Toward a Pragmatist Philosophy of Science [*]

Philip KITCHER

ABSTRACT: This three-part essay begins with a diagnosis of the career of general philosophy of science in the tradition begun by Hempel's reform of logical positivism. Since 1950, Anglophone philosophy of science has largely sought general accounts of confirmation, theory and explanation. It has not found them. Instead it has assembled some valuable tools for exploring problems in the natural and social sciences, and has put them to work in a range of useful studies. The second part of the essay pursues the question of whether there is a successor project to the general investigations that Hempel inspired. I argue that there are difficult methodological questions about collective inquiry that have largely been neglected. Attempts to answer these questions should expand the resources for philosophical understanding of scientific knowledge. The third part focuses on the situation of the sciences within society. Can scientific research be divorced from the interests of a broader public? Can its conclusions be generated in neglect of the evidential standards accepted by non-scientists? I argue for a development of the ideal of well-ordered science, and for the understanding of science (singular!) as one institution among others. Conceiving matters in this way recovers important themes from the neglected pragmatist tradition.

Keywords: logical empiricism; pragmatism; social epistemology; science and society; well-ordered science; science and values.

RESUMEN: Este ensayo en tres partes comienza con un diagnóstico de la trayectoria de la filosofía general de la ciencia, tal como se cultiva en la tradición que se origina con la reforma hempeliana del positivismo lógico. Desde 1950, la filosofía anglófona de la ciencia ha buscado enfoques generales sobre teorías, confirmación y explicación. Ha reunido, en cambio, algunas herramientas valiosas para explorar problemas en las ciencias naturales y sociales, aplicándolas en un buen número de estudios útiles. La segunda parte del ensayo se ocupa de la cuestión de si hay algún proyecto continuador del tipo de investigación general inspirada por Hempel. Sostengo que se han ignorado algunas cuestiones metodológicas difíciles sobre la investigación colectiva. Los intentos para responder a estas cuestiones debieran ampliar nuestros recursos para comprender filosóficamente el conocimiento científico. La tercera parte se ocupa de la la situación de las ciencias en la sociedad. ¿Puede la investigación científica divorciarse de los intereses del público general? ¿Pueden sus conclusiones ignorar el tipo de evidencia que aceptan los no científicos? Defiendo un desarrollo del ideal de una ciencia bien ordenada y una concepción de la ciencia (¡en singular!) como una institución entre otras. Este enfoque recuperar temas importantes de la olvidada tradición pragmatista.

Palabras clave: empirismo lógico; pragmatismo; epistemología social; ciencia y sociedad; ciencia bien ordenada; ciencia y valores.

PART 1: PROBLEMS AND CHANGES

I

It is a commonplace that contemporary work in the philosophy of science is dominated by ideas that descend from a general philosophical movement of the early twentieth century. Inspired by a variety of intellectual developments and sources—not only the new formal logic of Frege, Russell and Hilbert, but also the articulation of empiricism by Hume, strands in Kant and in neo-Kantianism and the scientific reflections of Mach—young philosophers in Berlin and Vienna proposed a revolution in philosophy. The history of philosophy, as they saw it, was dominated by protracted discussions about propositions lacking cognitive content, *pseudoproblems* as they called them. Senseless debates were to be excised with a surgical instrument, the criterion of cognitive significance, and, once the operation had been successfully completed, philosophy's task was to identify, with maximal clarity and explicitness, the logical and epistemological structure of the sciences. Because that project was to articulate the logic of all genuine knowledge, philosophy of science was to become the center, if not the whole, of philosophy.[1]

Yet the members of the Vienna and Berlin Circles were not alone in yearning for change. After the rise of fascism forced them out of central Europe, many of them settled in the United States, where they allied themselves with an indigenous revolutionary movement. The American pragmatists, too, particularly James and Dewey, had sought a "reconstruction in philosophy" (Dewey 1920), and, in their homespun and less formal ways, had apparently groped towards a similar conception of "cognitive significance"—the pragmatist thought that "a difference, to be a difference, must make a difference" (James 1907 [reprint 1995], 20). The symbol of the positivist-pragmatist alliance was the Encyclopedia of Unified Science. Yet it was clear, relatively early, and certainly by the 1940s, that the partnership was an unequal one. The positivists appeared to have sharp instruments, and their precision tools seemed to signal a more decisive break with the high-flown senselessness of past philosophy than the wordy and woolly formulations of their pragmatist comrades. Pragmatism gradually faded away, and, after Dewey's death in 1953, it was pursued only by a handful of scholars, apparently benighted figures who worked at the margins of the profession.

By 1953, however, positivism, too, was dead—or, perhaps better, had undergone metamorphosis. Although the obituary is often credited to Quine (1951), the transformation of logical positivism into logical empiricism was carried through a little earlier in a pair of articles by Hempel (1950; see Hempel 1965, ch. 4). Those articles set the agenda for a new philosophy of science, now less imperialistically ambitious. Hempel's agenda has largely defined philosophy of science, from the early 1950s to the present—and I have stolen a phrase from one of his titles, "Problems and Changes", for the first part of this essay.

---

[1] The narrative I sketch briefly here has been elaborated with subtlety and nuance by the principal contributors to the history of analytic philosophy and the history of the philosophy of science (including Richard Creath, Michael Friedman, Don Howard, Alan Richardson, Thomas Ricketts and others). For my purposes, only the broadest-brush treatment is necessary.

With the lucidity for which he is rightly revered, Hempel's critique of logical positivism chronicled the various attempts to present a sharp criterion of cognitive significance, and exposed the systematic problems to which they were subject. On the basis of his diagnosis, he proposed that the central negative positivist project, that of identifying exactly where traditional philosophy had lapsed into meaninglessness, should be abandoned. Nevertheless, the spirit of that project could be preserved. Philosophy, and other areas in which intellectual progress seemed elusive, could still learn from those intellectual ventures that proceeded most swiftly, surely, and reliably. Reflection on the natural sciences could hold up ideals at which less fully-developed investigations might aim. Specifically, philosophical explication of important meta-scientific concepts—and Hempel mentioned four: *confirmation, theory, explanation, and simplicity*—could help the nascent social sciences and provide a critical perspective on those long-winded humanistic discussions that circle in endless disagreement. Three of Hempel's central concepts have been at the center of decades' worth of philosophical investigation; only simplicity has remained relatively unpopular.[2] As the years have passed, other concepts have joined the Hempelian trinity: *law, reduction, progress,* even that empiricist bogey *causation*.[3] The task of general philosophy of science has, however, continued to be the Hempelian one, namely to provide explications of these concepts, or, as we might now say, to provide general accounts of them by specifying the necessary conditions for their application across the entire range of possible cases.[4]

Hempel's new agenda scaled back the positivist revolution. It also departed completely from the revolution envisaged by James and Dewey. Although there is a superficial similarity between the pragmatist principle (whose formulation by James is probably the most quoted sentence in any classical pragmatist work) and the criterion of cognitive significance, the architects of pragmatism had no serious interest in issues of content or linguistic meaning. What concerned them was significance in a different sense, that of *importance*. They saw philosophy as tending to lose sight of the really burning issues, and they wanted to reconnect it with human conduct, with human aspirations and with human life (Kitcher 2012, ch. 1). Here are two passages from the many in which Dewey attempted to characterize the reform of philosophy for which he hoped.

> Unless a philosophy is to remain symbolic—or verbal—or a sentimental indulgence for a few, or else mere arbitrary dogma, its auditing of past experience and its program of values must take effect in conduct. (Dewey 1916, 338)

> The problem of restoring integration and cooperation between man's beliefs about the world in which he lives and the values and purposes that should direct his conduct is the deepest problem of modern life. It is the problem of any philosophy that is not isolated from that life. (Dewey 1929, 204)

---

[2] The most prominent logical empiricist to take up the problem of explicating simplicity was Nelson Goodman (see, for example, Goodman 1959, 1961, 1963).

[3] Hempel emphasized the need for an analysis of the concept of law (a concept fundamental to his account of scientific explanation); like other logical empiricists of his generation, he supposed (on broadly Humean grounds) that the notion of causation was not available in explicating explanation.

[4] This is already a divergence from Carnap's own explication of explication, given at the beginning of (Carnap 1950), which included a pragmatic requirement: explicata must be fruitful for science. See (Kitcher 2012, ch. 8).

Like Neurath, Dewey envisaged philosophical work as embedded in the reform of society and the improvement of human life. He might have sympathized with the positivist program of excising the festering places in our corpus of attitudes, or with Hempel's agenda of clarifying the methods of rigorous inquiry, but he could not have viewed this as the end of the story, even if philosophical attention were focused on inquiry. He would have wanted to understand how the projects Hempel commended would have related to the lives of human beings and to the structures of their societies. Neurath, of course, died young—and so too did any hope for the ambitious revolution in philosophy that Dewey envisaged.

My aim in this essay is to revive the Deweyan program, with a specific focus on the philosophy of science. The first task, however, is a Hempelian one: to try to come to terms with the logical empiricist project as it has evolved since 1950.

II

Although the two revolutionary movements had distinct aims, it would be wrong to overlook some important shared features. From the early 1950s on, the new logical empiricist program harbored pragmatic hopes.[5] Clarification of metascientific concepts was viewed as valuable in the resolution of difficulties and controversies within the flourishing natural sciences (particularly in dissolving apparent problems in the foundations of the physical theories, relativity and quantum mechanics, that so inspired the early logical empiricists), and, even more significantly, it was supposed to aid in the development of nascent areas of science, especially in the social sciences. Hempel and his colleagues were concerned to forge concepts of confirmation, theory, and explanation that could help anthropology, archeology, economics and history attain the same status as the physical sciences. Early logical empiricism embodied the Deweyan impulse to the improvement of inquiry.

Sixty years on, it should be clear that the program has failed. We have no general accounts of confirmation, theory, explanation, law, reduction, or causation that will apply across the diversity of scientific fields or across different periods of time. Nor, I shall argue, are we likely to get any. Moreover, the successive failures of philosophical ventures in explication have been accompanied by a substantial weakening of the demands imposed on them. The early logical empiricists would bridle at the modal and causal notions contemporary philosophers allow themselves in their renewed attempts to deliver the desired explicata. Perhaps they would be wrong to do so, for one can argue that logical empiricism inherited an over-strict and over-simple epistemology (largely derived from Hume's foundationalism). I see systematic reasons for supposing that even the laxer requirements now in vogue cannot be satisfied.

So, like Hempel in 1950, I want to identify recurrent difficulties that beset the enterprise of providing the general explications. I'll focus on just two instances— confirmation and explanation—in part because these seem to be the most important of the metascientific concepts, in part because of the amount of labor that has been

---

[5] Sympathies with pragmatism are also evident in many of the writings of the first-generation logical empiricists: in Quine, Ernest Nagel, Israel Scheffler, Goodman, and even in Carnap (in his discussions of the utility of linguistic frameworks, and of explication [see fn. 4]).

expended, and in part because the troubles that beset these explicatory projects appear to be shared in other instances.

What are the sources of trouble? One difficulty arises from the diversity of scientific practice. In different fields of the natural sciences, different kinds of questions come into prominence, and those questions are typically addressed by forging methods that are adapted to a specific type of problem. It is no accident that the instruction given in "methodology" within the practice of training young scientists varies quite widely from domain to domain. The "methods" courses for particle physicists are quite different from those taken by geneticists, which differ further from the training received by climate scientists or by archeologists. Perhaps there is a "thin" general conception that picks out what is common to the diversity of fields, but that turns out to be too attenuated to be of any great use.[6] A second difficulty arises from the historical record. Here it is not just that historical practices further increase the diversity just noted—although they do—but also that the processes through which major scientific changes come about seem to involve considerations that logical empiricism, in its original stringent form or in its more relaxed modern version, cannot readily accommodate.

As we shall now see, these considerations of diversity and historical development tell differently in the two cases I have selected. Let us begin with the notion of confirmation.

III

In the early days of logical empiricism, there were a number of different target notions: *qualitative* confirmation, *relative* confirmation, *quantitative* confirmation. With respect to each of these (or, more exactly, with respect to the well-studied ones, the first and the third) there were several proposals whose merits were debated. In recent years, by contrast, there has been considerable unanimity: we are (almost) all Bayesians now. So doesn't Bayesianism provide just the explication of confirmation logical empiricism has yearned for?

If we are to judge the virtues of the Bayesian explication, it is worth asking whether it can meet Carnap's requirement that explicata must be fruitful: what kinds of tasks can the clarification of confirmation discharge? There's no doubt that thinking about some situations in Bayesian terms is helpful: a boringly familiar example is the welcome illumination probabilistic thinking brings in connection with diagnostic tests for which there are available statistics about types of error. If what is needed, however, is a clarification that will enable the practicing scientist, in any field, to sort out instances in which it is currently unclear which (if any) among rival hypotheses is well-confirmed by the available evidence, or which is best confirmed, then it is plain that, for virtually all problematic situations in all fields, Bayesianism delivers nothing useful.

---

[6] It will not serve the pragmatic purposes of improving inquiry. A telling illustration of this occurs in (Achinstein 2001), which seeks to respond to the challenge to show how a philosophical theory of evidence can be relevant to scientific practice. As I argue in (Kitcher 2011a), for all his care and technical sophistication, Achinstein fails to meet the challenge.

So if you emphasize the pragmatic purposes the logical empiricist program was originally supposed to fulfill, Bayesian confirmation theory is a thorough failure.

This judgment is likely to sound harsh and unwarranted. Bayesianism supplies a framework within which scientists, whatever their questions and whatever their ways of observing or experimenting, can use to draw conclusions about the probability of the hypothesis in light of the available evidence. Undergoing the process of reaching such conclusions will, of course, require the specification of certain probabilities—the prior probability of the hypothesis, the probability of the evidence conditional on various assumptions, etc. etc.—but it is not required of a *general* theory of confirmation that it identify how those needed probabilities are to be identified. That can, and should, be left to the practitioners of the special disciplines.[7] So, you might suppose, Bayesianism operates at just the right level to give the general account that was required.

The defense of Bayesianism just outlined is familiar, and it faces objections that are equally well-known, and also some that are underappreciated. Start with the old point that, strictly speaking, Bayesianism allows investigators to insert whatever prior probabilities they choose, so long as they are strictly between zero and one. Ignore complications that arise from the fact that there might be cases in which the choice concerns a parameter that could be assigned any of continuum many values. The standard response to worries about the range of choices of the priors is to appeal to coherence results: if you choose a very low value and I choose a high one, the evidence will eventually lead our estimates of the probability of the hypothesis to differ by as little as one might please. The point can be inverted. For any sequence of evidence that comes in at a prescribed rate, for any choice of a difference that we are prepared to specify as convergence, and for any specification of a period within which convergence is to be obtained, it's possible to choose prior probabilities strictly between zero and one such that convergence cannot be attained within the chosen period. If Bayesianism is to help with actual scientific decision-making, it will require supplements that declare how the choice of priors is to be restricted. Is there any general way to elaborate that declaration?

But the worries about priors are just the beginning. Unless you are sure that you have a partition of the possible hypotheses, for each of which you can identify the probability of the evidence, conditional on it, the Bayesian computation will need a probability that is not only typically hard to judge but arguably ill-defined: you have a number of hypotheses with which the evidence is to be confronted, and you know that the right answer may be "none of the above"; so the calculation requires the probability of the evidence on what Wesley Salmon called the "catchall" hypothesis (W. Salmon 1990a). To illustrate the point, consider one of the apparent great suc-

---

[7] It should be emphasized that there are extremely healthy projects in formal philosophy of science that attempt to extend probabilistic and statistical analyses into domains in which they have not previously been available: a prime example is the work of Clark Glymour and his colleagues (Spirtes et al. 2000). These efforts to extend methodology focus on particular types of disciplines and problems, rather than attempting to adumbrate a general account of scientific reasoning. I am indebted to Jan-Willem Romeijn and Gregory Wheeler for pressing me to make this explicit.

cesses of Bayesian confirmation, the boost Fresnel's wave hypothesis received from the observation of the Poisson bright spot. Bayes' Theorem gives us:

$$\Pr(F \mid P) = \frac{\Pr(F)\Pr(P \mid F)}{(\Pr(F)\Pr(P \mid F) + \Pr(-F)\Pr(P \mid -F))}$$

Suspend worries about the constraints on the prior probabilities—suppose, for example, that $\Pr(F)$ is 0.001 and $\Pr(-F)$ is 0.999. The substantial value of $\Pr(F \mid P)$ results from the supposed fact that the second term in the denominator is taken to be small, maybe even negligible. But that holds only if $\Pr(P \mid -F)$ is extremely small. Why should that be? True, Poisson didn't *expect* there to be a bright spot—he surely hadn't ever thought in terms of that possibility—but what does that tell us about the probability in question? Indeed, is the probability even well-defined? Why isn't it up to a scientist, even to Poisson, to declare that, now that the issue has come up, his original expectations were simply unfounded, conjectures based on very little—that, for all he knows, the probability may be substantial? Are there rules that govern this kind of game?

In my view, probabilistic concepts are wonderful tools, apt for application under conditions in which we have a well-specified sample space. Bayesianism attempts to export those concepts into situations where there are no such sample spaces, where subjects are forced to make arbitrary choices about the values to be assigned. Wise investigators should be aware, in advance, that nature is full of surprises, and hence diffident about committing themselves to conditional probabilities based on propositions they only vaguely understand. They should resist the forcing, withhold the probabilities—and thus give up on Bayesian computation. If they do that, of course, the alleged general theory of confirmation becomes inapplicable.

These preliminary points are deepened by lessons from the history of the sciences. In the first sentence of the last monograph of the Encyclopedia of Unified Science (a slim book now 50 years old) Thomas Kuhn suggested that history might change prevalent images of science. He was right. Strictly speaking, Bayesian analysis, like other theories of confirmation before it, takes over the Principle of Total Evidence: we're supposed to compute the probabilities conditional on all the evidence there is. Of course, it's a good question whether this is ever possible—Does anyone ever have access to all the evidence? If so, could that person know that that was all the evidence available?—but let's set that aside. A more important point, deriving from Kuhn and his successors, is that, normally, the evidence is not consistent. There are lots of "findings" around, and they don't all fit. A complete account of confirmation would have to explain just how scientists manage to sort out something that *reasonably* counts as the available evidence. Here, Bayesianism is completely silent. Consistency is not investigated as an *achievement*, but simply taken for granted.

Indeed, the normal situation in the many instances in which a theory of confirmation might be called on to do useful work is more complicated in ways Kuhn's discussions of scientific revolutions have made familiar. Fields that are embroiled in serious debate (and that might welcome the philosophical clarification logical empiricism hoped to provide) typically contain several clusters of hypotheses, each cluster being incompatible with its rivals, with each being able to explain and predict parts of a mass of "findings" that, taken as a whole are inconsistent, and each able to help in solving

different sorts of explanatory and interventional problems. Detailed historical studies can show us how these fields evolve, how the fortunes of different clusters of hypotheses wax and wane, how eventually a *reasonable* consensus can emerge. They also reveal how, even at the end, there are hypothetical ways of picking the evidence or of choosing which problems count as most important that could resist the final consensus. Phlogiston theorists could have insisted, to the end and beyond, that no theory that failed to account for the similarities of combustible materials in terms of a shared component substance could be accepted as satisfactory. (Contrary to historical folklore, the most prominent phlogistonians did not take that line.) Similarly, anti-Darwinians could—and still do—contend that the theory of evolution by natural selection cannot be counted as properly confirmed before it has an explanation of the origins of the cell. Bayesianism has no tools that enable us to identify these forms of opposition as unreasonable—and hence no way of reconstructing two of the greatest advances in the history of the sciences (as well as many other episodes) as reasonable transitions.

The heart of the problem, as Kuhn saw, and as Pierre Duhem had already glimpsed before him, is that there are styles of judgment (uses of *bon sens*) that are pervasive and crucial in scientific practice, and that cannot be formalized in the idioms logical empiricism allows itself. Whether there is some further style of formalization that logical empiricists would accept and that could provide a satisfactory analysis of scientific judgment is highly unclear—but I see no more reason to pin my hopes on finding one than Hempel found in 1950 for supposing some future piece of logical apparatus would enable the precise specification of cognitive significance. *Provided the task of explicating confirmation is viewed as one with pragmatic import, specifically as potentially leading to the improvement of inquiry,* the case for the failure of logical empiricism (in this domain at least) is as strong today as was the argument for the failure of logical positivism sixty odd years ago.

Of course, most philosophers of science who work on confirmation theory do not explicitly think in terms of that proviso.[8] Instead, they suppose that, independently of any application to actual scientific practice, there is a theoretical—"philosophical"—task of clarifying metascientific concepts. So their focus turns inward, to philosophy, and to a sequence of problems spun off from the initial enterprise: there are puzzling toy examples, and general problems (how to account for old evidence, for instance). The conversation continues, although the topics may constantly diminish in size and in importance, without attention to the pragmatic motives that—rightly—concerned the founders of logical empiricism. Like them, and like Dewey, I view this as a desertion of philosophy's proper task. As with other forms of inquiry, so too with philosophy: it can degenerate

> […] into sterile specialization, a kind of intellectual busy work carried on by socially absent minded men. Details are heaped up in the name of science, and abstruse dialectical developments of systems occur. Then the occupation is "rationalized" under the lofty name of devotion to truth for its own sake. (Dewey 1920, 164)

---

[8] Here, there is a striking contrast with the ventures commended in fn. 7, where the formal work is inspired by and responsive to the problems that arise in specific types of scientific investigation.

IV

The search for a general account of explanation has reached the same end by a different route. In this instance, there is no single explication of the original concept that dominates the contemporary scene. Once, of course, there was. Until the early 1960s, many philosophers of science believed the covering law models, articulated by Hempel from 1948 on, constituted a clear success for the logical empiricist program. In the course of the subsequent decade, however, it became evident that those models faced a large number of difficulties. They were too lax in allowing derivations that apparently run against the grain of causal dependence (and other forms of "objective dependence") to count as explanatory—you can't explain the height of a tower by computing it from the length of the shadow it casts, nor can you explain a man's failure to become pregnant by appealing to his scrupulous taking of birth-control pills, nor can you explain Boyle's Law by deducing it from the conjunction of Boyle's Law and Kepler's First Law. They faced severe difficulties in articulating a defensible concept of probabilistic explanation. They encountered trouble in application to historical sciences—evolutionary biology, paleoanthropology, and human history itself. They depended on a notion of scientific law that resisted explication. They were at odds with the thought that one can sometimes explain quite improbable conclusions (phenomena of quantum tunneling, for example). And they were vulnerable to the complaint that explanations are not—or do not have to be—arguments.[9]

The list of complaints was sufficiently extensive to convince most philosophers of science, including the admirably open-minded architect of the covering-law models, to abandon what had seemed a successful explication. Most reacted by seeking accounts that would solve some particular subset of the difficulties that had emerged—with different subsets occupying different philosophers. So there grew up a variety of causal accounts of explanation, approaches based on the idea that explanations raise the probability of the explanandum, mechanist proposals, interventionist theories, appeals to the thought that good explanations unify the phenomena, explicitly pragmatic theories of explanation, and so forth. *None* of these is adequate to capture all the successes and avoid all the failures. *All* of them come with promises that particular notions will (someday) be explicated, or that apparent counterexamples will (eventually) be overcome. Within each of the traditions, it's easy to discern just that focus on artificial philosophical puzzles that I've suggested affects the study of confirmation. There are cottage industries devoted to the famous troublesome cases: the tower and the shadow, the mayor's paresis, the lucky hole-in-one, and so forth. All too frequently, the pragmatic impulses that moved Hempel and his contemporaries seem to have been forgotten, the question of how to deal with genuine scientific disputes about explanatory power left behind.

I'll elaborate this diagnosis by focusing on one particular tradition, and, since I am among the guilty, my example will be the approach to explanation as unification. The inspiration for the Unification Church is twofold: first, a notion of unification seems particularly valuable in accounting for theoretical explanation; second, it reflects the

---

[9] The various difficulties are lucidly explained and chronicled in (W. Salmon 1989).

remarks of many practitioners who have defended the explanatory power of their hypotheses and theories by pointing to an ability to "unify the phenomena" (Lavoisier, Darwin, and Huxley are prominent examples). Perhaps the central puzzle for unificationists is that of distinguishing genuine unification from mere conjunction: showing why the derivation of Boyle's Law within the kinetic theory of gases is explanatory, while deducing it from the conjunction of Boyle's Law and Kepler's First Law is not.[10]

I proposed to address this difficulty (Kitcher 1981, 1989) by offering an account of what scientists acquire when they become proficient in a theory. To be an expert in classical electrodynamics or classical genetics (say) is to internalize some patterns that can be instantiated again and again to yield the answers to questions about the interactions among charged bodies or about the distributions of phenotypes in breeding populations. One virtue of the proposal is that it captures Kuhn's important insight that successful training in a scientific field must deliver more than the ability to affirm a list of pertinent propositions—properly educated students are not baffled when they turn to the problems at the end of the chapter, but can put their understanding to work.

Roughly, theories unify when a small set of patterns yields a large number of conclusions: the smaller the explanatory basis and the more diverse the collection of explananda, the better. Yet if, in traditional logical empiricist fashion, you identify a pattern with a formal derivation in which some of the non-logical vocabulary is subjected to empirical constraints, it's easy to see how flabby patterns, those that are very relaxed about the constraints, might prove very successful: consider the pattern that deduces C from C, or C from "God wills that C". So unification theory requires a notion of *stringency*, one that measures the ways in which an argument pattern sets severe constraints. Unification, properly understood, consists in deriving a large set of diverse conclusions by means of a small number of stringent patterns.

At this point, it's quite easy to see that several tasks need to be carried out to deliver the precise explication for which logical empiricism hoped. How exactly is stringency to be defined? How are explananda to be counted? How is the diversity of the explananda to weigh against the sheer number? (Would it be better to have more diversity, at the cost of having a smaller total number?) How is the number of patterns in the explanatory basis balanced against their stringency? (Is it preferable to have a somewhat larger set of more stringent patterns than a somewhat smaller collection of less stringent patterns?) What kinds of predicates are allowed in the patterns? (Are there restrictions against the kinds of predicates that figure in Goodman's New Riddle, and, if so, how are such restrictions to be formulated?)

Someone could spend a career trying to devise solutions to these technical difficulties. Fortunately, I didn't. One warning signal came from considering types of explanation at a far remove from those places at which unification appears most promising. If the explanatory power of high theory is, at least sometimes, plausibly explained in terms of the theory's unifying power, there are many cases of everyday causal explana-

---

[10] This problem inspired Michael Friedman's seminal article (Friedman 1974), although, as I argued in (Kitcher 1976) his solution was unsatisfactory.

tion, for which the thought that unification is the deep source of explanatory success is highly implausible. Consider one of the celebrated examples: the possibility of explaining the lengths of shadows by the heights of the pertinent objects, but the impossibility of explaining the heights by the lengths of the shadows. With a fair bit of strain, you can try to show that the height-to-shadow pattern fits into some unifying set of patterns and that the shadow-to-height pattern does not (Kitcher 1981). Even if that is successfully worked out, however, there's a residual question about how to relate the approach to the more straightforward idea that shadows causally depend on heights but heights don't causally depend on shadows. Adhering to logical empiricist scruples, and filled with Humean zeal for avoiding the epistemologically problematic concept of causation, I suggested that causal dependence was derivative from explanatory dependence—that the relation of causal dependence is the projection (or, if you like, the "shadow") of relations of explanatory dependence (Kitcher 1986). I am grateful to Nancy Cartwright for showing me how desperate this idea was.

Cartwright's first observation was that the logical empiricist scruples about causation rested on adopting a piece of Humean epistemology, long after other bits had been abandoned. She noted that almost all logical empiricists had given up the thesis that our scientific knowledge is founded in the apprehension of sense-data (or other surrogates for Hume's "impressions")—and that most had abandoned the idea of *any* foundation for knowledge. Why then suppose that some special foundation for causal knowledge was required? More pointedly, she asked me how I proposed to account for people's abilities to make causal judgments about ordinary events. When I replied that I supposed that part of our early learning consists in absorbing a set of patterns that we deploy to explain everyday affairs, she raised an eyebrow (I understate). I still recall her sentence: "But even Lucy [her then-young daughter] can tell me the cat is lapping up the milk."

An obvious response to Cartwright's trenchant criticisms would be to partition the cases of successful explanation. For theoretical explanation, or at least for some instances of theoretical explanation, the unification approach supplies a promising account. In dealing with the explanation of everyday events, however, the appropriate model of explanation is some causal one: explanations work by tracing causes. An ecumenical approach along these lines was already suggested by Wesley Salmon (1990b). Although it's a progressive move in the theory of explanation, it does not go far enough.

One reason for trouble comes from the fact that singular events are not always best explained causally. Take any large human population, for example the population of London in the early eighteenth century. In any given year, more boys than girls will be born in this population. But we don't explain the fact that, in 1700, more boys than girls were born in London, by elaborating the full causal details: all those copulations, the competition among sperm, the fertilization of eggs, their survival or premature abortion, and so forth. Instead, the particular fact is embedded within a unifying theory. Fisher's sex-ratio theory tells us that if a human population departs from a 1-1 sex ratio at sexual maturity, there will be evolutionary advantages for those with a propensity to generate the underrepresented sex; so the population will be expected to

approximate a 1-1 sex ratio at puberty; if boys are significantly more likely to die between birth and sexual maturity, more of them will have to be born. Sometimes, then, a particular fact or set of facts calls for equilibrium explanation, not for grubbing in the causal details.

Another, more significant, reason comes from any immersion in the historical sciences. Consider, for example, the explanations given by historians. They are remarkably diverse, sometimes focusing on particular aspects of causal structure (background social conditions, the geographical properties of a region and their effects on possibilities for trade, intellectual sources of political movements, and many other features), sometimes on the doings of prominent individuals. When historians offer causal explanations, they select from an array of causes. So far as I can tell, the selections are diverse, and defy any extant philosophical account that would try to reduce them under a set of manageable heads. Further, a lot of history is not causal at all. Chronicle is sometimes important. So too are styles of history that aim to show what it was like to be in a particular place at a particular time—to conjure up the experience of a battle (Keegan 1976), or the predicament of the convicts who were sent to colonize Australia (Hughes 1986), or to be part of particular small communities (Le Roy Ladurie 1990, Ginzburg 1980). Historians ask all sorts of questions, and deliver a wide diversity of types of answers to them.

So, too, do scientists. If you think of scientific explanation as what the sciences supply when we aren't interested in prediction or intervention or control, there are many kinds of inquiries that seek explanation. Here's a small sample. "What are the constituents of eukaryote cells?", "Will the universe continue to expand indefinitely?", "Is there intelligent life elsewhere in the universe?", "When did human language evolve?", "How many species of australopithecines were there, and how are they related?", "To what extent can one form a range of silicon compounds that rivals the diversity of carbon compounds?", "What is the natural host organism for the Ebola virus?", "Can non-human animals count?". I suggest that all these questions are significant, that their significance is quite independent of any practical use we might make of the answers, that in none of the instances does the significance arise because we seek some general law or theory, and that none of them is naturally formulated as a why-question. The work of scientific explanation is far more diverse and complex than philosophers have appreciated, and contemporary theories, taken in isolation or in combination, are quite inadequate to capture it.

You may object that I started with an extremely unpromising theory—the unification account—and that the logical empiricist project would look a lot better had I begun with some alternative (a causalist, mechanist, or interventionist approach, say). But I think the troubles are general and the situations symmetrical. Unificationists manage to capture one sort of explanatory virtue, a virtue important to some contexts of scientific work. The rivals succeed in covering other contexts. None of them does well in those domains where explanation as unification is a plausible idea—in high theory, and in parts of mathematics, for example, as well as in some equilibrium explanations. Moreover, once we appreciate the many neglected styles of scientific work, it's evident that there are types of explanation unaddressed by any extant account.

The position I want to recommend is intermediate between the one adopted by almost all theorists of explanation in the logical empiricist tradition and one resembling the stance once defended by Bas van Fraassen. Since 1948, philosophers of science have yearned for a completely general account of explanation—typically, a single conception, occasionally, as with Wesley Salmon, a small package of disparate models. Van Fraassen, by contrast, proposed a completely pragmatic account, conceiving explanations as answers to why-questions (van Fraassen 1980, ch. 5). Amend his proposal in two stages. First, following my claims about the diversity of explanation-seeking questions, view explanations as answers to significant questions (where the notion of significance is pragmatic—more on this in Part 3). Second, in line with a critique Salmon and I once offered of van Fraassen's suggestion (Kitcher and Salmon 1987), and in accordance with some of the guiding thoughts of logical empiricism, recognize that there are constraints on the possible explanatory answers beyond those van Fraassen acknowledges: in particular, although there is no *single* relation of relevance, not every possible relation counts as a relevance relation. Relevance relations form an untidy collection, one we don't know how to define. Traditional theories of explanation sometimes succeed in identifying one, or a few, of the relevance relations. Their insights can be viewed as capturing explanatory virtues; their failures lie in the pretense that theirs are *all* the explanatory virtues.

V

So far, an apparently sad story. Yet it would be wrong to suppose that the efforts of the past decades, both with respect to confirmation and with respect to explanation have been entirely wasted. Only if we think of the goal of philosophy of science in terms of the provision of completely general explicata need we view the program Hempel announced in 1950 as a failure.

In his discussion of Carnap's philosophy of science, Hempel ended with a well-known parable. A dying farmer calls his three sons to his bedside, to tell them of a treasure hidden in the vineyard that is their joint inheritance. After his death, the sons dig eagerly, in hopes of becoming wealthy. Their efforts stimulate the growth of the vines. "The rich harvest they reap proves to be the true and only treasure in the vineyard" (Hempel 1963, 707). I suggest that there have been parallel benefits from the logical empiricist program.

Just as the sons might have been spoiled for serious work had they actually discovered a chest full of gold coins, so too the achievement of some general philosophical account might actually have been counterproductive. Ironically, the success of the covering-law models of explanation, during their heyday, achieved one of the goals for which Hempel had hoped while frustrating his larger purpose. Social sciences eagerly seized on this piece of philosophy, and, for decades after philosophers had given up on covering-law models, practitioners earnestly criticized proposed explanations that failed to appeal to general laws. Archeology was particularly afflicted, and it took sustained philosophical effort—primarily on the part of Merrilee Salmon and Alison Wylie—to liberate it from the Hempelian shadow (M. Salmon 1990, Wylie 2002).

The yearning for a complete general theory led to the premature acceptance of methodological requirements that *retarded* work in areas logical empiricism hoped to advance. On the other hand, the *partial* proposals achieved with respect to confirmation, explanation, theory, and other metascientific notions have helped realize some of the pragmatic goals of the program—to help resolve intra-scientific disputes and to help in the development of nascent sciences. Concurrent with repeated failures to fulfill global hopes have been any number of local successes. Adopting a thoroughly pragmatist perspective, we can see that the metascientific analyses supplied a toolkit for thinking about particular areas of scientific research. Philosophers of science have carried that toolkit into areas of inquiry that have faced specific challenges, and, choosing judiciously from it, they have helped the practitioners to tackle those challenges. That toolkit and its uses constitutes the "true and only" value in the logical empiricist tradition.

Consider briefly a few examples. Philosophers and physicists have often wondered whether Einstein's decision to set the parameter ε (the ratio of the time taken by a one-way light signal to the time for a round trip) at ½ was essential to the special theory of relativity. John Winnie showed how use of the philosophical conception of a theory as an axiomatic system could resolve the question: he demonstrated that all the consequences about motions that can be obtained by making a specific choice of value can be obtained if the parameter is left indeterminate (Winnie 1970). This is not because all scientific theories are *really* axiomatic systems, but because parts of scientific practice can be illuminatingly reconstructed *for particular local purposes* by constructing judicious axiomatizations. I doubt that there is any single conception (or "essence") of the many things investigators refer to as "theories"—seeking an explication of this notion is simply a bad idea—but various possibilities offered by philosophers (axiomatic systems, families of models, systems of patterns of problem-solving) lend themselves to use in addressing specific issues. Judicious philosophy of science carries all the potential tools along, and selects whichever is best suited to the purposes at hand.

Similarly for discussions of intelligence and heredity. Since the eruption, during the early 1970s, of claims that intelligence is heritable and that the differences in IQ scores among racial groups therefore testify to innate distinctions of cognitive ability, philosophers have used ideas about testing and confirmation to disentangle what has actually been supported by the available evidence. They have attended to the calculation of the heritability estimates, to the relation between those estimates and causal claims about the determination of intelligence, and about the connection between test measurements and cognitive capacities. The result has been a socially important illumination of a controversy that has the potential to do much practical harm—even if the message has not always been heard, and the errors diagnosed have been repeated (for probing analyses and reanalyses, see Block and Dworkin 1976, Block 1996, Glymour 1998).

My last example comes from a controversy in which I have participated, the use of evolutionary biology to advance provocative conclusions about human behavior and human psychological traits (Kitcher 1985, Vickers and Kitcher 2005). One strand in

this debate concerns the propensity of enthusiasts for evolutionary explanations of complex human traits to appeal to "general laws" about particular kinds of behavior—mate choice, or "forced copulation", say—that allegedly apply across biological taxa. The enthusiast points to evidence supporting the hypothesis that a behavioral trait is adaptive in some group of organisms (or in several different species), and then challenges opponents to explain why the "law" so supported does not apply to human beings. Whether or not the evidence actually supports the claims about the nonhuman animals (and sometimes there are ample grounds for skepticism), a careful analysis of the practice of evolutionary theory shows both that "laws" are hard to come by, and that it is better to think of families of models, each subject to investigation of the precise conditions and constraints. With this analysis at hand, it is possible to show just how the preconditions of the models in the successful non-human cases are quite different from those that obtain for our own species: the enthusiastic generalization simply collapses. Once again, philosophical reconstruction of scientific practice, obtained using the toolkit we have inherited, enables us to see very clearly and precisely where an apparently seductive line of reasoning for provocative conclusions goes astray.

Over twenty years ago, in his Presidential Address to the *Philosophy of Science Association*, Arthur Fine argued that the main achievements of philosophy of science have been local (Fine 1988). I am renewing his theme, with a few important differences. First, I believe we can now see clearly that the logical empiricist project is highly unlikely to achieve its anticipated general accounts of metascientific concepts—indeed that the contemporary situation in general philosophy of science bears an uncanny resemblance to that Hempel diagnosed for logical positivism in 1950. Second, the ability to engage in the projects of local clarification that figured in the pragmatic hopes of logical empiricism is not impeded by the lack of those general accounts. Third, the local successes are not simply the result of philosophers immersing themselves in the practices of particular sciences, but are, in part, traceable to their use of tools furnished by the logical empiricist tradition.

VI

How then to go on? For all the reasons advanced by Dewey, it seems unlikely that much profit will come from attending to the technical puzzles spun off from the failed general analyses, the toy problems that fill so much journal space. Better to continue to sharpen and refine our tools—and, perhaps, forge new ones—through a wider range of local ventures. The emergence of philosophy of biology, philosophy of economics and other specialized disciplines during the past decades shows convincingly how that can be done.

But is that all? Do philosophers of science become handmaidens of the special sciences, or are there enduring broader questions? In the next two parts of this essay, I shall suggest that there is further general work to be done. So far I have taken an *internal* look at logical empiricism, examining the ways in which its official projects collapsed under the burdens unwisely assumed at the beginning. Yet part of the story of the last decades is one of *external* challenges. From 1962 on, history and sociology of

science have questioned the images of science by which logical empiricism has been possessed, inspiring many philosophers of science to take up the cudgels in angry defense. I commend a different reaction. The insights of Kuhn, and of others who have gone on from his work in ways that sometimes distressed him, lie in reminding philosophy of science of perspectives and questions the agendas of the 1920s, the 1930s and the 1950s completely neglected. Specifically, the internal processes through which scientific projects are pursued and new conclusions certified are thoroughly *social*, and the accounts of "individual method" around which logical empiricism has been built should be *supplemented* by philosophical analyses of the norms of collective inquiry. Further, science as an institution is placed in broader societies, interacting with other social institutions and with the lives of countless individuals. The primary tasks of general philosophy of science today are to make normative sense of these two types of social embedding. To undertake those tasks would constitute a return to the revolution that never occurred, the reconstruction of philosophy Dewey (and maybe Neurath too) once envisaged—a venture, then, toward a pragmatist philosophy of science.

PART 2: THE LOGIC OF COLLECTIVE DISCOVERY

I

To talk of particular scientific beliefs as being well-supported, or as being genuine knowledge, is apparently to make a claim about what a properly-constituted scientific subject would think, were she to be given particular opportunities to be confronted with different aspects of the world. We imagine a careful thinker moving from consideration of "the evidence" to the claims in question, by cogent reasoning. Of course, if our rational subject had begun in the wrong places (with incorrect, or incomplete, reports of experimental findings) even if her reasoning were cogent as could be it might still lead her to faulty conclusions. So we tacitly suppose that she has traced a complicated path through a selection of episodes from the history of science visiting all those places at which some piece of relevant evidence was originally established. Is even that enough? Surely our agent must also be confident that she has been exposed to all the relevant evidence, that there is not some finding overlooked in her peregrinations that would undermine her reasoning. Very well, then, let's drop the idea of a selection: our imaginary knowing subject embodies the entire history of inquiry.

There is an alternative to this silly fantasy, one recognized in Newton's famous image of himself as standing on the shoulders of giants. The alternative was elaborated in suggestive ways in a book whose fiftieth anniversary is—rightly!—now being celebrated (Kuhn 1962). To think of scientific knowledge is to envisage a chain of communities, along which scientific knowledge grows. Ancestral communities transmit findings to their descendants, who add, subtract, and modify their inheritance, passing on an amended set of findings to their descendants. The well-supported beliefs of *individuals* are parasitic on the findings that *properly-constituted communities* have transmitted. Although the individual may add some new well-grounded claim to the corpus, her con-

tribution is small—and if it is to be taken over as a starting point for her successors, it must be certified by a properly-constituted community to which she belongs.

The importance of the historical and sociological turn, inaugurated by Kuhn, has been underestimated, perhaps even by Kuhn himself. Too much contemporary writing about the traditional questions of logical empiricism is pervaded by the image of an isolated "rational subject", and all-too-frequently scientific communities are tacitly regarded as if they were "rational subjects" writ large. Among the earliest claims of Kuhn's monograph is the proposal that mature science begins when a heterogeneous community gives way to a group that is bound together by shared commitments. Later, in the discussion of the resolution of scientific revolutions, readers are told that there is no single argument that convinces all members of the pertinent community. I interpret this passage not as a form of relativism, as the claim that reasons break down completely—although one of the two major voices of the text is fond of talking of "conversion" and of "faith". Instead, I listen to the other major voice, the Kuhn who thinks that scientific reasoning is far more complicated than philosophical accounts have appreciated, and who claims that different members of a scientific community are *reasonably* persuaded by different arguments and by different types of evidence. On this reading, the adoption of Lavoisier's new chemistry, in the 1780s and 1790s, came about in stages, as chemists with different backgrounds and different bodies of special expertise found particular experimental results and analyses of them compelling.

Scientific communities thrive because all their members share a common core of commitments and because they elaborate this shared perspective in different ways. Collective pursuit of inquiry has two facets: consensus and diversity. A major task for general philosophy of science is to abandon the fantasy of the "rational subject" who somehow recapitulates the history of science and to take collective inquiry seriously. Once that task is set, some new questions arise: What kinds of consensus are needed for successful investigation? How much disunity of opinion is valuable? How does broad consensus on some things, a consensus articulated in different and incompatible ways by different individuals, result in an enlarged consensus at later stages? To pose these questions is to ask for the logic of collective discovery—and my allusion to Popper's title (Popper 1959) differs from his in that my attention is not confined to the "context of justification". My aim in this part of my essay is to take some first steps toward answers.

II

I'll start with a concrete illustration, based on an approach to collective inquiry that I have previously elaborated (Kitcher 1990; 1993, ch. 8), but whose shortcomings I hope to remedy. Diversity of attitudes has two obvious advantages, so that there are two clear respects in which a community might be expected to do better than a single "rational subject", however widely-traveled along the paths of history. One gain stems from *the division of epistemic labor*: different people can come to know different things and their knowledge can be pooled. I suspect that a division of epistemic labor has pervaded human societies for tens of thousands of years (Kitcher 2011b, ch. 4). Another arises from *the division of cognitive labor*: different people approach unanswered

questions in different ways, selecting different strategies and, typically, holding different not-fully-supported beliefs.[11] My illustration concerns the division of cognitive labor.

Suppose a group of investigators agrees that a particular question is important; for concreteness, let the problem be that of finding the structure of a very important molecule (VIM). Two methods, M_1 and M_2, are available. Assume that each method is associated with a number, p_i, that measures the chance that that method will succeed, given that at least one person works on the method. Moreover, it's impossible that both methods succeed—if one works, the other won't. Finally, $p_1 > p_2 > 0$. Obviously, if everyone in the group pursues M_1—the more promising method—the chance of community success will be p_1. On the other hand, if the community divides the cognitive labor, with some working on M_1 and some on M_2, the chance that the group will find the structure of VIM is $p_1 + p_2$, which, by the assumptions, is greater than p_1. So diversity is good for the community, although not apparently for those who spend their time working on M_2 (the inferior method).

The supposition that each method is associated with a fixed value that measures its chances of success is implausible, of course. We'd expect that, at least up to a point, the more investigators pursue a method, the higher the chances of that method's success (although those chances would always be bounded above by a fixed number between zero and one that represents the chance given the ideal assignment of investigators). The algebra is more complex if you amend the story in this way, but it generates more fine-grained recommendations about the division of cognitive labor (Kitcher 1990). In the original scenario, any distribution of effort that assigned someone to each method would be equally good; under the modified conditions, there will often be a single optimum distribution that, given the available pool of investigators, maximizes the chance of success.

Turn now to a different question. Whether there are lots of good distributions or a single optimum, how could a group of investigators manage to achieve it? Notice that if each member evaluates the methods correctly—if each behaves as a "rational subject"—all will agree that M_1 is superior. So if each wants to maximize his individual chances of finding the structure of VIM, each should adopt M_1. That will lead to an inferior distribution of effort. The community would do better if some of its members were high-minded. Looking around them, they see that all their fellows are pursuing M_1, and altruistically decide to adopt M_2 "for the good of the community". That might require a considerable sacrifice (years of hard work and effort, with little chance of individual success).

The path to a good division of labor can be smoothed if there are appropriate incentives. Imagine that finding the structure of VIM is deemed sufficiently important that the first discoverer can expect to win a noble reward, a prestigious prize. Imagine yourself to be a newly-minted chemist, attracted to this problem. Ten people are al-

---

[11] I should note that the labels are arbitrary. I talked of "division of cognitive labor" in (Kitcher 1990), and have retained this usage, using "division of epistemic labor" for a related but quite distinct condition.

ready working on it, and all are pursuing M_1. You think that, if you join them, your chances of being first discoverer are no better and no worse than any of the others. The probability you assign to winning the race is thus computed by dividing the chance that M_1 succeeds—p_1—by the number of competitors (11 if you join the others). Since nobody is using M_2, similar reasoning will lead you to estimate your chances of winning if you adopt that (inferior) method as $p_2/1$. So it would be in your interest to pursue M_2, given that what you care about is the prize, provided that

$$11p_2 > p_1.$$

Once this point is appreciated, it's not hard to see that, whatever the ratio of p_2 to p_1, if the community of inquirers is large enough, this incentive structure can lead it to a valuable division of cognitive labor.[12]

As many scholars have noted, philosophers as well as historians and sociologists, scientific research functions as a credit economy (see, for example, Hull 1988). Whether or not there are prizes, receiving credit is important to scientists. What my illustrative example reveals is that this credit economy can *sometimes* work to generate states that, from the point of view of knowledge-seeking, are good for the community of investigators. It can serve as an example of a general approach to the logic of collective discovery—a paradigm, if you like—that might be extended to further attempts at modeling the structures and relations within community inquiry that are adapted to reaching good epistemic ends.

III

But there are reasons to tread carefully. As students of evolutionary theory are well aware, the strategy of identifying the optimal form of a particular trait as the inevitable result of natural selection can easily go astray (Gould and Lewontin 1979, Dupré 1987). The genetic and developmental basis of a trait may well issue in other aspects of the phenotype whose disadvantages outweigh the benefits of attaining the alleged "optimum". An analogous point holds in the case at hand. Before celebrating the ability of the credit economy to guide scientists to good distributions of effort, we should consider whether the supposedly beneficial forces have problematic side effects.

Only a little reflection suggests obvious dangers. If scientific teams are in competition, some of the benefits of cooperative work may disappear: for example, rivals who have obtained partial results may be reluctant to share their findings. Competition can easily create pressures to cut corners: you know your conclusions won't be accepted unless other researchers take you to have performed a series of experiments, and you think you know already how these experiments will come out—so, in the interests of "winning the race", you spend an evening manufacturing some "data". Another possibility, one I'll explore in Part 3, is that if you fail to convince your peers, you may

---

[12] It's plainly implausible to suppose that scientists can make exact estimates of the probabilities I've casually discussed. But I think it's entirely reasonable to suppose they can approximate the reasoning ascribed in my story. They can recognize that M_1 is more likely to work than M_2, but that the difference in prospects of success is not so large as to outweigh the effect of the greater competition—they see that the chances of being first to the discovery are greater with M_2 because, even though it's less likely to work, they are more likely to be the one who makes it work.

seek credit by "going public" and trying to persuade a wider audience that important issues have been foreclosed.

The general difficulty of preserving the benefits of competition without sacrificing epistemically valuable forms of cooperation should be evident, and should block the immediate conclusion that the credit economy is good for inquiry. Yet there is a more basic trouble, less evident than the dangers just reviewed. Diversity of opinion itself has mixed consequences. If a group of inquirers has a spectrum of views about which problems are important, about which methods are appropriate for tackling them, about the prospects of different methods, that would seem to promote useful divisions of cognitive labor. Yet, when some members of the community announce their putative results, heterogeneity of opinion can set obstacles in the way of reaching consensus: you think you have constructed a fully rigorous experimental case for your new finding, but, from my perspective, on which factors not varied in your samples are likely to be crucial, you have not done enough. Given a sufficiently wide spread of opinions, any new "discovery" may be hard put to convince enough members of the scientific community—subgroups may view others as neglecting significant factors, and thus as unreliable investigators. The diversity that facilitates division of cognitive labor can thus undermine division of epistemic labor.

Scientific communities require *balanced* dissent. As Mill eloquently argued long ago, it is intellectually healthy to have some dissenters around to challenge received opinion. Using the ideas of the last section, we can easily recognize how dissent works in the credit economy: those who joust with orthodoxy almost invariably lose, and are viewed as cranks (or worse); but when they succeed, the rewards are very large, bigger even that the noblest prizes, for successful dissenters become the heroes of history. Yet, while it is valuable to maintain some dissent, too much of it slows down the processes of consensus-formation, and, in the limit, might even doom a scientific community to perpetual disagreement and stagnation. What exactly is the right balance? And how, exactly, is it to be generated and sustained?

Posing such questions leads us back to Kuhnian themes. In stark contrast to Feyerabend—often regarded by philosophers incensed by "historical relativism" as Kuhn's ally[13]—Kuhn insisted on substantial agreement within a community practicing "normal science". His historical examples suggested the kinds of agreement he took to be crucial (although he gave no general analysis of them), and he tacitly supposed that mature areas of science had found a good level of agreement: they had discovered the right balance of consensus and dissent. Even if it proves impossible to say in general how much and what kind of shared commitment is required for collective inquiry to proceed efficiently, the logic of collective inquiry should supply tools with which to appraise particular fields at particular times.[14]

---

[13] It's important to recognize how the initial philosophical reaction to (Kuhn 1962) was so fixated on attacking "Kuhnian relativism" that philosophers failed to recognize the many important innovations of Kuhn's discussion—including some that his scientific readers found perceptive, and some that set him quite starkly at odds with Feyerabend. Fifty years on, I hope we can manage a more nuanced reading of the book.

[14] Here I apply to the present context a moral drawn in Part 1.

IV

Let us take stock. I have proposed a way of modeling some of the processes that go on in collective inquiry, and conceded that normative analyses of any of those processes must recognize the possibility that what seems useful in the specific context studied might create havoc elsewhere. To make further progress on the problem, it is necessary to have an overview of the course of inquiry, one that makes evident the kinds of situations in which an apparently valuable, or apparently counterproductive, property of collective investigation might have an effect. Again, evolutionary studies provide a useful parallel. To scotch suspicion that the genotype underlying a seemingly fitness-enhancing trait could also have negative effects, you need enough grasp on the life-history of the organism to recognize the kinds of context in which the genotype might influence reproductive success.

Already in §I, my alternative to the well-traveled "rational subject" sketched a picture of collective inquiry, as a sequence of episodes, all having a shared form. Using the evolutionary analogy, I'll call each episode a *generation*. At the beginning of each generation, there's a community of researchers who share certain attitudes: all of them agree on a set of propositions, on the significance of a set of hitherto unresolved questions, on a set of available methods for seeking answers to questions, on certain claims about the relative chances of those methods' succeeding, on a set of standards for accepting putative answers, and on a division of epistemic labor (they know who to go to for information about specific issues). Call this the *consensus practice*. They also disagree about many things. Some of them accept propositions denied by others, their judgments about the degrees to which questions are significant vary, some of them trust methods that others dismiss and the assessments of chances of a method's success are sometimes quite various, and some would only accept new "results" if they were to satisfy conditions that others deem irrelevant (no single argument convinces all scientists). Hence while each *individual practice* contains the consensus practice, the various individual practices go beyond consensus practice in divergent ways.

At the beginning of each generation, our scientists distribute themselves across the questions viewed as significant: that is, each scientist addresses some question picked out as significant within her individual practice—the question chosen may or may not belong to consensus practice. For each question actively pursued, there's a subcommunity of researchers who address that question, and this subcommunity divides itself among some subset (possibly improper) of all the methods that someone or other views as suitable (again, the method may be one that figures in consensus practice). All the investigators then go to work. During the course of the generation, researchers may inform one another of conclusions they take to be supported by the work they have so far done (where, of course, they apply the standards of their individual practices). At the end, all the conclusions reached are subjected to the scrutiny of all, and each judges by the standards of her individual practice. There are rules that determine how the consensus practice should be modified in light of the distribution of individual judgments. That modification is made, and the next generation begins.

The philosophical task is to understand the effects of various types of normative rules and incentive structures with respect to the achievement of the purposes of collective inquiry. Plainly, this presupposes an initial account of what those purposes are. Given that first specification, you might then try to find the optimal way of organizing collective research: which combinations of norms and incentive structures maximize the expected epistemic utility (where that utility is derived from the account of the purposes)? The lessons of the general theories sought by logical empiricism strongly suggest that any attempt to tackle that question in full generality would be too ambitious. Different areas of research are likely to call for alternative analyses. Furthermore, rather than seeking the *optimum*, a more useful first venture might be to try to understand how collective inquiry could go badly wrong. In my original discussion of diversity (the simple model of §II), the focus was on the trap of homogeneity. So a better way to orient the study of collective discovery is to ask whether various decisions for the social organization of research might make it more difficult to obtain the desired goals – does this apparently attractive norm or this incentive structure place "obstacles in the path of inquiry" (to express the worry in terms of a familiar pragmatist phrase)?

My reflections on the simple model of the credit economy answered one version of this question in an impressionistic and unrigorous way. I argued that, although the awarding of credit to the first discoverer might foster welcome diversity, through motivating investigators to explore research strategies that are currently underrepresented, it was likely also to interfere with the sharing of partial results and to tempt scientists to cut corners (even to manufacture data). To determine whether or not worries of this sort are well-founded, a more precise specification of effects is needed: the account offered of "generations" makes it possible to identify places at which trouble might occur, and we might then hope to understand the effects on attainment of epistemic goals (or, more precisely, the contributions to epistemic utility). Notice, however, that a fully rigorous analysis can only be provided if an entire package of norms and incentive structures is specified, for otherwise there is no way of telling just how attributions of credit will affect individual decisions. The approach must be holistic, both in considering all phases of inquiry and in attending to the interactions among potential norms.

One last preliminary point needs to be developed. Although social structures can play an important causal role in the course of inquiry, they aren't omnipotent. The interactions of individual scientists with the world they are investigating will make important and unpredictable differences to how collective research goes. Nature sometimes speaks in different ways to different scientists, generating diversity independently of whatever mechanisms for avoiding an unhealthy heterogeneity of opinion are in place. Similarly, however powerful the pressures to abide by consensus practice, the emphases on accepting some body of shared commitments, nature can sometimes question the most central presuppositions of a field—this, after all, was Kuhn's point about the emergence of anomaly. Just when, and to what extent, this can happen cannot be determined in advance. Choices of social structures for inquiry have to be made in ways that permit an *as yet unknown* world to guide research decisions.

The task is to set things up so that a scientific community is not so dominated by its preconceived ideas that it can never shift to a different perspective, nor so easily diverted from its consensus practice that it degenerates into a cacophony of voices with inconsistent messages.

I'll now try to forge some tools for working out a logic of collective discovery in particular instances.

V

Start with the specification of the purposes of inquiry. As Part 3 will propose, those purposes are often linked to broader goals of the societies within which the practice of science is embedded. Sometimes, however, perhaps most of the time, whatever the ends of the ambient societies, what scientific investigation is required to provide is truth. So we come to an obvious proposal: a generation counts as successful insofar as it concludes with the certification of new truths.[15]

This proposal requires modification in several respects. First, *mere* truth is readily achieved—all of us can rack up new truths at a wonderful rate, simply by using our senses and by applying the elementary parts of deductive logic. What research aims to deliver are *significant* truths, correct answers to significant questions (Kitcher 2001 Chapters 5-6). Second, you can't cheat by lowering the standards of certification: even if you were lucky, and your undemanding requirements only let in truths, it's important that the newly-adopted conclusions be recognized as stemming from reliable methods (where the standards of reliability are often indexed to the sources of significance, sometimes set by the internal needs of research, sometimes by the external factors to be considered in Part 3). Third, as pragmatists rightly remind us, inquiry aims both to acquire truth and avoid error (James 1896 [reprint 1956]). It's important not only to add true answers to significant questions, but also to avoid adding false answers, especially when those falsehoods would block the ability to pursue further significant issues. Moreover, a generation might count as successful not in virtue of adding any new truth at all, but through its eradication of a fundamental error, one that stood in the way of addressing significant questions. Fourth, the success of a field of inquiry is sometimes dependent on the timing of its discoveries. Although a long-term perspective may be appropriate for some questions, there are problems that need to be solved within a particular time frame, or even as fast as possible. Finally, just as there can be divergence between what a scientific community believes and the truth, so too the judgments of significance can be at odds with what is genuinely significant. One epistemic benefit that a generation might deliver is the progressive revision of ideas about what questions are significant.

---

[15] Does my casual reference to truth here signal a departure from pragmatism? Or does it involve commitment to a special pragmatist theory of truth? I have attempted to show how pragmatists can reconcile a modest scientific realism with a naturalist stance and a modest correspondence theory of truth: see (Kitcher 2012, chs. 3-5). I am grateful to José Miguel Sagüillo for raising this issue. This is not to deny that pragmatists might choose to elaborate a notion of truth in other ways, as for example in (Frápolli 2013).

Is there a single way to specify the proposal that inquiry seeks significant truth so that it will apply to all domains? I think not. Some areas of research may be dominated by a single problem, whose perceived significance dominates all others; they may be mature enough to be confident that this judgment of significance is correct; further, because the problem is so central to any future investigations, the standards of certification must be set as high as possible—all the professionals, with their disparate standards for adopting new conclusions must be convinced. Perhaps particle physics at the moment is like this, and the search for the Higgs boson is legitimately given overwhelming priority (the discovery of the Higgs particle fell between the preparation of the lectures and the writing of this essay). Other areas may be dramatically different. Their recent difficulties may suggest that existing perspectives need to be enlarged, and that something may be wrong with present assumptions. So the primary epistemic benefit sought at once is the eradication of erroneous presuppositions, and possibly the refinement of judgments of significance. Major parts of economics may currently be in this condition.

Hence my distinctions about significance, certification, the balance between attainment of truth and avoidance of error, and the time-rate of problem-solutions need to be articulated in the context of different fields. Yet, even without specific elaborations, some traps are readily recognizable. For those areas that are expected to deliver results in time to solve some practical problem (climate science, parts of biomedical research), it's imperative to avoid social structures that would be expected to preclude the possibility of achieving the goal in the appropriate time-frame. For any domain of research, social arrangements that make the uncovering of fundamental errors impossible are to be avoided.[16] I shall now consider a few scenarios highlighting recurrent ways in which collective inquiry might fail to achieve some valuable epistemic goals.

VI

Say that a community is *Kuhnian* just in case all of its members, at all times, pursue projects that are framed by consensus practice: in the design and execution of research, as well as in the certification of results, no claim at odds with the doctrines of consensus practice is admitted or considered. Different members of Kuhnian communities may structure their research by commitments that go beyond the elements of consensus practice—their investigations rest on accepting answers to questions consensus practice leaves open—but their work embodies nothing that contradicts consensus practice. Does a Kuhnian community fall victim to dogmatism? Although Kuhn once (perhaps incautiously) wrote of "the function of dogma in scientific research" (Kuhn 1963), the view of *Structure* is that Kuhnian communities are not doomed to repeating their initial prejudices interminably. *Anomalies* can be recognized—as in Renaissance astronomy or in Röntgen's discovery of an unsuspected form of radiation.

---

[16] Again, there are connections to a fundamental pragmatist insight, encapsulated in Peirce's review of methods of belief-formation (Peirce 1877 [reprint 1992, ch. 7]).

Not all philosophers of science have agreed that the intense programs for training apprentice researchers so that they share a rich body of commitments are harmless. Feyerabend proposed a "principle of proliferation" because he believed that self-conscious pursuit of alternatives was required to escape the confinements of dogma (Feyerabend 1975, 1978). That principle is not refuted by citing the familiar examples of Kuhnian anomaly. It would be possible to concede that some few instances of subversion of framing commitments can occur without the elaboration of alternative perspectives, while insisting on the inability to discern the errors of many of our deepest presuppositions unless some effort is dedicated to developing rival approaches. Of course, any argument for this must be indirect: you can't show how difficult it is for us to identify our mistaken assumptions by pointing them out—they are *ex hypothesi* unidentifiable; the best that can be done is to use historical examples to argue that problematic framing assumptions start to become visible when alternatives to them are proposed and articulated (this is essentially Feyerabend's strategy). Hence it's extremely hard to offer a defensible thesis about the extent to which we're held captive by framing presuppositions.

It may seem, then, that the Kuhn-Feyerabend debate about the value of consensus is frustrating and irresoluble. I propose a different judgment. The debate calls our attention to an empirical question we can't currently settle: how much, and how dangerously, are we confined by the assumptions that guide research? Even though we can't know the answer, it's possible to appreciate the potential danger. Equally, you can recognize the danger of forswearing all consensus in favor of a massive proliferation of alternative points of view: that strategy might lead to interminable disagreement. Inspired by both threats to successful inquiry, a safe strategy might be to supplement whatever subversion might directly and accidentally come from the emergence of Kuhnian anomalies by institutionalizing a dedicated collective policy of fostering dissent.

> *Valuable Dissent*: Scientific communities should contain a small subgroup of researchers who devote themselves to exploring alternatives to the basic assumptions that guide the research of the other investigators. Their task is to formulate challenges to consensus practice.

We can imagine that the principle is enunciated as an explicit norm for the community. It might also be supplemented by awarding credit for effective forms of dissent.

Even though the principle is deliberately vague—just what is the right amount of dissent?—it's possible to confront it with an obvious difficulty. Consider that stage in each generation at which new findings are certified. A potential norm for governing the certification is:

> *Total Agreement*: New statements are added to the body of doctrine in consensus practice just in case all members of the community judge that the evidence adduced in their favor conforms to the standards they would apply.

If *Valuable Dissent* is conjoined with *Total Agreement*, trouble is likely to ensue. For, given that the new "results" are obtained on the basis of studies that presuppose parts of consensus practice, some members of the community will probably find the evi-

dence brought forward inadequate. More exactly, if the framing assumptions are the statements in a set $\{F_1, F_2, \ldots, F_n\}$, and if each piece of new research requires at least one of the F_i to be true if it is to be well-supported, and if, for each i, there is some member of the group of dissenters who challenges F_i, then no piece of new research can obtain unanimous support. The only way new research could be certified under *Total Agreement* would be for it to be free from presupposing any of the F_i (in which case the framing assumptions would be playing no substantive guiding role) or for it to depend on some F_i not questioned by any of the dissenters (in which case the dissent would be incomplete). A community that adopted *Valuable Dissent* and *Total Agreement* would only be able to achieve incremental additions to its body of accepted statements if the statements were independent of its framing assumptions or if its forms of dissent were inadequate to the scrutiny of all the presuppositions.

There's an obvious remedy for this predicament. Dissenters play a special role in the community, that of trying to eradicate the errors that are taken for granted in attempts to accumulate further knowledge. That role cannot and should not be combined with the active task of appraising research that's framed by the sorts of assumptions the dissenters are re-examining. So it's appropriate to limit *Total Agreement* by a rule.

> *Exclusionary Rule*: For the purposes of applying *Total Agreement*, dissenters don't count as members of the community.

Now, of course, the trouble from conjoining the two principles is resolved. You might wonder, however, if the resolution is purchased at cost of undercutting the benefits of dissent. I suggest that it is not. The rationale for *Valuable Dissent* is that articulating alternative perspectives to the framing assumptions may expose problems with those assumptions—effectively generating analogues to Kuhnian anomalies, which arise without the need for rival points of view. *Once* the dissenters have produced such quasi-anomalies, providing grounds for dissatisfaction with a particular presupposition, they have done their work, and that assumption can no longer be unquestioningly used in the accumulation of further research results: in that case, *Total Agreement* with the *Exclusionary Rule* will debar the accumulation (since non-dissenters now have grounds for suspending the crucial assumption). *Before* the quasi-anomalies are at hand, however, there's no harm in continuing to deploy the assumption, since no grounds for doubt have been provided.

VII

I've started with the example of encouraging challenges to framing assumptions because it exposes an important feature of collective inquiry, the possibility that a variety of points of view might combine the benefits of research guided by strong presuppositions with the possibility of escaping dogmatism. Moreover, the relatively simple problem of squaring the encouragement of dissent with a plausible idea about certification prepares the way for a more interesting challenge.

In parallel to *Valuable Dissent*, you might favor a principle encouraging members of a scientific community to extend their individual practices beyond the shared commitments enshrined in community practice.

> *Valuable Diversity*: It is good for a scientific community if its members differ in their opinions about which unsolved problems are most significant and differ in their factual beliefs on open questions whose answers bear on research strategies for addressing unsolved problems the scientists take to be significant.

Valuable Diversity is motivated by the wish to focus research on a range of issues, and to pursue investigations through the use of a variety of strategies. In accordance with my earlier remarks, I don't suppose that this principle is a good norm for all fields at all times—it may, for example, be quite inapt for the present state of particle physics—but some of the considerations about diversification of approaches I suggested in §II show that it would be valuable in some contexts. But the worries that surfaced in §III provoke a question: can *Valuable Diversity* be combined with *Total Agreement* (subject to the *Exclusionary Rule*)?

Consider a pair of examples. First, imagine a community of biologists who want to resolve a question about the effects of a drug on mammalian behavior. The species of principal interest to them is *Homo sapiens*, but they all agree that injecting human subjects with the drug would be dangerous unless there had been prior trials on non-human animals. Some of them think mice are the appropriate experimental animals, others prefer dogs, and yet others select monkeys. Those who favor mouse models are criticized by their rivals on the grounds that mice cannot reveal certain potential behavioral effects. Those who prefer investigating monkeys face challenges on the grounds that their samples do not exhibit a sufficiently large range of genetic and environmental variation. Dogs are judged to be superior to mice in their ability to express behavior, and superior to monkeys in the ability to control for potential interfering factors, but they aren't completely satisfactory in either respect. Hence, *because of individual heterogeneity in belief*, which plays a valuable role in suggesting rival strategies for experimentation, none of the bodies of experimental findings is judged sufficient by all of the members of the community. Benefits in experimentation are paid for with difficulties in achieving certification.

Second, consider a community of investigators into the history of life on earth. Many members of this community suppose that explanations of particular transitions are acceptable if it can be shown that no rival account can be given by appealing to the causal agents recognized in orthodox neo-Darwinism. Some other members, however, hold that the failure to give precise neo-Darwinian accounts of particular ancient events—the origins of the cell, for instance—demonstrates that the neo-Darwinian inventory of causes is incomplete, and that explanations must rule out rival "forces" (perhaps even an "intelligent designer"). They reject every explanation offered by the orthodox as unsupported. Although the diversity in exploring "expansions" of Darwinism is valuable, it brings stagnation to the project of elaborating paleontological knowledge.

The structure of the problems is as follows. Suppose valuable diversity in research strategies is attained by means of a distribution of beliefs, $\{B_1, B_2, \ldots, B_n\}$, where the-

se beliefs go beyond the corpus of statements accepted in community practice. Let the experimental results achieved by groups with these beliefs be $\{E_1, E_2, \ldots, E_n\}$. For each E_i there is some j such that, from the perspective of B_j, E_i is inadequate to support the conclusion for which it is alleged to be evidence. Then *Total Agreement* supplemented with the *Exclusionary Rule* will not allow for any of the experimental results to certify any of their intended conclusions. (Variants of the problem are easily generated by concentrating on alternative judgments of significance, rather than on differences in factual belief.)

The obvious way to address this problem is to extend the *Exclusionary Rule*. Now it wouldn't be promising to make that extension by taking those outside the group actually performing the experiment to count, for the present purposes of certification as quasi-dissenters: that would effectively forego the advantages different perspectives might bring in the appraisal of new "results", by making a privilege of the idiosyncratic beliefs different groups have brought to their experimental design. A more promising suggestion is to combine *Total Agreement* with a restriction on challenges to experimental studies.

> *Legitimate Challenge*: Total Agreement is only vitiated if the evidence adduced is subject to legitimate challenge, where legitimate challenge to an experimental study must be based only on parts of consensus practice.

Effectively, *Legitimate Challenge* aims to differentiate the acceptable use of idiosyncratic commitments (parts of individual practice that go beyond consensus practice) in the context of pursuit from attempts to import those commitments into community-wide discussion of which new findings are to be accepted.

Yet there are grounds for thinking that *Legitimate Challenge* is too weak. For, if the questions resolved by the additional beliefs (the B_i) are genuinely open, then, from the perspective of consensus practice, for each B_i it's possible that B_i is true. Hence the critique may be rephrased so that it depends only on consensus practice. If B_j casts doubt on the validity of E_k, then given the possibility that B_i is true, there's a possibility that this study has failed to exclude, and a fully rigorous case for the conclusion must go further.

If this is correct, then combining *Valuable Diversity* with *Total Agreement* is troublesome. That is readily seen in the second example, where diversity appears to generate a veto on accepting any research. With respect to the first example, however, the result might appear to embody an admirable caution—maybe the right thing to do is to suspend belief about the effects of the drug. Notice, however, that the result is generated quite independently of the needs the drug would address: it holds if the drug is potentially enormously beneficial, and when its function is to remedy some minor discomfort. Moreover, the blanket dismissal of the individual research findings blocks the possibility of making a preliminary case for limited human testing on the basis of suggestive, but inconclusive, results that reinforce one another.

Comparison of the two examples is interesting, in showing how different types of diversity can have more or less sweeping effects. Fostering some types of diversity allows the questioning of a certain set of cherished assumptions, while permitting some significant problems to be resolved. Broadening the scope of self-criticism yields a

more searching attempt to eradicate error, but at cost of limiting the extension of knowledge. There is a structural analogy with Lakatos' methodology of research programmes: consensus practice corresponds to the "hard core", and the forms of diversity that are tolerated to the "protective belt" (Lakatos 1970).[17] The community's choice of what forms of diversity are to be tolerated (and which count as heresies to be cast out) encapsulates a decision about the importance of eradicating error (and where the likely errors lie) and about the benefits of adding new information (and with respect to which open questions).

Behind that decision, of course, lies James' fundamental point. Strategies for inquiry rest on value-judgments: how is the obtaining of new information (potential truth) to be weighed against the avoidance of error? No general answer to that question can be given. My exploration of some extremely simple norms of collective investigation is intended to prepare the way for detailed examination of actual research communities, to expose the value-judgments implicit in their practices and (eventually) to assess those judgments.

VIII

In conclusion, I want to return to some issues about the credit economy that I posed earlier, and to relate them to the points about diversity just adduced. Imagine a scientific community whose educational system encourages a few bold (and, perhaps, ambitious) dissenters, who are playing for the really big prizes in the history of their science. For the rest, the training they receive inculcates a very strong commitment to "the fundamentals"—that is to consensus practice—but also imparts the idea that researchers must strike out on their own, that they should make conjectures that go beyond what is regarded as firmly established, design experiments partially in the light of those further idiosyncratic beliefs, and not be too easily discouraged if their efforts don't immediately succeed. If one of the various ventures actually resolves some significant open question, by allowing for certification of the answer (under *Total Agreement*), then the first researcher (or team) to answer that significant question will receive credit; and the more significant the question, the larger the reward.

One issue that arises under these envisaged circumstances concerns the possibility of cooperative research.[18] Suppose one member of the community has arrived at an experimental finding that takes a step towards solving the larger problem. Should that finding be shared with others? Should it be published? The questions are distinct. There are circumstances under which two scientists (or teams of researchers) might pool their efforts, without disclosing their results on a broader scale. Each regards the other as having taken a step towards the ultimate goal, and they decide to go on as a team and (if they are fortunate) to divide the credit; information is withheld from the broader community (which contains their competitors). Publishing would be a good strategy only if the awarding of credit is proportionally based: if you solve what is

---

[17] Some perceptive remarks of Bence Nanay suggested this connection to me.

[18] In the discussion that follows, I'm indebted to discussions with Thomas Boyer, who has developed some interesting models concerning the advantages or disadvantages of publishing partial results.

agreed on as 0.3 of the problem, you get 0.3 of the credit. When the attribution of credit is massively skewed towards finding the answer to a significant question, the rewards come at the end and preliminary results are given very little credit, those who care about credit would not do well to divulge their intermediate findings. The most they might be expected to do would be to find some potential allies with whom they could enter into profitable informational exchange.

The earlier worry about combining the incentive structures of a competitive credit economy with the free flow of information could be addressed by introducing norms recognizing partial contributions. That could be done by retreating from the current practice of celebrating the big achievements that serve as the culmination of a line of research, and settling for a system of proportional credit. Or it could be achieved in a different way by imposing a team structure where none has been explicitly agreed. The credit economy might be regulated by a norm.

> *Recording Debts*: It is the responsibility of those who answer significant questions to include those on whose findings they depend as virtual co-authors and to make fully explicit the exact nature of the contributions made. Reports on such discoveries must preserve the structure of the teamwork.

Recording Debts aims at restoring equity, and so encouraging the free flow of information.

Yet this is to make the problem of integrating competition and cooperation appear too easy, for the focus is on a *public act*, the announcing of an answer to a significant question. Assuming the presence of some norm like *Recording Debts*, those who publish partial results are protected. If their rivals build on their achievements, those rivals must either abide by the norm, or they can be called to account. There is a public forum in which claims to contribution can be evaluated. Matters are different, when those who intend to exploit their competitors can hide.

Return, for the last time, to the phase at which results are certified. Actual processes of certification involve the scrutiny of the evidence submitted. Potential reviewers have two opportunities for taking advantage of those who put forward their partial results. First, they have access to information their competitors do not yet have. Second, by preventing, or delaying, the publication of the research submitted, they can avoid being held to a norm like *Recording Debts*. In cases where the stakes are high, and when the competition for credit is intense, it's quite possible that the certification process will contribute to an unjust assignment of credit, and that scientists who appreciate this will be reluctant to make their intermediate findings public.

I have set up the scenario in stark terms, inviting you to consider a thoroughly unscrupulous gatekeeper, who abuses the role to secure a competitive advantage. Perhaps those who would act this way are very rare—although who is in any position to know that?—but it's quite easy to conceive of episodes in which a more complex set of judgments brings about similar consequences. Recall the earlier points about the effects diversity of belief might have in the context of certification. Our reviewer has an individual practice that goes beyond consensus practice. He has additional commitments that guide his own research, and that diverge from those figuring in the experimental work he is evaluating. Instead of regarding the work that has been done as

completely convincing, and then going on to do what he can to suppress it, to his own selfish advantage, he may just see it as deficient, as having ignored some important possibility. He now sees how to do the work properly—and that will occupy him during the coming weeks—but there is no need to acknowledge a debt to the wrongheaded, blundering efforts he has been appraising.

IX

To sum up—a few conclusions. First, there are serious questions about the logic of collective inquiry, some of which I've been trying to illustrate in a preliminary fashion. Second, in accordance with the first lecture, I don't see those questions as opportunities for constructing some large general theory—How to organize an ideal scientific community—but as prompting analyses and concepts that might then be further elaborated with respect to particular areas of inquiry at particular stages. Third, my discussions have not proceeded from any thorough scrutiny of the actual social organization of any of the sciences, but have been inspired by reflection on well-known episodes and situations from historical and contemporary investigations: that should be amended, by looking carefully at the actual ways in which research strategies are generated and findings collectively certified. (As far as I can tell, contemporary sociology of science has not focused on these questions, although they present opportunities for useful collaborations between philosophers and sociologists.) Finally, I have insulated inquiry from society, looking at collective inquiry as if it were carried on in a closed community. That is a piece of fiction, now to be replaced by a wider, more realistic, perspective.

PART 3: SCIENCE AMONG OTHER INSTITUTIONS

I

Although there are many sciences, each with their own proprietary questions, techniques and methods, it is plausible to think of them as parts of a single institution, Science, that plays an important role within broader societies. It was not always so. For centuries, in many western nations, the Church fulfilled double duty, serving both as the locus at which individuals received spiritual guidance and as the center of worldly learning. Part of the revolution initiated by the "new science" of the seventeenth century was a sequence of institutional changes: first, a few ventures in seeking knowledge of nature outside the traditional places from which authoritative pronouncements were delivered (the "schools"); second a flourishing of inquiry in the new centers, in the scientific societies that had been constituted; third, a reaction to the successes that members of such societies had achieved, and a transformation of universities to reabsorb both the fruits and the processes of research; fourth, the recognition of the importance of the enterprise, whether carried out in university settings or in government laboratories, to the health and wealth of nations—the deep embedding of Science within public policy and within political debate. Nor is this the end, for, in our own

times, it is clear that scientific research is finding new homes and new patrons, that it is becoming increasingly integrated with the capitalist marketplace.[19]

Those who took the first small steps in the 1660s could not have anticipated these developments. In London and in Paris, scientific research was to satisfy the disinterested curiosity of a few.[20] Membership in the Royal Society was restricted to "gentlemen free and unconfin'd"—tradesmen, whose worldly interests might corrupt the course of inquiry were excluded, and women were not deemed suitable for admission (only in 1945 would come the first female fellows). The rules originally laid down for governing the choice of investigations, the discussions, and the certification of new ideas were adapted to a quite different context, designed to achieve correct answers to questions the gentlemen-scientists viewed as significant. Three and a half centuries on, the intellectual descendants of the pioneers work in a very different social environment, one in which their decisions can have immediate impact on social policy. It is worth asking whether an institution that had no inkling of its future influence and that has evolved by happenstance is well-designed for the functions it now fulfills.

Devout believers in an orderly cosmos, and typically confident that a wise creator was behind it, seventeenth-century "natural philosophers" took the ultimate sources of significance to lie in the general rules imprinted upon the creation: their task was to fathom those rules and so "think God's thoughts after him". Even when the questions they posed might bear upon policy-decisions, they did not think in terms of pressing deadlines for the resolution of the questions that occupied them. What counted was getting things right, not falling into the "vain speculations of the schools", and so they emphasized the importance of painstaking experiments and a wide range of observations. They were to be on guard against premature hypothesizing, and concerned with the eradication of error.

There is no harm in adapting something close to this older conception of the goals of inquiry, and using it as the standard of individual investigation or as the source of norms for collective discovery. In a secular age, the religiously-inspired notion of significance might give way to something different: a search for principles that disclose the organization of nature, or answers to questions that express human curiosity, say. Building on a concept of significance of that kind, it is possible to explore the social procedures that are well-adapted to attaining significant truth (where significance is a matter for the expert investigators to judge)—or, perhaps, to identify which social procedures might not foreclose the possibility of achieving that goal—as Part 2 did. That is *one* way to consider the social pursuit of research, but it should not be the only one. Precisely because Science now plays so much larger a role in human life, the analysis of an institution that seeks significant truth (again: as defined by the expert community) should be supplemented with a broader vision, one that sees how the sciences are embedded in human life and in human societies, that pays attention to their ex-

---

[19] The evolution of public knowledge is discussed in (Kitcher 2011b), particularly in ch. 4.

[20] As Lorraine Daston reminded me, many of the members of the first scientific societies also hoped that their activities would generate "useful knowledge". Even though their curiosity was disinterested, it was sometimes focused on what they took to be "good for mankind".

tended functions, and that asks how they might be well-suited to a larger collection of tasks. Science should be considered as part of a framework of social institutions.

II

What are the goals Science should deliver, if it is to play its crucial part in the division of epistemic labor that has evolved in the past centuries? A proposal recommends itself. Inquiry generally aims at delivering reliably-based answers to significant questions, answers that are true enough to satisfy the purposes of those to whom the questions are pertinent and that are able to be disseminated to the people who need them. All the unanalyzed notions that figure in this characterization are pragmatic. What makes a question significant, what counts as a reliable basis, what constitutes an answer as "true enough", and what conditions suffice for accessibility are all dependent on characteristics of people, their aims and values, and their social organization. Although much of traditional philosophy of science comes, reluctantly and with embarrassment, to a final, cursory, discussion of "science and values", that theme must enter at the beginning of any serious approach to science within society. The topic is better conceived as "values and science".

Normative analysis of institutions is concentrated on asking how a given *framework* of institutions promotes or retards collective human well-being. The central questions that immediately arise concern how we should conceive well-being for individuals and how distributions of individual well-being (including, perhaps, relational properties) affect the well-being of a group. I shall summarize, but not defend, some proposals I have made elsewhere (Kitcher 2011c).

Following Mill – not the author of the familiar reading of *Utilitarianism* – but the philosopher who wrote a whole sequence of other major works, I suppose that well-being is not reducible to momentary states of subjects (pleasures and pains) but is a holistic property of a human life.[21] Our lives go well when they are founded in a conception of who we are and of what is important for us, a conception I'll refer to as our *life theme*. Reflective people can sometimes be fully explicit about what their life theme is, but you don't need that kind of clear self-knowledge to have one: the theme might be expressed in the pattern of your activities and daily projects, or you might only be able to identify it, if some keen observer of you offered you a characterization of it. It's important that your theme be the product of your own choice, that you have not been coerced into it or simply fallen into the categories to which the traditions of your society have assigned you. That theme marks out particular goals for you, specifies some central desires that it's important for you to satisfy. It's necessary for you to live well that your pursuit of these aims should not be too unsuccessful: you need to realize some of the ends that matter most to you.

Although Mill sometimes writes as if the conditions on life themes are minimal—we are to pursue our own good in our own way, so long as our doing so doesn't interfere with the like activities of others—I hold that there are important constraints on the choice of an adequate theme. To Mill's negative requirement that we mustn't get

---

[21] For further discussion of this approach to Mill, see (Kitcher 2011c, ch. 8).

in one another's way I'd add the demand that worthwhile themes must involve relations with others and their projects. As Dewey saw (Dewey 1927), it's not primarily individuals, but rather groups of individuals engaged in joint enterprises that need protection from outside interference. Further, life themes should be directed towards goals with effects beyond the lifetime of the individual or the group. We cannot ask that the impact of what we do is permanent, but part of our living well is to achieve something that, for some short time, makes a difference in the lives of others.

Plainly what I have said here is schematic and rough, and a lot more can and should be said to develop some of the details. But I hope a relatively brief outline will yield an account of the epistemic good sufficient for my present normative purposes. The next task is to generate the notion of collective well-being. Let's say that a *collective future* is based on the individual well-being of all the people who actually exist in it. A first constraint on valuable collective futures is that they allow each of those people to enjoy serious and roughly equal opportunities for developing and pursuing worthwhile life themes: that is, they must have the chance to make uncoerced choices, and real prospects of satisfying their aspirations. The second condition attempts to specify how collective futures in which different individuals thrive to different degrees are to be compared. Here, I propose, the right tradeoffs are to be those that would be endorsed in an ideal discussion among representatives of the full range of points of view found among the people who figure in the collective future, where an ideal discussion would be one in which identifiable factual errors were eliminated, and in which each discussant was concerned to arrive at an outcome in which each of the others was able to realize his prospects to the greatest possible degree. This condition, too, involves an ideal of equality. (As will appear, the conception of deliberative discussion it favors will figure prominently in later discussions.)

It's now possible to begin the final project of this essay, the articulation of standards and norms for Science, no longer conceived as a collective process of generating significant truth, but as an institution that promotes human well-being. I'll proceed in two stages, first by relating the search for information directly to the idea of collective well-being just sketched, without considering other social institutions that might modify the impact on human lives, and second (more briefly) by explicitly placing Science in a matrix of other institutions.

III

Collective inquiry generates answers to questions. To which questions? And must the answers be true? How reliable must the generating procedures be? What are the requirements on the modes of presentation of the answers?

Start with a very simple point. There are some questions for which *public* answers would be disastrous. If those questions are ever answered, we would fervently hope the deliverances were not socially certified and widely disseminated. Consider, for instance, the question "How can one build an explosive device, equivalent to a nuclear bomb, using ingredients that anyone can acquire?" If that question were answered directly (that is, not by denying the presupposition—saying "You can't do it"), public certification of the answer would be expected to have disastrous consequences. At

least, that is true, given the variation in temperament among members of our species, and given other institutions we have: were there to be a fullproof system of detecting and preventing all forms of criminal and anti-social behavior, or if there were an educational regime capable of turning all pupils into angels, the pressure on Science would be different. I shall not speculate on the probability of our ever achieving institutions like that.

There are many other questions so irrelevant to human well-being that pursuit of them would be a frivolous waste of time and energy. Part 1 raised the possibility that philosophical programs can degenerate into preoccupation with such questions. Perhaps that is also true for parts of science. Those who think of themselves as trying to satisfy "disinterested curiosity", researchers "free and unconfin'd" might fall into this trap.

So imagine that the disastrous questions and the irrelevant ones are discarded. Plenty will remain, almost certainly too much that the entire collection can be thoroughly investigated. How, then, should the truly significant questions be selected from this extensive remainder?

I propose that the significant questions are those that would emerge under *Well-ordered Science*.[22] Well-ordered science envisages an ideal discussion, in which representatives of all points of view, tutored to expunge identifiable errors, aim to advance each of the human projects to the greatest possible extent. In recapitulating the notion of ideal discussion I already introduced in generating the idea of collective well-being, I intend to link the ideal of well-ordered science to the promotion of collective well-being, to give content to the idea that the aim of scientific research is to advance the collective human good.

More precisely, the notion of significance is grounded in that of an ideal plan of research, itself generated from the ideal discussion. The ideal discussants start by reformulating their own individual projects in light of the information they receive from people who convey the current state of knowledge: in this phase they are tutored, so that identifiable factual errors are corrected. In consequence of the tutoring, they can recognize all sorts of opportunities for research, places in which further effort might resolve outstanding questions. They work out an ideal plan of research by seeking a combination of projects that would do as well as possible in promoting human projects, perhaps transforming some people's chances of success, while simultaneously endeavoring to ensure that nobody's interests are slighted, so that all those projects could be sustained, with reasonable chances of success, given the resources available. As Part 2 noted, the likelihood of a strategy's delivering the information sought will typically increase (at least up to a point) if more investigators pursue it. The deliberators thus have to balance their desire to endorse a wide range of lines of investigation against the possibility that resources will be spread so thin, that none of them has much chance of success. In doing this, it's possible that they might prefer a risky venture with potential for massive impact, to a safer bet whose consequences for human well-being were less far-reaching.

---

[22] I introduced this idea in (Kitcher 2001). The present account of it reflects the subsequent elaboration undertaken in (Kitcher 2011b).

Well-ordered science generates not only standards of significance, but simultaneously links them to considerations of approximation and of reliability. If a question wins a place on the ideal plan of research, the discussants will already have had to consider how close to the truth an answer would have to be to meet the needs of those whose projects would be advanced by receiving information about the topic, and they will also have had to think about issues of reliability. In their budgeting of resources, these issues automatically arise, for it must be decided if it is worth aiming at a higher degree of reliability or a closer approximation to the truth, rather than creating opportunities for lines of investigation that could bring quite different benefits. Further, the linkage of the questions endorsed to the lives of specific types of people usually determines the constraints on the necessary mode of presentation of any answers that are forthcoming: the information has to be accessible by and intelligible to those populations through whose needs it earns its place on the agenda.

Some clarificatory remarks are surely in order. First, the perspectives included should be those likely to appear in our collective future—not only those of *all* our contemporaries, but also those we estimate as present among our descendants. That is as it should be. If you are inclined to think that the international scientific community should be divided up into national groups, with each nation determining the ideal plan of research for meeting the needs of its citizens, you need some basis for judging that *this* division of the population is the correct one. Why not, instead, apportion representation by contribution to national productivity, or by the size of an individual's tax bill? Further, it is hard to imagine how you could make a compelling case for slighting the pressing needs of people from parts of the world too poor to support local programs of research, especially if that case were to be made in a forum in which those people were represented. So my ideal of well-ordered science is internationalist and cosmopolitan.

Second, there are obvious worries about the possibility that certain kinds of questions would be slighted. Should the focus of scientific research be so thoroughly practical as to ignore all large considerations of pure curiosity? Should it turn its back on those apparently risky proposals offered by investigators whose track record marks them out as brilliant and innovative? Can you really foresee the potential course of inquiry well enough to make judgments about the likely impact of different lines of inquiry? All these questions deserve answers (if only brief ones).

First, there may be good evidence that investment in certain forms of theoretical research will reap rich dividends in practical solutions in the future: sometimes, even when problems are urgent, the better strategy is to move patiently ahead with the elaboration of "basic science" (Morgan's strategy of focusing on the fruit fly eventually led to the molecular biology that now underlies much biomedical research—part of the tutoring of the deliberators under well-ordered science should acquaint them with parts of the history of science that expose possibilities like this one). Furthermore, even when a question cannot be expected to prepare the way for applications that would be of practical benefit to many, answering it may be worthwhile. Ideal deliberators might well agree that a question of purely theoretical interest was so interesting for so many people, that the epistemic benefits of knowing the answer outweighed the

practical payoffs of other inquiries. Well-ordered science does not eliminate the purest forms of inquiry. It does ask, however—and I think rightly—whether from a range of mutually-engaged, well-informed human perspectives, satisfying pure curiosity can outweigh attending to other, more mundane, human predicaments.

Similar considerations support the possibility that well-ordered science might sometimes endorse a line of research merely on the stellar track record of its champions. The fundamental aim of the ideal deliberation is to pick out the best ways of contributing to the collective good, and among the grounds for judging are facts about past patterns of success. Just as those patterns sometimes recommend a patient and indirect approach (witness Morgan's decision not to tackle human genetics head on), so too ideal discussants should be aware of generalizations about the qualities associated with large breakthroughs. The important point is that deliberation should be informed by everything that is known about research, including not only substantive details about the promises of particular fields but also the social and personal characteristics that make a difference to success and failure.

Yet, despite all the information that might ideally be garnered, you might still wonder whether serious judgments about significance could, even in principle, be made. For nature is surely unpredictable: nobody could have foreseen that efforts to reform the church calendar might have led the canon lawyer who sought improvements to offer a radical revision of astronomy. Even if they are made under ideal conditions, judgments of significance would often turn out to need revision by the next generation of deliberators—but that point should be no more troubling than the everyday fact that planning for the future, whether undertaken in families or in larger units, is a continual process, one in which the best designs of yesterday need to be modified today, and those now advanced will surely require further amendment tomorrow. The concept of significance behind well-ordered science insists that there is no outside source of value, that all people have to go on is our own recognition of alternative points of view, our appraisals of admittedly uncertain prospects of advancing various projects, and our best efforts to engage with the perspectives of all. Despite the lack of certainty endemic to any such procedure, its deliverances should no more be slighted than those made in humbler situations, in the planning with which small groups attempt to prepare for an uncertain future. It would be folly for such groups to fall into a forlorn skepticism, throwing up their hands in despair and electing some random process for choosing their options. By the same token, the admittedly revisable ideal discussion provides an account of significance that is worth having.

Many scientists bridle at the thought that attributions of significance depend on the judgments of outsiders. They are quite right to protest the potential tyranny of ignorance, generated from the most vulgar form of democracy, in which the uninformed opinions of the many are given weight. Well-ordered science, however, proposes that the full range of human perspectives be considered only after identifiable factual errors have been corrected. The democracy it introduces corrects a form of bias that would be present if matters of significance were simply the domain of elite researchers, bringing into clear view human needs that inquiry can all too easily overlook. Insistence on the autonomy of research, on the authority of scientists to decide for

themselves what questions are important to pursue, fails to take into account the differences science-shaped policies are expected to make, not just to a handful of people, but to the collective good. That insistence is a throwback to an era in which research into nature was a diversion (or a sacred duty) for a privileged few, to an age in which the "gentlemen" could explore what they pleased and govern themselves as they thought fit.

IV

Philosophy of science has often regarded itself as explaining scientific practices without subverting them: the philosophers have frequently sought to interpret Science, but not to change it. By contrast, my proposed ideal of well-ordered science is intentionally revisionary. As I have just remarked, common attitudes among researchers stem from the persistence of old ideas, customs and conceptions of the investigator's role, long after the institution has evolved into a radically different social environment. By placing Science within society, I intend to scrutinize rules, conventions, habits and self-images that are taken for granted. It should not come as any surprise then that I suppose the current state of Science to be at odds with the ideal.

So far, I have developed well-ordered science solely within the context of agenda-setting. That was an appropriate beginning, because, as argued, well-ordered science furnishes the pragmatic concepts required for articulating the epistemic goals (significance, true enough, reliable enough). As we shall shortly discover, however, some of the failures to achieve well-ordered science suggest extending the ideal to other contexts (certification, transmission of information). I'll begin, however, with a striking example of failure to pursue inquiries well-ordered science marks out as significant.

Consider research into human diseases.[23] That research is primarily carried out in a small number of affluent countries. It should come as no surprise that the principal foci of research are conditions that afflict the citizens of those countries, and that diseases by which they are not likely to be affected are neglected. Many of the world's poor fall victim to a variety of infectious diseases, suffering disability and premature death. The fatalities are often disproportionately skewed towards children.

Well-ordered science envisages an ideal deliberation, and, in the case at hand, it is reasonable to expect that the outcome would be to apportion resources to diseases, insofar as they were judged equally tractable, according to the burdens they produced. Call this the *Fair Share Principle*. Disease burden might itself be measured in terms of lives lost, or, more inclusively, through factoring in the hardships that result from disabling conditions. However you decide to do the measurements, similar results are forthcoming. Actual disease budgets for nationally-funded research are radically at odds with the Fair Share Principle: a look at the figures reveals that many of the conditions that affect the world's poor are underfunded by a factor of 100 or more. Although quite minor complaints from which wealthy people suffer sometimes receive serious attention, many conditions that cripple and truncate the lives of poor children go almost unstudied.

---

[23] For more detailed discussion of the issues outlined here, see (Flory and Kitcher 2004, Reiss and Kitcher 2009).

Perhaps this is the most dramatic divergence between the judgments of significance that would flow from well-ordered science and those that emerge from negotiations between national policymakers, intent on shaping research to fit the demands of powerful constituencies, and scientists who are—understandably—bent on pursuing the questions that fascinate them. There are surely others of a similar character: failures to extend parts of biology relevant to agriculture in ways that would best help people constantly vulnerable to drought and crop failure, failures to develop areas of materials sciences to promote more stable forms of shelter in places where there are special environmental challenges, failures to focus research in the earth sciences and in oceanography on problematic phenomena in the poorer parts of the globe. It is important to expose places in which this type of mismatch between scientific potential and human needs is occurring, and even philosophers might consider practical means of correcting such situations.

Well-ordered science is an ideal, but the value of ideals lies in calling for plans to realize them better. Plainly it would be absurd to suppose that better judgments of significance should be generated by assembling representatives of all the world's many points of view, tutoring them, motivating them to engage with one another, and then drawing up an ideal plan for research. Yet, without envisaging any such ludicrous venture, it's possible to specify concrete steps. First, a systematic drawing up of the state of scientific inquiry as we have it, one that shows clearly what has been achieved, and what lines of research currently have exciting prospects. Second, an equally encompassing review of human needs on which current research might bear, one that did not simply leave out the lives of the many whose countries are too poor to invest heavily in inquiry. Third, a representative group of people, fully tutored with respect both to needs and prospects, who could attempt to balance the available resources among the world's diverse constituencies. Any practical venture along these lines could not be expected to deliver the judgments that would ideally result from well-ordered science, but it would surely correct for current biases by responding to those whose voices are not currently heard. I strongly suspect it would also provide a more balanced approach to those research areas in which a chorus of enthusiastic voices starts a bandwagon.

Well-ordered science is also a valuable ideal for exposing, and trying to correct, some further shortcomings of contemporary scientific practice. These attend the contexts of certification and transmission. Recall the basic picture of how Science advances the collective good: scientific communities need to pursue questions that bear on human prospects, but they must also certify conclusions and transmit information. Matters can go awry in these latter endeavors, and they may do so in several distinct ways. One possibility is for "findings" to be certified too easily, for appropriate standards of reliability not to be honored, and for the information then transmitted in ways that are humanly and socially damaging. A converse trouble occurs when collective discovery is too cautious, when the adoption of valuable results is postponed until "more evidence" comes in, and, in consequence, people's projects suffer and their aspirations are defeated. Yet a further difficulty arises when certification is properly carried out—that is the standards for collective inquiry are appropriately formulated and

appropriately applied—but there is significant public doubt either about the standards or concerning their application: under these circumstances, valuable, even crucial, information may fail to flow where it is urgently required. Finally, there can simply be parts of scientific discovery that would be immensely useful to particular groups of people, but which they do not know how to find or which they cannot comprehend when they have access to them: the success of Science depends on clear channels for *transmitting* information.

It is not hard to point to examples of all of these deficiencies. The history of attempts to introduce safeguards against health hazards and environmental challenges is, in part, the record of cases in which those who crafted and applied research standards failed to think systematically enough about the potential consequences of allowing the certified results to frame policy. Of course, there are also instances in which nobody could have known, in which the investigators did the best they could and were simply unlucky. On the other hand, some programs for the release of substances to solve an environmental problem have failed to appreciate elementary facets of evolution and thus have not foreseen easily-anticipated difficulties. A less well-known example of overly lenient standards of certification occurs in the reiteration of claims about cognitive, temperamental, and behavioral differences allegedly derived from evolutionary considerations, during the past few decades. Originally in human sociobiology, and more recently in the most visible types of evolutionary psychology, investigators have announced "findings" about all sorts of human traits and their distributions, on grounds that would never be taken to suffice if the subject were some experimentally tractable organism (ants, or fish, say). Although the conclusions trumpeted in the press often create serious disadvantages for some groups of people—often for groups of people who have suffered discrimination in the past—those possibilities seem to play little role in focusing attention on the need to acquire serious evidence. It is quite possible that the public thirst for sensational news in this area gives rise to media pressure to provide it, thus stimulating the internal credit economy of studies of animal behavior and lowering the standards of certification.[24]

Insistence on a high standard of evidence, with consequent delay in certifying conclusions that would be expected to deliver important relief, probably occurs at a lower rate: a famous example was the hesitation in validating medicines for the treatment of HIV (in which researchers who were committed to continuing a sequence of trials with placebo controls eventually gave way to the pleas of AIDS activists). A far more common cause of blocking the flow of important information to the people who need it comes in examples where the outcome of research has already been fully certified, in accordance with appropriate standards, although suspicion about the certification procedures lingers and prevents the conclusions from being widely believed. These situations are poignant, in that there is knowledge "on the books" that is pertinent to people's life themes, even though those people remain in ignorance because they "do not know what is known".

---

[24] For critiques of work of these types, see (Kitcher 1985, Vickers and Kitcher 2005, Haufe forthcoming).

One iconic example, much discussed by scientists and philosophers who have wished to remedy the situation is the acceptance of Darwinian evolutionary theory. In the countries that harbor suspicion about "evilyoushun", skepticism is usually sparked by a sense that a supposed piece of science is at odds with religious claims of overriding importance. Skeptics are thus stimulated to place very high demands on evidence, far beyond those that would be introduced in any other area of research: a typical requirement would be that evolutionary theory must provide a detailed explanation for some particular set of transitions in the history of life (for example, the origin of the cell).[25] Yet, for all the energy that has been devoted to this example, it is far less important, or poignant, than other instances of the same phenomenon. Resistance to conclusions about anthropogenic global warming, engendered by just the same mistrust of the standards that have been employed in certifying the conclusions of climate scientists, strikes centrally at most human projects.

Many people's life themes involve doing things that will prepare or sustain parts of the world for others who will come after them. Caring for the future your children or grandchildren will have is a common central value—and even those who have no direct descendants of their own are likely to share the concern for the conditions under which future people will live. Thousands of climate scientists, from many different parts of the world, and with a wide range of professional and social perspectives, have come to consensus on the minimal claim that, even if steps are now taken to halt the emission of greenhouse gases, the mean temperature of our planet will increase by 2-3°C by the end of the century, with an increase of between 3 and 7 degrees if nothing is done. Despite the overwhelming agreement, and despite the consequences of the predicted effects for almost all people's life themes, many countries continue to contain a large number of skeptics—climate deniers—and many of those who pay lip service to the consensus fail to be moved to action by their alleged belief. Attitudes to the findings are shaped by distrust: there are doubts that appropriate standards of certification have been deployed, often expressed as the suggestion that "liberal ideology" has entered and runs ahead of the objective evidence.

In this important case, the trouble might be traced either to a lack of transparency in the procedures of certification or to deficiencies in the media charged with transmitting valuable information from the scientific community to those parts of the public where it is needed. Arguably both facets of the current condition need correction (and, with respect to the diagnosis of the transmission failure, you might already see the need to consider Science in relation to other institutions, a perspective I shall suggest shortly). Similarly, with respect to another major example, research into the safety of genetically modified organisms (GMOs), there's a mismatch between the certified results that win broad acceptance within a scientific community and the beliefs of members of ambient societies, a mismatch that might be traced either to worries about procedures of certification or to biased transmission through the institution of the media (an institution that deserves critique in virtue of that fact). In the latter ex-

---

[25] As noted in Part 2, allowing demands like this into the community practice of certification would bring research to a halt. Even when the skeptics are outside the community, their participation in public debate can easily affect the transmission of knowledge.

ample, yet another part of the framework of societal institutions comes into the picture, for the entanglement of scientific research with economic life also plays a role. Beyond that, however, with respect to the debates over climate change and about the safety of GMOs, an internal pathology of collective inquiry expresses itself. The pressures of the credit economy can drive those who fail to convince their peers to seek credit in other ways, by "going public" and portraying themselves as brave dissenters, whose responsible but unpopular views should not be drowned out by the agreement among their fellow investigators. Their endeavors may muddy the waters, so that an important message can no longer be discerned.

Let me sum up my cursory review of these examples in some conclusions I draw from my framing of them. First, there are lapses from well-ordered science that occur not only in the context of agenda-setting (in the identification of significance) but also in certification and transmission. Second, *some* of the difficulties to which I have pointed might be overcome by extending the approach I recommended earlier, namely forming small groups of representative people, who might be taken behind the scenes, tutored in the ways in which new findings have been grounded, and given a mandate to report to the broader public—effectively, these groups would be conduits between Science and the ambient societies, charged with the task of restoring trust. Third, analyses of the credit economy of a scientific community might reveal further pressures that should be resisted, for example the pressure to flourish disagreements in a public forum; those analyses would, in turn, call for the exploration of norms for the conduct of investigators (and, for example, a new self-image as a member of a complex team that has the function of coordinating activities to the public good); thus they would give rise to a generalized version of the explorations of Part 2, in which the collective good, as I have now understood it, was viewed as the ultimate goal. Fourth, some of the difficulties with respect to disputes over proper certification can be approached in the ways recommended at the end of Part 1 (philosophically informed analyses of the evidential basis of anthropogenic global warming or philosophical critiques of the evidential weakness of evolutionary psychology, say); I am convinced, however, that full resolution of these important lapses from well-ordered science must at some points engage with the collective character of inquiry and with the character of the social connections between the scientific community and the human population whose well-being it is supposed to promote.

V

Earlier, I remarked that a satisfactory account of the effects Science exerts on human well-being must attend to the interactions among a framework of social institutions, but, for almost all the previous discussion, I considered Science in relation to society by abstracting from the details of other entrenched parts of the social matrix. You can take some steps towards analyzing collective inquiry by supposing that truth is typically a good thing to have and that significance is recognized by scientists (as in Part 2), and you can make some progress in broadening the focus by linking significance to the ideal of well-ordered science and assuming that reliably certified information will eventually find its way to those who need it. The largest task of socially-oriented phi-

losophy of science is, however, to consider Science as one institution among many, and to seek a full package of norms for the complex that will enable each of them to discharge the tasks required of it for the promotion of the collective human good.

Let us look very briefly at four further societal institutions, each of which plays a role in the impact Science has on overall well-being. The first of these is the institution through which information is widely reported, the media. With respect to debates about climate change, newspapers and magazines, radio and television, internet sites and blogs have all played a significant part in the current distribution of public opinion. In some places, a commitment on the part of the media to so-called "balanced treatment" has provided opportunities for minority points of view, often voiced by spokesmen whose professional credentials were not in climate science, to portray themselves as equally legitimate to the thousands of professionals who share a consensus position. The problem of this kind of reporting is evident, and it is tempting to overreact to it by campaigning for a norm according to which journalists must report the position of any large scientific majority. A moment's reflection, however, should remind us that to insist on a norm like that would deprive the media of potentially valuable critical functions, as well as possibly provoking widespread distrust of "the organs of orthodoxy". Furthermore, in those parts of the world where investment in media ventures is open to interested entrepreneurs, and in which the operation of a particular television station can be seen as legitimate free expression, it would be hard to resist market pressures that promise large profits for those who tell a sizeable subpopulation just what its members want to hear.

Second, there is an urgent need for efforts to prepare citizens better, so that they can more easily and fluently receive new scientific information that bears on their life themes. Scientific education, throughout the world, is largely concerned to treat the young as though they were starting a full career as researchers. There is some reason for that, since the autonomous choice of a life theme is fostered by acquainting students with potential careers, with respect to which they can judge their own potential. No adequate system of scientific education should block off the path to participation in inquiry. Yet once it's clear to those being educated that this isn't the route they intend to follow, it's appropriate to appreciate that there are other valuable goals of a scientific education, and that these include not only a particular form of literacy—the ability to read, and thus keep up with important discoveries—but also a curiosity about nature and its workings. These goals are ill-served by programs demanding repeated exercises in rote memorization and continued wrestling with the problems of application on which beginning researchers cut their teeth. From the perspective of the educational institution, one goal is to produce citizens who have a commitment to keeping up with the picture of the world articulated by the various research communities and who possess a concomitant skill to grasp the concepts involved in formulating new and exciting results. Those citizens will be aided, of course, if members of scientific communities are willing to take time writing for a broader public in ways that are entertaining and accessible—and it's a welcome trend of recent decades that researchers who are willing to develop such skills and to put them to work are no longer seen as doing something thoroughly second-rate.

Third, the economic institutions that sometimes distort the role of the media are themselves coming to form a closer relation with scientific research. Increasingly, large corporations take an interest in particular fields of science, sometimes through setting up their own laboratories, sometimes through developing relationships with university scientists (often with the encouragement of the universities, who are excited by the prospect of financial support). It's easy to make some conjectures about the effects of this trend. First, research undertaken for the purposes of profit is highly unlikely to validate questions whose potential for practical products is uncertain: to the extent that "basic" research will thrive under such regimes, it will do so because it is expected to be the key to later applications. Second, the practical difficulties encountered by those who do not have the means to pay, are likely to be further neglected, so that the departures from well-ordered science along this dimension will be exacerbated—diet medicines and skin creams will be more attractive than methods of stopping the world's most deadly infections. Third, the privatization of research requires obstacles to the free flow of scientific information, and it is more than probable that studies of significant questions will be handicapped because investigators cannot combine the results that would yield solution. Finally, the "safety studies" issuing from laboratories with clear interests in settling questions in a particular direction will undermine confidence in standards of scientific certification, perhaps across the board, thereby contributing further to the distrust that has prevented the public from coming to know what the experts have already established.

Fourth, since the sixteenth century the institution of inquiry into the natural world has been at odds with another institution, whose (vague) function was to offer understanding, consolation, and comfort. The clash between Science and Religion is best seen as a conflict between two important social institutions. Both have roots in our deep human past, in our ways of trying to fathom the world around us and in our efforts to make peace with ourselves (individually and collectively). The pious men who developed the first scientific societies of the seventeenth centuries believed in the harmony of these institutions, but, for a wide variety of reasons, that belief turned out to be wrong. Today it should be completely clear that large aspects of the institution for Spiritual Well-being (usually some form of religion) are often called into question by the march of inquiry.[26] Effectively, a spectrum of investigations has challenged the claims about supernatural beings offered by all the world's religions. One reaction, an unhealthy one, manifested in the iconic example of resisting evolution, is to preserve the Spiritual institution intact, at the cost of incapacitating Science for fulfilling its proper role. Equally unhelpful is the thought that it is simply possible to sweep away the Spiritual institution, and leave societies to cope without any substitute (Dawkins 2006). The right answer, I believe, is to develop an analysis of both institutions that will enable a myth-free refashioning of the Spiritual institution through a positive version of secular humanism.

---

[26] Here, as I've suggested elsewhere, the trouble doesn't come so much from the natural sciences, cosmology and evolutionary biology, but from the rigorous studies of the origins and growth of religious belief—see ch. 5 of (Kitcher 2007).

In adducing four institutions that interact with Science, I've tried to show, very sketchily, how the image of Science as an information-producing system directed at promoting the quality of human lives poses general problems about Science in relation to other social institutions. I have simply and bluntly gestured towards the ways in which the problems arise, and the kinds of steps that might be taken to address them. This survey of the ground is intended to be in the spirit of Hempel's revision of the positivist agenda. Here, I suggest, are some important problems that ought to concern us, that is to preoccupy the general philosophy of science. To make serious progress with them, you'd have to do what the best philosophy of science has so often done, to wit to explore empirical details. This time, however, the bodies of material we need are not only, perhaps not even primarily, bits and pieces from various natural sciences (physics or biology, say), but a more accurate and detailed understanding of the social processes that go on, both within the sciences and between scientific communities and the ambient human societies. Philosophy of science so envisaged might learn much from historians and sociologists of science, from economists, political theorists, sociologists and anthropologists too—perhaps even so much from so many different people that philosophy of science, done in this way, would no longer be a subject for individuals but for teams—teams on which the philosopher might be cast as "normative analyst and coordinator". That development is not necessarily to be despised or rejected.

VI

I have wandered a long way from my beginning, from the clear and orderly world of logical positivism, and the equally clear, if somewhat smaller, world of logical empiricism. In the end, I have surely found myself in wilder territory, in which issues of large scope prove hard to define. I have seen it as a continuous journey, one in which later steps were natural extensions of those I had taken earlier. Perhaps I am wrong to think so—maybe the ending point is so strange and uncomfortable that it would be better to remain at the beginning?

Why are these steps towards a *pragmatist* philosophy of science? In part, because there are obvious pragmatist themes, in the emphasis on local studies, on introducing considerations of value, in treating Science (and other institutions) in relation to their goals. Beyond that, however, is a more general perspective. I began by pointing to a revolution that never happened, the pragmatist desire to reconstruct philosophy so that it reconnected with life. Pragmatism asks us to concentrate on the problems that matter, and a first philosophical task is to understand what that slogan might amount to: mattering is significance, and significance is generated from an account of collective human well-being. Pragmatic philosophy explores those places at which change in view or commitment or in understanding could promote well-being, and part of its mission lies in exposing the ways in which the customs and institutions of the past confine people's lives. Social diagnostics becomes a centerpiece of pragmatist philosophy generally. A pragmatist philosophy of science, then, is just what you get when you take that perspective on the achievements and processes of the sciences and on the institution of Science.

REFERENCES

Achinstein, Peter. 2001. *The Book of Evidence*. New York: Oxford University Press.

Block, Ned. 1996. How Heritability Misleads about Race. *Boston Review*, January 6, 30-35.

—. and Gerald Dworkin. 1976. *The I.Q. Controversy*. New York: Pantheon.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Dawkins, Richard. 2006. *The God Delusion*. Boston: Houghton Mifflin.

Dewey, John. 1916. *Democracy and Education* (reprinted as volume 9 of *John Dewey: The Middle Works*. Carbondale IL: University of Southern Illinois).

—. 1920. *Reconstruction in Philosophy* (reprinted as volume 12 of *John Dewey: The Middle Works*. Carbondale IL: University of Southern Illinois).

—. 1927. *The Public and its Problems* (reprinted as volume 2 of *John Dewey: The Later Works*. Carbondale IL: University of Southern Illinois).

—. 1929. *The Quest for Certainty* (reprinted as volume 4 of *John Dewey: The Middle Works*. Carbondale IL: University of Southern Illinois).

Dupré, John, ed. 1987. *The Latest on the Best*. Cambridge MA: MIT Press.

—. 1993. *The Disorder of Things*. Cambridge MA: Harvard University Press.

Feyerabend, Paul K. 1975. *Against Method*. London: Verso.

—. 1978. *Science in a Free Society*. London: New Left Books.

Fine, Arthur. 1988. Interpreting Science. In *PSA* 1988, ed. A. Fine and J. Leplin, East Lansing: PSA.

Flory, James and Philip Kitcher. 2004. "Global Health and the Scientific Research Agenda." *Philosophy and Public Affairs* 32: 36-65.

Frapolli, Maria José. 2013. *The Nature of Truth*. Dordrecht: Springer.

Friedman, Michael. 1974. Explanation and Scientific Understanding. *Journal of Philosophy* 71: 5-19.

Ginzburg, Carlo. 1980. *The Cheese and the Worms*. Baltimore: Johns Hopkins.

Glymour, Clark. 1998. What Went Wrong? Reflections on Science by Observation and the Bell Curve. *Philosophy of Science* 65: 1-32.

Goodman, Nelson. 1959. Recent Developments in the Theory of Simplicity. *Philosophy and Phenomenological Research* 19: 429-46.

—. 1961. Safety, Strength, Simplicity. *Philosophy of Science* 28: 150-151.

—. 1963. Science and Simplicity. In *Philosophy of Science Today*, ed. S.Morgenbesser. New York: Basic Books (1967).

Gould, Stephen Jay and Richard Lewontin. 1979. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B* 205: 581-98.

Haufe, Christopher. Forthcoming. *The Sound of Science* (ms.)

Hempel, C.G. 1963. Implications of Carnap's work for the philosophy of science. In *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp. Library of Living Philosophers. Vol. XI. La Salle IL:Open Court.

—. 1965. *Aspects of Scientific Explanation*. New York: Free Press.

Hull, David. 1988. *Science as a Process*. Chicago: University of Chicago Press.

James, William. 1896. The Will to Believe. In *The Will to Believe and other essays*, 1-31. New York: Dover, 1956.

—. 1907. *Pragmatism*. Reprint New York: Dover, 1995.

Hughes, Robert. 1986. *The Fatal Shore*. New York: Vintage.

Keegan, John. 1976. *The Face of Battle*. Harmondsworth: Penguin.

Kitcher, Philip. 1976. Explanation, Conjunction, and Unification. *Journal of Philosophy* 73: 207-212.

—. 1981. Explanatory Unification. *Philosophy of Science* 48: 507-31.

—. 1985. *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge MA: MIT Press.

—. 1986. Projecting the Order of Nature. In *Kant's Philosophy of Physical Science*, ed. R.Butts, 201-235. Dordrecht: Reidel.

—. 1989. Explanatory Unification and the Causal Structure of the World. In *Scientific Explanation*, ed. P.Kitcher and W.Salmon, 410-505. Minneapolis: University of Minnesota Press.

—. 1990. The Division of Cognitive Labor. *Journal of Philosophy* 87: 5-22.

—. 1993. *The Advancement of Science*. New York: Oxford University Press.

—. 2001. *Science, Truth, and Democracy*. New York: Oxford University Press.

—. 2007. *Living with Darwin*. New York: Oxford University Press.

—. 2011a. On the very Idea of a Theory of Evidence. In *Philosophy of Science Matters: The Philosophy of Peter Achinstein*, ed. G.Morgan. New York: Oxford University Press.

—. 2011b. *Science in a Democratic Society*. Amherst NY: Prometheus Books.

—. 2011c. *The Ethical Project*. Cambridge MA: Harvard University Press.

—. 2012. *Preludes to Pragmatism*. New York: Oxford University Press.

—. and Wesley Salmon. 1987. Van Fraassen on Explanation. *Journal of Philosophy* 84: 315-330.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

—. 1963. The Function of Dogma in Scientific Research. In *Scientific Change*, ed. A. C. Crombie, 347-69. New York: Basic Books.

Lakatos, Imre. 1970. Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge,* ed. I. Lakatos and A. Musgrave. Cambridge UK: Cambridge Univ. Press.

Le Roy Ladurie, Emanuel. 1990. *Montaillou*. Harmondsworth: Penguin.

Peirce, C.S. 1877. The Fixation of Belief. In *The Essential Peirce,* ed. N. Houser and C. Kloesel, 109-23. Volume 1. Bloomington IN: University of Indiana Press (1992).

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Quine, W.V. 1951. Two Dogmas of Empiricism. In *From a Logical Point of View*, W.V. Quine. Cambridge MA: Harvard University Press (1964).

Reiss, Julian and Philip Kitcher. 2009. Biomedical research, Neglected Diseases, and Well-Ordered Science. *Theoria* 66: 263-82.

Salmon, Merrilee. 1990. On the Possibility of Lawful Explanation in Archeology. *Critica* 22: 87-114.

Salmon, Wesley. 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

—. 1990a. The Appraisal of Theories: Kuhn meets Bayes. In *PSA 1990*, 325-32, East Lansing: Philosophy of Science Association.

—. 1990b. Scientific Explanation: Causation and Unification. *Critica* 22: 3-23.

Spirtes, Peter, Clark Glymour and Richard Scheines. 2000. *Causation, Prediction, and Search*. Cambridge MA: MIT Press.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Vickers, A. Leah and Philip Kitcher. 2005. Pop Sociobiology Reborn: The Evolutionary Psychology of Sex and Violence. In *In Mendel's Mirror*, Philip Kitcher. New York: Oxford University Press.

Winnie, John. 1970. Special Relativity without One-Way Velocity Assumptions. *Philosophy of Science* 37: 81-99.

Wylie, Alison. 2002. *Thinking from Things*. Berkeley: University of California Press.

PHILIP KITCHER is the John Dewey Professor of Philosophy at Columbia University. He is the author of many books and articles, among them *The Advancement of Science* (1993), *Science, Truth, and Democracy* (2001), *Science in a Democratic Society* (2011), and *Preludes to Pragmatism* (2012).

ADDRESS: 717 Philosophy Hall, 1150 Amsterdam Avenue, New York, New York 10027, USA. E-mail: psk16@columbia.edu