) Type C hypotheses are false and compatible with all previous observations, but not with O. Type D hypotheses are false and incompatible with some observation made prior to the current research project.

There is some fundamental randomness involved in any scientific endeavor, and we can therefore view the outcome of the theorizing process as analogous to the random selection of a ball from an urn. Each ball in the urn is labelled with a hypothesis of type A, B, C or D, and the outcome of the scientist's research is determined by the ball that he draws.

In proposing this model, we do not mean to suggest that the actual process of scientific research in any way resembles the process of drawing a ball from an urn. We mean only to assert that the outcome of the research process is the same as the outcome of our imaginary ball-drawing process, if the contents of the urn are specified correctly. That is, the actual research process and the process of drawing a ball are observationally equivalent from the viewpoint of an observer who does not know in advance what hypothesis the scientist will construct.

We assume that researchers are capable of avoiding hypotheses that are incompatible with previously established observations. This is tantamount to assuming that there are no balls of type D in the urn.

Thus the balls in the urn of are types A, B and C, and these occur in some proportions. We write p for the proportion of type A balls and q for the proportion of type B balls; then the proportion of type C balls is $1 - p - q$. The numbers p and q can be thought of as characteristic of the researcher's 'abilities'. For example, if p is high then the researcher's hypotheses are often true.

Consider a scientist who theorizes first and then makes an observation that turns out to be compatible with his hypothesis. Bayes's formula gives

(5) $$Pr(H \mid T \& E) = \frac{Pr(H \mid T) \times Pr(E \mid H)}{Pr(E \mid T)} = \frac{p \times 1}{p + q}.$$

In this computation, we used the following observations: first, $Pr(H \mid T)$ is the probability that a theorist draws a type A ball, which is p. Second, $Pr(E \mid T \& H) = 1$, since the truth of hypothesis H ensures the truth of statement E. Finally, $Pr(E \mid T)$ is the probability that a theorist draws a ball of type A or B, which is $p + q$.

Next we consider a scientist who looks first. By looking first, he is able to avoid constructing a type C theory, which amounts to discarding all of the type C balls from the urn. The balls of types A and B remain, in proportions $p/(p + q)$ and $q/(p + q)$. It is now easy to see that $Pr(H \mid L \& E)$ is the probability that this scientist draws a type A ball, or $p/(p + q)$.

In summary, we have found that $Pr(H \mid T \& E)$ and $Pr(H \mid L \& E)$ have the same value, namely $p/(p + q)$. That is: conditional on the observation's consistency with the hypothesis, the scientist's research strategy has no bearing on the probability that his hypothesis is correct.

We have sometimes spoken in this section as if the scientist makes a voluntary decision about whether to look first or to theorize first. In practice, this might not be the case. For example, for technological reasons, Einstein could not have chosen to observe the bending of light rays by the sun prior to constructing the general theory of relativity. But this makes no difference to our argument. We can still compare the probability that general relativity is true given the novelty of light-bending with the probability that general relativity is true in the counterfactual situation where Einstein 'looks first'. Our calculations still apply and our conclusion still holds.

2 SCIENTISTS OF UNCERTAIN ABILITIES

In the model of Section 1, we found that

$$Pr(H \mid T \,\&\, E) = Pr(H \mid L \,\&\, E) \tag{6}$$

The common value of these expressions, which we can simply denote $Pr(H \mid E)$, is $p/(p+q)$. To calculate this value requires knowledge of p and q. But to derive the equality (6), which asserts the irrelevance of novelty, it is necessary only to know that some values of p and q exist, and that they do not differ across scientists.

However, when scientists differ in their abilities (as measured by p and q), and when the abilities of a given scientist are unknown to the observer, the argument of Section 1 breaks down. We shall demonstrate that the conclusion breaks down as well; that is, equation (6) fails to hold; that is, novelty matters.

We shall demonstrate the failure of equation (6) by providing a simple counterexample. The ideas underlying the counterexample suffice to demonstrate that the equation will fail in any example where scientists have differing, unobservable abilities.

We suppose that there are two types of scientists: Type i scientists draw from urns which contain A, B and C balls in the proportions p, q and $1-p-q$, while type j scientists draw from urns which contain A, B and C balls in the proportions r, s and $1-r-s$. The proportion of type i scientists in the population is i and the proportion of type j scientists in the population is $j = 1 - i$. We assume that

$$p/q > r/s \tag{7}$$

$$p + q > r + s. \tag{8}$$

These conditions can be interpreted to mean that type i scientists are more proficient than type j scientists, in two senses. Equation (7) says that when scientists look first, a type i scientist is more likely than a type j scientist to produce a true theory. Equation (8) says that when scientists theorize first, a type i scientist is more likely than a type j scientist to produce a theory that survives testing. Now suppose that a scientist produces a hypothesis H

and a compatible observation O. We want to compute the probability that his hypothesis is correct, first assuming he used the 'theorize first' strategy and then assuming he used the 'look first' strategy. As in Section 1, E denotes the statement that O is compatible with H, T denotes the statement that the scientist theorized first, and L denotes the statement that the scientist looked first. We also let I denote the statement 'The scientist is of type i' and let J denote the statement 'The scientist is of type j'.

To compute $Pr(H \mid T \,\&\, E)$, we begin with two preliminary calculations. We have

$$\begin{aligned} Pr(I \,\&\, H \mid T \,\&\, E) &= \frac{Pt(E \mid T \,\&\, I \,\&\, H) \times Pr(I \,\&\, H \mid T)}{Pr(E \mid T)} \\ &= \frac{Pr(E \mid T \,\&\, I \,\&\, H) \times Pr(I \,\&\, H \mid T)}{Pr(I \,\&\, E \mid T) + Pr(J \,\&\, E \mid T)} \\ &= \frac{1 \times ip}{i(p+q) + j(r+s)} \end{aligned}$$

and similarly

$$\begin{aligned} Pr(J \,\&\, H \mid T \,\&\, E) &= \frac{Pr(E \mid T \,\&\, J \,\&\, H) \times Pr(J \,\&\, H \mid T)}{Pr(E \mid T)} \\ &= \frac{Pr(E \mid T \,\&\, J \,\&\, H) \times Pr(J \,\&\, H \mid T)}{Pr(I \,\&\, E \mid T) + Pr(J \,\&\, E \mid T)} \\ &= \frac{1 \times jr}{i(p+q) + j(r+s)} \end{aligned}$$

so that finally

$$\begin{aligned} (9) \qquad Pr(H \mid T \,\&\, E) &= Pr(I \,\&\, H \mid T \,\&\, E) + Pr(J \,\&\, H \mid T \,\&\, E) \\ &= \frac{ip + jr}{i(p+q) + j(r+s)}. \end{aligned}$$

Next, we compute $Pr(H \mid L \,\&\, E)$, again beginning with two preliminary calculations. Recall first that $L \,\&\, E$ is equivalent to the simpler statement L. We invoke the assumption that a scientist's research strategy conveys no new information about his type (this is the second of the assumptions in the first paragraph of Section 1). This assumption can be expressed formally as $Pr(I \mid L) = Pr(I) = i$; it is used in the third line of the following calculation:

$$\begin{aligned} Pr(I \,\&\, H \mid L \,\&\, E) &= Pr(H \mid I \,\&\, L \,\&\, E) \times Pr(I \mid L \,\&\, E) \\ &= Pr(H \mid I \,\&\, L \,\&\, E) \times Pr(I \mid L) \\ &= \frac{p}{p+q} \times i. \end{aligned}$$

A similar calculation reveals that

$$Pr(J \,\&\, H \mid L \,\&\, E) = \frac{r}{r+s} \times j$$

so that finally

$$(10) \qquad \begin{aligned} Pr(H \mid L \,\&\, E) &= Pr(H \mid I \,\&\, L \,\&\, E) + Pr(H \mid J \,\&\, L \,\&\, E) \\ &= \frac{ip}{p+q} + \frac{jr}{r+s}. \end{aligned}$$

A comparison of expressions (9) and (10) reveals that, in contrast to what we found in Section 1, the two conditional probabilities are not equal. In fact, it is a straightforward exercise in algebra, using equations (7), (8), (9) and (10), to verify that $Pr(H \mid T \,\&\, E) > Pr(H \mid L \,\&\, E)$. This means that in the model of this section, novel evidence carries more weight than non-novel evidence.

We want to comment briefly on the intuition underlying this result. Speaking very loosely, we can think of the type i scientists as more talented, with respect to the problem at hand, than the type j scientists. Prior to successfully making a novel prediction, a scientist has probability $i = Pr(I)$ of being 'talented'. Once he has made a novel prediction, the probability of his being talented increases to

$$\begin{aligned} Pr(I \mid T \,\&\, E) &= \frac{Pr(E \mid T \,\&\, I) \times Pr(I \mid T)}{Pr(E \mid T)} \\ &= \frac{(p+q) \times i}{i(p+q) + j(r+s)} \\ &> i. \end{aligned}$$

Since the scientist who has successfully made a novel prediction is more likely to be talented, it follows that his hypothesis is more likely to be correct. Indeed we could calculate $Pr(H \mid T \,\&\, E)$ by using the same formula as in (10) with i and j replaced by i' and j', here $i' = Pr(I \mid T \,\&\, E)$ and $j' = Pr(J \mid T \,\&\, E)$. This gives the same result as (9) and reveals that the new information about the probabilities of I and J is the sole source of the discrepancy between (9) and (10).

3 ABILITIES GUIDE RESEARCH STRATEGIES

In Section 2, we assumed that $Pr(I\,|\,L) = Pr(I\,|\,T) = Pr(I)$ (though we found that $Pr(I\,|\,T\,\&\,E)$ does *not* equal $Pr(I)$. That is, we assumed that the scientist's choice of research strategy does not by itself convey any new information about his abilities.

There are a number of circumstances in which this assumption is a reasonable one. For example, the choice of research strategy is often dictated by considerations completely distinct from the abilities of the scientist. We have already mentioned one example: for technological reasons, Einstein could not possible have chosen to 'look first' at the bending of light by the sun before formulating his general theory of relativity. Some scientists may work in environments where the research strategies are mandated by supervisors or grant administrators.

It is also possible that some scientists are 'born theorists' or 'born empiricists' whose internal constitution dictates their research strategies. If the proportion of type i scientists is the same among theorists and empiricists, then the arguments of Section 2 apply.

Yet it is easy to envision situations in which research strategies do reveal abilities. Suppose, for example, that each scientist knows his own type, even though this information is not available to anyone else. For example, a scientist may be aware that he lacks a strong insight into a particular problem, making him type j in this instance even though he might be well known for being type i when he works on other problems.

Suppose also that society offers high rewards (money, prestige, etc.) for successful novel predictions, low rewards for unsuccessful novel predictions, and middling rewards for non-novel predictions. Then the talented type i scientists, knowing that they are often successful, might be willing to risk making novel predictions, while the less talented type j scientists would elect the safer course of looking first. In this case, the scientist's research strategy reveals his type, and this should be taken into account in assessing the probable truth of his hypothesis. Indeed, we would have $Pr(I\,|\,T) = 1$ and $Pr(J\,|\,L) = 1$, necessitating a modification of the results in Section 2.

At this point, it might seem that one could make any arbitrary assumption about the way society rewards its scientists, and thereby achieve any arbitrary conclusion about the import of novelty. To avoid such empty theorizing, we want to argue that the rewards themselves should be determined within the model, as the outcome of a competitive process among scientists and other members of society. This leads us to the realm of economic theory.

It has been suggested by others (see *e.g.* Nickles [1985]) that economics has a role to play in this analysis, but the tools of economics have not previously been brought to bear on the matter. Radnitzky [1987] attempted to demonstrate the gains from cost-benefit analysis in questions of methodology. While noting the possibility of a discrepancy between a researcher's private goals (*e.g.* 'an increase in one's reputation') and those of a benevolent social planner, he ignored this distinction in his analysis. By contrast, the conflict between individual and social welfare plays a key role in our own research, as we shall shortly indicate.

There is a general principle in economics to the effect that under certain conditions, competitive processes tend to lead to the same outcomes that would be chosen by a benevolent social planner seeking to maximize economic welfare. (We do not propose to explain here the precise meaning of economic welfare, but we want to emphasize that the term *has* a precise meaning that can be found in economics textbooks. See, for example, Chapter 8 of Landsburg [1989].) This tendency is sometimes expressed through the metaphor of an 'invisible hand' guiding the actions of individuals in a competitive economy to have in a way that promotes the general welfare. The existence of such an invisible hand is by no means axiomatic; it is a deep mathematical theorem that holds in a variety of economic models.

Invisible hand theorems are powerful tools in economic theory. It is often extremely difficult to calculate directly the behavior of a competitive system, but easy to calculate the optimal strategy for a benevolent social planner. The solution to the social planner's problem then reveals how the competitive economy will behave.

We have not been able to prove an invisible hand theorem in the context of our model of scientific research. However, perhaps overly optimistically, we have proceeded to solve the corresponding social planner's problem, hoping that this will provide insight into how a competitive system might function. Even in the absence of an invisible hand theorem, the solution to the planner's problem is of interest in its own right, as a potential source of advice to the institutions that guide science policy.

By way of example, we will describe an extremely simple version of a social planner's problem. Imagine a planner who would like to build a bridge, and is seeking a scientific theory to guide its design. If the theory is true, the bridge will stand and if the theory is false the bridge will fall. The planner can direct the activities of a fixed population of scientists, and can require them all either to look first or to theorize first. What should he do?

The advantage of 'theorize first' is that any theory it produces and which survives testing has an enhanced probability of being true, as was shown in Section 2. The disadvantage of 'theorize first' is that it might produce a theory that is ultimately rejected by the new observation, and is therefore useless.

If there are many researchers, each working independently and each theorizing first, then the probability that all of their theories will be rejected is very small. Therefore, since the planner requires only one surviving theory, he should order researchers to theorize first. This increases the probability that the bridge will be built and stand, at the cost of only a very small risk that no bridge can be built at all.

If there are relatively few researchers, however, then the planner might prefer either of his options, depending on the value of having a bridge that stands and the cost of having a bridge that falls. By ordering researchers to look first, he decreases the risk of being unable to build the bridge at all but increases the risk that the bridge will be built and collapse.

In the example just described, scientists are never given the opportunity to choose their own research strategies and thereby reveal private information about their abilities. In Kahn, Landsburg and Stockman [1992], we have posed and solved a far richer social planner's problem in which such revelation plays a central role.

In our formulation, the planner is allowed to set up different reward systems for different research strategies and to make a scientist's reward contingent on the outcome of his research. In so doing, the planner must balance several goals: first, he would like the choice of research strategy to reveal the scientist's type, so he can make more accurate estimates of whose theories to believe. Second, all other things being equal, he would prefer not to have many scientists theorizing first and then having their theories rejected, leaving nothing of scientific value. Third, he must consider the effects of his reward structure on the incentives to become a scientist in the first place. If the rewards are too small, there will be too little scientific activity. If the rewards are too great, people will be tempted into science even when they would be more socially useful in some other occupation (such as philosophy).

The precise formulation and solution of the social planner's problem is a rather intricate exercise in economic theory. We have carried out this exercise and reported our results in detail in Kahn, Landsburg and Stockman [1992]. Here we give a brief summary of our conclusions.

At the planner's optimum, all scientists are offered a choice between flat salary and a contingent fee that depends on the outcome of their research. (For example, the reward structures could differ at various types of research institutions, and scientists could be given a choice among institutions.) The rewards are structured so that type j scientists choose the flat salary, while type i scientists choose the contingent fee. Thus by choosing his reward structure, each scientist publicly reveals his type. Depending on the values of parameters like p, q, r and s, as well as the value of scientists in occupations other than science, type i researchers might be directed either to theorize first or to look first. Type j scientists look first.

Type j scientists are typically 'overpaid' in that their flat salary can far exceed the social value of their research. Nevertheless, it is in the planner's interest to overpay them, so that they are not tempted to choose the contingent fee, which would make it impossible to distinguish them from the type i scientists. This in turn would be detrimental to the planner's ability to distinguish hypotheses that are probably true from those that are probably false.

We believe that this solution bears some resemblance to our casual observations of the real world. For example, it is true that scientists can pursue careers either at high-powered research institutions where rewards are heavily contingent on outcomes, or at other institutions where salaries and advancement are distributed more equally. We are cautiously optimistic that further research will illuminate the relationship between the solution to the planner's problem and the outcome of the competitive process.

4 OBSERVATIONAL ERRORS

We have seen in Sections 2 and 3 that novelty can become important as a result of unobservable differences in scientists' abilities. In this section, we shall argue that novelty can become important for an entirely different reason: the possibility of errors in observation.

We return to the model of Section 1, in which the scientist's abilities are known in advance. He draws theories of types A, B and C with probabilities p, q and $1 - p - q$. Now, however, we drop the assumption that observations are always accurate. When the scientist makes an observation, he interprets it correctly with probability i and incorrectly with probability $j = 1 - i$ (thus the latters i and j are used differently here than in Sections 2 and 3).

Our goal is to demonstrate that in this situation, the result of Section 1 (novelty is irrelevant) is overturned. We shall do so by means of a counterexample. For purposes of constructing the counterexample, we need to make some additional assumptions.

First, we suppose that if a theory is of type A or B (so that a correctly interpreted observation is always consistent with the theory) then an *incorrectly* interpreted observation always appears *in*consistent with the theory. Second, we assume that if a theory is of type C (so that a correctly interpreted observation is never consistent with the theory) then the probability that an incorrectly interpreted observation appears consistent with the theory is given by some number x which is substantially less than 1. This assumption can be justified by the intuition that there are many ways to misinterpret an experiment, so that the chance of a given misinterpretation confirming a given false theory is rather small.

In the interests of precision, we can assume that type C theories are divided into N subtypes $C_1, C_2, \ldots, C_N$, each equally represented in the urn. There are N ways to misinterpret an experiment, the nth misinterpretation being consistent only with theories of type C_n. In that case, a given misinterpretation is consistent with a given type C theory with probability $x = 1/N$, which is small if N is large.

Let H be the scientist's hypothesis, O his observation, E the statement 'observation O appears to be consistent with hypothesis H', I the statement 'the observation was interpreted correctly', and J the statement 'the observation was interpreted incorrectly'.

Then

$$\begin{aligned} Pr(E \mid T \,\&\, H) &= Pr(E + I \mid T \,\&\, H) + Pr(E \,\&\, J \mid T \,\&\, H) \\ &= i + 0 \\ &= i \end{aligned}$$

and

$$\begin{aligned} Pr(E \mid T) = {} & Pr(E \,\&\, I \,\&\, H \mid T) + Pr(E \,\&\, I \,\&\, \textit{not } H \mid T) \\ & + Pr(E \,\&\, J \,\&\, H \mid T) + Pr(E \,\&\, J \,\&\, \textit{not } H \mid T) \\ = {} & ip + iq + 0 + xj(1 - p - q) \end{aligned}$$

so that

$$Pr(H \mid T \,\&\, E) = \frac{Pr(E \mid T \,\&\, H) \times Pr(H \mid T)}{Pr(E \mid T)} = \frac{ip}{i(p+q) + xj(1-p-q)}. \tag{11}$$

When x is small, the latter expression is approximately equal to $p/(p + q)$; that is, we approximately recover equation (5), which gives $Pr(H \mid T \,\&\, E)$ in a model where all observations are accurate. The reason is that the agreement between theory and observation implies that the observation is very likely to be accurate, so that we are almost certain to be in the situation described by the earlier model. Next we compute

$$Pr(I \,\&\, H \mid L \,\&\, E) = p/(p + q)$$

because in the event that I is true (which occurs with probability i), the 'look first' scientist discards all of the C balls and then has probability $p/(p + q)$ of drawing an A ball. (As in Section 1, the '& E' is redundant.) Also

$$Pr(J \,\&\, H \mid L \,\&\, E) = 0$$

since in the event that J is true, the scientist begins by discarding all of the A and B balls (as well as most of the C balls, namely those that are inconsistent with the misinterpreted observation).

Thus we have

$$(12) \qquad Pr(H \mid L \,\&\, E) = Pr(I \,\&\, H \mid L \,\&\, E) + Pr(J \,\&\, H \mid L \,\&\, E)$$
$$= ip/(p+q).$$

Comparison of (11) and (12) reveals a discrepancy between $Pr(H \mid T \,\&\, E)$ and $Pr(H \mid L \,\&\, E)$; the latter is approximately i times the former. Thus in this model, as in the model of Section 2, equation (6) is overturned and a novel confirmation is worth more than a non-novel confirmation.

The intuition here is as follows: by virtue of theorizing first and then achieving a novel confirmation, the scientist is more likely than otherwise to have made an accurate observation. Thus the confirmation is less likely to be spurious, and the hypothesis is more likely to be true.

5 CONCLUSIONS

Our purpose has been to demonstrate that the Bayesian approach to confirmation theory requires a precise specification of the process by which theories are generated. Under certain very strong hypotheses, we saw in Section 1 that the novelty of a confirmation does not affect its strength. By weakening various hypotheses, we saw in Sections 2 and 4 that novelty can be relevant, and we have been able to pinpoint precisely the sources of that relevance.

We have also argued that the very fact that a scientist chooses to make novel versus non-novel predictions may reveal some of his private information, and that it would be wrong to ignore that information in assessing the strength of a confirmation. We have discussed this briefly in Section 3, and extensively in another paper that is cited in the references.

One important feature of real-world science that is entirely absent from our models is this: sometimes the hypothesis that a scientist constructs can suggest an experiment to be performed. Thus the observation O that would be made by a 'theorize first' scientist is not the same as the observation O' that would be made by a 'look first' scientist. In our other paper, we have shown how to incorporate this effect into the models.

The reader can no doubt think of other aspects of scientific method that are ignored in our discussion. Indeed, all of our models, by virtue of being models, abstract from many important features of the real world. Nevertheless, we believe that only in the context of a carefully specified model is it possible to settle questions such as those we have been addressing. The models presented here can be taken as a first step toward a more detailed and realistic model capable of addressing subtler questions in the philosophy of science.

Department of Economics
University of Rochester
Rochester, NY 14627
USA

REFERENCES

CAMPBELL, R. and VINCI, T. [1983]: 'Novel Confirmation', *British Journal for the Philosophy of Science*, **34**.

GARDNER, M. [1981]: 'Predicting Novel Facts', *British Journal for the Philosophy of Science*, **32**.

HOWSON, C. [1984]: 'Bayesianism and Support by Novel Facts', *British Journal for the Philosophy of Science*, **35**.

KAHN, J., LANDSBURG, S. and STOCKMAN, A. [1992] 'The Positive Economics of Methodology', preprint.

LANDSBURG, S. [1989]: *Price Theory and Applications*. Chicago: The Dryden Press.

NICKLES, T. [1985]: 'Beyond Divorce: Current Status of the Discovery Debate', *Philosophy of Science*, **52**.

NIINILUOTO, I. [1984]: 'Novel Facts and Bayesianism', *British Journal for the Philosophy of Science*, **34**.

RADNITZKY, G. [1987]: 'The "Economic" Approach to the Philosophy of Science', *British Journal for the Philosophy of Science*, **38**.