Foregrounding the Background

Helen Longino*†

Practice-centric and theory-centric approaches in philosophy of science are described and contrasted. The contrast is developed through an examination of their different treatments of the underdetermination problem. The practice-centric approach is illustrated by a summary of comparative research on approaches in the biology of behavior. The practice-centric approach is defended against charges that it encourages skepticism regarding the sciences.

1. Introduction

1.1. Theory Centrism and Practice Centrism. Let me begin my remarks with some situating moves. First, I think of philosophy of science as an interpretive and critical engagement with the sciences. The philosopher's interpretive tools, of course, are logical and conceptual, not literary, tools (although, it must be said, attention to metaphors in science can be very revealing). Within this engagement, it is possible to distinguish at least two major forms. Some of us focus on the content of the sciences, in what we might call a "theory-centric" approach; others focus on practice in and of the sciences, in what we can call a "practice-centric" approach. Issues of content and practice are different; aspects of the philosophical approaches are different, but they are entangled. Arguments in one approach do have a bearing on positions and perspectives in the other—hence the label "-centric," to convey that the interests of these approaches ripple out from different focal points, but overlap, creating lovely, and sometimes unlovely, interference patterns.

*To contact the author, please write to: Department of Philosophy, Stanford University, 450 Serra Mall, Stanford, CA 94305; e-mail: hlongino@stanford.edu.

†A heartfelt thank-you to Jessica Pfeifer for her energetic and effective activities as executive director of the PSA, to Chris Smeenk for organizing an excellent program, and to the members of the PSA Governing Board for their commitment to our profession and our organization. I am also grateful to the Brown Foundation for a residency at the Dora Maar House, during which I prepared a draft of this address.

Philosophy of Science, 83 (December 2016) pp. 647–661. 0031-8248/2016/8305-0001$10.00

What questions interest a philosopher coming from a content- or theory-centric approach? Here are a few:

- What claims is a theory or hypothesis making about natural or social worlds?
- How do these relate to claims made/supported in other theories?
- How do they relate to broader metaphysical questions?
- And, what is there really?
  - Particles in motion?
  - The midsize objects of everyday experience?
  - The observable world? The directly measurable world? The entities and processes postulated in successful scientific theories?
  - And, while we're at it, should we think that reality consists of entities, of processes, of structures?

These are questions prompted by considering the content of scientific theories, of quantum and relativity theory, and of evolutionary theory and genetics. Practice (practice-centered) questions, on the other hand, start this way:

- What are the questions of a specific scientific subdiscipline or approach within a subdiscipline?
- What methods and strategies are employed to answer those questions?

Before we go further, let me address a question that immediately comes to mind. Isn't asking about the actual methods and strategies scientists use in the laboratory or in the field sociology rather than philosophy?

I think not. There is a delicate relationship between social science scholars who study scientific practice and philosophers who do so. As philosophers, we have different questions than the sociologists have. Our questions are in one way or another related to the supposition that the sciences create knowledge. We begin our investigations with some more or less elaborate sense of what counts as a cognitive practice, what counts as a good reason, and with the expectation of evidence. We then use what we discover in the bit of science under investigation both to analyze and interpret what is going on in that bit, making it available for criticism, and to refine, or even challenge, the philosophical notions. Sociologists and anthropologists of science explicitly eschew any supposition about knowledge or truth and represent themselves as describing what they find upon observation and incorporating that into preferred theories of social structure or interaction.

Social structure and interaction—in science, the organization of a laboratory or research community, the institutions within and for which scientific inquiry is conducted, and the forms of communication within and among communities—constitute scaffolding for the cognitive activities of data assembly, modeling, model assessment, and theory and hypothesis assessment that interest philosophers epistemologically. This social scaffolding and interaction constitute support for and resources for the cognitive, not deviations from it. As such, they are epistemologically relevant and, for some at least, arguably constitutive of the cognitive.

We can expand the practice-centered questions:

- What are the questions of a specific scientific subdiscipline or approach within a subdiscipline?
- What methods and strategies are employed to answer those questions?
- What do those methods and strategies produce or count as evidence?
- Are they different from the methods and strategies employed in neighboring or competing approaches?
- How is what counts as evidence brought to bear on the substantive claims being made (or "suggested")?
- What are the standards of acceptability or acceptance employed in a given approach or subdiscipline?
- How do these practices relate to epistemological concepts such as rationality, objectivity, and evidence?

1.2. Theory Centrism and the Underdetermination Problem. As an example of the difference between taking a theory-centric and practice-centric approach, consider the underdetermination problem. The two philosophical approaches will treat the issue differently, and in ways that are differently relevant to larger metascientific or philosophical questions.

From a theory-centric perspective the underdetermination problem amounts to the empirical equivalence of two or more theories making incompatible claims about the same domain. The Schrödinger and the Heisenberg interpretations of quantum theory are frequently taken as examples of such empirical equivalence. Taken literally, they make different claims about the quantum world, but they have the same empirical consequences.

If we go back to the first articulation of the underdetermination thesis (as a problem distinct from the problem of induction), we find the following statement by Pierre Duhem. Suppose a physicist decides to conduct an experimental test of a hypothesis, he says. "In order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced, [the physicist] . . . does not confine himself to making use of the proposition in question; he makes use also of a whole group of theories accepted by him as beyond dispute" (Duhem 1954, 185). The result is that when the predicted phenomenon is not produced, the experimenter does not know which of the set, proposition plus the "whole group of theories," is at fault. Furthermore, if the phenomenon is produced, the experimenter cannot conclude unequivocal or exclusive confirmation of the hypothesis, as an alternate group of theories, together with some other hypothesis, might support the same prediction. In the decades following Duhem's argument, underdetermination can be seen to have received two different treatments.

Quine (1951) emphasized the holist character of the Duhemian argument. He expanded its reach to the entirety of science. For Quine, a "recalcitrant experience," that is, the experimental contradiction of prediction, does not confront a single hypothesis, or even a single theory, but the whole "web of belief." The web is so underdetermined by its boundary conditions that there is no uniquely correct or optimal adjustment. What became known as the "Duhem–Quine" thesis was understood as a thesis about theories, namely, the possibility of empirically equivalent theories (theories with exactly the same empirical consequences). So understood, it became entangled with the proposal associated with Reichenbach that, given a theory, it's always possible to construct an alternative to that theory that has exactly the same empirical consequences, for example, by the postulation of a universal or absolute force.

The holist interpretation of underdetermination has been subject to two main criticisms. One is that in cases of genuine empirical equivalence of two theories, the two are just notational variants of one another, and so not rival representations of the state of the world underlying the shared observational data. The second is that the claim of empirical equivalence is relative to the state of science at any given time and may be subject to revaluation as science progresses. Hence, there can be no argument from the present empirical equivalence of a set of theories to their persistent or permanent empirical equivalence.

These argumentative strategies for defanging the underdetermination argument were recently undermined by Kyle Stanford (2006). Stanford has argued for a new version of underdetermination, what he calls "the problem of unconceived alternatives." He shows that there were credible and empirically supported alternatives to features of Darwin's evolutionary theory throughout the nineteenth century that were not taken up in the debates about evolution that ran through that period. If there were (realistic and available) alternatives then, he asks, how do we know that the same isn't true now for us? Like the Quinean version of underdetermination, this promotes general skepticism about the deliverances of scientific research. Interestingly, Stanford's argument is parallel to that of the second kind of response mentioned above but comes to a contrary conclusion. It is still, however ingenious, theory centered and about empirically equivalent (or equivalent enough) alternative theories.

1.3. Practice Centrism and the Underdetermination Problem. Suppose we start, instead, not from questions about theories and general theses about scientific theories, from concerns about holism, scientific realism versus scientific antirealism or the cognitive authority of science, and so on, but from questions arising from the examination of particular episodes of scientific investigation, from examination of practice. Underdetermination then takes on a different aspect.

Underdetermination now occurs in the context of investigating particular research programs. And in looking at particular research programs, the question is not whether auxiliary assumptions may in the future be augmented or discarded, but this: Are auxiliaries at work here, in this investigative situation? And what are they? These are questions about practice. How is inquiry proceeding in this research project? In this family of projects (approach)? The practice-centric approach focuses on the evidence part of the underdetermination of theory by evidence, rather than on the theory component of the relation. Directed to practices, the underdetermination problem constitutes an opening for deeper investigation of the constitution and organization of data so that they may enter into an evidential relationship with some hypothesis.

Questions about auxiliaries foreground the background and open to more questions. The background includes auxiliaries, but we want to know the following:

- What are they?
- What grounds support them?
- Are there alternatives in active (or available for active) play?
- How are the alternatives attended to?

The auxiliaries or background assumptions provide intellectual scaffolding, and questions about their status prompt questions about the social scaffolding. Hence, we may also find ourselves asking about

- the social practices that play a role in sustaining a given set of auxiliaries;
- intellectual social relations: other approaches/research projects bearing on the phenomena under investigation (may not need to reach for the unconceived);
- the social relations among scientists; and
- the social institutions that make certain scientific questions salient at a particular time (and others not).

2. An Example of Practice-Centric Analysis. But, except for the second, these are questions, largely, for social scientists. I return to this point below. For now, I turn to an extended example of the kind of investigation of the construction and organization of evidence that I am proposing. For this I draw from my recent analysis of sciences of human behavior (Longino 2013).

2.1. A Multiplicity of Causal Factors. This study was a comparative analysis of (biological or biologically related) approaches to studying human behavior. These were all proximal, rather than distal (i.e., evolutionary), approaches and encompassed

- quantitative behavior genetics;
- molecular behavior genetics;
- neurophysiology and neuroanatomy of behavior;
- social environment–oriented developmental psychology; and
- a variety of approaches attempting integration of two or more of the above.

I also included in the study population-level (ecological) approaches, although these contrasted with those just listed with respect to the object of explanation and do not figure in the discussion until later. The listed approaches are focused on individuals and variation among individuals. The comparative approach involved reviewing many empirical studies, identifying for each type, the questions being asked, the size and character of study and control populations, the methods used, the operationalizations employed, and the hypotheses tested or considered. This kind of dissection facilitated epistemological, ontological, and social analysis of the research.

Efforts to understand human behavior scientifically arise because we are curious or concerned about the variation in humans' expression of any number of behaviors or behavior patterns. Thus, we find studies of novelty seeking, risk taking, nurturant behavior, addictive behavior, divorce, aggressive behavior, sexual orientation, and many more. Any given study focuses on one behavior type, operationalized appropriately for measurement. For now let us call it *T*. Then the phenomenon to be explored is the variation in the expression of trait *T* in a given population *P*.

Applying the comparative methodology reveals an array of potential causal factors. Here they are arrayed in what I call the potential causal field. This is the set of causal factors whose influence on *T* is measured in the various approaches. It is not fixed, but can change as additional factors are sought.

Genotype 1	Genotype 2	Intrauterine environment	Physiology	Nonshared environment	Shared (intrafamily) environment	Socioeconomic status
[allele pairs]	[whole genome]		[hormone secretory patterns; neurotransmitter metabolism]	[birth order; differential parental attention; peers]	[parental attitudes re discipline; communication styles; abusive/nonabusive]	[parental income; level of education; race/ethnicity]
			Anatomy [brain structure]			

The task is to associate measured variation in factors in the potential causal field with measured variation in *T*. This presupposes measurability and meth-

https://doi.org/10.1086/687990 Published online by Cambridge University Press

odologies for conducting measurement. Measurability requires access to the phenomena or to surrogates. Application of the methodologies and their associated measurement strategies to the task de facto parses the potential causal field into different fields.

Take, for example, molecular genetics. Molecular genetics has developed several strategies to identify genetic variation and associate that with phenotypic variation. Linkage analysis uses polymorphic alleles, called markers, to associate regions of the genome with *T*, the trait under investigation. These markers are not themselves the causal factors, but concordance of a particular one of the forms the allele takes with the trait is taken as evidence that a gene in the vicinity of the marker is a factor involved in *T*. Genome-wide association studies, made possible only recently, associate genome-wide variation with *T*. That is, the entire genome is canvassed, and elevated frequency of genic forms that correlate with elevated frequency of expression of the trait is also taken as evidence for a causal influence of those genes on the trait. What is canvassed by molecular genetic methods is variation in the genome, not variation in any of the other possible factors. But, of course, the genic variation does not completely match the trait variation; allowance is made by including a category of "error" or "other" to represent those unidentified causal factors that may account for the variation unaccounted for by the factors under study. So the causal field is reparsed:

Genotype 1 [allele pairs]	Genotype 2 [whole genome]	Error/other

Quantitative behavior genetics utilizes the methods of classical genetics, genetics before molecular techniques made access to structural features of DNA possible. Its strategies of identifying genetic contribution to a trait or its variation involve identifying biological relatedness and comparing the expression of *T* in biologically related or less biologically related individuals. In humans this means using twin or adoption studies, on the assumption that one can hold either biological relatedness or environment constant. The environmental contribution is measured as a degree of similarity not accounted for by biological relatedness. Thus, both the genetic contribution and the environmental contribution are measured indirectly. Here, again, the potential causal field is reparsed:

Genotype 2 [whole genome]	Nonshared environment [measured by nonheritable trait variation unaccounted for by environmental variation]	Shared environment [includes socioeconomic status, parental attitudes, other shared features]	Error/other

https://doi.org/10.1086/687990 Published online by Cambridge University Press

Social-environmental approaches, by contrast, measure environmental factors directly. Thus, the genetic factors, to the extent that they are playing a role in differentiating one segment of a study population from another in the expression of *T*, now fall into the category "other," assumed not to be playing a role or to be so randomly distributed as not to play a significant role.

Nonshared environment [birth order; differential parental attention; peers]	Shared (intrafamily) environment [parental attitudes re discipline; communication styles; abusive/nonabusive]	Socioeconomic status [parental income; level of education; race/ethnicity]	Error/other

Neurobiological approaches focus neither on genes nor on environment, but on neurobiological processes or structures. Researchers can associate signatures of physiological processes (e.g., metabolites of neuropeptides in blood or urine) with expression of *T*, or through various imaging techniques correlate structural variation with variation in *T*. Here, the category "other" will represent the quantity of expression or variation that cannot be accounted for by the quantity or variation of a neurobiological factor.

Physiology [hormone secretory patterns; neurotransmitter metabolism] Anatomy [brain structure]	Error/other

If the relation between these various causal factors were additive, then the role of "other" would be nonproblematic. It ought to be possible to simply take the sum of (average) correlation values across the different approaches and discharge or significantly reduce that category. The difficulty is that the relation is not additive. Additivity would require the influence of any factor to be independent of the influence of any other factor. But genetic effects vary by environment, environmental effects vary by genetic and/or physiological status of subjects, physiological processes are affected by many other concurrent processes in the organism, and so on. Nor is there a single method of measuring the role of all possible factors. The methods for measuring variation in a given causal space and hence for measuring the strength of association of variation in that space and variation in the trait vary, and they shape the field of investigation and the association measures possible in that field. Hence, the results of the different approaches cannot be combined or compared in any straightforward way.

Furthermore, focusing just on present features, the classification methods of the different approaches place the 'same' factor in different categories.

For example, genetic and environmental approaches each will put intrauterine effects in the other category. Parental divorce is classified as part of the shared environment for one approach and part of the unshared environment in another approach. But the values assigned to the set of purported causal factors must sum to 100%, because the observed values for *T*/not-*T* also sum to 100%. When values assigned to a set of factors must sum to 100%, different parsings of the causal space will yield different values for the 'same' factor.

What this shows, however, is not that single-factor investigative approaches cannot show us anything. Molecular genetic investigation, for example, can produce data sufficient to distinguish between alternative molecular genetic hypotheses, for example, that mutation μ_1 versus mutation μ_2 is more highly associable with trait *T*, or that a particular genetic mutation is associable with trait *T* in contrast with the nonmutated gene. Molecular methods can also tell us something about epigenetic processes that may be involved. But the methods we have for determining the degree of association of one genetic mutation with a trait of interest cannot determine the degree of association of variation in a neurophysiological factor with a trait of interest or the degree of association of a social environmental factor with such trait.

The same holds, mutatis mutandis, for the other approaches. Investigation within an approach cannot produce data sufficient to establish the empirical superiority of hypotheses articulable within that approach to hypotheses articulable within another approach. For example, investigation within quantitative behavior genetics cannot produce data sufficient to establish the superiority of quantitative behavior genetic hypotheses to social environmental hypotheses, and vice versa. Each requires as a background assumption a particular parsing of the causal space. The data, the measured relationships, function as evidence for hypotheses within an approach on the assumption that the parsing is correct.

How is this underdetermination? These approaches are all trying to account for the same phenomenon: observed variation in expression of *T* in population *P*. What are in contestation are pairs like the following (as all approaches acknowledge the influence of other factors):

> **H1.** Behavior genetics explains more of the variation in *T* than does environmentally oriented developmental psychology.

> **H2.** Environmentally oriented developmental psychology explains more of the variation in *T* than does behavior genetics.

These hypotheses are not empirically equivalent. They do not have exactly the same empirical consequences, nor do they call on the same data. But each can call on data generated through its specific investigative methods. So we might say that they are (or can be) equally empirically adequate, meaning that they do just as well as one another with the data their methods generate. Equal empirical adequacy with respect to a common question constitutes local underdetermination. Not only does it constitute a local underdetermination, but it also constitutes one that is unresolvable both currently and in the foreseeable future. This situation supports taking a pluralist stance toward the plurality on display. Moreover, that (as-yet-unspecifiable) methods might be developed in the future that resolve this particular underdetermination situation does not resolve the situation now, nor does that possibility obviate the possibility that such new methods might not emerge, or the possibility that, if they do, they would enter into a new underdetermination situation. A practice-centric approach to evidential relations and underdetermination must be open to plurality being a permanent possibility.

2.2. Defining Behavior. There's more to the background than relations between data and hypotheses. There's the construction of the data in the first place. The behavioral categories with which research starts are drawn from our everyday experience and are, like other concepts of ordinary language, vague when taken out of context. Any particular behavior whose expression will be the object of scientific investigation must be measurable. This requires separation from the context in which it occurs, that is, individuation, and specification in a way that permits identification and reidentification of a given behavior as this rather than that. In the project from which I am drawing, I focused on empirical research on two families of behavior: aggressive behavior and sexual behavior, especially sexual orientation. Here I will focus on issues in the measurement of aggression. What enables a particular behavioral manifestation to serve as a criterion of aggression is a specification of the criterion at a level of abstraction that permits identification and reidentification, that is, that permits counting. To put it differently, aggression must be operationalized, meaning that measurable criteria or indices of aggression must be specified. How is aggression operationalized for purposes of measurement? To be a little more precise, what the research seems to seek to understand are dispositions to aggressive (or violent) behavior. Looking through the literature, we see this disposition measured via

i. conviction of violent crime;
ii. fighting in prison;
iii. delinquency (including truancy and drug use);
iv. violent rage (verbal or physical);
v. anger, irritability, verbal aggression;
vi. hitting a doll;
vii. diagnosis of antisocial personality disorder or oppositional defiant disorder or childhood conduct disorder;
viii. score on a personality inventory (self or other report on Buss–Durkee, Cloninger, others).

https://doi.org/10.1086/687990 Published online by Cambridge University Press

There are some obvious issues affecting our judgments of the adequacy of these indices as criteria of aggression. They are prejudicially selective (in the United States, at least, arrest is racially biased, and conviction and incarceration are racially and economically biased); delinquency and antisocial personality are menu categories that include nonaggressive, nonviolent behavior. It is quite possible for an individual to qualify as delinquent or antisocial without injury or harm to anyone. It is not clear what is the relation of rage/irritability to aggression or violence (here one might want to distinguish physical from psychological aggression/violence). And finally, there is systematic variation in third-party answers to psychological inventories. Parents give different answers to personality inventory questions than do teachers. Even supposing that these current shortcomings could be satisfactorily addressed, it is striking to notice what is missing from the list: state-sanctioned aggression or violence (military), state-overlooked aggression or violence (correctional officers), or the violence to communities caused by corporate negligence, exploitation of vulnerable populations, or the deliberate prioritization of certain values over others.

What does the operationalization of aggression in the ways indicated reveal about the way aggression and violence are understood? In spite of the problems with the categories of delinquency and antisocial personality disorder, the disposition that emerges from the operationalizations is the tendency of an individual to inflict (unsanctioned) harm on another individual. Aggression and violence are represented and measured as individual behaviors and treated as individual dispositions to behave internal or proper to the individual. In spite of the effort to decontextualize and objectify the object of inquiry, its conceptualization retains the negative moral valence aggression has in ordinary discourse. Placing these operationalizations in the context of the investigative approaches reviewed, it seems that to understand the etiology of behavior is to understand how individual dispositions are inculcated. When aggression research is made relevant to crime (as it is), criminality is also understood as the tendency of individuals to act in a way injurious to other individuals. The research question becomes, why do some individuals manifest these kinds of behaviors, while others do not? Thus, whatever answers the investigative approaches using these operationalizations generate about the etiology of aggressive dispositions may only generalize over a limited set of such dispositions.[1] These difficulties raise a natural question: is there any other way to conceptualize behavior, or even just aggressive behavior?

1. This point about limited generalizability is independent of the earlier point about the pluralism indicated by the multiple parsings of causal space. Those issues hold no matter what phenotype we are considering, as long as its etiology is complex enough.

2.3. Behavior beyond Individuals. Here, too, the comparative method can be revealing. I mentioned above a population approach that takes as its explanandum not differences between or among individuals in a population, but differences between populations. Population-level analysis (from neighborhood to nation) asks about the distribution of violent or aggressive or antisocial acts and the dependence of variation in that distribution on population-level factors. Such factors can include social features of a population such as employment opportunities, access to resources, and policing/law enforcement practices, as well as physical environmental factors such as climate and terrain. What is measured in such an approach is not behaviors of individuals, but frequency of episodes or interactions. Researchers can study the frequencies of violent interactions and distributions of violent interactions across different regions, the frequencies of arrests and distributions of arrests across different regions, the frequencies and distributions of recidivism, and so forth. The investigative task then becomes to associate variation in these frequencies and distributions (and variation among them) with variation in other population properties such as

- distribution of police resources;
- patterns of neighborly interaction;
- distribution of employment opportunities;
- distribution of income levels or wealth;
- access to housing;
- age structures;
- kinship structures; and
- features of or changes in physical environment, such as climate stable or changing and in what direction, water availability, soil quality, and so on.

What's the point of exposing these multiple ways of representing the phenomena? It is not to debunk the particular sciences supporting the investigations or the possibilities of gaining knowledge from scientific investigation. Nor is it to engage in debates about scientific realism versus antirealism or instrumentalism. It is rather to point out the multiplicity of kinds of questions we may ask about a complex phenomenon and the variety of perspectives from which we can approach it. It is also to direct attention to the challenges of producing data, the relativization of evidential status to the assumptions/presuppositions involved in treating the data as evidence. For example, the parsings of potential causal space implied by the different approaches will yield different and different kinds of statistical associations to serve as evidence, even when the phenomenon studied remains the same. Assumptions about the representativeness or comprehensiveness of the operationalizations are required to move to hypotheses of broader scope or higher levels of generalization or articulated in a language that outruns the data language. And different conceptualizations of a problem (individual, population) can support different measurement targets and differently constrain the actual (as distinct from the intended) scope of the investigation. Whatever knowledge is generated through one approach remains partial, and its integration with other partial knowledge of the phenomenon cannot be additive, but requires a different kind of understanding.[2]

3. Discussion: Values, Complexity, and Pluralism. Foregrounding the background displays evidential structure in a way that opens the structure to critical/normative engagement. It makes perspicuous the values (scientific and social) that may be informing the constitution of data in a particular way or informing the assumptions licensing inferences from the data to hypotheses. Attending to the difference in targets of measurement also enables us to ask how conceptualizations are preserved or modified as they travel from the laboratory to the broader research world, and then to policy and general knowledge, as well as how they reflect and, by traveling through the laboratory or research design, reinforce social preoccupations. Again, this is not to dismiss the forms of genetic investigation or neurobiological investigation or psychological developmental investigation, but to ask what questions a given approach can answer (as well as what it can't) and to what broader concerns such questions or their answers might (or might not) be relevant.

The example I've offered comes from a particular topical area of behavioral science. My guess, however, is that philosophical investigation of other complex phenomena will show similar complexities in the construction and organization of data. One task for practice-oriented philosophers, then, is to dig into the evidential structure of the sciences in which they are interested. Many philosophers of science (Parker 2011; Lloyd 2012; Winsberg 2012, among others) are already exploring climate research and biomedical research. But I think that it would be fruitful to take this comparative approach to other contemporary controversies that involve science. In educational research, there is a debate as to how to measure effectiveness of instruction. Should it be measured by average scores across an entire school population (Chapman et al. 2011) or by mean scores in different demographic groups (Coulson 2013)? A live question in economic research concerns the financial crisis of 2008. Was the collapse caused by an out-of-control (i.e., too laxly regulated) financial system, as Joseph Stiglitz (2008) contends, or by the mortgage lending practices of the federal programs Fannie Mae and Freddie Mac, which encouraged borrowers to take loans for sums and at interest rates that exceeded their ability to repay, as Myron Scholes contends (Tricks 2008)?

2. For arguments about existing efforts at integration, see Longino (2013).

In some cases positions on the socially relevant matters in a given area of controversy are reflected in or disproportionately supported by some aspect of or approach in the research. To understand the possible role of social values or interests in a given research project, that is, to see where exactly they might be playing a role, it's necessary to look not just to the "external" agents supporting one or another of different approaches to the problem under investigation, but to the structure of proposed evidential relations, to see what data are taken as relevant, and in light of what considerations they are relevant and how they become data. In the case of the behavioral work, such an inquiry reveals that the target of political criticism should not be the use of genetic methodologies, but the ways in which the target behavior is operationalized and conceptualized.

As we thus foreground the background, it becomes evident that some approaches are more visible than others. Might we ask why? Some assumptions are more available, more acceptable, than others. We might ask what factors play a role in the availability of assumptions. Some conceptualizations seem deeply entrenched. We might ask how they become entrenched.

Philosophical analysis can make some headway, not least by illuminating the structures of investigation that inspire such questions, but these are also questions about the social scaffolding of research that sociological investigation might answer. The sociological investigation needs the interpretive-cognitive investigation of the philosopher that can actually generate those questions in a form that makes them relevant to answering questions about evidence, relative credibility, and so on. Philosophy and sociology/anthropology need each other to fully understand the socio-techno-scientific (to coin a phrase) world that is our shared concern.[3]

4. Conclusion. Does this pluralist stance emerging from the focus on scientific practice undermine the cognitive authority of science?

To the contrary! It is overestimations of the degree of certainty we can expect from scientific research that lead to general skepticism about the sciences. The everyday, commonsense concept of evidence draws from the legal context, "beyond reasonable doubt." But reasonable doubt is the watchword of science. Oreskes and Conway (2010) have shown how, in recent decades, clever rhetoricians have undermined confidence in the established results of science by exploiting their inherent uncertainties. The more modest understanding of evidential relations coming from the practice-focused pluralist view supports not skepticism but closer scrutiny of the arguments offered, including statements of purported evidence and of the assumptions in light of which the purported evidence functions as evidence. Our work as

3. For a similar proposal, focused on assumptions in economic modeling, see Bandelj et al. (2016).

philosophers ought to include promulgating noninflationary concepts of scientific acceptability. Such noninflationary concepts are more likely to facilitate appreciation of the complexity of factors going into scientific assessment and interpretation of data than rejection of a hypothesis simply because there are also data serving as evidence for an alternative. This realization is the beginning, not the end, of analysis.

Philosophy of science has an important role to play in our complex technoscientific culture. It plays that role effectively when it attends not just to the cognitive outcomes of scientific inquiry, "theory," but to the practices and social conditions that generate those outcomes, to the rich background that makes scientific inquiry possible.

REFERENCES

Bandelj, Nina, Julia Elyachar, Gary Richardson, and James Owen Weatherall. 2016. "Comprehending and Regulating Financial Crises: An Interdisciplinary Approach." *Perspectives on Science* 24 (4): 443–73.

Chapman, Chris, Jennifer Laird, Nicole Ifill, and Angelina KewalRamani. 2011. *Trends in High School Dropout and Completion Rates in the United States: 1972–2009* (NCES 2012-006). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Coulson, Andrew. 2013. "New NAEP Scores Extend Dismal Trend in U.S. Educational Productivity." *Cato at Liberty*, June 28. http://www.cato.org/blog/new-naep-scores-extend-dismal-trend-us-education-productivity.

Duhem, Pierre. 1954. *Aim and Structure of Physical Theory*. Trans. Philip Wiener. Princeton, NJ: Princeton University Press.

Lloyd, Elisabeth A. 2012. "The Role of 'Complex' Empiricism in the Debates about Satellite Data and Climate Models." *Studies in History and Philosophy of Science Part A* 43 (2): 390–401.

Longino, Helen E. 2013. *Studying Human Behavior*. Chicago: University of Chicago Press.

Oreskes, Naomi, and Erik Conway. 2010. *Merchants of Doubt*. New York: Bloomsbury.

Parker, Wendy. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78 (4): 579–600.

Quine, W. V. O. 1951. "Two Dogmas of Empiricism." In *From a Logical Point of View*, 2nd ed., 20–46. Cambridge, MA: Harvard University Press.

Stanford, Kyle. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.

Stiglitz, Joseph. 2008. "The Opposition's Closing Statement." *Economist*, October 22. http://www.economist.com/node/12411048?zid=295&ah=0bca374e65f2354d553956ea65f756e0.

Tricks, Henry. 2008. "The Moderator's Opening Statement, *Economist* Debate Series: The Financial Crisis." *Economist*, October 16. http://www.economist.com/node/12411051?zid=295&ah=0bca374e65f2354d553956ea65f756e0.

Winsberg, Eric. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22 (2): 111–37.