Rochester Center for Economic Research

The Positive Economics of Methodology

Kahn, James A., Steven E. Landsburg, and Alan C. Stockman

Working Paper No. 165
November 1988

University of
Rochester

THE POSITIVE ECONOMICS OF METHODOLOGY*

James A. Kahn

Steven E. Landsburg

Alan C. Stockman

University of Rochester

Working Paper No. 165

November 1988

*We thank Nicholas Rowe and numerous colleagues at Rochester and elsewhere for informal (and often lively) discussions of this topic. The title of this paper was almost "Should We Take the 'Econ' out of Econometrics?"

Suppose that you pick up the latest issue of the Journal of Political Economy 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 fact, 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 fact at the time when he constructed his theory. Does—and should— the third section contribute to the degree of belief that you attach to the theory?

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; while others prefer to develop a "structural" model and then test it. We think of the first type as "looking first" and the second type as "theorizing first". Under what circumstances, and in what senses, does it matter which strategy the researcher pursues? How would a social planner structure rewards to scientists to induce them to choose optimal strategies?

We will address this issue within the context of an explicit model of scientific research. In Section 1, we will examine the question of the appropriate degree of belief for a reader to attach to a theory. We will argue that in the case where the characteristics of the researcher are known in advance, any two theories that are compatible with the same set of facts

are equally likely to be true, regardless of how those theories were arrived at. Thus a theory that was arrived at by "looking first" is neither more nor less believable than a theory that has survived the testing process in "theorize first". In this case, if there is any cost to theorizing, researchers should look at all data in advance, to avoid the cost of constructing theories that are ultimately discarded.

Suppose, on the other hand, that the characteristics of the researcher—in particular, his level of "intuition"—are not known. In that case, the research strategy can be important information. The researcher's ability to construct a theory that survives the testing process leads us to update our assessment of his intuition. If scientists' abilities to construct theories that fit currently undiscovered facts are correlated with their abilities to construct true (or useful) theories, then the researcher who is successful at theorizing first sends a signal of his research ability an so is more likely to be rationally believed than one who looks first.

In Section 2, we ask which strategy is more desirable from the point of view of a social planner who wants to use the research as a guide to policy. To the planner, "theorize first" has the disadvantage that resources are used to produce theories that are ultimately rejected and hence have no social value. Under the conditions mentioned above, however, it has the advantage that those theories which survive hypothesis testing are more likely to be true than those that would be arrived at under "look first". When the costs of producing theories are small (for example if there is a large number of researchers with no opportunity costs), the planner will prefer that all researchers look first. But when the cost of theorizing is substantial (for example, if there is only one researcher, who can only produce one theory in a given time period) then either strategy might be preferred.

In Section 3, we present a complete model in which different researchers have different characteristics. A social planner, who does not know either the level of intuition or the opportunity cost of any individual researcher, must announce rewards for successful and unsuccessful attempts at research under either strategy. In doing so, he affects both the strategies chosen by existing researchers and the incentives for entry into research. We examine the optimal reward structure for researchers. We address questions such as whether we should refuse to reward researchers who look first—say, by refusing to publish their papers? When should the researcher's compensation be tied to the outcome of the policy that he recommends, and when should he receive a guaranteed salary? We answer these questions under a variety of assumptions about parameter values. One surprising result is that it can be optimal to reward researchers for producing theories even when it is known in advance that their theories will have no social value.

Throughout Section 3 we maintain the assumption that the planner is able to verify that researchers have actually used the strategies that they claim to have used. In Section 4 we assume that research strategies are private information, not directly verifiable by the planner. We show that there can be significant costs associated with the planner's inability to verify strategies which may justify costly procedures designed to keep researchers honest.

Throughout all of this discussion, we assume that the researcher has available exactly one set of data to look at either before or after theorizing. In Section 5, we modify the model to incorporate another important aspect of research: the choice of data sets. Theorizing first can be useful because it helps the researcher choose a data set to look at or an experiment to perform.

Section 6, containing empirical analysis consistent with our model, has been deleted to avoid problems of self-reference. Instead, we merely state some conclusions. We state our results starting from section 2 as normative, that is, as the solution to a social planner's optimizing problem. However, many of our results could be reinterpreted as positive rather than normative; hence our title.

1. Degree of Belief

We suppose that a body of facts is known, and that there is one "fact" that is not yet known (this might be an experiment that has not yet been conducted, or a data set that has not yet been examined). There are three types of theories, and an infinite number of possible theories of each type. Type A theories are consistent with all of the facts, and are also true. (The meaning of "truth" is deliberately left somewhat nebulous in this section. Roughly, a true theory is one that the reader would want to believe if he could foresee all of the consequences of believing it. In subsequent sections, where we formulate precise social planner's problems, we shall reveal the precise meaning of "truth".) Type B theories are consistent with all of the facts, but are not true. Type C theories are consistent with all of the known facts, but not with the unexamined one.

1.1 A Researcher with known abilities.

A single researcher theorizes without looking at the unexamined fact. The outcome of his research process is a theory of type A with probability p, a theory of type B with probability q, and a theory of type C with probability 1-p-q. After formulating a theory, he tests it against the unexamined fact. If it fails to conform (which happens with probability

1−p−q) then he discards the theory and does not publish a paper. If it does conform, then he knows that his theory is either of type A or of type B, and he publishes the paper. The probability that a published theory is true is given by

$$\text{Prob [Theory is type A} \mid \text{Theory is type A or type B]} = \frac{p}{p+q}.$$

Alternatively, suppose that the researcher first examines the remaining fact, and then selects among the theories that are consistent with it. This enables him to discard theories of type C. We imagine the theorizing process as drawing a ball from an urn; now all balls marked "C" have been removed, so that of the remaining balls, the fraction p/(p+q) are marked "A". Thus this researcher will always publish, and his theory will be true with probability p/(p+q).

We think of the parameters p and q as descriptive of the researcher's "abilities". This allows us to state our first result:

If the researcher's abilities are known, then the probability that his theory is true, conditional on its being published, is independent of the research strategy he uses.

All published papers have the same probability of being correct. But a researcher who looks first always publishes, while one who theorizes first may not. So, under these assumptions, "theorize first" is never good advice.

1.2 A researcher with unknown abilities.

The condition in 1.1 that the researcher's abilities be known is crucial. To see why, imagine two types of researchers, whose probabilities of selecting theories of various types are 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

One might think of these two types as differing in their level of intuition.

We make the following assumptions:

$$p + q > r + s \qquad (*)$$

$$p/(p+q) > r/(r+s)$$

This is, a type–i "theorizer" is more likely than a type–j theorizer to select a theory that is consistent with the unexamined fact, and a type–i "looker" is more likely than a type–j looker to select a true rather than a false theory.

The assumptions (*) will be in force through the remainder of the paper. Note that they imply that

$$p/q > r/s. \qquad (**)$$

We also suppose that a given researcher has probability i of being type–i and probability $j = 1 - i$ of being type–j.

Suppose now that a given researcher looks first and then chooses a theory consistent with the facts. If he is type–i, he selects 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 $r/(r+s)$. A reader who does not know this particular

researcher's ability calculates that the theory is true with probability

$$\gamma = i \cdot \frac{p}{p+q} + j \cdot \frac{r}{r+s} . \qquad (1.1)$$

Suppose alternatively that the researcher theorizes first and produces a theory that survives testing. (That is, he produces a theory that is known to be of either type A or type B.) Given this, we update the probability of his being type–i as follows:

$$i' = i \cdot \frac{p+q}{i\cdot(p+q) + j\cdot(r+s)} . \qquad (1.2)$$

We write $j' = 1 - i'$ for the updated probability that the researcher is type–j. Using the first part of condition (*), we see that $i' > i$. Given that the researcher produces a publishable paper (i.e. one that survives the testing process), the probability that his theory is true is given by

$$\gamma' = i' \cdot \frac{p}{p+q} + j' \cdot \frac{r}{r+s} . \qquad (1.3)$$

Since $i' > i$, condition (*) guarantees that this expression is greater than the expression (1.1) for probability of truth under the look–first regime. Thus we have our next result:

Assume that the researcher's characteristics are <u>not</u> known and that (*) holds. Then the probability that a theory is true, conditional on its being published, is greater if it was produced under a "theorize first" strategy than if it was produced under a "look first" strategy.

1.3. Alternative Interpretations and Discussion.

The model of Section 1.2 allows alternative interpretations. First, suppose that there is only one type of researcher, who sometimes interprets his experiment incorrectly (or whose experimental results are altered by sampling error). With probability 1−i, he inadvertently reverses the experimental conclusion, and so believes that his experiment is consistent with theories of type C but inconsistent with theories of types A and B. In that case, we set r/(r+s) = 0 in equations (1.1) and (1.3) to get the probability of truth for a theory developed under "look first" and "theorize first" strategies.

Suppose the experiment is subject to sampling error. The researcher is interested in the population value of a parameter B, and his sample produces an estimate b. If he theorizes first then he chooses a rule of inference to determine if the estimate b is consistent or inconsistent with the implication of his theory. If he looks first, he constructs a theory to be consistent with his estimate b (by this rule of inference). Let 1−i denote the probability that the inference is incorrect. Then the model of section 1.2 applies with r/(r+s) = 0 in equations (1.1) and (1.3).

Second, 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 1.2, where researchers are endowed with the given quantitites of intuition.) If the effort is unobserved by the reader, then equations (1.1) and (1.3) apply.

Under any of these interpretations, the difference between expressions (1.1) and (1.3) measures the extent to which the reader should discount

evidence that was examined by the researcher prior to theorizing.[1] This "pretesting discount" depends on the parameters i, p, q, r, and s, which describe the scientific community as a whole.

Suppose researchers have a set of experiments from which they costlessly and randomly choose, and suppose that researchers do not always report all of the experiments they performed. In the context of the model of section 1.2, 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 looks first might repeat the experiment 20 times, take the result of one repetition and construct a theory consistent with it, reporting only this experiment in his paper. 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 researchers who theorize first follow this practice of repeating the experiment until it yields results consistent with his theory, and reporting only this result, then the research strategy yields no information to the reader: the equivalence result from section 1.1 is obtained.

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 "fact" 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.

---

[1]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.

2. A Social Planner's Problem

Consider a social planner who 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.

We say that a theory is "true" if a bridge built using that theory will stand, and "false" if a bridge built using that theory will fall. We suppose that there are type-i and j researchers, and that their probabilities of selecting various theories are as in the table from Section 1.2. We continue to assume the conditions (*). What research strategy should the planner command the researchers to use?

The advantage of "theorize first" is that any theory it produces has an enhanced probability of truth. The disadvantage is that it might produce no theory at all. (That is, it might produce a theory that is rejected in the testing stage.) When there are many researchers, the chance that none will produce a theory is very small. Thus we expect that "theorize first" is the right strategy in the case of many researchers, but not necessarily in the case of few researchers. In this section, we will confirm that expectation.

Assume that researchers look first. In that case, every researcher picks a theory consistent with all of the facts, and the planner selects one of those theories at random. The theory is true with probability γ as defined in equation (1.1), and the expected utility to the planner if he builds a bridge using this theory is

$$\gamma G + (1 - \gamma) L .$$

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

$$\text{Max}(\gamma\, G + (1 - \gamma)\, L,\ 0). \qquad (2.1)$$

Now we ask what happens when researchers are told to theorize first. Each researcher selects a theory and tests it against the facts. Then the social planner chooses randomly among those theories (if any) that survive the testing. There are two extreme subcases: that in which there are sufficiently many researchers to virtually guarantee that at least one theory survives testing, and that in which there is only one researcher.

2.1 Many researchers

With many researchers, any of the surviving theories is true with probability γ' as defined in equation (1.3). The expected utility from using this theory is

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

Since $\gamma' > \gamma$ (by condition (*)), it follows that this expression is always greater than (2.1), so that all researchers should be told to theorize first. So when there are many researchers whose individual characteristics are unobservable, and a single research project, it is always preferable for everyone to theorize first.

2.2 A single researcher

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 (2.1). 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 can be built. With probability

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

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

$$[i\ (p + q) + j\ (r + s)]\ \text{Max}\ \{\gamma'\ G + (1 - \gamma')\ L,\ 0\}. \qquad (2.2)$$

This can be either greater or less than (2.1), so either strategy might be preferred. Experiments with parameter values indicate that there is no simple characterization of when one or the other is superior. Note, however, that when i is either 0 or 1, γ' is less than γ, so that (2.2) is less than (2.1) and looking first is preferred to theorizing first. This is because the researchers' characteristics are known in advance, so a successful test conveys no information.

We note that if the planner expects to use the researcher's services on future projects, 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 rejected.

2.3 The value of type–j researchers' theories

Would the planner want to build a bridge if he knew in advance that the only available theory had been constructed by a type–j researcher? By the results of Section 1, the strategy used by the researcher when he constructed the theory is irrelevant. The planner would build if and only if

$$r\,G + s\,L > 0. \tag{2.3}$$

When this inequality holds, we will say that type–j bridges are worth building.

With one researcher (the model of Section 2.2), the planner might be willing to incur the costs associated with the "theorize first" strategy in order to gain information. This information is valuable if type–j bridges are <u>not</u> worth building. If, on the other hand, inequality (2.3) holds, then a researcher's characteristics would be irrelevant to the planner's decision. This suggests that when type–j bridges are worth building, "look first" is always preferred, and this is the case. It is easy to see that in the presence of (2.3), expression (2.1) is always greater than expression (2.2). We have therefore shown that when type–j bridges are worth building and there

is a single researcher with unknown characteristics, the "look first" strategy is always preferred.

The more general observation is that when projects guided by type-j research are valuable (though less valuable than projects guided by type-i research) then researchers should look first. But when projects guided by type-j research are actually harmful, theorizing first might be preferred.

3. Mechanism Design by a Social Planner

An unresolved issue from the previous sections 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 2 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. In this section we 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 some 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

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. In this section we continue to assume that the planner can verify whether an agent actually 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.

Having 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 announcement. In this way we avoid taking a stand on what particular device or institution is in practice used 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 represent 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 if no bridge is built. So, for example, in a particular allocation ($\mathcal{A}$) a type-i researcher who theorizes first gets an expected reward of $py_A^{i\mathcal{A}} + qy_B^{i\mathcal{A}} + (1-p-q)y_C^{i\mathcal{A}}$, while a type-j researcher who looks first gets $(ry_A^{j\mathcal{A}} + sy_B^{j\mathcal{A}})/(r+s)$.

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

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 2. 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 would like to set up a reward structure to get 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. (This implies certain incentive-compatibility constraints.) These additional constraints will generally imply a second-best outcome.

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 therefore a 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

$$(\bar{y}^i/\alpha)\bar{U}^i - \bar{y}^{i2}/2\alpha + (\bar{y}^j/\beta)\bar{U}^j - \bar{y}^{j2}/2\beta, \tag{3.1}$$

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 via large negative penalties for failures. Second, there are the incentive-compatibility constraints that require that agents have no incentive to lie about their type.

3.1. Type-j Bridges Not Worth Building

In this section we will assume that $rG + sL \leq 0 < pL + qG$ so that 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 3.2, 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.

Allocation L: All Look First

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

$$(py_A^i + qy_B^i)/(p+q) \geq (py_A^j + qy_B^j)/(p+q) \tag{3.2}$$

$$(ry_A^j + sy_B^j)/(r+s) \geq (ry_A^i + sy_B^i)/(r+s) \tag{3.3}$$

These conditions imply that agents will not have any incentive to claim to be the other type so as to get the other type's reward structure. Once the planner can identify the agents' types, he will discard 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 true. This imposes the additional constraint that type–j agents are paid a flat fee,

$$y_A^j = y_B^j \tag{3.3A}$$

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

$$\bar{U}^i = (pG + qL)/(p+q) \tag{3.4}$$

$$\bar{U}^j = 0 \tag{3.5}$$

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

$$y_A^j = y_B^j = ry_A^i/(r+s) \tag{3.6}$$

because the incentive-compatibility constraint (3.3) is binding. So

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

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

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

and y_A^i is the solution to

$$\max_{y \geq 0} \ \frac{p}{p+q}\, y\left[\frac{p}{p+q}\, G + \frac{q}{p+q} L\right] - \left[\frac{p}{p+q}\right]^2 \frac{y^2}{2\alpha} - \left[\frac{r}{r+s}\right]^2 \frac{y^2}{2\beta}\,. \tag{3.7}$$

Therefore

$$y_A^i = \frac{\beta\left[\frac{p}{p+q}\right]^2}{\beta\left[\frac{p}{p+q}\right]^2 + \alpha\left[\frac{r}{r+s}\right]^2}\left(G + \frac{q}{p}\, L\right). \tag{3.8}$$

This implies that there are too few type-i researchers relative to the full-information first-best, because the first-best would require (given that $y_B = 0$) $y_A^i = G + \frac{q}{p}\, L$.

There are obviously too many type–j researchers in the optimal constrained allocation, since $\bar{y}^j > 0$. The reason is that we must give type–j agents enough to keep them honest. Consequently we attract $\frac{\gamma y_A^i}{\beta(\gamma+s)}$ of them into research, though we ignore their research in practice and we know this in advance. (The authors have met some of these people.)

Allocation T: All Agents Theorize First

The possible gain from an allocation in which everyone theorizes 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

$$py_A^i + qy_B^i + (1-p-q)y_C^i \geq py_A^j + qy_B^j + (1-p-q)y_C^j \tag{3.9}$$

$$ry_A^j + sy_B^j + (1-r-s)y_B^j \geq ry_A^i + sy_B^i + (1-r-s)y_B^i \tag{3.10}$$

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

$$py_A^i + qy_B^i + (1-p-q)y_C^i \geq \max\{y_C^i, y_C^j\} \tag{3.11}$$

$$ry_A^j + sy_B^j + (1-r-s)y_B^j \geq \max\{y_C^i, y_C^j\} \; . \tag{3.12}$$

In addition, we continue to have the constraint (3.3A). The gross social gains are

$$\bar{U}^i = pG + qL \tag{3.13}$$

$$\bar{U}^j = 0 \ . \tag{3.14}$$

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

$$\max_{y \geq 0} \ p \frac{y}{\alpha}(pG + qL) - \frac{p^2y^2}{2\alpha} - \frac{r^2y^2}{2\beta} \ . \tag{3.15}$$

So we have

$$y_A^i = \frac{\beta p^2}{\beta p^2 + \alpha r^2}\left(G + \frac{q}{p} L\right). \tag{3.16}$$

Conditional on everyone theorizing, the first-best number of type-i researchers would require $y_A^i = G + \frac{q}{p} L$, so the constrained allocation gets too few type-i and too many type-j researchers. On the other hand, type-i agents have a comparative advantage at theorizing, so that type-j agents bear more of the incidence of the reduction in research activity in this arrangement.

Allocations TL,LT: Different Strategies for type–i and type–j Researchers

Now 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 determines the expected reward to type–j agents, not the value of what they produce or the strategy they choose. 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, 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.

3.1.1 Welfare Analysis and Discussion

Welfare under allocations L and T is given by

$$W_L = \frac{\left[\frac{p}{p+q}\right]^2}{2\alpha}\left\{\frac{\beta\left[\frac{p}{p+q}\right]^2}{\beta\left[\frac{p}{p+q}\right]^2 + \alpha\left[\frac{r}{r+s}\right]^2}\right\}(G + \frac{q}{p}L)^2. \qquad (3.19)$$

$$W_T = \frac{p^2}{2\alpha}\left[\frac{\beta p^2}{\beta p^2 + \alpha r^2}\right](G + \frac{q}{p}L)^2 \qquad (3.20)$$

These expressions were derived by plugging the optimal reward structures into the objective function (3.1). From these expressions it is clear that either allocation may be preferred to the other. For example, if p+q=1, r>0, and r+s<1 then $W_T > W_L$, whereas if r=0 and p+q < 1 we have $W_L > W_T$. The intuition is that theorizing brings about a reduction in both types of research, but a proportionately greater reduction of bad (type-j) research. For p+q near one the cost of theorizing in terms of the reduction of good (type-i) theories is small, so the benefit outweighs the cost. For smaller p+q the cost grows, while for r near zero the benefits from theorizing are small because very few type-j agents are attracted to research anyway. (The number of type-j researchers 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.)

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 for not doing research while falsely claiming to be type-i.

Unfortunately, in this case <u>all</u> 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 3. Thus 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 <u>at a research institution</u> (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 effected 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 are not all in research activities. 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.

3.2. Type–j Bridges Worth Building

We now turn to the case in which $pG + qL > rG + sL > 0$ so that bridges build with theories constructed by type–j agents are worth building, thought not as valuable as those from type–i agents. In this case, the first–best solution involves a positive number of type–j agents doing research.

The incentive–compatibility constraints are the same as before. The planner's expected utility (3.1) now includes a positive value for $\bar{U}^j$. So the optimum is easily calculated for each allocation rule.

Allocation L: All Look First

As before, $y_B^i = 0$ and y_A^j and y_B^j are given by equation (3.6). Now y_A^i is the solution to

$$\max_{y\geq 0} \frac{p}{p+q}\frac{y}{\alpha}\left[\frac{p}{p+q}G + \frac{q}{p+q}L\right] - \left(\frac{p}{p+q}\right)^2 \frac{y^2}{2\alpha} \qquad (3.7')$$

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

which is

$$y_A^i = G + \frac{pq\beta(r+s)^2 + rs\alpha(p+q)^2}{p^2\beta(r+s)^2 + r^2\alpha(p+q)^2} L \; . \qquad (3.8')$$

As before, this solution has the characteristic that

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

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.

Allocation T: All Agents Theorize First

With the incentive compatibility constraints unchanged from before and equation (3.14) replaced by

$$U_j = r\,G + s\,L \; , \qquad (3.14')$$

the optimum requires (as before) $y_B^i = y_C^i = 0$ and $r\,y_A^j + s\,y_B^j + (1-r-s)\,y_C^j = r\,y_A^i$. Now y_A^i is the solution to

$$\max_{y \geq 0} \; p \frac{y}{\alpha} (p\,G + q\,L) - \frac{p^2 y^2}{2\alpha} + r \frac{y}{\beta} (r\,G + s\,L) - \frac{r^2 y^2}{2\beta} \qquad (3.15')$$

which is

$$y_A^i = G + \frac{\beta pq + \alpha rs}{\beta p^2 + \alpha r^2} \cdot L \; . \qquad (3.16')$$

Allocations TL and LT

When bridges based on theories constructed by type–j agents are worth building, there allocations cannot be handled as easily as above. We consider each allocation in turn.

Type–i Agents Theorize First, Type–j Look First

The incentive–compatibility constraints are

$$py_A^i + ay_B^i + (1-p-q)y_C^i \geq \frac{p}{p+q} y_A^j + \frac{p}{p+q} y_B^j \qquad (3.21)$$

$$\frac{r}{r+s} y_A^j + \frac{s}{r+s} y_B^j \geq ry_A^i + sy_B^i + (1-r-s)y_C^i \qquad (3.22)$$

ensuring that no type–i researcher will play type–j and vice–versa.

The first best outcome is achieved if every researcher receives the expected social value of his bridge, i.e.

$$\bar{y}^i = \bar{U}^i \quad \text{and} \quad \bar{y}^j = \bar{U}^j$$

or

$$py_A^i + qy_B^i = pG + qL \quad \text{and} \quad ry_A^j + sy_B^j + (1-r-s)y_C^j = \frac{rG+sL}{r+s} \ . \tag{3.23}$$

One way to achieve this is

$$y_A^i = G + \frac{q}{p} L, \ y_B^i = y_C^i = 0, \ y_A^j = y_B^j = \frac{r}{r+s} G + \frac{s}{r+s} L \ . \tag{3.24}$$

which satisfies the incentive constraints (9) if and only if

$$pG + qL \geq \frac{r}{r+s} G + \frac{s}{r+s} L \geq r(G + \frac{q}{p}L) \ . \tag{3.25}$$

In order to avoid a proliferation of cases, we assume that (3.25) holds. There are two other possibilities in which one or the other of the two inequalities fails. (It is not possible for both to fail.) The reader can solve these cases for himself.

Type-i Agents Look First, Type-j Theorize First

The incentive compatibility constraints are now

$$\frac{p}{p+q} y_A^i + \frac{q}{p+q} y_B^i \geq py_A^j + qy_B^j + (1-p-q)y_C^j \tag{3.26}$$

$$ry_A^j + sy_B^j + (1-r-s)y_C^j \geq \frac{r}{r+s} y_A^i + \frac{s}{r+s} y_B^i \ . \tag{3.27}$$

We can argue as in the preceding subsections; here we just list conclusions.

The optimal outcome is achieved by $y_B = 0$,

$$y_A^i = \frac{p^2\beta(r+s)^2+r^2\alpha(p+q)^2(r+s)}{p^2\beta(r+s)^2+r^2\alpha(p+q)^2} G + \frac{pq\beta(r+s)^2+rs\alpha(p+q)^2(r+s)}{p^2\beta(r+s)^2+r^2\alpha(p+q)^2} L, \quad (3.28)$$

$$y_A^j = y_B^j = y_C^j = \frac{r}{r+s} y_A^i . \quad (3.29)$$

3.2.1 Welfare Analysis

Now we can compute the planner's optimal strategy when type-j bridges are worth building and (3.25) holds.

We first show that the allocation in which all agents look first is preferred by the planner to the case in which type-i agents look first but type-j agents theorize first.

Define $f(y) = \frac{1}{\alpha} \frac{p}{p+q} (\frac{p}{p+q} G + \frac{q}{p+q} L) y = \frac{1}{2\alpha} (\frac{p}{p+q})^2 y^2$

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

$$h(y) = \frac{1}{2\alpha} (\frac{r}{r+s})^2 y^2.$$

Then the planner's objective function when all agents look first is

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

while in the other case his objective function is the smaller expression,

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

Therefore we can ignore the allocation in which type–i agents look first and type–j theorize first.

Similar considerations show that the allocation in which type–i agents theorize first and type–j agents look first is preferred to that in which all agents theorize first. So the latter allocation can be ignored. So the planner will choose an allocation in which type–j agents look first. We have only to determine whether type–i agents will look first or theorize first.

Social welfare when all agents look first is

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

Social welfare when type–i agents theorize first is

$$\frac{(pG+qL)^2}{2\alpha}+\frac{(rG+sL)^2}{2\beta(r+s)^2}.$$

Either of these could be larger, and the social planner chooses the better of the two equilibria.

We conclude that when type–j bridges are worth building and (3.25) holds, type–i researchers receive fees contingent on whether their bridges stand, while it suffices for type–j researchers to receive a flat salary independent of the outcome of their research. As in Section 3.1, type–i researchers could either theorize first or look first. In contrast to Section 3.1, where the research strategies of type–j agents are irrelevant, here the type–j agents always look first.

The intuition is similar to that of Section 3.1; the difference is that here the type-j bridges are worth building. Hence the type-j research strategy becomes relevant.

3.3 Discussion

The results from the previous sections 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 3 it will frequently be in the interest of a researcher to claim to have theorized first while actually having looked first.[2] 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

---

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

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 3.2), it could be useful to separate the theorizers from the lookers at the outset. Thus we could have 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 would be no problem verifying that 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.

4. 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 have an incentive to look first if $y_C = 0$. So when researchers are supposed to look first the problem of verifiability does not arise. Thus the solution under allocation L from Section 3 applies regardless of verifiability, whereas the solution under allocation T must be modified.

We also allow a researcher to submit no theory at all, which is equivalent to coming up with a theory of type C. Since $y_C^i > 0$, any researcher can guarantee himself a positive payoff by not producing a theory. So for theorizing first to be viable we must ensure that agents whose bridges are worth building actually attempt to construct theories that work.

In addition to the incentive compatibility constraints (3.9) and (3.10), which ensure truth about type, and (3.11) and (3.12), which ensure truth about type-C theorizes, we have the constraints

$$py_A^i + qy_B^i + (1-p-q)y_C^i \geq \frac{p}{p+q}\, y_A^i + \frac{q}{p+q}\, y_B^i \tag{4.1}$$

$$ry_A^j + sy_B^j + (1-r-s)y_C^j \geq \frac{r}{r+s}\, y_A^j + \frac{s}{r+s}\, y_B^j. \tag{4.2}$$

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

$$py_A^i + qy_B^i + (1-p-q)\, y_C^i \geq \frac{p}{p+q}\, y_A^j + \frac{q}{p+q}\, y_B^j \tag{4.3}$$

$$ry_A^j + sy_B^j + (1-r-s)\, y_C^j \geq \frac{r}{r+s}\, y_A^i + \frac{s}{r+s}\, y_B^i. \tag{4.4}$$

We will consider first the case in which type-j bridges are not worth building, and then the case in which they are. Without verifiability of

research strategy, payments are contingent on the outcome A, B, or C (as before) and on the announced research strategy.

4.1 Type–j Bridges Not Worth Building, No Direct Verifiability

As in Section 3.1, if type–j bridges are not worth building then type–j agents 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. Conditions (3.10), (3.12), and (4.4) then simplify to

$$\bar{y}^j \geq r y_A^i + (1-r-s) y_C^i \tag{4.5}$$

$$\bar{y}^j \geq y_C^i, \tag{4.6}$$

$$\bar{y}^j \geq r y_A^i/(r+s). \tag{4.7}$$

and condition (4.2) 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 (4.1) for type–i agents simplifies to

$$y_C^i \geq p y_A^i/(p+q). \tag{4.8}$$

As suggested above, (4.8) is a binding constraint and therefore holds with equality. This implies that of the three conditions (4.5)–(4.7), (4.6) 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 optimum can be found by solving the unconstrained problem

$$\max_{y} \frac{p}{p+q}\frac{y}{\alpha}(pG+qL) - \left[\frac{py}{p+q}\right]^2 \frac{1}{2\alpha} - \left[\frac{py}{p+q}\right]^2 \frac{1}{2\beta} \qquad (4.9)$$

for y_A^i and then using $\bar{y}^j = y_C^i = py_A^i/(p+q)$. The solution to (4.9) is

$$y_A^i = \frac{\beta}{\alpha+\beta}(p+q)(G+\frac{q}{p}L) \qquad (4.10)$$

which implies that

$$y_C^i = \bar{y}^j = \frac{\beta}{\alpha+\beta}p(G+\frac{q}{p}L). \qquad (4.11)$$

Comparison with the results from section 3.1 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 (3.17) gives the planner's welfare. When all researchers theorize first the planner's welfare is

$$W_T = \frac{p^2}{2\alpha}\left[\frac{\beta}{\beta+\alpha}\right](G + \frac{q}{p}L)^2. \qquad (4.12)$$

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

4.2 Type–j Bridges Worth Building, No Direct Verifiability

We now return to 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 4.1, 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 (3.2) and (3.3), and the optimal payments are given by equations (3.6) and (3.8').

All Theorize First

The planner maximizes expected utility subject to (3.9)–(3.12) and (4.1)–(4.6). This implies a solution in which

$$y_B^i = 0,\ y_A^j = y_B^j = y_C^j = \frac{p}{p+q}\, y_A^i = y_C^i \tag{4.13}$$

so that constraints (3.10) and (4.4) are slack, while the other six constraints bind. The solution for y_A^i is

Then the planner chooses y_A^i to maximize

$$y_A^i = \frac{\beta\left[pG + qL\right] + \alpha\left[rG + sL\right]}{\left[\frac{p}{p+q}\right]\left[\alpha+\beta\right]} . \tag{4.14}$$

Type–i Agents Theorize First, Type–j Look First

Obviously this case involves $y_B^i = 0$ and $y_C^j = 0$. Then the incentives compatibility constraints are (3.9), (3.11), (4.13), and (4.15) for the type–i agents and (3.3),

$$\frac{r}{r+s}\, y_A^j + \frac{s}{r+s}\, y_B^j \geq r y_A^i + (1-r-s)\, y_C^i \tag{4.15}$$

and

$$\frac{r}{r+s}\, y_A^j + \frac{s}{r+s}\, y_B^j \geq y_C^i \tag{4.16}$$

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

$$y_A^i = \frac{\beta\left[pG + qL\right] + \alpha\left[\frac{r}{r+s}\, G + \frac{r}{r+s}\, L\right]}{\left[\frac{p}{p+q}\right]\left[\alpha+\beta\right]}\,. \tag{4.17}$$

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 (3.2) and

$$\frac{p}{p+q}\, y_A^i + \frac{q}{p+q}\, y_B^i \geq p y_A^j + q y_B^j + (1-p-q)\, y_C^j \tag{4.18}$$

and

$$\frac{p}{p+q}\, y_A^i + \frac{q}{p+q}\, y_B^i \geq y_C^j \tag{4.19}$$

for the type–i agent, and (3.10), (3.12), (4.2), and (4.4) for the type–j agent. Then (3.2), (3.12), (4.1), and (4.4) bind, $y_B^i = 0$, $y_A^j = y_B^j = y_C^j = \frac{r}{r+s}\, y_A^i$, and

$$y_A^i = \frac{\beta\left[\frac{p}{p+q}\right]^2 \left[G + \frac{q}{p} L\right] + \frac{\alpha r^2}{r+s}\left[G + \frac{s}{r} L\right]}{\beta\left[\frac{p}{p+q}\right]^2 + \alpha\left[\frac{r}{r+s}\right]^2} . \tag{4.20}$$

Welfare and Discussion

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

$$W^L = \frac{\left\{\beta\left[\frac{p}{p+q}\right]^2 \left[G + \frac{q}{p} L\right] + \alpha\left[\frac{r}{r+s}\right]^2 \left[G + \frac{s}{r} L\right]\right\}^2}{2\beta\left[\frac{p}{p+q}\right]^2 + 2\alpha\left[\frac{r}{r+s}\right]^2} . \tag{4.21}$$

When all researchers theorize first, welfare is

$$W^T = \frac{1}{2}\left[\frac{\alpha\beta}{\alpha+\beta}\right]\left\{\frac{p}{\alpha}\left[G + \frac{q}{p} L\right] + \frac{r}{\beta}\left[G + \frac{s}{r} L\right]\right\}^2 . \tag{4.22}$$

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

$$W^{TL} = \frac{1}{2}\left[\frac{\alpha\beta}{\alpha+\beta}\right]\left\{\frac{p}{\alpha}\left[G + \frac{q}{p}L\right] + \frac{r}{\beta(r+s)}\left[G + \frac{s}{r}L\right]\right\}^2 . \qquad (4.23)$$

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

$$W^{LT} = \frac{\left\{\beta\left[\frac{p}{p+q}\right]^2\left[G + \frac{q}{p}L\right] + \alpha\left[\frac{r^2}{r+s}\right]\left[G + \frac{s}{r}L\right]\right\}^2}{2\beta\left[\frac{p}{p+q}\right]^2 + 2\alpha\left[\frac{r}{r+s}\right]^2} . \qquad (4.24)$$

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 (4.21) and (4.23). With some tedious algebra one can show that

$$W^L > W^{TL}. \qquad (4.25)$$

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 4.1, 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 3 shows. In those

circumstances, the inability to verify research strategies reduces welfare. This could justify costly procedures designed to keep researchers honest.

5. Choice of Experiments to Perform

Until now, we have assumed that researchers take their data sets as given, and that the only decision they have is whether to look at the data before or after theorizing. In doing so, we have ignored the possibility that the researcher may have discretion over what facts to examine (i.e. what experiments to perform), and that theorizing may help him to choose which facts are most relevant.

In this section, we return to the simpler environment of a single type of researcher. In section 5.1 we present a very simple model of scientific research that incorporates both the "look first–theorize first" decision and the researcher's decision over what experiment to perform. This embodies the idea that one of the benefits of theorizing first is that it enables the researcher to choose more wisely what data to examine or what experiment to run. In section 5.2 we extend the analysis to the choice of experimental design.

5.1 Theories Suggest Experiments

We now assume that the universe of theories consists of two types, called A and B. Type A theories are in fact correct and type B are incorrect. The key assumption in this section is that the process of theorizing tells you what experiment to perform. In other words, the process of theorizing both produces a theory *and* suggests an experiment. With probability p, a

researcher constructs a theory of type A and performs an associated experiment. Because theory A is true, it is never rejected by that experiment. With probability $q = 1 - p$, he constructs a theory of type B and performs an associated experiment. This experiment rejects the theory with probability ρ.

A researcher who looks first chooses an experiment randomly. With probability $\pi < 1$ the experiment tests an implication of type B theories. That experiment rejects those theories with probability ρ. The fact that $\pi < 1$ builds in the aforementioned advantage to theorizing first. The disadvantage to theorizing first, however, is that the researcher uses less information than one who looks first. In short, theorizing first provides information about what experiment to perform; looking first provides information about what theory to choose.

The social gain from a true theory is $G > 0$, the gain from a false theory is $L < 0$, and the gain from no theory at all is 0. A research strategy j that leads to a true theory with probability γ_G^j and a false theory with probability γ_L^j yields $W_j = \max \{\gamma_G^j G + \gamma_L^j L, 0\}$. The question is which research strategy is preferred.

Suppose a researcher looks first. With probability $\pi\rho$, he rejects theories of type B, and he will then choose a correct theory with probability one. With probability $1 - \pi\rho$, his experiment will not reject B and he will then choose a correct theory with probability p. This implies that the probability of coming up with a correct theory is $p(1 - \pi\rho) + \pi\rho = p + q\pi\rho$. So we have

$$W_{look} = \max \{(p + q\pi\rho)G + q(1 - \pi\rho)L, 0\}$$

The corresponding calculations for theorizing first lead to

$$W_{\text{theorize}} = \max \{pG + q(1 - \rho)L, 0\}$$

Several points are worth noting. First, as in the previous sections of the paper, the probability of coming up with the correct 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 an incorrect theory is *also* greater under looking first, because the experiment chosen is not expected to be as informative.

Whether theorizing first is preferred to looking first depends only on π, G, and L. In fact theorizing dominates looking if and only if $\frac{G}{|L|} < \frac{1-\pi}{\pi}$. It is remarkable that neither p nor ρ plays any role in determining which strategy is preferred. 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 theory that is true (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 an incorrect 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 might 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 might 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.

When researchers theorize first, they are guided by previously known facts. When researchers look first, they are pesumably guided by previously constructed theories. In the simple model we have just presented, there are no previously constructed theories. When there is a previously constructed theory, a new question arises. A researcher who theorizes first can either perform an experiment to test the common implications of the new and previously constructed theories or perform an experiment to distinguish between them. This choice would also arise in any dynamic extension of our model. In the next section we present a model that enables us to discuss these issues can either perform an experiment to test the common implications of the new and previously constructed theories or perform an experiment to distinguish between them. This choice would also arise in any dynamic extension of our model. In the next section we present a model that enables us to discuss these issues.

5.2 Theories Guide the Choice of Experiments

As in earlier sections, we assume there are three types of theories, A, B, and C, consistent with the previously known facts. A is true while B and C are false.

One theory, called the "old theory," has already been proposed.[3] Researchers do not know if it is true, or if it will be consistent with results from experiments that have yet to be performed. So they do not know if the old theory is type A, B, or C. The probabilities that it is type A, B, or C are p, q, and (1–p–q).

There are four experiments, a, b, c, and bc, suggested by the old theory. These experiments, when performed, result in facts that are consistent with theories as in the following table:

	Theory		
Fact	A	B	C
a	x		
b	x	x	
c	x		x
bc	x	x	x

x = fact is consistent with theory
A blank space indicates the fact is inconsistent with the theory.

Because theory A is true, all experiments yield facts that are consistent with it. Experiment b yields a fact that is also consistent with theory B; experiment c yields a fact that is consistent with theories A and C; the fact from experiment bc is consistent with all three theories.

All researchers are alike. There is one researcher per project. As before, the social planner decides whether to build a bridge and gets payoffs

---

[3]The old theory can be thought of as coming from a previous round of this game.

G, L, or zero. If a researcher looks first, he researcher chooses an experiment randomly: π_i is the probability that he chooses experiment i. He then chooses a theory consistent with the fact that results from the experiment. The planner then chooses whether to build, and which theory, old or new, to use as a guide.

If a researcher theorizes first, he chooses a new theory randomly: he chooses a theory of type A, B, or C with probabilities p, q, and 1-p-q. He then examines the implications of the new theory and contrasts them with those of the old theory. It is possible that the new and old theories are observationally equivalent: They have identical implications if both theories are the same type. In that case, it is impossible to perform an experiment to discriminate between the two theories, but it is possible to test their common implications.

If the new theory is not the same type as the old theory, then an experiment can be performed to discriminate between them. Alternatively, an experiment could test their common implications.[4] If the new and old theories differ, a researcher who theorizes first must choose between these two types of experiments: "type-one" experiments that distinguish between the two theories and "type-two" experiments that test their common implications. The researcher chooses which type of experiment to conduct and the chooses randomly among experiments of that type. We assume that the researcher acts to maximize the planner's expected utility.

We prove the following results in the Appendix.

---

[4]For example, if the old and new theories are types A and B, they have common predictions about the outcomes of experiments b and bc, but different predictions about the outcomes of experiments a and c.

If the researcher looks first, the social planner's expected utility is

$$W^L = \max\{(G-L)[q\pi_a+q\pi_c p/(1-q)+(1-p-q)\pi_a+(1-p-q)\pi_b p/(p+q)] + L[q(\pi_a+\pi_c)+(1-p-q)(\pi_a+\pi_b)], 0\} + \max\{(G-L)p + L[p+q(\pi_b+\pi_{bc})+(1-p-q)(\pi_c+\pi_{bc})\}, 0). \quad (5.1)$$

If the researcher theorizes first and chooses a type-one experiment — that tests the new against the old theory — the planner's expected utility is

$$W^T(exp=t1) = [p^2+q^2(\pi_b+\pi_{bc})+(1-p-q)^2(\pi_c+\pi_{bc})] \max\{0, [G-L]p^2/[p^2+q^2(\pi_b+\pi_{bc})+(1-p-q)^2(\pi_c+\pi_{bc})] + L\} + [2p(1-p)+2q(1-p-q)] \max\{0, (G-L)2p(1-p)/[2p(1-p)+2q(1-p-q)] + L\}. \quad (5.2)$$

If the researcher theorizes first and chooses a type-two experiment — that tests the common implications of the new and old theories, the planner's expected utility is

$$W^T(exp=t2) = [p^2+q^2(\pi_b+\pi_{bc})+(1-p-q)^2(\pi_c+\pi_{bc})] \max\{0, [G-L]p^2/[p^2+q^2(\pi_b+\pi_{bc})+(1-p-q)^2(\pi_c+\pi_{bc})] + L\} + [2p(q+(1-p-q))+2q(1-p-q)\pi_{bc}/(\pi_a+\pi_{bc})] \max\{0, (G-L)p(q+(1-p-q))/[2p(q+(1-p-q)+2q(1-p-q)\pi_{bc}/(\pi_a+\pi_{bc})] + L\} \quad (5.3)$$

A researcher first decides whether to look first or theorize first. If he theorizes first, the probability that the new theory has the same implications as the old theory is $\xi = p^2 + q^2 + (1-p-q)^2$. In this case, the theorist cannot perform a type-one experiment, so he chooses a type-two experiment. With probability $1 - \xi$ the two theories have different implications and the researcher chooses a type-one experiment if $W^T(exp = t^1) > W^T(exp = t^2)$ and a type-two experiment otherwise. So if a researcher theorizes first, the planner's expected utility is

$$(5.4) \qquad W^T = \xi W^T(exp = t2) + (1 - \xi) \max \{W^T(exp = t1), W^T(exp = t2)\}$$

The planner's expected utility is then the maximum of W^L and W^T.

The optimal research strategy depends in a complicated way on the model parameters and can take either sign. For example, if G = 10 and L = −15, p = .2, q = .1, and $\pi^a = \pi_b = \pi_c = \pi_{bc} = .25$, then $W^L = 2.05$, while $W^T|exp=t1$ is only 1.10 and $W^T|exp=t2$ is zero. With these parameters, a researcher should look first. But with these parameter values and p = .5 the researcher should theorize first and perform a type-one experiment (if it is feasible): $W^L = 3.06$, while $W^T|exp=t1 = 5.02$ and $W^T|exp=t2 = 1.02$, and $W^T = 3.34$. Although we have not been able to prove it, we conjecture that, in our model, a researcher would alsways prefer a type-one to a type-two experiment if the former is feasible. This is a remarkable result. The question of its applicability to more general models will have to await further research.

Would information about whether a researcher theorized first or looked first convey any information to a person reading a research report in a scientific journal? To make sense of this question in the present model we

must assume that the reader is uncertain about the underlying parameter values. The research strategy can only convey information about the reward structure established by the social planner, which in turn conveys some information about parameter values. These parameter values may in turn convey information about the likelihood that a theory resulting from the optimal strategy is true. (For example, parameter values for which the social planner chooses rewards to make researchers theorize first might tend to be parameter values for which there is either a very high or a very low probability that the optimal strategy leads to a true theory.) Whether a theory is more likely to be true if one research strategy rather than another is followed depends on the probability distribution of these parameter values.

6. 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[5] 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

---

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

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 3 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 5 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 3 shows that, for some range of parameters, a subset of the most valuable scientists should be assigned to the theorize–first strategy. However, if research strategies are private information, then at the optimum all scientists look first and welfare is lower. The model of Section 5 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 the questions we have posed.

APPENDIX

1. Proof of equation (5.1):

Suppose the researcher looks first. We first derive the probability that the new theory is A conditional on the old theory being rejected. The old theory is rejected in four possible states of nature, where the state consists of the pair (old theory, fact). For example, in state Ba the old theory is B and the fact generated by the experiment is a. The probabilities of these states and the probabilities that the new theory is A in each is given in the following table:

state	prob. of state	prob. (new theory is A)
Ba	$q\pi_a$	1
Bc	$q\pi_c$	$p/(p+(1-p-q))$
Ca	$(1-p-q)\pi_a$	1
Cb	$(1-p-q)\pi_b$	$p/(p+q)$

So the probability that the new theory is A (NT=A) given that the old theory has been rejected (OTR) is

$$P(NT=A|OTR) = \frac{q\pi_a+q\pi_c p/(p+(1-p-q))+(1-p-q)\pi_a+(1-p-q)\pi_b p/(p+q)}{q\pi_a+q\pi_c+(1-p-q)\pi_a+(1-p-q)\pi_b}$$

The numerator can be rewritten as

$$[(q+(1-p-q))(p+q(1-p-q))\pi_a + p(1-p-q)(1-q)\pi_b$$

$$+ pq(1-\pi_c)(1-p-q)]/[(1-q)(1-(1-p-q))].$$

Next, suppose the old theory is <u>not</u> rejected (OTNR). This can happen in eight possible states. Let NT=A mean that the new theory is A, and OT=A mean that the old theory is A. Then the analogous table is

state	prob. of state	prob. (OT=A)	prob (NT=A)
Aa	$p\pi_a$	1	1
Ab	$p\pi_b$	1	p/(p+q)
Ac	$p\pi_c$	1	p/(p+(1−p−q))
Abc	$p\pi_{bc}$	1	p
Bb	$q\pi_b$	0	p/(p+q)
Bbc	$q\pi_{bc}$	0	p
Cc	$(1-p-q)\pi_c$	0	p/(p+(1−p−q))
Cbc	$(1-p-q)\pi_{bc}$	0	p.

So the probability that the old theory is A (OT=A) given that the old theory is not rejected (OTNR) is

$$\text{Prob}(\text{OT=A}|\text{OTNR}) = \frac{p\pi_a+p\pi_b+p\pi_c+p\pi_{bc}}{p\pi_a+p\pi_b+p\pi_c+p\pi_{bc}+q\pi_b+q\pi_{bc}+(1-p-q)\pi_c+(1-p-q)\pi_{bc}}$$

$$= \frac{p}{p+q(\pi_b+\pi_{bc})+(1-p-q)(\pi_c+\pi_{bc})} .$$

This also equals the probability that the new theory is A (NT=A) given that the old theory is not rejected, i.e. Prob(OT=A|OTNR) = Prob(NT=A|OTNR), and we can assume w.l.g. that the planner always uses the old theory to build (if he builds) when both theories are supported by the facts.

The planner will build if the expected return from building exceed zero. So his utility is W^{look}

$$= \text{prob(OTR) max\{Gprob(NT=A|OTR) + L[1-prob(NT=A|OTR)], 0\}}$$
$$\text{+ prob(OTNR) max\{Gprob(NT=A|OTNR) + L[1-prob(NT=A|OTNR)], 0\}}$$

$$= \max \{(G-L)[q\pi_a+q\pi_c p/(p+(1-p-q))+(1-p-q)\pi_a+(1-p-q)\pi_b p/(p+q)]$$
$$+ L[q\pi_a+q\pi_c+(1-p-q)\pi_a+(1-p-q)\pi_b], 0\}$$
$$+ \max \{(G-L)p + L[p+q(\pi_b+\pi_{bc})+(1-p-q)(\pi_c+\pi_{bc})], 0\},$$

which is the result (5.1).

2. Derivations of equations (5.2) and (5.3).

Consider the planner's utility if the researcher theorizes first. There are six possible combinations of old theory and new theory. In three cases (in which the old and new theories are both A, both B, or both C), the two theories have the same implications about the outcomes of all experiments. In this case, the researcher either rejects both theories or fails to reject both. If he fails to reject both, the probability that the old theory is A given that the old and new theories have identical implications and that the theories were not rejected equals the probability that the new theory is A given those same conditions. So, without loss of generality, suppose the planner uses the old theory to build if he chooses to build. The probability that the old theory is A under these conditions is

$$\text{prob(OT=A|OT=NT, not rejected)} = \frac{\text{prob(OT=A, OT=NT, not rejected)}}{\text{prob(OT=NT, not rejected)}}$$

$$= \frac{p^2}{p^2 + q^2(\pi_b+\pi_{bc}) + (1-p-q)^2(\pi_c+\pi_{bc})}$$

where the probability that the old and new theory have the same implications and are not rejected is the denominator.

Suppose next that the two theories differ in their implications. Suppose first that the researcher limits himself to type-one experiments, for which the two theories have different implications. Denote this strategy by exp=t1. Then one theory will always be rejected, and one theory not rejected. Call the latter (not-rejected) theory the established theory and denote by est=A the event that the established theory is A. The possible states are (where, for example, the state AB means that the old theory is A and the new theory is B),

state (OT,NT)	prob. of state	prob. (est=A|exp=t1)
AB	pq	1
AC	p(1-p-q)	1
BA	qp	1
BC	q(1-p-q)	0
CA	(1-p-q)p	1
CB	(1-p-q)q	0

So the probability that the established theory is A given that the old and new theories differed and that a type-one experiment was chosen so that one theory was rejected is

$$\text{prob}(\text{est=A}\,|\,\text{OT}\neq\text{NT},\ \text{exp=t1}) = 2p(1-p)/[2p(1-p)+2q(1-p-q)]$$

where the denominator (which could be rewritten as $(1-p^2-q^2-(1-p-q)^2)$ is the probability that old theory and the new theory differ in their implications.

Then the planner's utility is

$$W^{th}|exp=t1 =$$

$$prob(OT=NT, \text{ not rej}) \max\{0, prob(OT=A|OT=NT, \text{ not rejected})G$$
$$[1-prob(OT=A|OT=NT, \text{ not rejected})]L\}$$
$$+ prob(OT=NT, \text{ rejected}) \max(0, L)$$
$$+ prob(OT\neq NT) \max\{0, prob(est=A|OT\neq NT, exp=t1)(G-L) + L\}.$$

which equals the expression in equation (5.2).

Finally, suppose that the researcher limits himself to type-two experiments, for which the two theories have the same implications. The possible states (in which the theories differ) are

state (OT,NT)	prob. of state	type-two experiments	prob. (not reject|state)
AB	pq	b,bc	1
AC	p(1-p-1)	c,bc	1
BA	qp	b,bc	1
BC	q(1-p-q)	a,bc	$\pi_{bc}/(\pi_a+\pi_{bc})$
CA	(1-p-q)p	c,bc	1
CB	(1-p-q)q	a,bc	$\pi_{bc}/(\pi_a+\pi_{bc})$

The planner does not build if the theories are both rejected because U_c is negative. The probability that the theories differ and are not rejected, given that the researcher is limiting himself to type-two experiments, is

$$prob(OT\neq NT, \text{ neither rej})$$
$$= 2p(q+(1-p-q)) + 2q(1-p-q)\pi_{bc}/(\pi_a+\pi_{bc}).$$

The probability that the old theory is A given that the theories differ but neither is rejected is

$$\text{prob}(OT=A \mid OT \neq NT, \text{neither rej}) = \frac{p(q+(1-p-q))}{2p(q+(1-p-q))+2q(1-p-q)\pi_{bc}/(\pi_a+\pi_{bc})} .$$

So the planner's utility if the researcher theorizes and performs a type-two experiment is $W^{th}|exp=t2$ =

$$\begin{aligned}
&\text{Prob}(OT=NT, \text{not rej}) \max\{0, \text{prob}(OT=A|OT=NT, \text{not rejected})G \\
&\qquad\qquad [1-\text{prob}(OT=A|OT=NT, \text{not rejected})]L\} \\
&\quad + \text{prob}(OT=NT, \text{rejected}) \max(0, L) \\
&\quad + \text{prob}(OT \neq NT, \text{neither rej}) \\
&\qquad \max\{0, \text{prob}(OT=A|OT \neq NT, \text{neither rej})(G-L) + L). \\
&\quad + \text{prob}(OT \neq NT, \text{both rej}) \max\{0, L\}
\end{aligned}$$

which is equal to the expression in equation (5.3).

REFERENCES

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

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.

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