NBER TECHNICAL WORKING PAPER SERIES

THE POSITIVE ECONOMICS OF METHODOLOGY

James A. Kahn

Steve Landsburg

Alan C. Stockman

Working Paper No. 82

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1989

We thank Jim Dolmas, Milton Friedman, Nicholas Row, and numerous colleagues at Rochester and elsewhere for informal (and often lively) discussions of this topic. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Technical Working Paper #82
November 1989

THE POSITIVE ECONOMICS OF METHODOLOGY

ABSTRACT

Does an observation constitute stronger evidence for a theory if it was made after rather than before the theory was formulated, when it may have influenced the theory's construction? Philosophers have discussed this question (of "novel confirmation") but have lacked a formal model of scientific research and incentives. The question applies to all types of research. One example in economics involves evaluating models constructed on the basis of VARs (where a researcher looks at evidence and then constructs a theory) versus structural models with formal econometric tests (where a model is constructed before some of the evidence on it is obtained).

This paper develops a simple model of scientific research. It discusses the issues that affect the answer to this question of the timing and theory-construction and observation or experimentation. We also address issues of social versus private incentives in the choice of research strategies, and of socially optimal rewards for researchers in the presence of information and incentive constraints.

James A. Kahn
Department of Economics
University of Rochester
Rochester, NY 14627
(716)275-5781

Steve Landsburg
Department of Mathematics
Colorado State University
Ft. Collins, CO 80523
(303)491-6573

Alan C. Stockman
Department of Economics
University of Rochester
Rochester, NY 14627
(716)275-7214

Suppose that you pick up the latest issue of the *American Economic Review* or the *New England Journal of Medicine* and find an article with three sections. The first section presents a theory that is consistent with some well-known facts. The second section presents an entirely new observation, discovered by the author of the article, that is also consistent with the theory. In the third section the author argues convincingly that he was not aware of the new observation at the time when he constructed his theory. Should the third section contribute to the degree of belief that you attach to the theory? Does an observation constitute stronger evidence for a theory if it was made after the theory was proposed rather than earlier, when it may have influenced the theory's formation?[1]

To put the issue another way: When a researcher has a body of data at his disposal, he can follow either of two research strategies. The first is to examine only a portion of the data before formulating a theory, and then use the remainder of the data to test the theory. The second is to examine all of the data and then construct a theory that fits. We refer to these as the "theorize first" strategy and the "look first" strategy. For example, some researchers in macroeconomics estimate vector autoregressions and then determine which prevailing theories are and are not consistent with their findings. Others prefer to develop a "structural" model and then test it.

---

[1]Musgrave (1974, 1978) poses the question in the following way. Suppose that two theoriests, A and B, devise identical hypotheses. A's hypothesis is constructed so as to account for known facts e_1 and e_2, while B's hypothesis—owing to his unfamiliarity with recent results—is devised to account for e_1 alone. Does e_2 then lend corroborative support to the hypothesis as proposed by B, but not as proposed by A? Less realistically, suppose that the sinister researcher C, out of whatever personal motivation, begins a campaign to end the careers of A and B. Is it possible that he can succeed merely by anonymously mailing to his rivals observations detailing hitherto unknown facts, thus effectively silencing them? Can researchers be made worse off as their background knowledge grows?

We think of the first type as "looking first" and the second type as "theorizing first". More generally, given a set of data to be explained and lists of exogenous and endogenous variables, a researcher could either estimate a reduced form before selecting among those theories that are consistent with it, or alternatively he could first propose an (identified) theory and then test it on the reduced form. Under what circumstances, and in what sense, does it matter which strategy the researcher pursues? How should society structure rewards to scientists to induce them to choose the right strategies?

In the literature of the philosophy of science, there has been a considerable debate on the subject of "novel confirmation". Novel evidence for a theory is evidence that is obtained *after* the theory is proposed. In a survey of this literature, Campbell and Vinci (1983) write that

> Philosophers of science generally agree that when observational evidence supports a theory, the confirmation is much stronger when the evidence is 'novel'.

Nevertheless,

> The notion of novel confirmation is beset by 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.

Moreover, while it has been suggested that economics has a role to play in this analysis (see for example Nickles, 1985), the tools of economics have not previously been brought to bear on the matter.

Much of the philosophical literature relies on various forms of Bayesian analysis. For example, given a hypothesis H, evidence e, and background information b, one Bayesian measure of the support that e lends to H is

$$\text{Degree of support} \equiv P(H|e \,\&\, b) - P(H|b).$$

Other authors (e.g. Howson (1984)) have proposed alternative Bayesian measures of the degree of support. One problem with these debates is that the definition of degree of support appears arbitrary. What basis can there be for preferring one definition to another?

It is our position that this question cannot be satisfactorily addressed in the absence of an explicit model of the process by which theories are generated.[2] Our section 1 provides such a model. When scientists have access to private information, we also require an explicit model of the process by which scientists select their research strategies. The reason for this is that the choice of research strategy can reveal some of the scientist's private information. In the simple models of Sections 1 and 2, we take the research strategies to be determined exogenously. In later sections, we assume more realistically that scientists respond to an existing incentive structure. With these preliminaries in place, we return to the question of how to choose a definition for the degree of support. Our solution (beginning in Section 3) is to examine a well-defined social planning problem and ask what information would be useful to the planner. This, of course, is simply the standard economic approach to such a problem.

1. A Simple Model

Science progresses through a sequence of theories and observations. We present in this section a very simple model that allows us to address the

---

[2]In a recent paper titled "The 'Economic' Approach to the Philosophy of Science", Gerard Radnitzky (1987) attempted to demonstrate the potential gains from applying cost-benefit analysis to rational theory-preference. Radnitzky notes that researchers' private goals (e.g. "an increase in ones' reputation") may differ from those of a benevolent social planner, but he ignores this distinction in his analysis. In our model, the distinction leads to incentive-compatibility constraints.

issues of interest even though it abstracts from many features of real-world science. We make the following assumptions in this simple model. A theory is a set of statements that predicts that under certain circumstances, certain events will occur. We call a theory true if its predictions are always accurate.[3] We assume that a single researcher works on a problem, and that the researcher has time to do exactly two things: make one new observation and construct one new theory.[4] The observation can be thought of as the result of an experiment. For now, we abstract from the choice among experiments and take the experiment performed as exogenous. In Section 3 we will show how to relax this assumption. Experiments that have been performed in the past, and theories that have been constructed in the past, guide the researcher in ways to be discussed below.

We assume that theories can be divided into four mutually exclusive sets, each with infinitely many elements. Theories of type D are inconsistent with previous observations. Theories of type C are consistent with previous observations but inconsistent with the new observation to be made. Theories of types A and B are consistent with all the old observations and with the new observations to be made. However, theories of type A are true and theories type B are not.[5]

---

[3]We abstract from several things here. Instead of truth, we could speak about usefulness, or degrees of usefulness, so that one theory is more useful than another if its predictions have a higher probability of being accurate, or are closer to being accurate in some appropriate sense. Alternatively, a theory could predict not an event but a probability distribution over events.

[4]More generally, we could assume that he has a certain amount of time available for various activities, and chooses how much time to devote to research. He may then choose various amounts of time to allocate to making observations and to constructing theories, and he might devote all his time to only one of these tasks. Our assumptions are made solely in the interest of simplifying the problem.

[5]We make the simplifying assumption that the only true theories are those consistent with all the observations. This involves two simplifications.

This categorization of theories can be illustrated by considering the identification problem in econometrics. Suppose a government is considering some kind of policy change. Prior to any new data analysis there is enough information to rule out some models' implications. These are the type D models. Given the available data, reduced form estimation can rule out some other models (type C), but cannot distinguish between some others (A and B). However, once a theory is selected (and perhaps used as a guide to policy), subsequent events will make it possible to distinguish type A models from type B.

We want to emphasize that it is in no way our intention to take a position on the philosophical nature of "truth". We simply define a true theory, for the purposes of this paper, to be one whose predictions are always accurate. Some readers will object that by this definition, no real-world theory is ever true. Again, what we really have in mind is a notion of usefulness.

Researchers construct theories in a way that involves some fundamental randomness. We assume that theory-construction is analogous to drawing balls from an urn, with the balls representing the possible theories. Observations made in the past and theories constructed in the past affect the number of balls of each type in the urn.

One immediate issue is whether a researcher can "build in" previous observations when he constructs a theory. In other words, will a researcher

---

First, some theories may be useful even though they are inconsistent with facts. Theories of how to build a bridge, for example, may be based on Newtonian mechanics even though Newtonian mechanics is inconsistent with some facts. We abstract from this possibility in our simple model, though the model could easily be modified to incorporate it. Second, some observations may involve errors (perhaps of interpretation): experiments can involve mistakes, and statistical work can be subject to important sampling error. We abstract from that possibility in this section but we discuss this issue in section 2.2.

ever construct a type D theory? We will assume that researchers can build in previous observations when they construct theories.[6,7] Then we can assume that there are no type D balls in the urn, i.e. the probability that the researcher will construct a type D theory is zero. (The Appendix discusses the other case in which researchers cannot build in previous observations when they construct theories.)

Let p, q, and 1−p−q be the probabilities that a researcher constructs a theory of type A, B, or C. Generally, the sizes of p and q depend on the results of past observations and past theorizing as well as the quality of the researcher and the difficulty of the problem.

We consider two research strategies, theorize-first and look-first. In this section we regard the research strategy as exogenous; we examine the optimal choice of research strategy in sections 3 and 4. Suppose a researcher theorizes first — he constructs a new theory, then performs an experiment (or makes a non-experimental observation).[8] The researcher's

---

[6]An alternative interpretation of this assumption might be that a researcher can keep constructing theories until he comes up with one that is consistent with previous observations. That is, he can costlessly discard a type D theory and choose another one. This is not the interpretation that we pursue.

[7]We also ignore incentive problems, assuming that researchers do in fact choose to build in previous observations. Such incentive problems are explicitly addressed in later sections of the paper.

[8]We have not modeled the process by which an investigator chooses experiments. The probabilities of the various outcomes of this process (i.e. p, q, etc.) are simply taken as primitive. Nonetheless these probabilities can be given richer interpretations. Rather than treating the experiment or data as given, we can assume that there are many possible experiments. Each experiment can be associated with probabilities of outcomes. We assume that no experiment can refute a true theory, but experiments do not necessarily reject a false theory.

For a researcher who theorizes first, let $F(\theta)$ be the distribution function over theories, and let $\rho_k(\theta)$ be the probability that experiment k *rejects* theory θ. If theory θ is true ($\theta \in A$), then $\rho_k(\theta) = 0$ $\forall k$ by assumption. Hence p is still just the probability the researcher chooses a true theory,

theory is true with probability p. The theory is consistent with the new observation if and only if it is type A or type B, which occurs with probability p+q. So the probability that the theory is true conditional on its being consistent with the new observation is p/(p+q).

Suppose alternatively that a researcher looks first — he makes the new observation and then constructs a new theory to be consistent with it. Because he can build all observations into his theory and he seeks a true theory, all balls marked C are removed from the urn prior to his drawing. So the conditional and unconditional probabilities that his theory is type A are both p/(p+q).

Our first result, then is a *neutrality result*. In this model, *the probability that a new theory is true, conditional on its being consistent with all the (old and new) observations, is independent of the research strategy.* This result should not be surprising: There is nothing in classical

---

i.e. $\int_{\theta \epsilon A} dF(\theta)$. If θ is not true, then $\rho_k(\theta)$ is equal either to zero or one depending on θ. (More generally $p_k(\theta)$ could be anywhere in the [0,1] interval, but this would complicate the analysis under looking first). We assume that some process (that we do not model) leads the researcher to adopt experiment k with probability π_k independent of θ (in Section 4 we will relax this assumption). Then $q = (1-p)\int_{\theta \notin A} \sum_k \pi_k[1-\rho_k(\theta)]dF(\theta)$. Thus whether a particular theory ends up as type B or type C may depend on the experiment chosen to test it. Later when we let p and q differ across researchers this gives us the interpretation that researchers differ in the ability both to choose correct theories and to choose informative experiments.

For a researcher who looks first, the interpretation in this framework is that the researcher first chooses an experiment (according to the distribution π_k). The experiment points the way toward a particular set of possible theories. Let $\Omega_k = \{\theta | \rho_k = 1\}$. The researcher chooses one theory out of this restricted set according to the distribution $F_k(\theta) = F(\theta)/\int_{\Omega_k} dF(\theta)$, $\theta \epsilon \Omega_k$, zero otherwise. Hence the probability that the researcher who looks first comes up with a true theory is $p/(p+q) = \sum_k \pi_k \int_{\theta \epsilon A} dF_k(\theta)$.

statistical inference that allows the timing of observation relative to the formation of a hypothesis to matter. The data themselves are a sufficient statistic.

2. Modifications of the Simple Model

2.1. Private Information about Researchers' Characteristics

We now modify the model in a way that makes the research strategy matter. Assume there are two types of researchers, type i and type j. Only the researcher knows his type, but everyone knows that the population of researchers consists of a proportion i who are type i and j = 1-i who are type j. Researchers differ only in their probabilities, given by the following table.

Theory Type	Type i researcher	Type j researcher
A	p	r
B	q	s
C	1-p-q	1-r-s

We make the following assumptions throughout the paper:

$$p + q > r + s \tag{2.1}$$

$$p/(p+q) > r/(r+s) \tag{2.2}$$

Condition (2.1) says that if a researcher theorizes first, he is more likely to construct a a theory that is consistent with the new observation if he is type i. Condition (2.2) says that if a researcher looks first, he is more likely to construct a true theory if he is type i.

Suppose now that a given researcher looks first and then constructs a theory consistent with the observations. If he is type i, he constructs a theory of type A with probability p/(p+q), and if he is type j, he selects a theory of type A with probability r/(r+s). Someone who does not know the researcher's ability calculates that the theory is true with probability

$$\gamma = i \cdot \frac{p}{p+q} + j \cdot \frac{r}{r+s} . \qquad (2.3)$$

Suppose alternatively that the researcher theorizes first and produces a theory that survives testing, that is, it turns out to be consistent with the new observation. The fact that the theory survives testing conveys information about the researcher's type. The updated probability that he is type i is

$$i' = i \cdot \frac{p+q}{i\cdot(p+q) + j\cdot(r+s)} . \qquad (2.4)$$

Let $j' = 1 - i'$ be the updated probability that the researcher is type j. Condition (2.1) implies that $i' > i$. The probability that the new theory is true is

$$\gamma' = i' \cdot \frac{p}{p+q} + j' \cdot \frac{r}{r+s} . \qquad (2.5)$$

Since $i' > i$, condition (2.2) guarantees that $\gamma' > \gamma$. This expression is greater than the expression (2.3) for probability of truth under the look-first regime. Thus the neutrality result of Section 1 is overturned and we have:

Assume that the researcher's characteristics are unknown, and that his theory is consistent with the new observation. Then the probability that the theory is true, conditioned on its being consistent with all the observations, is greater if it was produced under a "theorize-first" strategy than if it was produced under a "look-first" strategy.

2.2. Alternative Interpretations and Discussion

The model of Section 2.1 allows alternative interpretations, some of which we pursue here.

2.2.1. Sampling Error

Suppose that there is only one type of researcher, who sometimes interprets his experiment incorrectly or whose experimental results are tainted by sampling error. With probability 1-i, he misreads the experimental conclusion, and so mistakenly believes that his experiment is consistent with theories of type C but inconsistent with theories of types A and B.

We want to admit the possibility that a misread experiment is compatible with some kinds of type C theories but not others. Therefore, we subdivide the class of type C theories into subclasses $C_1,\ldots,C_N$ where N is large. For concreteness, assume that a researcher who constructs a type C theory is equally likely to choose any of the types $C_1,\ldots,C_N$. We assume also that there are N different ways to misinterpret the experiment, the n^{th} misinterpretation being consistent only with theories of type C_n. Each misinterpretation is assumed to be equally likely, so that any given misinterpretation occurs with probability $(1-p-q)/N$.

Suppose that a researcher theorizes first. We compute the probablity that his experiment appears to support his theory. With probability $i(p+q)$, the researcher constructs a theory of type A or B and correctly interprets his experiment as confirming the theory. With probability $(1-i)(1-p-q)/N^2$, the researcher constructs a theory of type C_n and misinterprets his experiment so as to support the theory. Thus the probability that the theory and the experiment appear compatible is

$$i(p+q) + N\,\frac{(1-i)\,(1-p-q)}{N^2} = i(p+q) + \frac{(1-i)(1-p-q)}{N} \qquad (2.6)$$

Therefore, if after theorizing first a researcher's observation and theory agree, the updated probability that he is type i is

$$i' = \frac{i(p+q)}{[i(p+q) + (1-i)(1-p-q)/N]} \,. \qquad (2.7)$$

The probability that the new theory is true is $i'p/(p+q)$. When N is large, i' is close to one. This is because it is very unlikely that a researcher who theorized first would construct an incorrect theory and misinterpret his experiment <u>in just the right way</u> as to lend his theory support. So if a researcher theorizes first and his experiment is consistent with his theory, the updated probability that he interpreted the experiment correctly, i', exceeds the original probability, i. This raises the probability that his theory is true from

$$\frac{ip}{(p+q)} \qquad (2.8)$$

before knowing the outcome of the experiment to

$$\frac{i'p}{(p+q)} \tag{2.9}$$

after observing the consistency of the experiment with the theory.

Now suppose a researcher looks first. Then the probability that his theory is true is ip/(p+q). Since i' exceeds i (which is less than one) for large N, we find that the possibility of experimental or sampling error implies the following result: *Conditional on its consistency with all the observations, a theory constructed under a theorize–first strategy is more likely to be true than one constructed under a look–first strategy.* That is, even when all scientists are alike, the possibility of sampling error overturns the neutrality result of section 1.

The italicized result does depend on specific assumptions about the probabilities: We assumed that the theory types C_n are all equally likely and that the various experimental misinterpretations are all equally likely. The same result could be obtained under much more general assumptions. It is possible, however, to specify the probabilities in such a way that the agreement between theory and experiment is strong evidence that the experiment was misinterpreted. In such an example, a theorize–first theory that is supported by evidence is actually less likely to be true than a look–first theory supported by the same evidence.[9]

[9]This sometimes appears to be the case in economics: If evidence supports a theory, this fact alone makes economists suspicious of the evidence or its interpretation.

2.2.2. Moral Hazard and Related Issues

We can interpret a type i researcher as one who has exerted additional effort, at some cost to himself, to develop intuition into the phenomenon he is investigating. (This contrasts with the interpretation of section 2.1, where researchers are endowed with the given abilities.) If the effort is unobserved by the reader, then equations (2.3) and (2.5) apply.

Under this interpretation, as well as those of Sections 2.1 and 2.2.1., the difference between expressions (2.3) and (2.5) measures the extent to which the reader should discount evidence that was examined by the researcher prior to theorizing.[10] This "pretest discount" depends on the parameters i, p, q, r, and s, which describe the scientific community as a whole.

So far, we have assumed that researchers honestly report all of their work. Suppose instead that there are several experiments from which researchers can costlessly and randomly choose, and that researchers do not always report all of the experiments they performed. In the context of the model of section 2.1, think of a single type of researcher and an experiment that is subject to sampling error or other error, as discussed above. A researcher who theorizes first might continue experimenting until he obtains a result consistent with his theory, and report only this result. It is interesting to note that if a researcher who theorizes first follows this practice of repeating the experiment until it yields results consistent with his theory, and if he reports only this result, then the research strategy

---

[10]The only treatment of a related problem that we have found is in Leamer (1976, Chapter 9), who addresses the different question of whether research strategies affect the researcher's posterior probability distribution because they reflect his prior distribution.

yields no information to the reader: the neutrality result from section 1 is obtained.

2.2.3. Non-empirical Criteria for Judging Theories

The models discussed above do not deal explicitly with theoretical criteria for evaluating theories. It is easy to reinterpret the model to allow for this by reinterpreting the observation in the models of section 1 as an "a priori criterion." One can think of this criterion as indicating whether a theory is consistent with other theories in the discipline. A researcher who looks first builds this consistency into his theory; one who theorizes first checks for this consistency afterwards.

3. A Social Planner's Problem

Until now, we have treated the choice of research strategy as exogenous. As a first step towards endogenizing the decision, we consider the problem of a social planner who seeks to maximize social surplus and who can mandate research strategies. In subsequent sections, the planner will set rewards for researchers, and the researchers themselves will choose their strategies.

Suppose the social planner would like to build a bridge. He asks his researchers to produce a theory of bridge building to guide the construction. If the bridge stands, then the planner receives utility $G > 0$. If it collapses, he receives utility $L < 0$. If he elects not to build the bridge at all, he receives zero utility. A true theory tells how to build a bridge that will stand. We return to the model of Section 2.1, with type i and j researchers and private information about type. We ask which research strategy the planner should command the researchers to use.

The advantage of "theorize first" is that any nonrejected theory it produces has an enhanced probability of truth. The disadvantage is that it might produce a theory that is rejected by the new observation, and is therefore useless. If many researchers work independently on the same problem, the chance that everyone will produce a rejected theory is very small. We will show that "theorize first" is the optimal strategy when there are many independent researchers, but not necessarily when there are few researchers.

Assume that researchers look first. Every researcher picks independently a theory consistent with all of the observations, and the planner selects one of those theories at random. The theory is true with probability γ as defined in equation (2.3), and the expected utility to the planner if he builds a bridge using this theory is

$$\gamma\ G + (1 - \gamma)\ L\ . \tag{3.1}$$

The planner builds the bridge if and only if this expected utility is positive. So when researchers look first, the planner's expected utility is

$$\mathrm{Max}(\gamma\ G + (1 - \gamma)\ L,\ 0). \tag{3.2}$$

This result is independent of the number of researchers.

Now suppose the planner tells researchers to theorize first. Each researcher selects a theory, then tests it against the new observation. The social planner chooses randomly among those theories (if any) that survive testing. There are two extreme cases, (a) one researcher, and (b)

sufficiently many researchers to virtually guarantee that at least one theory survives testing.

With many researchers, any of the surviving theories is true with probability γ' as defined in equation (2.5). The expected utility from using this theory is

$$\text{Max}\{\gamma' G + (1 - \gamma') L, 0). \tag{3.3}$$

The expression (3.3) exceeds (or equals) (3.2) because $\gamma' > \gamma$, so the planner prefers that all researchers theorize first. So when there are many researchers with unobservable characteristics and a single research project, they should all theorize first.

The case of a single researcher is quite different. When a single researcher of unknown characteristics looks first, the planner's expected utility is given by (3.2). If he theorizes first, there are two possibilities. With probability

$$i (1 - p - q) + j (1 - r - s)$$

his theory is rejected and no bridge is built. With probability

$$i (p + q) + j (r + s)$$

his theory is not rejected and so has probability γ' of being true. So the planner's expected utility is

$$[i (p + q) + j (r + s)] \text{ Max } \{\gamma' G + (1 - \gamma') L, 0\}, \tag{3.4}$$

which can be either greater or less than (3.2). So either strategy might be preferred.

Keynes called theories "induction" and observations "cases;" he wrote in his Treatise on Probability:

> If all our inductions had to be thought before we examined the cases to which we apply them, we should, doubtless, make fewer inductions; but there is no reason to think that the few we should make would be any better than the many from which we should be precluded.

The idea that non-novel evidence offers zero support for a hypothesis, as Nickles (1985) notes, renders efficiency decisions quite complex. "For scientists must now balance efficient testing against efficient generation. . . . And will this loss of generational efficiency—a problem that has always plagued Popper's view of science—be offset by gains at the testing and justification stage?" This is the tradeoff at issue in comparing (3.2) and (3.4). Experiments with parameter values indicate that there is no simple characterization of when one strategy or the other is superior. Note, however, that when i is either 0 or 1, γ' equals γ and (3.4) is less than (3.2), so looking first is preferred to theorizing first. This is because the researchers' characteristics are known in advance, so a successful test conveys no information.

In a multiperiod model the planner might expect to use the researcher's services on future projects. Then theorizing first becomes more attractive because information gained about the researcher's type can be used repeatedly. In a multi-period model, there is a range of parameter values that imply that researchers should theorize first when they are young in order to signal their abilities, and then look first when they are old,

because by then their abilities have been largely revealed, leaving no reason to waste resources producing theories that are immediately rejected.

Finally, note that information about a researcher's type has no value to the planner if

$$rG + sL > 0. \tag{3.5}$$

In that case, the planner would use an unrejected theory to build the bridge even if he knew that the theory was constructed by a type j researcher. The theorize-first research strategy then conveys no useful information, but has a cost because it may produce a rejected theory. So, in this case, the planner prefers the look-first strategy. On the other hand, if $rG + sL < 0$, then the planner would not build the bridge with a nonrejected theory if he knew it was constructed by a type j researcher, so information about the researcher's type is valuable. If this information is sufficiently valuable, then the planner prefers the theorize-first strategy.

4. Theories Suggest Experiments

Until now, we have assumed that there is only one feasible new observation for a researcher to make. In so doing, we have ignored the possibility that researchers have a choice of several possible experiments to perform, and that theories may help guide the choice of experiments by suggesting which ones are most useful.

In this section we return to the simpler environment of a single researcher of a known type. We examine a very simple model that incorporates both the choice of research strategy and the choice of experiment. To further simplify the issue, we assume that there are only two kinds of

theories that are consistent with the previous observations. Type A theories are true and type B theories are not. There are also two feasible experiments that the researcher could perform. Type A theories suggest an experiment E_A, which tests an implication of type A theories, and type B theories suggest an experiment E_B, which tests an implication of type B theories. Because type A theories are true, we assume that the experiment E_A never rejects its implications.[11] The experiment E_b rejects the implications of type B theories with probability ρ. A social planner who maximizes expected utility gets gain G from using a true theory, loss L from using a theory that is not true, and zero utility if he takes no action at all.

A researcher who theorizes first constructs a type A theory with probability p and a type B theory with probability $q = 1-p$. So his theory is true with probability p, not true but not rejected with probability $q(1-\rho)$, and rejected with probability $q\rho$. Then the planner's expected utility if the researcher theorizes first is

$$\text{Max } \{pG + q(1-\rho)L, 0\}. \tag{4.1}$$

A researcher who looks first chooses randomly between the two feasible experiments. Let $1-\pi$ and π denote the probabilities that the researcher chooses experiments E_A or E_B. If he chooses E_A, then he makes an observation that is consistent with both theories. In that case, he constructs a true (type A) theory with probability p. If he chooses E_B, then with probability ρ, the observation rules out type B in which case he constructs a true theory

[11]As noted in an earlier footnote, we could easily extend our model so that useful theories have some implications that are rejected. This would not affect the main issues we focus on here.

with probability 1. With probability $1-\rho$, the observation from E_B fails to rule out any theory, in which case the researcher constructs a true theory with probability p. So the a priori probability that he constructs a true (type A) theory is

$$(1-\pi)p + \pi\rho + \pi(1-\rho)p = p + q\pi\rho. \tag{4.2}$$

The probability that he constructs a type B theory is $1-\{p + q\pi\rho\} = q(1-\pi\rho)$. So the planner's expected utility if the researcher looks first is

$$\max \{ (p + q\pi\rho)G + q(1-\pi\rho)L, 0\}. \tag{4.3}$$

Several points are worth noting. First, as in the previous sections of the paper, the probability of coming up with a true theory is greater under looking first than under theorizing first. Looking does provide information, at least with some probability, that can be used to help select among theories. On the other hand, the probability of coming up with a false theory is *also* greater under looking first, because the experiment chosen is not expected to be as informative.

It is remarkable that neither p nor ρ plays any role in determining which strategy is preferred. In fact theorizing dominates looking if and only if $\frac{G}{|L|} < \frac{1-\pi}{\pi}$. A smaller (absolute) value of L relative to G makes looking first more attractive, whereas theorizing first will be preferred when failure is very costly. This accords with the intuition that if the primary goal is to come up with a true theory (i.e. $G/|L|$ is big) then one should look at all the data, whereas if one is more concerned about not believing a false theory ($G/|L|$) is small), one should theorize first.

The basic point is that in a setting where the costs of proceeding on the basis of a bad theory are high relative to the benefits from successful research, theorizing first should be encouraged, whereas if the benefits from success are high relative to the costs from a mistake, looking first is preferred. An application of this principle might be the debates over activist discretionary macroeconomic policy versus fixed rules. Proponents of the latter argue that discretionary policies aimed at "fine-tuning" based on observed regularities in data have small potential benefits and large potential costs (e.g. the Great Depression), and therefore that activist policymaking should await the outcome of theorizing first and testing the theories, even if it means foregoing the benefits of stabilization in the meantime. Proponents of activist policy argue, on the other hand, that the costs of doing "nothing" are significant, while the risks are not that great given the ability to react to any mistakes that might arise. Thus perhaps it is no coincidence that those who tend to argue in favor of the slower process of theorizing and testing (e.g. Lucas in "Understanding Business Cycles") also are likely to believe that the costs of fluctuations are not that great.

5. Mechanism Design by a Social Planner

An unresolved issue from section 3 is whether a social planner would choose to have all researchers look, all theorize, or perhaps signal their private information in a separating equilibrium. The discussion in Section 3 clearly indicates that either of the two symmetric allocations ("all look" and "all theorize") could be preferred, depending on the values of the parameters p, q, r, s, and i. We now analyze whether the social planner would prefer a separating equilibrium to the better of the pooled equilibria, and whether the choice of research strategy plays any important role in the

separation. In order to analyze welfare issues we also generalize the model by giving agents an alternative activity, thereby making endogenous the number of agents of each type that choose to engage in research. We assume that agents are risk-neutral, that agents' types are private information, and that payments to researchers must be non-negative (though any lower bound that is a binding constraint would suffice).

Note that if the planner is able to make large lump sum transfers to every agent in the economy, then the non-negativity constraint is non-binding. The reason is that the planner could announce a lump sum payment to everyone except certain researchers; this is equivalent to giving those researchers a negative reward. Thus we assume that while the planner can transfer income in moderate amounts (enough to make appropriate payments to all researchers, who constitute a small part of the economy), he can not make massive transfers.

The specifications of project and agent characteristics are the same as in Sections 1 and 3. Now, however, we suppose a linear upward-sloping supply of each type of agent into the research market, with each agent's decision determined by his opportunity cost. The distribution of opportunity costs across agents is such that to get, for example, M type i agents and N type j agents into research requires that type i agents expect to receive αM and type j agents expect to receive βN.

The question is what sort of mechanism brings about the most desirable allocation of resources. There are four basic scenarios: the planner could choose an allocation in which

i. All researchers theorize first.

ii. All researchers look first.

iii. type i researchers theorize first and type j researchers look first.

iv. type i researchers look first and type j researchers theorize first.

The social planner wants to build a bridge, as in section 3, and has the payoffs G and L discussed in that section. The social planner chooses a reward structure to induce the optimal number of each type of agent into research (this may be zero in the case of type j agents), and to induce agents to choose the most effective research strategies. However, he must take into account the non-negativity constraints on rewards. In addition, agents' types are private information, which implies certain incentive-compatibility constraints.

The setup is as follows: a social planner announces a reward structure that consists of non-negative contingent payments to researchers depending on their *announced* type, their research strategy, and on the outcome of their research including whether bridges built with these theories actually stood or collapsed. In this section we continue to assume that the planner can verify whether an agent actually looked or theorized first. Each agent decides whether to engage in research and, if so, what to announce as his type and what strategy to pursue. The research takes place, bridges are built, they either stand or collapse, and contingent payments are distributed to researchers accordingly.

Our assumption that researchers literally announce their types may appear somewhat artificial, but we invoke the so-called Revelation Principle to argue that any allocation achievable by indirect announcements is also achievable by direct announcements. In this way we avoid taking a stand on

what particular device or institution is used in practice as a sorting device. The point is that the reward structure is designed so that agents sort themselves, and therefore have no incentive to misrepresent their types; the simplest way to model this is just to have agents announce their types, and to have the rewards satisfy incentive-compatibility constraints.

The planner sets the following general reward structure: Researchers who claim to be type k (k=i or j) receive y_A^k for submitting a theory of type A, y_B^k for submitting a theory of type B, and y_C^k for submitting a theory of type C, in which case no bridge is built. So, for example, in a particular allocation (A) a type i researcher who theorizes first gets an expected reward of $py_A^{iA} + qy_B^{iA} + (1-p-q)y_C^{iA}$, while a type j researcher who looks first gets $(ry_A^{jA} + sy_B^{jA})/(r+s)$.

Let $\bar{y}_k$ denote the expected reward to a type k agent if he goes into research and pursues the most rewarding strategy. Then if the agent's opportunity cost is z, he will choose to engage in research if $\bar{y}_k > z$. Any incentive-compatible social planner's allocation rule implies values for $\bar{y}^i$ and $\bar{y}^j$, and thereforea supply of $\bar{y}^i/\alpha$ type i researchers and $\bar{y}^j/\beta$ type j researchers. The rule also implies values for the expected social gain from a researcher's activities, which we denote by $\bar{U}_k$ (k=i,j). The social planner chooses an allocation rule to maximize

$$(\bar{y}^i/\alpha)\bar{U}^i - \bar{y}^{i2}/2\alpha + (\bar{y}^j/\beta)\bar{U}^j - \bar{y}^{j2}/2\beta, \qquad (5.1)$$

which represents the total welfare gain from research activity net of private opportunity costs. The planner faces two sets of potentially binding constraints. First, the non-negativity constraints on the payments prevent

the achievement of the first-best through large negative penalties for failures. Second, the incentive-compatibility constraints require that agents have no incentive to lie about their type.

5.1. Information About Type is Valuable

In this section we will assume that $rG + sL < 0 < pL + qG$ so that bridges based on theories of type j researchers are not worth building, and the best number of type j agents in research is zero. Under this assumption we will compare the benefits of various allocations of research types to strategies induced by appropriate reward structures. In Section 5.2, we will repeat this exercise under the alternative assumption that $0 < rG + sL < pG + qL$, so that a bridge is worth building even if the theory was known to be constructed by a type j researcher.

Our analysis consists of two stages. First, we examine the social planner's optimal choice of a reward structure conditional on his decision on the optimal research strategies. We consider each of the four scenarios listed above for research strategies. Then we optimize over the choice of research strategies.

Our first result is this:

Suppose that $rG + sL < 0 < pG + qL$, *so that information about a researcher's type is useful to the planner. Suppose also that the planner can verify the research strategy. Then:*

a) The planner offers researchers a choice between a flat salary and a contingent fee.

b) Type i researchers choose the contingent fee.

c) Type j researchers choose the flat salary.

d) The planner requires researchers who choose the contingent fee to theorize first if $W_T^1 > W_L^1$, *where*

$$W_L^1 = \frac{\left[\frac{p}{p+q}\right]^2}{2\alpha}\left[\frac{\beta\left[\frac{p}{p+q}\right]^2}{\beta\left[\frac{p}{p+q}\right]^2 + \alpha\left[\frac{r}{r+s}\right]^2}\right](G + \frac{q}{p}L)^2. \tag{5.2}$$

$$W_T^1 = \frac{p^2}{2\alpha}\left[\frac{\beta p^2}{\beta p^2 + \alpha r^2}\right](G + \frac{q}{p}L)^2 \tag{5.3}$$

e) The planner allows researchers who choose the flat salary to use either research strategy. He does not use their theories to build bridges.

f) If $W_T^1 > W_L^1$, *then researchers of type i are paid*

$$y_A^i = \frac{\beta p^2}{\beta p^2 + \alpha r^2}(G + \frac{q}{p}L) \tag{5.4}$$

if they construct a theory of type A and

$$y_B^i = y_C^i = 0 \tag{5.5}$$

otherwise. The planner's expected utility is W_T^1.

g) If $W_L^1 > W_T^1$, *then type i researchers are paid*

$$y_A^i = \frac{\beta\left[\frac{p}{p+q}\right]^2}{\beta\left[\frac{p}{p+q}\right]^2 + \alpha\left[\frac{r}{r+s}\right]^2}(G + \frac{q}{p}L) \tag{5.6}$$

if they construct a theory of type A and

$$y_i^B = y_i^C = 0 \tag{5.7}$$

otherwise. The planner's expected utility is W_1^L.

h) The flat salary chosen by type j researches is

$$y_j^A = y_j^B = y_j^C = ry_i^A / (r+s). \tag{5.8}$$

i) These are too few type i agents and too many type j agents relative to the first best.

These results follow easily from our two-stage procedure. Suppose first that the planner wants all researchers to look first. Call this Allocation L.

In this case the planner does not use the research strategy as a sorting device. All researchers look first, but they still may self-select according to the reward structure. The incentive-compatibility constraints are

$$(py_A^i + qy_B^i)/(p+q) \geq (py_A^j + qy_B^j)/(p+q) \tag{5.9}$$

$$(ry_A^j + sy_B^j)/(r+s) \geq (ry_A^i + sy_B^i)/(r+s) \tag{5.10}$$

These conditions imply that agents will not have any incentive to lie about their types. Once the planner can identify the agents' types, he discards the theories of type j agents because bridges built with them have negative value. Consequently the planner never learns whether a type j agent's theory

is useful. This imposes the additional constraint that type j agents are paid a flat fee,

$$y_A^j = y_B^j \tag{5.11}$$

The gross social gains from research activities (ignoring agents' opportunity costs) are then

$$\bar{U}^i = (pG + qL)/(p+q) \tag{5.12}$$

$$\bar{U}^j = 0 \tag{5.13}$$

Without the non-negativity constraint the optimum would be to have $\bar{y}^i = \bar{U}^i$ and $\bar{y}^j = \bar{U}^j$, which implies $y_A^i = y_A^j = G$, $y_B^i = y_B^j = L$. The constrained optimum requires $y_B^i = 0$ because the non-negativity constraint is binding and

$$y_A^j = y_B^j = ry_A^i/(r+s) \tag{5.14}$$

because the incentive-compatibility constraint (5.10) is binding. So

$$\bar{y}^i = \frac{p}{p+q}\, y \tag{5.15}$$

$$\bar{y}^j = \frac{r}{r+s}\, y \tag{5.16}$$

$$\bar{U}^i = \frac{p}{p+q}\, G + \frac{q}{p+q}\, L, \tag{5.17}$$

and y_A^i is the solution to

$$\max_{y \geq 0} \frac{p}{p+q} y \left[\frac{p}{p+q} G + \frac{q}{p+q} L \right] - \left[\frac{p}{p+q} \right]^2 \frac{y^2}{2\alpha} - \left[\frac{r}{r+s} \right]^2 \frac{y^2}{2\beta} . \tag{5.18}$$

Therefore

$$y_A^i = \frac{\beta \left[\frac{p}{p+q} \right]^2}{\beta \left[\frac{p}{p+q} \right]^2 + \alpha \left[\frac{r}{r+s} \right]^2} (G + \frac{q}{p} L). \tag{5.19}$$

By substituting this reward structure into expression (5.1), we find that the planner's expected utility is W_L^i given by equation (5.2).

Suppose instead that the planner wanted all researchers to theorize first. Call this Allocation T.

The possible gain from an allocation in which everyone theorizes first arises from the fact that type i agents have a comparative advantage at theorizing. Consequently, although requiring all agents to theorize first will reduce the number of agents who choose to do research, the incidence of that reduction will tend to fall more heavily on the undesirable type j agents.

The incentive-compatibility constraints are

$$py_A^i + qy_B^i + (1-p-q)y_C^i \geq py_A^j + qy_B^j + (1-p-q)y_C^j \tag{5.20}$$

$$ry_A^j + sy_B^j + (1-r-s)y_B^j \geq ry_A^i + sy_B^i + (1-r-s)y_B^i \tag{5.21}$$

and because researchers must have incentives not to claim falsely to have a type C theory,

$$py_A^i + qy_B^i + (1-p-q)y_C^i \geq \max \{y_C^i, y_C^j\} \tag{5.22}$$

$$ry_A^j + sy_B^j + (1-r-s)y_B^j \geq \max \{y_C^i, y_C^j\} . \tag{5.23}$$

In addition, we continue to have the constraint (5.11). The gross social gains are

$$\bar{U}^i = pG + qL \tag{5.24}$$

$$\bar{U}^j = 0 . \tag{5.25}$$

As before, the non-negativity constraints and the second incentive-compatibility constraint are binding. The optimum requires $y_B^i = y_C^i = 0$, and $ry_A^j + sy_B^j + (1-r-s)y_C^j = ry_A^i$, and y_A^i is then the solution to

$$\max_{y \geq 0} \; p \frac{y}{\alpha} (pG + qL) - \frac{p^2 y^2}{2\alpha} - \frac{r^2 y^2}{2\beta} . \tag{5.26}$$

So we have

$$y_A^i = \frac{\beta p^2}{\beta p^2 + \alpha r^2} (G + \frac{q}{p} L). \tag{5.27}$$

The second planner's expected utility in this case is W_T^1 given by equation (5.3).

Allocation T, like Allocation L, involves too few type i researchers and too many type j researchers, relative to the first best.

Finally suppose the planner allows the research strategy itself to be a signal of underlying quality. In order to accomplish this he allows researchers to choose whether to look or to theorize and sets the payment in such a way that type i researchers choose one strategy while type j choose another. We assume for now that the researcher's strategy is verifiable by the planner.

It is not necessary to solve the planner's problems in these two possible scenarios because it is clear that they will be equivalent to the two cases described above. The case where type i looks first and type j theorizes first is equivalent to Allocation L, while the case where type j looks first and type i theorizes first is equivalent to Allocation T. The reason is that in both cases the incentive-compatibility constraint, not the value of what type j agents produce or the strategy they choose, determines the expected reward to type j agents. So the social gains are determined only by what type i agents produce (i.e. whether they look or theorize), and the number of each type that choose to engage in research. The possibility of signaling by looking or theorizing is of no value to the planner (or the market) because researchers can be sorted by the reward structure. So the question is simply whether it is better to have type i agents look or theorize first, which we can answer simply by comparing welfare under the T and L allocations. Only if announcement mechanisms were for some reason costly, and if research strategy is the least costly such mechanism, does the choice of strategy do anything useful. But nothing in our basic setup implies that choice of strategy has any value as a signal under the present set of assumptions.

5.1.1 Discussion

It is clear from the expressions (5.2) and (5.3) for W_L^1 and W_T^1 that either can be larger, so that either allocation can be preferred to the other. In either allocation, these are too few type i researchers relative to the first best, since $\bar{y}^i = py_A^i < \bar{U}^i$, whereas the first best would require $\bar{y}^i = \bar{U}^i$. There are obviously too many type j researchers relative to the first best, since $\bar{y}^j > 0 = \bar{U}^j$. The reason is that we must give type j researchers enough to keep them honest. Consequently, we attract a positive number of them into research, though we ignore their research in practice and we know this in advance. (The reader can supply his own real-world examples.)

When agents theorize first, there are fewer researchers of each type. The reduction in the number of type i researchers is a cost, while the reduction in the number of type j researchers is a benefit. In addition, theorizing is costly because some type i agents produce type C theories. When p + q is large, the latter cost is small, and consequently theorizing is preferred ($W_T^1 > W_L^1$). When r is small, the benefits from theorizing are small since very few type j agents are attracted into research anyway. (The number of type j agents depends on r through the incentive compatibility constraint: The smaller is r, the less must be paid to type j agents to keep them from pretending to be type i.) Thus when r is small, looking first is preferred to theorizing first ($W_L^1 > W_T^1$).

To summarize, the planner may prefer to have researchers theorize first, despite the cost of having fewer good bridges, so as to discourage type j agents from going into research. This will be the case so long as the cost in terms of the reduction in type i theories is not too great.

Perhaps the easiest way to envision the sorting mechanism is to consider the choice of academic jobs. Some agents choose jobs in which there is a high payoff to successful research, while others choose jobs where the rewards do not depend much on the quality or success of research. According to the model, the distinguishing features of the latter set of agents are, first, the poor quality of their research; and, second, a low opportunity cost of doing research (so that either they are not very good at anything else either, or they get some enjoyment from engaging in research, even if no one pays any attention to it).

In any of the allocations considered here, there is a set of agents who get rewarded for producing theories that are known in advance to have no social value. Because these agents have positive opportunity costs, social welfare could be improved by freeing them from the obligation to perform research. However, it would still be necessary to reward these agents to prevent them from doing research while falsely claiming to be type i.

Unfortunately, in this case all agents (including those who, because of high opportunity costs, would never really enter research) would claim to be potential researchers in order to collect the rewards. This would require the planner to make the sort of massive lump sum transfers that were ruled out in the second paragraph of Section 5. So he must accept the social loss inherent in requiring all "researchers" actually to perform research.

There is one partial solution to this dilemma: type j agents might be able to perform some socially valuable function at a research institution (perhaps teaching undergraduates?) where the planner could easily verify that they are not simultaneously pursuing other productive activities. In this case, the planner can require presence at the institution as a requisite for the rewards, but still assign type j agents to this alternative activity.

In fact, if the best alternative employment of type j agents can be affected at a research institution and monitored by the planner, then he can assign those agents to that activity and achieve the first best optimum. In that case, only type i agents do research, and the planner sets rewards so that they look first.

We have assumed that this first best outcome is not achievable; that is, we have assumed that the most valuable activities of type j agents can not all be performed in research institutions. In fact, our welfare analysis assumes that type j agents have no socially useful functions to perform at research institutions, but this extreme assumption is not required.

5.2 Information About Type is Valueless

Our next result deals with the case in which information about a researcher's type has no value to the social planner. That is, it deals with the case in which $rG + sL > 0$, so that the planner wants to build even the bridges designed by type j researchers. The results in this case turn out to depend on various inequalities involving the parameter values. To avoid a proliferation of subcases, we assume some of these inequalities. Our result is the following:

Suppose that $pG + qL > [rG + sL]/(r + s) > r(G + qL/p) > 0$. *Suppose also that the planner can verify the research strategy. Then:*

(a) The planner offers researchers a choice between a flat salary and a contingent fee.[12]

(b) Type i researchers choose the contingent fee.

[12]In this case, as in Section 5.1, the flat salary could be replaced by any stochastic payment of equal expected value that satisfies the constraints.

(c) Type j researchers choose the flat salary and are directed by the planner to look first.

(d) The planner requires researchers who choose the contingent fee to theorize first if $W_{TL}^2 > W_L^2$, *and to look first if* $W_L^2 > W_{TL}^2$, *where*

$$W_L^2 = \frac{\left[\frac{1}{\alpha}\left[\frac{p}{p+q}\right]^2 (G + \frac{q}{p} L) + \frac{1}{\beta}\left[\frac{r}{r+s}\right]^2 (G + \frac{s}{r} L)\right]^2}{\frac{2}{\alpha}\left[\frac{p}{p+q}\right]^2 + \frac{2}{\beta}\left[\frac{r}{r+s}\right]^2} \tag{5.29}$$

$$W_{TL}^2 = \frac{(pG + qL)^2}{2\alpha} + \frac{(rG + sL)^2}{2\beta(r+s)^2} \tag{5.30}$$

(e) If $W_{TL}^2 > W_L^2$, *then the reward structure to researchers satisfies*

$$py_A^i + qy_B^i = pG + qL \tag{5.31}$$

and

$$ry_A^j + sy_B^j + (1 - r - s)y_C^j = \frac{r}{r+s} G + \frac{s}{r+s} L. \tag{5.32}$$

One way to achieve this is the following: Type i agents are paid

$$y_i^A = G + \frac{q}{p} L \tag{5.33}$$

if they construct a theory of type A and

$$y_i^B = y_i^C = 0 \tag{5.34}$$

otherwise. Type j agents receive a flat salary of

$$y_A^j = y_B^j = y_C^j = \frac{r}{r+s} G + \frac{s}{r+s} L . \tag{5.35}$$

The planner's expected utility is W_{TL}^2.

(f) If $W_L^2 > W_{TL}^2$, *then type i researchers are paid*

$$y_i^A = G + \frac{pq\beta(r+s)^2 + rs\alpha(p+q)^2}{p^2\beta(r+s)^2 + r^2\alpha(p+q)^2} L \tag{5.36}$$

if they construct a theory of type A and

$$y_i^B = y_i^C = 0 \tag{5.37}$$

otherwise. Type j researchers are paid a flat salary of

$$y_j^A = y_j^B = y_j^C = ry_i^A / (r+s). \tag{5.38}$$

The planner's expected utility is W_L^2.

(g) If $W_L^2 > W_{TL}^2$, *then there are too few type i researchers and too many type j researchers relative to the first best. If* $W_{TL}^2 > W_L^2$, *then the quantities of type i and type j researchers are optimal* <u>*subject to the constraint that type i researchers theorize first*</u>. *There are still too few type i researchers relative to the full-information first best, although the number of type j researchers is fully optimal.*

These results follow easily from our two-stage procedure; see the Appendix.

5.2.1 Discussion

We shall comment briefly on point (g) above.

If $W^2_{TL} < W^2_L$, then

$$\bar{U}^j < \bar{y}^j < \bar{y}^i < \bar{U}^i$$

which implies that the optimal equilibrium of this sort has too many type j researchers and too few type i researchers relative to the first best.

If $W^2_L < W^2_{TL}$, then

$$\bar{U}^j = \bar{y}^j \text{ and } \bar{U}^i = \bar{y}^i. \tag{5.39}$$

This implies that in this case, the first-best number of researchers of each type is achieved. (The reader should remember, however, that these "first bests" are only first best subject to the requirement that type i researchers theorize first.)

When $W^2_{TL} < W^2_L$, the planner accepts non-optimal numbers of researchers in order to allow everyone to look first, so that no useless type C theories are produced. When $W^2_L < W^2_{TL}$, the planner accepts the waste inherent in having type i researchers theorize first, in order to achieve optimal numbers of each type of researcher.

5.3 General Discussion

The results from sections 5.1 and 5.2 are hardly conclusive, but they do illustrate the manner in which scientific method (what we have been calling "research strategies") can play a role in the evaluation of the results of scientific research, and hence in the allocation of resources. The basic idea is as follows: Conditional on knowing the ability of an individual, it would not make sense to ask that he ignore any relevant information in the

process of coming up with theories. But because theorizing first has certain desirable selection or screening properties, it can be socially beneficial when underlying abilities are private information.

We can go beyond the framework of the model in thinking about how this applies to the real world. For example, we have assumed that all researchers can choose whether or not to look first or theorize first, and that the actions are publicly verifiable. In practice, though, it is very difficult to verify whether someone looked first or not, and in the model of Section 5 it will frequently be in the interest of a researcher to claim to have theorized first while actually having looked first.[13] On the other hand, it is probably not the case that all researchers can easily choose one strategy or the other. In fact, scientific training in many fields (economics included) tends to get individuals to commit themselves early on to be either a theorist or an applied scientist. Although one occasionally hears complaints about theorists' distance from the real world, the research market does not appear to discourage this type of specialization. While standard arguments about the gains from specialization can account for this, it bears mentioning that the analysis in this paper suggests another story. In a setting where theorizing first is a signal of ability (as in the TL allocation in Section 5.2), it can be useful to separate the theorizers from the lookers at the outset. The type i agents become theorists, incapable of doing empirical work, while type j agents become applied scientists, testing the theories of type i agents as well as their own. There is then no problem verifying that

[13]Interactions within the scientific community such as seminars and ongoing, informal discussions with colleagues may serve partially as a monitoring device.

the type i theories were arrived at without looking first, since type i agents would have demonstrated their inability to look at data.

The model also does not really deal with dynamic issues such as reputation. It does suggest, though, that agents might theorize first early in their careers, either to learn about themselves or to signal their private information to the market. Eventually, though, their reputation would be established, and there would be nothing more to gain by theorizing first. Thus an individual who establishes himself as a type i by successful theorizing early in his career might be observed changing his research strategy and starting to look first.

6. Non-Verifiable Research Strategies

Without direct verifiability of research strategies the planner must set rewards so that agents do not make false claims about their strategies. These "verifiability constraints" are generally binding when a researcher is supposed to have theorized first. If p+q and r+s are less than one, both types of researchers will want to look first if $y_C = 0$. So when researchers are supposed to look first the problem of verifiability does not arise. Thus in Section 5.1, if $W_L^1 > W_T^1$, the solution still applies regardless of verifiability; the same is true in Section 5.2 if $W_L^2 > W_{TL}^2$. If the opposite inequalities hold, then the solutions must be modified.

In addition to the incentive compatibility constraints (5.14) and (5.15), which ensure truth about type, and (5.16) and (5.17), which ensure truth about type C theories, we have the constraints

$$py_A^i + qy_B^i + (1-p-q)y_C^i \geq \frac{p}{p+q} y_A^i + \frac{q}{p+q} y_B^i \qquad (6.1)$$

$$ry_A^j + sy_B^j + (1-r-s)y_C^j \geq \frac{r}{r+s}\, y_A^j + \frac{s}{r+s}\, y_B^j. \qquad (6.2)$$

These verifiability constraints ensure truth about research strategy. In addition, we impose truth-telling about type and strategy jointly:

$$py_A^i + qy_B^i + (1-p-q)\, y_C^i \geq \frac{p}{p+q}\, y_A^j + \frac{q}{p+q}\, y_B^j \qquad (6.3)$$

$$ry_A^j + sy_B^j + (1-r-s)\, y_C^j \geq \frac{r}{r+s}\, y_A^i + \frac{s}{r+s}\, y_B^i. \qquad (6.4)$$

We will consider first the case in which information about a researcher's type is valuable, and then the case in which it is not. Without verifiability of research strategy, payments are contingent on the outcome A, B, or C (as before) and on the announced research strategy.

6.1. Information About Type Valuable; Research Strategy Unobserved

Our next result deals with the case in which information about a researcher's type is valuable but the planner cannot directly observe the research strategy.

Suppose that $rG + sL < 0 < pG + qL$, *so that information about a researcher's type is useful to the planner. Suppose also that the planner cannot verify research strategy (but can still make rewards contingent on announced strategy). Then:*

(a) The planner offers researchers a choice between a flat salary and a contingent fee.[14]

(b) Type i researchers choose the contingent fee.

---

[14]Footnote 9 applies here as well.

(c) *Type j researchers choose the flat salary.*

(d) *Researchers who choose the contingent fee always look first.*

(e) *Researchers who choose the flat salary might follow either research strategy. Their bridges are not built.*

(f) *The contingent fee to type i researchers is given by equations (5.6) and (5.7)*

(g) *The flat salary chosen by type j researchers is given by equations (5.8).*

(h) *The planner's expected utility is* W_L^1, *as given by equations (5.2).*

The proof is in the Appendix.

6.2 Information About Type Valueless; Research Strategy Unobserved

Finally, we present our results when information about a researcher's type is of no value to the planner, and when the planner cannot directly observe the research strategy.

Suppose that $pG + qL > rG + sL > 0$. *Suppose also that the planner cannot verify research strategy (but can still make rewards contingent on announced strategy). Then*

(a) *The planner offers a choice between a contingent reward given by (5.36) and (5.37), and a flat salary given by (5.38). Type i researchers choose the contingent reward scheme and type j researchers choose the flat salary.*[15]

(b) *All researchers look first.*

[15]Footnote 9 applies again.

(c) *The planner's maximized expected utility is* W_L^2, *given by equation (5.29).*

Again, the proof is in the Appendix.

6.3. General Discussion

For certain ranges of parameter values, the planner's choice of outcome is independent of whether he can verify research strategies. This occurs whenever a) information about type is valuable, as in sections 5.1 and 6.1, and expression (5.2) is greater than expression (5.3), and also whenever b) information about type has no value, as in sections 5.2 and 6.2, and expression (5.29) is greater than expression (5.30).

If the parameter values are in these ranges, then no costly procedures to verify research strategy are ever socially justified.

On the other hand, if either c) information about type is valuable and (5.3) exceeds (5.2) or d) information about type is valueless and (5.3) exceeds (5.29), then the planner's inability to verify research strategy leads to inferior outcomes. In these cases, costly verification procedures can be justified. The maximum amount that the planner is willing to spend on such procedures is given by the difference between (5.3) and (5.2), or between (5.30) and (5.29).

7. Conclusions

We can return to the question posed at the beginning of the paper: Does a reader's rational belief in the truth or usefulness of a theory depend upon whether the facts with which it is consistent were known to the researcher before he constructed his theory?

In a well-known text on the philosophy of science, Mary Hesse[16] states that "... apart from the psychological effect of a surprisingly successful prediction, that a fact was predicted before it was observed should not in itself affect the final judgment on a theory for which it is evidence." Our own informal survey of economists indicates that, by about 2 to 1, they think that a theory is more believable if some of the facts supporting it were unknown when the theory was constructed. Almost all responses were given quite forcefully. The authors of this paper originally disagreed with each other about the answer. Some economists follow the practice of deliberately hiding part of a data set from themselves and using only the other part to help formulate a theory. This paper has shown that these beliefs and practices can be rational with certain assumptions about the nature of scientific research. The model of Section 4 shows that for some range of parameters a theory is more believable if the theorize-first strategy was followed. This section requires that the scientist's research strategy is publicly observable. The model of Section 3 could explain a stronger belief in theories obtained by a theorize-first strategy, but only in a very roundabout manner.

This paper has also addressed the question of what research strategies are socially optimal, given information and incentive constraints. The model of Section 5 shows that, for some range of parameters, a subset consisting of the most valuable scientists should be assigned to the theorize-first strategy. Less valuable scientists, however, should always be assigned to look first, regardless of parameter values. If research strategies are

---

[16]The Structure of Scientific Influence, University of California, 1974, page 207.

private information as in Section 6, then at the optimum all scientists look first and welfare can be lower. The model of Section 4 provides an alternative explanation of why scientists might optimally follow the theorize-first strategy.

The questions raised in this paper typically elicit strong opinions but poorly articulated reasons. This paper offers a coherent analysis of the issues. Our results should challenge those who have taken for granted some particular answer to these questions.

APPENDIX TO SECTION 1

1. Section 1 of the paper makes the assumption that researchers can build previous observations into their theories, so that a researcher can always construct a theory that is consistent with these observations. We now take up the opposite case. Suppose that a researcher cannot determine the implications of his theory until after he has written it down and fully worked out the theory. Then a researcher may find that he has constructed a type D theory. And looking first does not guarantee that a researcher will not construct a type C theory. Denote the probabilities of constructing theories of types A, B, C and D as p, q, $1-p-q-\delta$, and δ; earlier we had $\delta=0$.

Let UPr(A) denote the unconditional probability that a theory is type A, and let CPr(A) denote the probability that a theory is type A conditional on the event that it is consistent with the new observation (that is, conditional on its being in the set AUB). If a researcher theorizes first, then UPr(A) = p and CPr(A) = p/(p+q). The Same probabilities apply to a researcher who looks first: UPr(A) = p and CPr(A) = p/(p+q). The only difference from the case in the text (in which $\delta=0$) is that UPr(A) = p for a researcher who looks first. This reflects the researcher's inability to rule out a type C theory even though he already has made the new observation. Because the forms of the conditional probabilities are unchanged by this modification of the model, the neutrality result of section 1 continues to apply. These probabilities, UPr(A)=p and CPr(A)=p/(p+q), continue to apply even if a research can build old observations into his theory but not the new observation. So, even in that case, the neutrality result of section 1 applies.

APPENDIX TO SECTION 5.2

First, consider Allocation L, where all look first. The analysis is just as in Section 5.1, except that the maximization problem (5.18) is replaced by

$$\max_{y \geq 0} \ \frac{p}{p+q} \frac{y}{\alpha} \left[\frac{p}{p+q} G + \frac{p}{p+q} \right] - \left[\frac{p}{p+q} \right]^2 \frac{y^2}{2\alpha}$$

$$+ \frac{r}{r+s} \frac{y}{\beta} \left[\frac{r}{r+s} G + \frac{s}{r+s} L \right] - \left[\frac{r}{r+s} \right]^2 \frac{y^2}{2\beta} .$$

This leads to the solution (5.36).

Next, consider Allocation TL, where type i agents theorize first and type j agents look first. The incentive compatibility constraints are

$$py_A^i + qy_B^i + (1 - p - q)\, y_C^i \ \geq \ \frac{p}{p+q} y_A^j + \frac{q}{p+q} y_B^j$$

$$\frac{r}{r+s} y_A^j + \frac{s}{r+s} y_B^j \ \geq \ ry_A^i + sy_B^i + (1 - r - s)\, y_C^i .$$

The solution given by the equations (5.33)-(5.35) satisfies these constraints in view of the inequalities assumed at the beginning of the section.

Next, consider Allocation LT, where type i agents look first and type j agents theorize first. We will show that Allocation L always dominates Allocation LT. Define

$$f(y) = \frac{1}{\alpha} \frac{p}{p+q} \left[\frac{p}{p+q} G + \frac{q}{p+q} L \right] y = \frac{1}{2\alpha} \left[\frac{p}{p+q} \right]^2 y^2$$

$$g(y) = \frac{1}{\beta} \frac{r}{r+s} \left[\frac{r}{r+s} G + \frac{s}{r+s} L \right] y$$

$$h(y) = \frac{1}{2a}\left[\frac{r}{r+s}\right]^2 y^2.$$

Then the planner's objective function in Allocation L is

$$f(y_A) + g(y_A) - h(y_A)$$

whereas his objective function in Allocation LT is the smaller expression

$$f(y_A) + (r + s)g(y_A) - h(y_A).$$

Thus Allocation LT can be ignored.

Similar considerations show that Allocation TL dominates Allocation T, in which all agents theorize first. Thus Allocation T can be ignored as well.

APPENDIX TO SECTION 6.1

As in Section 5.1, iagents can be given a flat reward $\bar{y}^j$ sufficient to satisfy both the incentive-compatibility and verifiability constraints and y_B^i can be set to zero.

The constraints simplify to

$$\bar{y}^j \geq ry_A^i + (1-r-s)y_C^i \qquad \text{(A6.1.1)}$$

$$\bar{y}^j \geq y_C^i, \qquad \text{(A6.1.2)}$$

$$\bar{y}^j \geq ry_A^i/(r+s), \qquad \text{(A6.1.3)}$$

and condition (6.2) is trivially satisfied. These conditions say that a type j agent must do at least well by announcing that he is type j and accepting $\bar{y}^j$ as if he claims to be a type i and either theorizes, looks first, or does not submit a theory. The verifiability constraint (6.1) for type i agents simplifies to

$$y_C^i \geq py_A^i(p+q). \qquad \text{(A6.1.4)}$$

As suggested above, (A6.1.4) is a binding constraint and therefore holds with equality. This implies that of the three conditions (A6.1.1)-(A6.1.3); (A6.1.2) is the one that applies. In other words, we have to pay enough for unsuccessful theories to keep type i agents from looking first. Consequently the options can be found by solving the unconstrained problem

$$\max_{y} \frac{p}{p+q} \frac{Y}{\alpha}(pG+qL) - \left[\frac{py}{p+q}\right]^2 \frac{1}{2\alpha} - \left[\frac{py}{p+q}\right]^2 \frac{1}{2\beta} \qquad \text{(A6.1.5)}$$

for y_A^i and then using $\bar{y}^j = y_C^i = py_A^i/(p+q)$. The solution to (A6.1.5) is

$$y_A^i = \frac{\beta}{\alpha+\beta}(p+q)(G+\frac{q}{p}L) \qquad \text{(A6.1.6)}$$

which implies that

$$y_C^i = \bar{y}^j = \frac{\beta}{\alpha+\beta} p(G+\frac{q}{p}L). \qquad \text{(A6.1.7)}$$

Comparison with the results from section 5.1 shows that non-verifiability of research strategies has a real social cost: the number of type i researchers is reduced while the number of type j researchers is increased.

The question is then whether theorizing is viable at all without direct verifiability. When researchers look first, the verifiability constraints are non-binding, so equation (5.2) gives the planner's welfare. When all researchers theorize first the planner's welfare is

$$W_T = \frac{p^2}{2\alpha}\left[\frac{\beta}{\beta+\alpha}\right](G + \frac{q}{p}L)^2. \qquad \text{(A6.1.8)}$$

It is clear that without direct verifiability, theorizing first is not viable: W_L is strictly greater than W_T.

APPENDIX TO SECTION 6.2

We now consider the case in which bridges based on theories of type j agents are worth building,

$$pG + qL > rG + sL > 0.$$

All Look First

As in Section 6.1, nonverifiability of research strategies poses no problem in this case. The payments y_C^i and y_C^j can be set to zero. Then the incentive-compatibility constraints are given by equations (5.9) and (5.10), and the optimal payments are given by equations (5.8) and (5.36).

All Theorize First

The planner maximizes expected utility subject to (5.20)-(5.23) and (6.1)-(6.4). This implies a solution in which (A6.1.1), and (A6.1.2)

$$y_B^i = 0,\ y_A^j = y_B^j = y_C^j = \frac{p}{p+q}\, y_A^i = y_C^i \qquad \text{(A6.2.1)}$$

so that constraints (5.20 and (6.4) are slack, while the other six constraints bind. The solution for y_A^i is

$$y_A^i = \frac{\beta\left[pG + qL\right] + \alpha\left[rG + sL\right]}{\left[\frac{p}{p+q}\right]\left[\alpha+\beta\right]}. \qquad \text{(A6.2.2)}$$

Type i Agents Theorize First, Type j Look First

Obviously this case involves $y_B^i = 0$ and $y_C^j = 0$. Then the incentive compatibility constraints are (5.20), (5.22), (A6.2.1), and (A6.2.2) for the type i agents and (5.10),

$$\frac{r}{r+s}\, y_A^j + \frac{s}{r+s}\, y_B^j \geq r y_A^i + (1-r-s)\, y_C^i \qquad \text{(A6.2.3)}$$

and

$$\frac{r}{r+s}\, y_A^j + \frac{s}{r+s}\, y_B^j \geq y_C^i \qquad \text{(A6.2.4)}$$

for the type j agents.

Then we have $y_C^i = y_A^j = y_B^j = \frac{p}{p+q}\, y_A^i$ and

$$y_A^i = \frac{\beta\left[pG + qL\right] + \alpha\left[\frac{r}{r+s}\, G + \frac{r}{r+s}\, L\right]}{\left[\frac{p}{p+q}\right]\left[\alpha+\beta\right]} . \qquad \text{(A6.2.5)}$$

Type i Agents Look First, Type j Theorize First

This is the final case to consider. Clearly $y_C^i = 0$. The incentive-compatibility constraints are then (5.9) and

$$\frac{p}{p+q}\, y_A^i + \frac{q}{p+q}\, y_B^i \geq p y_A^j + q y_B^j + (1-p-q)\, y_C^j \qquad \text{(A6.2.6)}$$

and

$$\frac{p}{p+q}\, y_A^i + \frac{q}{p+q}\, y_B^i \geq y_C^j \qquad \text{(A6.2.7)}$$

for the type i agent, and (5.21), (5.23), (6.2), and (6.4) for the type j agent. Then (5.9), (5.23), (6.1), and (6.4) bind, $y_B^i = 0$, $y_A^j = y_B^j = y_C^j = \frac{r}{r+s} y_A^i$, and

$$y_A^i = \frac{\beta\left[\frac{p}{p+q}\right]^2 \left[G + \frac{q}{p}L\right] + \frac{\alpha r^2}{r+s}\left[G + \frac{s}{r}L\right]}{\beta\left[\frac{p}{p+q}\right]^2 + \alpha\left[\frac{r}{r+s}\right]^2}. \qquad \text{(A6.2.8)}$$

Welfare and Discussion

When all researchers look first, the planner's welfare is

$$W^L = \frac{\left\{\frac{1}{\alpha}\left[\frac{p}{p+q}\right]^2 \left[G + \frac{q}{p}L\right] + \frac{1}{\beta}\left[\frac{r}{r+s}\right]^2 \left[G + \frac{s}{r}L\right]\right\}^2}{2\frac{1}{\alpha}\left[\frac{p}{p+q}\right]^2 + 2\frac{1}{\beta}\left[\frac{r}{r+s}\right]^2}. \qquad \text{(A6.2.9)}$$

When all researchers theorize first, welfare is

$$W^T = \frac{1}{2}\left[\frac{\alpha\beta}{\alpha+\beta}\right]\left\{\frac{p}{\alpha}\left[G + \frac{q}{p}L\right] + \frac{r}{\beta}\left[G + \frac{s}{r}L\right]\right\}^2. \qquad \text{(A6.2.10)}$$

When type i agents theorize first but type j look first, welfare is

$$W^{TL} = \frac{1}{2}\left[\frac{\alpha\beta}{\alpha+\beta}\right]\left\{\frac{p}{\alpha}\left[G + \frac{q}{p}L\right] + \frac{r}{\beta(r+s)}\left[G + \frac{s}{r}L\right]\right\}^2. \qquad \text{(A6.2.11)}$$

Welfare when type i researchers look first and type j theorize first is:

$$W^{LT} = \frac{\left\{\beta\left[\frac{p}{p+q}\right]^2\left[G + \frac{q}{p}L\right] + \frac{r}{\beta(r+s)}\left[G + \frac{s}{r}L\right]\right\}^2}{2\beta\left[\frac{p}{p+q}\right]^2 + 2\alpha\left[\frac{r}{r+s}\right]^2}. \qquad \text{(A6.2.12)}$$

It is easy to show that W^L . W^{LT} and $W^{TL} > W^T$. So type j researchers will definitely look first. We must determine the optimal strategy for type i agents. This involves a comparison of the welfare expressions (A6.2.9) and (A6.2.11). With some tedious algebra one can show that

$$W^L > W^{TL}. \qquad \text{(A6.2.13)}$$

So the social planner will, when he cannot verify research strategies, choose rewards to induce <u>all</u> researchers to look first. Given the results of Section 6.1, this conclusion holds regardless of whether bridges based on theories of type j agents are worth building.

To summarize: theorizing first is never a solution when research strategies are private information. Theorizing first is sometimes a solution when the research strategy is verifiable, as Section 5 shows. In those circumstances, the inability to verify research strategies reduces welfare. This could justify costly procedures designed to keep researchers honest.

REFERENCES

Richmond Campbell and Thomas Vinci, "Novel Confirmation", The British Journal for the Philosophy of Science 34 (1983),315-341.

Mary Hesse, The Structure of Scientific Influence, University of California Press, 1974.

C. Howson, "Bayesianism and Support by Novel Facts", The British Journal for the Philosophy of Science, 35 (1984)

Larry Laudan, "Why Was the Logic of Discovery Abandoned?", in Scientific Discovery, Logic and Rationality (T. Nickles, ed.), Dordrecht: Reidel, 1980.

Edward Leamer, Specification Searches, New York: John Wiley & Sons, 1978.

Robert E. Lucas, Jr., "Understanding Business Cycles," Carnegie-Rochester Conference Series on Public Policy 5, 1977, 7-30.

R. McLaughlin, "Invention and Induction: Laudan, Simon and the Logic of Discovery", Philosophy of Science 35 (1984).

A. Musgrave, "Logical versus Historical Theories of Confirmation", British Journal for the Philosophy of Science 22 (1974).

A. Musgrave, "Evidential Support, Falsification, Heuristics and Anarchism" in Progress and Rationality in Science (Radnitzky and Andersson, eds.), Dordrecht: Reidel, 1978.

T. Nickles, "Beyond Divorce: Current Status of the Discovery Debate", The British Journal for the Philosophy of Science 52 (1985).

M. Pera, "Inductive Method and Scientific Discovery", in On Scientific Discovery (R.D. Grmek, R.S. Cohen, and G. Cimino, eds.), Dordrecht: Reidel, 1981.

Karl R. Popper, The Logic of Scientific Discovery, New York: Harper and Row, 1959.

Gerard Radnitzky, "The 'Economic' Approach to the Philosophy of Science", The British Journal for the Philosophy of Science 38 (1987), 159-179.

W. Whewell, The Philosophy of the Inductive Sciences, London: John W. Parker, West Strand, 1847.