Economic Approaches to Understanding Scientific Norms

Michael Strevens

To appear in *Episteme*

Abstract

A theme of much work taking an "economic approach" to the study of science is the interaction between the norms of individual scientists and those of society at large. Though drawing from the same suite of formal methods, proponents of the economic approach offer what are in substantive terms profoundly different explanations of various aspects of the structure of science. The differences are illustrated by comparing Strevens' explanation of the scientific reward system (the "priority rule") with Max Albert's explanation of the prevalence of "high methodological standards" in science. Some objections to the economic approach as a whole are then briefly considered.

Falling under the rubric of "economic approaches" to the study of science are a wide range of projects employing different methods to study diverse phenomena. The purpose of this paper is to sketch an enterprise that unifies much, though not all, philosophical work on this theme, namely, the investigation of the interface between social norms and individual interests in science; to examine what I take to be a major divergence of explanatory styles within the enterprise, using Albert's paper in this issue of *Episteme* and Strevens (2003) as exemplars of contrasting approaches; and briefly to defend the enterprise as a whole against some important objections.

I will not attempt a literature review, but readers new to the area might browse, in addition to the work discussed in this paper, Sarkar (1983), Gold-

man and Shaked (1991), the parts of Kitcher (1993, chap. 8) not directly concerned with the impact of credit allocation on the division of cognitive labor, Dasgupta and David (1994), Rueger (1996), Brock and Durlauf (1999), Strevens (2006), Weisberg and Muldoon (2009), Strevens (2010), and Zollman (2010)—to cite just a smattering of recent work taking an "economic approach" to science (not all of it concerned explicitly with norms).

1. Norms and Interests

We used to live in caves; now we live in condominiums. We used to huddle around the campfire; now we lounge in front of the HDTV. We used to be killed by the plague; now we are killed by partially hydrogenated vegetable oil. Science has brought us great success.

The old-fashioned philosophical explanation of this success invokes a special scientific method for truth-finding—in its most austere form, nothing more than a theory of confirmation, that is, a theory of what evidence supports what theories to what degree. There is much to be said for this explanation, but it seems almost certainly not to be the full story. Science is special not only in its attention to the evidence but in its social organization as well, and it is rather plausible, as a number of writers have noted, that science's social peculiarities are in part responsible for its epistemic prowess.

Merton (1942) argued for the importance of such norms as universalism, communism (common ownership of intellectual property), disinterestedness, and organized skepticism, to which he later added the sometimes apparently arbitrary norms governing the allocation of scientific credit (Merton 1957, 1968).

The non-epistemic norms of science entered mainstream philosophy with Popper and Kuhn. For Popper, norms were the meat on the bones of his exiguous account of evidential support, the falsificationist theory. It was the norms that enjoined scientists to conjecture boldly and refute tirelessly, to subject new auxiliary hypotheses introduced in ad hoc reasoning

to witheringly severe testing, or to avoid ad hoc reasoning altogether in favor of more virile choices. Scientists, Popper thought with some reason, would be eager to shoulder heroically burdens such as these.

Kuhn, by contrast, was happy to leave the logic of confirmation to others: his central thesis concerns the social organization of science, or more exactly, the organization of what he calls *normal science*. What distinguishes science from all other human endeavors, writes Kuhn, and in particular, from all other knowledge-seeking endeavors, is its psychology and sociology. Whereas a diversity of approaches is the usual state of affairs in philosophy, or theology, or literary criticism, science is not even possible without a narrow-minded methodological consensus inherent in what Kuhn calls a paradigm (there is no need here to unpack the many senses in which Kuhn used that word).

It was the 1960s' "new look" psychology that most saliently policed the borders of the paradigm for Kuhn, but it is clear that sociological factors—the whole apparatus of "professionalization", in Kuhn's terms—were crucial to carving normal science from the undifferentiated human substrate, and so to the success of science in exploring new theoretical possibilities:

> Professionalization leads, on the one hand, to an immense restriction of the scientist's vision and to a considerable resistance to paradigm change. . . . On the other hand, within those areas to which the paradigm directs the attention of the group, normal science leads to a detail of information and to a precision of the observation-theory match that could be achieved in no other way . . . [and without which the observations] that lead ultimately to novelty could not occur. (Kuhn 1996, 64–65)

Without conformist social norms, then, no intellectual innovation—no progress.

Kuhnian norms are, it seems, strictly top-down: the individual either meekly submits to social reality or becomes a scientific outcast. Yet it is all for the best. These are the norms that—however counterintuitive it may

seem—find the optimal scientific outcomes. *The Structure of Scientific Revolutions* was, after all, written in the heyday of group selection, functionalist explanation in the social sciences, and Organization Man.

Why suppose, though, that it will not occur to working scientists, when they see the prospect of some gain, to violate the norms, or to manipulate them to their own advantage? It is hardly possible to believe any more, if it ever was, that scientists are made of morally better stuff than the rest of us, and they have ample opportunity to exploit the system: a university scientist typically runs their own lab, determining what is to be explored there, which post-doctoral fellows to employ, and how to put these fellows and graduate students to work. They decide what to publish and what not to publish, and more generally which of their findings to share with their rivals and the world at large. They are intellectual entrepreneurs. Like all entrepreneurs, they must obtain funding from somewhere, but as in business, so in science: the funders play a largely passive role, choosing among proposals that are put before them, rather than actively determining the course of future research.[1]

It is at the interstices of social norms and individual interests that much work on the structure of science has grown. In the mid-twentieth century, Polanyi (1962) claimed that scientists' free individual choices led to a kind of tacit, unintended cooperation in the pursuit of research that was more efficient than anything within the reach of central planning. Later sociologists of science, especially those associated with the "Strong Program" (Barnes and Bloor 1996), have argued among other things that many epistemic and social features of science should be understood as consequences of self-interested, Machiavellian empire-building similar to that found in any human institution. Finally, philosophers of science have more recently introduced formal methods—computer modeling, decision theory, and other mathematical techniques—to understand the aggregate effect of individual scientists' deci-

1. "Largely", but not entirely: the NSF has its initiatives, as do venture capitalists and state and national governments dabbling in venture capitalism—the new industrial policy.

sions either in concert with or independent of social norms.

It is this last kind of research that is often considered to be taking an "economic approach" to the structure of science. What makes it redolent of economics, then, is on the one hand something it shares with Polanyi, the Strong Program, and so on, a concern with the macrolevel effects of individual choice, and on the other hand something that differentiates it, formal mathematical models and methods. It should be said, however, that the economic approach is typically concerned neither with markets nor with the place of the production of knowledge within the wider economy. Its relation to economics is less that of part to whole than of family resemblance.

There are many different schemes for taxonomizing work on scientific norms. You might, for example, classify according to moral tenor: corresponding respectively to Polanyi, the Strong Program, and economic philosophy of science you would then have the utopian, the cynical, and the technocratic approaches. Or you might more usefully classify according to the interaction envisaged between social norms and individual interests, resulting in the following three-way division.

First are the stories that posit a withering away of the scientific norm (if it ever existed). Where there appear to be norms governing the behavior of scientists as a whole, there are in reality only individual decisions adding up to something sufficiently beneficial for science or society as a whole as to create the illusion of active cooperation or the imposition of standards from above. Polanyi's "republic of science" fits this mold: it is a republic in the old-fashioned sense, an association of intellectual gentry whose unity stems not from one small group's political hegemony but from each individual's acting out of a natural sociability and integrity. Even the purely intellectual norms of science—the norms that determine a hypothesis's intrinsic interest or degree of evidential support—do not for Polanyi subsist in individuals' recognition of overarching rules but in the standards of individual groups of scientists, forming a patchwork of overlapping neighborhoods, this over-

lap being enough to establish "uniform standards of merit". The uniform standards, then, are explained by, rather than explaining, the opinions of individuals.

Under this heading I would also put Goldman and Shaked's (1991) explanation of science's truth-finding power and Albert's explanation of "high scientific standards", the latter because in Albert's view what explains high standards is not a norm enjoining scientists to aim high, but rather their personal interest in having their work used by other scientists.

In the second class are the stories that posit the alignment of norms and individual motivation. The norms are real and have motivating power, but they are not irresistible. The reason that scientists do not, on the whole, resist them is that they see that it is in their own interest to conform. Under this heading I would put Hull (1988)'s work explaining the relatively greater punishment accorded to scientific fraudsters as opposed to plagiarists. Hull does not deny that there are norms outlawing both fraud and plagiarism; what he wants to show is first, that even a scientist immune to the norms' motivating power has reason to fear and therefore to punish both plagiarism and fraud, and second, that this individual motivation can perhaps explain why the norm against fraud is enforced more brutally than the norm against plagiarism.

To achieve his second aim, Hull notes that fraud stands to harm scientists far more than plagiarism because it has a far greater number of victims: plagiarism harms only the plagiarized scientist, whereas fraud harms all who build on the fraudulent work. Your probability of being harmed by an act of fraud is much greater than your probability of being harmed by an act of plagiarism, therefore you have an interest in punishing the former act more severely than the latter. (I will not go into the complications in connecting possible harm and punishment here.) In Hull's explanation, then, interests not only reinforce, but also effectively *modulate* norms, that is, influence the parameters of the rules in question, in this case, the prescribed level of

punishment.

My third class of stories are those that posit a more complex interaction between norms and interests: the two work together, not merely by pulling in the same direction, but by, as it were, driving different parts of a single, cohesive motivational system. Hull's modulation is a step in this direction; my principal other examples are the work of Kitcher and Strevens, to be examined in the next section.

I should note that different scientific norms might be treated in different ways: some might turn out, on closer inspection, to be nothing more than the large-scale expression of individual interests (class one); some might be reinforced by individual interest (class two); and some might interact with interest in more sophisticated ways (class three).

2. The Division of Cognitive Labor and the Priority Rule

To provide a contrast with Albert, let me sketch Kitcher's (1990) and Strevens's (2003) investigations of the system of credit allocation in science, sometimes called the "priority rule". For reasons that will become clear, the discussion begins with the apparently separate issue of the division of cognitive labor—that is, of how best to allocate resources among different possible scientific projects.

Peirce (1879) examines the question of how limited resources should be divided between two scientific studies. For each study he assumes the existence of a cost-utility function representing the amount of knowledge, or utility, to be gained by allocating any particular quantity of resources to the study. He makes two further assumptions:

> *Additivity.* The utility produced by one study is independent of the utility produced by the other, so that the total utility derived from pursuing the two studies in concert is equal to the sum of the individual utilities.

> *Decreasing marginal returns.* Each additional unit of resources allocated to a program brings a smaller increment of utility than the last. Peirce justifies the assumption with these solemn words (p. 198): "When chemistry sprang into being, Dr. Wollaston, with a few test tubes and phials on a tea-tray, was able to make new discoveries of the greatest moment. In our day, a thousand chemists, with the most elaborate appliances, are not able to reach results which are comparable in interest with those early ones. All the sciences exhibit the same phenomenon, and so does the course of life. At first we learn very easily, and the interest of experience is very great; but it becomes harder and harder, and less and less worth while, until we are glad to sleep in death." On a lighter note, marginal returns never reach zero; there is always some benefit to investing more, however slight.

Call this the *Peirce model*.

From the Peirce model's assumptions and the familiar mathematics of constrained maximization problems, Peirce derives the conclusion that maximum utility will be realized when resources are distributed so as to equalize marginal returns, that is, so that spending one more resource unit on either of the studies would bring approximately the same return: "When new and promising problems arise they should receive our attention to the exclusion of the old ones, until their urgency becomes no greater than that of others" (198). Except in extreme cases this Peircean optimum will distribute resources to both studies.[2]

⋆ ⋆ ⋆

Peirce used his model to determine the best policy for some master distributor of resources—perhaps the superintendent of the US Coastal Survey—to follow. In academic science, there is no superintendent. As observed above,

2. Peirce's model represents only two possible avenues of research, but his conclusion is quite general, holding for circumstances in which any number of studies are competing for resources.

the course of future research is largely determined by the individual decisions of the world's myriad scientists.

Kitcher (1990, 1993) uses the Peirce model to raise the question whether these decisions achieve anything like Peirce's optimal distribution. To set up the problem, he supposes that the payoff from a scientific research program is stochastic: the program might succeed in providing knowledge, but then again it might not, in which case it provides nothing epistemically worthwhile. (Kitcher builds the probabilities into the Peirce model's cost-utility function; in Kitcher's hands, then, the function determines the *expected* utility returned for a given investment of resources.)

Now, writes Kitcher, consider scientists motivated only by the search for truth. Suppose that such scientists are choosing between two research programs. Suppose also that these programs' cost-utility functions do not intersect: there is one program that, for any given amount of resources, returns a higher expected utility than the other. Call the program that yields the consistently higher return the higher-potential program. A scientist who desires only to learn the truth ought to join the higher-potential program, because if joined, it is more likely than the other (if joined) to deliver the sought-after knowledge. Their joining the higher-potential program will only increase its prospects relative to those of the lower-potential program. Thus the next scientist who comes along will be even more motivated to join the higher-potential program. And so on: all of the truth-seeking scientists will join the higher-potential program; the lower-potential program will be left entirely unstaffed.

This is (except in extreme cases) very far from the Peircean optimum, which will allocate some resources to each program—hedging society's epistemic bets, as it were—though more to the higher potential program. We have a problem: researchers choosing their scientific approach solely on the grounds of personal knowledge maximization distribute themselves among approaches in a way that does anything but maximize the total production of

knowledge.

It is just as well, then, that scientists care about more than personally learning the truth. Following Merton, Hull, and other sociologically minded thinkers, Kitcher takes seriously the idea that scientists are principally motivated by the desire to receive credit for their ideas and discoveries—they are interested in their reputation, in scientific status.[3] Imagine a scientist who cares about nothing else: when deciding which research program to join, they consider only the credit they can reasonably expect to accrue, that is—given the uncertain nature of scientific success—their *expected credit* in the statistical sense. Such a scientist will take into account not only the probability of their program's succeeding in realizing its goal, but also the number of other scientists with whom they will have to share the credit if the program does succeed. (Here it is assumed that credit, like prize money, is divided among the participants in the winning program; the more participants, then, the smaller is each scientist's personal share.) There will come a point when so many scientists have joined the higher-potential program that the next scientist's expected credit is greater if they join the lower-potential program: if they do so, they have a lower chance of receiving any credit at all, but the amount they stand to receive more than makes up for the low probability.

Credit-driven scientists will, as consequence, tend to distribute themselves among the available programs, rather than concentrating their numbers in the highest-potential program. In so doing, they come closer to the Peircean optimum than they would if they were concerned only with truth—or at least, with their personal proximity to truth. Kitcher tentatively concludes that

> Social institutions within science might take advantage of our personal foibles to channel our efforts toward community goals

3. Kitcher (1990) at first talks of a "much-coveted prize" (15), alluding to the Nobel Prize that brings as much cash as credit, but later introduces a scientist's "desire to be singled out by posterity as an early champion of [some] accepted theory" (21).

> rather than toward the epistemic ends that we might set for ourselves as individuals (1990, 16).

That is, the social norm that determines who gets credit for what might arrange incentives so that scientists pursuing credit for their own gratification rather than the happiness of the many end up arranging themselves so as to promote the good of science or society as a whole. This would be the kind of interplay between—rather than alignment or inter-substitutability of—norms and interests introduced in section 1.

In order for "social institutions within science" to have this happy effect, two things would have to be manipulated: first, the social norm that determines who gets credit for what, and second, the individual scientist's value system, in order to ensure that researchers pursue credit rather than alethic intimacy. Or perhaps not—perhaps such norms and values are already in place. So Kitcher suggests.

⋆ ⋆ ⋆

Suppose that Kitcher is right: a social norm, the scientific credit system, directs naturally credit-hungry scientists to research programs so as to realize a socially beneficial distribution of cognitive labor. Putting aside the questions of how such a credit system evolved, or why scientists place so much value on its emoluments, you might ask: just how well tuned is the system? How close does it come to the Peircean optimum?

Very close, according to Strevens (2003). The justification for this answer has two steps. In the first, the additivity assumption made by Peirce and Kitcher is retained, meaning that the combined utility of the various available research programs, were they to realize their aims, would be equal to the sum of the utilities provided by their success in isolation.[4] Strevens shows that, given additivity, there is a simple scheme for assigning credit for scientific

4. Kitcher, recognizing that some research programs are engaged in what I call below "winner-contributes-all" competition, assumes only approximate additivity, which is good enough for his qualitative purposes.

success that will, provided that scientists seek to maximize their expected credit, result in distributions of labor that are optimal, that is, that maximize total expected utility.

The scheme, which Strevens calls *Marge*, is as follows: reward scientists in proportion to the marginal increase in a program's expected utility that they bring when they join. Reward them, that is, according to their personal contribution to the program's overall expected utility. Scientists will then join the program to which they make the greatest incremental contribution. Under the additivity assumption, total expected utility is the sum of these incremental contributions, so the reward scheme effects a simple hill-climbing algorithm which, as a consequence of additivity and decreasing marginal returns, is guaranteed to find the highest point attainable with any given quantity of resources.

Utopia, then: reward scientists according to *Marge*—according to the contribution they make to the probability of success, weighted by the value of that success—and you maximize total expected value. This goes for any conception of value you like: if you want to maximize pure, deep knowledge, then weight the success probabilities according to a program's goals' theoretical merit; if you want to maximize benefit to society at large, then weight according to practical merit instead. Mix both weightings to achieve a desirable combination of both goods.

The result also holds for whatever notion of probability you like. If you are lucky enough to know the objective probabilities of the available programs' success (whatever that means), then allocate credit according to these probabilities and you will get what is the objectively best distribution of scientific labor. If your probabilities represent only reasonable beliefs about the programs' chances of success given your current state of knowledge, then allocate credit according to *these* probabilities and you will get what is *as far as you can determine* the best distribution of labor. The credit system is not a crystal ball; it is not going to tell you things about research programs that

scientists do not already know. But it can take what scientists do know and efficiently distribute them among research programs so as to realize what is, in the light of their knowledge, the optimally productive organization—quite an achievement, given the scale of the coordination problem.

There are two worms in the apple of knowledge: scientific research programs do not always have additive utilities, and the credit system in science departs in significant ways from *Marge*. The second step in Strevens' argument seeks to show that two worms are better than one: the relevant features of the reward system compensate for the non-additivity.

Let me begin with non-additivity. Consider two research programs that are competing to make the same discovery, for example, Kitcher's case of X-ray crystallographers and mechanical model-builders seeking to discover the structure of DNA, or the two teams at Fermilab seeking to demonstrate the existence of the top quark. Suppose that in one of these scenarios, both teams succeed in making the discovery. Additivity implies that two such successes are twice as good as one. But this is clearly false on any conception of utility that matters to us. Discovering the structure of DNA twice is not much better, quite possibly no better, than discovering it once. The competition between the two programs has a "winner-contributes-all" aspect: the first discovery brings great benefits to science or society, while the second discovery brings nothing at all.

Strevens analyzes non-additive winner-contributes-all scenarios using a modified Peirce model. As with additive scenarios, the optimal allocation splits resources between two competing programs (except in extreme cases). Compared to the additive case, however, the optimal distribution in the non-additive case assigns relatively more resources to higher-potential programs. Even in the additive case the best allocation favors higher-potential programs; in the winner-contributes-all case, higher-potential programs are favored even more. Since the optimal allocations are different in the additive and winner-contributes-all cases, the *Marge* incentive scheme, which

is optimal for the additive case, creates a sub-optimal distribution for the winner-contributes-all case.

Now compare the *Marge* system of credit allocation to the actual reward system in science. There are two differences. First, *Marge* rewards scientists whether their research programs succeed in realizing their goals or not. By contrast, the actual reward system gives credit only for actual discoveries.

Second, building on the first difference, the actual system only rewards the first program to make a given discovery. If the crystallographers and the model-builders both succeed, independently, in discerning the structure of DNA, but the model-builders do so before the crystallographers, and so are the first to publish the double helix, then the model-builders get all the credit for the discovery, and the crystallographers, just for being late, get nothing.[5] Merton (1957), writing about the reward system, notes a number of cases of independent discovery where this "winner-takes-all" rule was strictly applied, often spurring bitter disputes about who was in fact first to make a discovery.

When the interval between discoveries is short, the winner-takes-all rule of priority may seem arbitrary and capricious, but it seems to be enforced ruthlessly:

> Questions as to priority may depend on weeks, on days, on hours, on minutes

wrote François Arago, the permanent secretary from 1830 to 1853 of the French Academy of Sciences—a dictum which, Merton remarks, "only [puts into] words what many others have expressed in behavior" (Merton 1957, 322).

Whereas on *Marge*, scientists are rewarded—are given credit—in proportion to their *expected* contribution to utility, on the priority system, they are rewarded in proportion to their *actual* contribution to utility. This reference

5. If the discoveries are not made independently then the rival programs share the credit, as happened in actuality.

to a scientist's "actual contribution" captures both differences between *Marge* and the priority rule: first, a scientist makes no actual contribution if their program fails to realize its goal, and second, in a winner-contributes-all scenario the realization of a program's goal makes no difference to total utility if the goal has been reached already by some other program.

What is the practical difference between *Marge* and the priority rule? How will the distribution of scientists resulting from the priority rule—the distribution effected by scientists' expectation that they will be rewarded only for actual contributions—differ from the distribution resulting from *Marge*? Strevens shows that the priority rule results in relatively more scientists joining higher-potential programs. *Marge*'s distribution already favors higher-potential programs; the priority rule's distribution favors higher-potential programs even more.

But this is just what is needed to optimize the distribution of labor in winner-contributes-all cases: in such cases, relatively more scientists should be allocated to higher-potential programs, and the winner-takes-all reward scheme has precisely this effect.[6] Perhaps more than Kitcher imagined, then, there is a subtle interplay of social norms and individual interests at work in science, solving with élan the difficult problem of coordinating the decisions of the many so as to maximize the good of all.

Give me scientists who crave credit, some voice on high seems to declaim, and I will give you a system for allocating credit that arranges those scientists so as to produce the greatest amount of pure knowledge, social benefit, or whatever other good you desire.

3. Two Explanatory Strategies

In section 1, I laid out three broad headings under which to classify economic approaches to understanding scientific norms: some argue that there are

6. Whether it has this effect to precisely the right degree depends on parameters such as scientists' degree of risk aversion; a simple optimality result is therefore not available.

no norms, just aggregate effects of individual decisions; some argue that norms persist and perhaps originate because they are aligned with individual interests; some argue that there is a more complex interaction between norms and interests (thereby postponing any explanation of the norms' origins).

Let me now give you two specific explanatory strategies: a strategy that supposes that norms are internally generated by the pursuit of science, and a strategy that supposes that norms are in some sense externally imposed. The "internally generated" strategy is characteristic of the first approach above, on which what appear to be norms are nothing more than patterns of individual choice, while the "externally imposed" strategy is characteristic of the third approach, on which norms and individual choices have distinct roles to play in explaining the structure of science.

Rather than attempting to define the difference between "internally generated" and "externally imposed" norms, I will illustrate the distinction by contrasting Strevens and Albert. Albert asks: "Why is the average quality of scientific research so high?". Strevens' research on the priority rule might be seen as an answer to a different question of the same form: "Why is the allocation of cognitive labor in science so efficient?". The two studies, then, have somewhat different subject matters, but both turn on scientists' choices among possible research strategies, and both suppose that these choices are driven by the desire for credit or status. They do not, however, have the same view of credit allocation, Strevens taking what I will call an external view and Albert an internal view. In what follows I compare and contrast the internal and external views about the scientific reward system, in order to better illuminate the distinction between internal and external approaches and to assess the advantages and disadvantages of each.

Begin with Strevens. Grant the presupposition inherent in the question "Why is the allocation of cognitive labor in science so efficient?"—grant that science does manage to allocate labor so as to efficiently produce a desirable mix of pure knowledge and societal benefit. Call this mix "social good" for

short (without forgetting that some of the good may be purely epistemic or intellectual). Then Strevens' explanation is as follows: scientists choose among research programs based on their prospects for earning scientific credit—status or reputation—and the reward system in science calibrates the allocation of credit so that the choices made lead to a distribution of labor that (roughly) maximizes the production of social good.

More specifically, the reward system allocates credit in proportion to a scientist's actual contribution to social good; as described in the previous section, this finds a distribution that efficiently produces social good. Had the system instead allocated credit according to a scientist's actual contribution to some other end—say, the creation of entertainingly wacky theories—then what would be efficiently produced would be wacky, rather than true and useful theories. Science serves socially beneficial ends, then, because those ends calibrate the reward system. This is what I call an external theory of the reward system, since it posits that the reward system is largely shaped by a standard for social good that is dictated from the outside. The theory does not explain, but merely assumes, the nature of the reward system.

An internal reward system is calibrated, by contrast, without reference to external standards; its shape is determined by other aspects of the practice of science. That does not, in principle, mean that the reward system does not encourage behavior that is advantageous to society at large. But what does it mean in practice?

Albert's paper provides two different versions of an internal reward system, one building on the other. The first version—not Albert's complete story—is taken from Hull (1988). According to Hull (who has in mind an evolutionary picture of science), scientists' ultimate aim is to have their research used by other scientists. Albert presents this thesis in a very pure form: scientists want their research to be put to work *and that is all*.

Why do scientists use one another's work? To answer this question, Albert introduces the idea that some properties of research are *hereditary*, where a

property is hereditary if it is easier to produce a new piece of work having that property if you use as your basis existing pieces of work also having the property.

Scientists use existing work, then, because they want to produce work that has (some of) the same hereditary properties as that work. But by assumption, their sole motivation in deciding what work to produce is the desire that it be used by others. They want to produce work with a certain hereditary property, then, only because they think work with that property will be used by others, that is, only because they think that others will want to produce work having that property. This motivational circle, Albert notes, makes science into a "beauty contest", in the economist's sense of the word: in order to decide what to produce, you must predict what others will decide to produce.

Life in such an endeavor will be simple and uncomplicated if everyone assumes that everyone else will try to produce work with some particular hereditary property (or property cluster) *P*. Then it is clear that you, too, should produce work with *P*, since that is the work that will get used. It does not matter at all what *P* is, provided that it is hereditary.

Suppose that there is a mechanism for producing this sort of *P*-centered equilibrium. Then what Albert's model can explain, if all the assumptions about scientists' motivations and so on are correct, is why scientists get fixated on some single property (or property cluster) rather than having diverse preferences. As Albert interprets it, it explains why there is methodological consistency in science, that is, why scientists agree on what constitutes a paper worth using and thus what constitutes a paper that is likely to secure its creator a good measure of status.

What cannot be explained is why, of all the hereditary properties out there, scientists fixed upon *P*. One hereditary property that scientists have adopted, according to Albert, is being well corroborated, or as non-Popperians might say, being empirically well tested or robust. But why this rather than, say,

being entertainingly wacky (if, as you might suppose, wackiness is hereditary, that is, you can build a wackier theory by learning from and building on the wackiness of others)?

Or consider the unflattering (and unfair, but in this context expositorily useful) picture that some writers paint of late twentieth century literary theory, which Albert alludes to in his paper's note 3. Literary theory, on this picture, rewards obscurity and opacity. One way to achieve such qualities is to scatter references and allusions to other obscure and opaque theorists throughout your work, so these qualities are hereditary. Albert's model might well be used to explain theorists' alleged intellectual preferences. Why is science not the same?

Albert presents the problem in more abstract terms. He posits two hereditary properties X and Y, and notes that a universal preference for X and a universal preference for Y are both stable equilibria.[7] Can anything break the symmetry between them?

One possible symmetry-breaker would be an externally imposed standard tipping the balance in favor of one hereditary property over the others. Suppose, for example, that scientists are rewarded not solely in proportion to the use made of their theories by other scientists, but in proportion to a mix of both use and the degree to which the theories withstand empirical testing, or the degree to which the theories prove useful to society at large. Then you would have reason to expect the equilibrium to settle on hereditary properties that are correlated with the satisfaction of these externally given standards—and quite possibly, to settle on the very properties rewarded by the standards. This would explain why scientists try to create empirically robust theories that serve a greater good (although the beauty contest and the matter of hereditariness would seem to take a back seat in such an explanation).

7. There is more structure to Albert's model than this: sometimes attempting to produce research with X results in research with Y instead. This does not in itself assuage the need for a symmetry-breaker.

Albert introduces his own symmetry-breaker as follows:

> Let us . . . assume that many researchers have a preference for using and producing *X*-papers but also try to be successful in the terms of the above model [i.e., having their work used as much as possible by other scientists] (§3.4)

This sentence might be read as suggesting the existence of an externally given standard favoring the property of *X*-ness in a paper. But although this is a legitimate explanatory option, it is not what Albert has in mind:

> Methodological standards are neither imported into science nor imposed from the outside (5)

Further, in his conclusion, he suggests that his model explains why, rather than assuming that, scientists prefer well-corroborated theories (where corroboration is the "*X*" factor in question).

An alternative explanatory option is suggested by Albert's posit that "researchers have a preference for *demanding* standards" (§3.4, my emphasis). Perhaps researchers have a general preference for properties that are hard to produce. This is an externally imposed standard, but a rather broad one.[8] What can be explained once the new externally imposed standard is introduced? That scientists do not take the easy way out, perhaps: the hereditary property that they end up by focusing on is difficult rather than straightforward to produce. What is not explained is why scientists focus on the particular demanding, hereditary properties that they do—such properties as empirical robustness, explanatory power, theoretical significance, and social benefit.

A final explanatory option, and what Albert actually intends, is that a significant number of researchers have a merely personal preference for

8. Or perhaps it is not externally imposed, but is rather an instance of "costly signaling". Then there must be some further quality that is signaled, a preference for which needs in turn to be explained.

certain properties that is not endorsed by any external norm. They happen to prefer empirically robust hypotheses, for example, but not because there exists an epistemological norm, prior to or outside of the institution of science, that puts a high value on empirical testing. On the contrary, insofar as we regard empirical robustness as a good thing, it is because the scientific establishment has settled upon it as a predictor of future scientific use, that is, because scientists expect future scientists to build on empirically robust hypotheses. This expectation arises in turn from an equilibrium in Albert's beauty contest reached largely because of an accidental concentration of corresponding non-normative preferences. The norm favoring empirical testing is explained, then, by a sociological "founder effect".

The contrast between Strevens and Albert demonstrates, I think, some of the advantages and disadvantages of giving a theory in which norms or standards are internally generated rather than externally imposed. On the one hand, when such a theory succeeds, the standards are themselves explained; externally imposed norms are by contrast unexplained explainers. Further, the theory is flexible: it can explain why radically different methodological standards exist in different fields, and perhaps how methodological standards can change (if some external force tips the equilibrium from one point to another). On the other hand, it is not easy for such a theory to succeed entirely unaided. Albert's theory seems to explain the particular methodological standards of science only so far as externally given norms or widely shared non-normative preferences are present to lend a hand.

I do not present this as an argument for favoring one approach in general over another. I do think that there is copious sociological evidence in favor of a strong external influence on the scientific reward system: profound truth, as measured by extra-scientific standards, is rewarded more (where these standards can be discerned in the enthusiasm of the educated public or the attention of philosophers from ancient times onwards), as is social benefit (think Pasteur). The existence of a distinctly normative system for

determining these rewards is clearly demonstrated, as Merton observes, by the keen interest taken by third parties in priority disputes, and more particularly, by the moral tenor of such disputes. Thus the picture found in Merton and endorsed by Strevens is, in this case, largely correct.

4. Conclusion

Many chapters remain to be written on the "economic" approach to scientific norms, and to those norms that mold science's social structure in particular. Let me conclude by defusing three objections to the entire enterprise.

The first objection is that the economic approach takes economics too seriously, that it leans too heavily on an analogy between markets and what goes on in science (the "marketplace of ideas") (Mirowski 1997). This is a misplaced worry; as observed in section 1, what the "economic approach" takes from economics is not its subject matter or its models, but its formal methods—decision theory, game theory, computer simulation, and so on—none of which are proprietary to economics.

Another objection is that the models used in the economic approach are too idealized. This is a reasonable concern, but all social science, indeed all science of anything much more complicated than a hydrogen atom, makes use of highly idealized models. Further, there are many ways in which the Peirce model and other simple models can be made more complex when the need arises. In any case, the economic approach's models are as complex as anything else in the philosophy of science; there is as much structure, if less anecdotal detail, to the Peirce model as there is to Kuhn's normal science or Lakatos's research programs.

The third and final objection is that the economic approach ignores the social, the human, the idiosyncratic side of science. This complaint, borrowed from critiques of economics itself, is based on a false premise. The social, the human, and even the idiosyncratic are amenable to mathematical modeling and so are in no way marginalized by the economic approach.

Consider Strevens' story: the priority rule catches our attention because it is idiosyncratic, it has its bite because human all too human scientists are bewitched by the prospect of scientific status, and it uses its power for social good rather than evil because its allocations of credit are modulated by a social norm that transcends the interests of individual scientists.

References

Barnes, B. D. and J. H. Bloor. (1996). *Scientific Knowledge: A Sociological Analysis*. University of Chicago Press, Chicago.

Brock, W. A. and S. N. Durlauf. (1999). A formal model of theory choice in science. *Economic Theory* 14:113–130.

Dasgupta, P. and P. A. David. (1994). Toward a new economics of science. *Research Policy* 23:487–521.

Goldman, A. I. and M. Shaked. (1991). An economic model of scientific activity and truth acquisition. *Philosophical Studies* 63:31–55.

Hull, D. (1988). *Science as a Process*. University of Chicago Press, Chicago.

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy* 87:5–21.

———. (1993). *The Advancement of Science*. Oxford University Press, Oxford.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*. Third edition. University of Chicago Press, Chicago.

Merton, R. K. (1942). The normative structure of science. *Journal of Legal and Political Sociology* 1:115–126. Originally titled "Science and Technology in a Democratic Order". Reprinted under the new title in Merton (1973).

———. (1957). Priorities in scientific discovery. *American Sociological Review* 22:635–659. Page references are to the reprint in Merton (1973).

———. (1968). The Matthew effect in science. *Science* 159:56–63.

———. (1973). *The Sociology of Science*. University of Chicago Press, Chicago.

Mirowski, P. (1997). On playing the economics trump card in the philosophy of science: Why it did not work for Michael Polanyi. *Philosophy of Science* 64:S127–S138.

Peirce, C. S. (1879). Note on the theory of the economy of research. In *Report of the Superintendent of the United States Coast Survey Showing the Progress of the Work for the Fiscal Year Ending with June 1876*, pp. 197–201. US Government Printing Office, Washington DC.

Polanyi, M. (1962). The republic of science. *Minerva* 1:54–73.

Rueger, A. (1996). Risk and diversification in theory choice. *Synthese* 109:263–280.

Sarkar, H. (1983). *A Theory of Method*. University of California Press, Berkeley, CA.

Strevens, M. (2003). The role of the priority rule in science. *Journal of Philosophy* 100:55–79.

———. (2006). The role of the Matthew Effect in science. *Studies in History and Philosophy of Science* 37:159–170.

———. (2010). Reconsidering authority: Scientific expertise, bounded rationality, and epistemic backtracking. *Oxford Studies in Epistemology* 3:294–330.

Weisberg, M. and R. Muldoon. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science* 76:225–252.

Zollman, K. J. S. (2010). The epistemic benefit of transient diversity. *Erkenntnis* 72:17–35.