Underdetermination
a Dirty Little Secret?

Helen Longino

STS Occasional Papers 4

Underdetermination

a Dirty Little Secret?

Helen Longino

Stanford University

occasional papers

Department of Science and Technology Studies

The Department of Science and Technology Studies publishes STS Occasional Papers to disseminate current research undertaken by our staff and affiliated scholars, and to publish the text of public lectures at STS. The aim is to inform, investigate, and provoke. The series covers the whole our diverse field: history, philosophy and sociology of science, science policy, and public engagement of science.

STS Occasional Papers no. 4

Longino, Helen, 2016. Underdetermination: a Dirty Little Secret (London: Department of Science and Technology Studies, UCL)

Published by Department of Science and Technology Studies, UCL, Gower Street, London, WC1E 6BT, United Kingdom

UCL

1. This year I have been revisiting the problems about evidential relations that I proposed to solve with the social account of knowledge developed in my *Science as Social Knowledge* (Longino 1990) and *The Fate of Knowledge* (Longino 2002). The problems start with the observation that, except in the case of empirical generalizations, there are no formal connections between theoretical hypotheses and the empirical data brought forward as evidence for them. Such formal connections (as articulated, for example, in the logical empiricist account of confirmation) would guarantee the relevance of data to hypotheses. In the absence of such formal, logical, connections, data acquire their status as evidence for some hypothesis or other in virtue of background assumptions that establish the relevance of the data to the plausibility or acceptability of the hypothesis. This is what is known as the problem of underdetermination: data underdetermine hypothesis evaluation. This is not a new problem. French physicist and philosopher Pierre Duhem articulated it in the early 20th century (Duhem 1956). But most philosophers of science taking up underdetermination have followed the American logician Willard Van Orman Quine, and represented the issue as the possibility of multiple empirically equivalent incompatible theories (1951). Thus, philosophers as otherwise different as Jarrett Leplin and Philip Kitcher write as though the problem of underdetermination can be solved by showing that such multiplicity is not possible, or that multiple empirically equivalent theories of the same phenomena are in the end the same theory. Kyle Stanford, writing more recently, expresses the problem as that of unconceived alternatives. His formulation resists the solutions Leplin and Kitcher are inclined to offer, but still treats the problem as the availability of multiple theories. We may call this group the holist interpretation of underdetermination. I will review these arguments, indicate why they do not solve the original problem of underdetermination, properly construed, describe some additional problems about evidence that can come under the label of underdetermination and propose some of the solutions that are suggested by the social account of scientific knowledge.

2. Some philosophers assimilate the underdetermination argument to the problem of induction. And indeed, both call into question the rational legitimacy of inferring from a limited sample to claims that go beyond the sample. Induction, that is, enumerative induction, however, still maintains a formal connection between evidence statements and the hypothesis, in that evidence statements are instances of the hypothesis. And hypotheses are just generalizations of the evidence statement(s), e.g., evidence: all birds I have observed fly; hypothesis: all birds fly. Pierre Duhem articulated the underdetermination argument as a problem distinct from the classic problem of induction.

Duhem asks us to suppose that a physicist decides to conduct an experimental test of a hypothesis. Of this proposed action, he says:

> In order to deduce from this proposition [the hypothesis] 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 1956, p. 185)

Duhem spends the rest of chapter VI making the following points. The first expands the comment about the "whole group of theories." The consequence of the necessity of relying on additional information simply to specify what will count as a test of a hypothesis is that the test is not a test of the hypothesis considered on its own, but of the hypothesis together with the additional information. Secondly, when the predicted phenomenon does not occur, the hypothesis itself is not falsified, only the conjunction of hypothesis and additional information; and when the predicted phenomenon does occur, the hypothesis itself is not confirmed, only the conjunction of the hypothesis and additional information. Thirdly, if different additional information is substituted for the original, the evidential relevance of the original testing phenomenon correspondingly change.

This, of course, spells dire consequences for the credibility of scientific investigation. Without some check on the additional information, the confirmation and disconfirmation of hypotheses seems arbitrary. Duhem's own proposal that invokes the "good sense" of the physicist seems at best underdeveloped, if not empty. For a time, Duhem's problem took a back seat to the innovations of the logical empiricists of the Vienna Circle. Working out the details of their formalist accounts of explanation, theory, and confirmation, was challenging and in the context of developments in mathematical logic, intellectually rewarding. But the neglect of Duhem did not last long. The logician, W. V. O. Quine, revived the underdetermination problem in his book of essays, *From a Logical Point of* View, published in 1951.

Quine emphasized what I am calling the holist character of the Duhemian argument by expanding its reach to the entirety of science, understood as our (or an individual's) complete belief set. As Quine put it with Joseph Ullian, it extends to the entire "web of belief" (Quine and Ullian 1970). In Quine's view, a "recalcitrant experience," that is, a phenomenon that contradicts a prediction or expectation does not confront a single hypothesis, or even a single theory, but the whole web of belief. This web is so underdetermined by its boundary conditions, that there is no uniquely correct or optimal adjustment. This position is an aspect of Quine's famous denial of any significant distinction between analytic and synthetic statements. Any beliefs can be revised to maintain the coherence of elements in the web. This view of Quine's came to be the primary interpretation of Duhem's argument, which thereafter came to be known as the Duhem-Quine thesis. But notice what has happened to Duhem's argument. The thesis is understood as a thesis about *theories*, whether we hold, with Quine, that our entire system of beliefs constitutes a single theory, or hold that there can be distinct theories. Quine, as part of his dissolution of the analytic-synthetic distinction, argued that one could make any adjustments in the web so long as all elements remained coherent. I label this a holist interpretation of Duhem (if it can be treated as an interpretation of *Duhem* at all), because it focuses not on individual hypotheses and their evidence, but on the entire system of propositions that constitute a theory and the evidence for such a system. Philosophers who took up the challenge of the Duhem-Quine thesis took it up as the possibility of empirically equivalent theories. By this was meant theories with exactly the same empirical consequences.

As such, it became entangled with the proposal that, given any theory, it's always possible to construct an alternative to that theory that has exactly the same empirical consequences. For example, one can add to classical physics a postulation of a Universal Force (as Reichenbach considered) that acts on everything in the same way and is thus undetectable, but nevertheless contradicts classical physics by introducing an additional element into the universe. Thus we have theory T and theory T+U incompatible but with the same empirical consequences, that is, empirically equivalent. If one holds, as do the logical empiricists, that all meaning derives from the empirical base of theories, then one could hold that T and T+U are identical. But if one holds that theoretical expressions are meaningful independently of their empirical base, then this possibility becomes a serious one. How can one justify accepting theory T over theory T+U, or vice versa? This twist on underdetermination moves attention away from issues of data and their evidential relevance to focus on issues of theory. That is, it is about creating alternative theories, not about interpreting data. The ingenious proposals of Nicholas Maxwell are of similar character: one can construct an aberrant theory T1, alternative to T, along with aberrant evidence e1, that includes but goes beyond the evidence e for T (Maxwell 1974). No purely empirical considerations will decide between T and T1 or between T and T+U. Maxwell, arguing at a high level of abstraction, took this to show that a purely empiricist account of scientific rationality was insufficient and that empiricism needed supplementation. Rather than talking, as Duhem did, of particular assumptions needed to secure the evidential relevance of data to theories, Maxwell was interested in a more global rescue of scientific rationality.

Underdetermination, labeled as the Duhem-Quine thesis, came to be understood as the proposal that given any scientific theory, it is possible to construct an alternative theory that has exactly the same empirical consequences, that is, an empirically equivalent theory, or a theory that could not be empirically differentiated from it.

This shift can be seen in one of the major papers that attempted to defeat the threat posed by underdetermination. Laudan and Leplin (1991), and many other philosophers of science, discussed underdetermination as the Duhem-Quine thesis, that is, as a thesis about the empirical equivalence of theories. As they approach the problem, empirical equivalence is the *source* of the underdetermination thesis, not the other way round. The focus is entirely on empirical theory equivalence, that is, on the prospect of creating theories with exactly the same consequences through devices such as a universal force acting in the same way on everything, or aberrant alternatives to a theory for which aberrant evidence can also be imagined, not on relations between theory and evidence.

Laudan and Leplin (1991) are concerned about the way that empirical equivalence can be used as an epistemological lever against realist interpretations about the unobservable. Contrary to skeptics, they claim that, properly understood, the thesis of empirical equivalence loses all significance for epistemology. By taming empirical equivalence, they claim to have defeated or set aside worries about underdetermination. Let us see how their argument proceeds and if it actually succeeds.

They begin by accepting what they term familiar theses:

> 1. the variability of the range of the observable (VRO): Any circumscription of the range of observable phenomena is relative to the state of scientific knowledge and the technological resources available for observation and detection.
>
> 2. the need for auxiliaries in prediction (NAP): Theoretical hypotheses typically require supplementation by auxiliary or collateral information for the derivation of observable consequences.
>
> 3. the instability of auxiliary assumptions (IAA): Auxiliary information providing premises for the derivation of observational consequences from theory is unstable in two respects: it is defeasible and it is augmentable.
>
> Auxiliary assumptions once sufficiently secure to be used as premises frequently come subsequently to be rejected, and new auxiliaries permitting the derivation of additional observational consequences frequently become available. (Laudan and Leplin, pp. 451-452)

By 1, the variability of the range of the observable, we can never be sure that the set of observation statements presently determined to be consequences of a theory will remain constant. Improved instrumentation may make observable in the future what is not observable now. Thus, the empirical equivalence of two theories is relativized to the present state of science, including what other theories are currently accepted and the instruments of observation and experimentation that are at the time available. You might object that empirical equivalence means that *all* the empirical consequences of the theories are the same, not just the ones we can establish now. The 2nd and 3rd theses address this objection. By 2, the need for auxiliaries, and 3, the instability of the auxiliaries, that a given theory has a particular set of observable consequences, whether or not they can currently be ascertained, is also a temporally indexed phenomenon. The consequences of theory T at time t1, being relative to the auxiliaries accepted at t1, may no longer be consequences of T at t2, when a different set of auxiliaries has displaced the originals. Because both the range of the observables may change and the auxiliaries may change, Laudan and Leplin point out, no purely logical or conceptual argument can show that the condition of empirical equivalence is a permanent condition. Empirical equivalence may be temporary. As temporary, it is not the epistemological threat it seems at first to be.

So, there can be no argument from the present empirical equivalence of theories to persistent or permanent empirical equivalence, and thus, present empirical equivalence offers no support for underdetermination as "the radical thesis that theory choice is radically underdetermined by any conceivable evidence." You might have noticed, however, that the "familiar theses" they employ to dismiss empirical equivalence are just theses that Duhem, himself, would endorse, especially the 2nd of the theses. Can underdetermination really be used to defeat underdetermination?

Kyle Stanford's (2006) develops a new version of the underdetermination problem, which Stanford calls the problem of unconceived alternatives. He demonstrates this new version with examples from the history of biology in the 19th century. Three case studies anchor his argument. 1) Charles Darwin failed to appreciate available alternatives to his theory of pangenesis even when those were brought to his attention. 2) Francis Galton failed to see alternatives to his view of invariant particulate inheritance, even when they were supported by the evidence he had available to him at the time. Similarly, August Weissman, although insightful about many aspects of inheritance, failed to grasp that the phenomena that led him to postulate a multiplicity of intracellular determinants of specific characters might as well be explained by the germ plasm's function as a factory for the production of those determinants. That is, that the germ plasm might itself be capable of variable responses in different contexts. Stanford draws the obvious implications for our epistemic situation: if there were (realistic and available) alternatives then, how do we know the same isn't true now for us? Just as the 19th century biologists either did not conceive or did not appreciate the already conceived alternatives to the views they advanced, so we today may be in the same position of not conceiving or not appreciating equally well supported alternatives to the theories we have adopted. Underdetermination, as the existence of alternative and incompatible theories supported by the available evidence, persists as a fundamental aspect of our epistemological condition.

There are three points worth noting about Stanford's argument here. First, Quine, Laudan and Leplin, and Stanford understand and are concerned with underdetermination as it bears on questions of scientific realism or anti-realism. They are concerned either to support scientific realism (Laudan and Leplin) or scientific anti-realism, in the guise of instrumentalism (Stanford). The issue for these thinkers is this very general question about the interpretation of theories and their potential connection to metaphysical claims about the nature of reality. This is very clear in Maxwell's treatment of underdetermination, which he sees as bearing on the metaphysical question of the intelligibility of the natural world. Secondly, Stanford's approach is still theory-centered. It proceeds from noting the availability, whether in principle or in actuality, of actual alternative theories. It differs from the Duhem-Quine versions only in foregoing fantasized alternatives such as a universal force. In this respect, Stanford's, as well as Laudan and Leplin's, version is still holist in character. This holism is to be distinguished from the holism of theorists of scientific knowledge such as Thomas Kuhn, who argued that observation and the description of evidence are theory laden and in some sense already presuppose the theory for which they will serve as evidence. In contrast both to the logical empiricist picture of the observational grounding of meaning and to the Kuhnian picture of theory grounding of meaning, underdetermination presumes the theory- independence of observation and of description. Its proponents then are faced with the problem of explaining what the relation between observation and theory is, if not formal.

Thirdly, and most importantly, the basic structure of Stanford's argument is similar to that of Laudan and Leplin. Laudan and Leplin argue essentially that we do not know whether the future will develop in such a way as to preserve or to dissolve a present situation of empirical equivalence, hence what they take to be the grounding of underdetermination – permanent empirical equivalence – is removed. Stanford argues that we do not know whether we are failing to conceive of an empirically supportable alternative to our present theories and accepted hypotheses. In both cases, the linchpin of the argument is that we do not know what we do not know. We do not know the full extent of the relationships and processes in the natural world. We do not know how our instruments and framing assumptions will change in the future. We do not know what alternatives we are ignoring or failing to think of now. It is the very contingency of the relationships between data and theory that undermine any kind of absolutist or atemporal claim about features of that relationship. But, just as we cannot argue that a current situation of underdetermination is permanent, so we cannot argue that a current situation of seemingly solid empirical support is permanent. The range of the observables may change and the auxiliaries in play may change. Indeed just because we cannot in the moment conceive of alternative auxiliaries does not mean that they would not be plausible if we were to conceive of them. So, the Laudan and Leplin strategy for defeating underdetermination is not successful.

Maxwell's challenge, furthermore, remains. It prompts two related observations. First, even though Laudan and Leplin take themselves to be addressing empirical equivalence, they see it as a matter of the relations between particular evidence and particular hypotheses or theories. Maxwell's aberrant evidence can persist through multiple developments in real science that change particular evidential relations. Second, whatever solution is proposed to the highly abstract problem he is proposing, underdetermination as a problem of the relations between actual observations and actual hypotheses, rather than a problem about empiricism, persists. That is, even if one were to adopt his aim-oriented empiricism as a shield against the aberrant, there is a continuing problem of securing adequate evidence to support choices among perfectly realistic hypotheses.

3. If we start not from questions about theories and general theses about scientific theories, concerns about realism, or the cognitive authority of science, but from questions arising from the examination of particular episodes of scientific investigation occurring now, what might be called underdetermination has a different cast. And, the theoretical situation with which underdetermination was linked takes a different form than empirical equivalence. Whether we are scientific realists or anti-realists, underdetermination poses a problem for how we analyze evidential relevance. As I represented the underdetermination problem in my earlier work, it concerns the semantic gap between descriptions of single observations (or of sets of observations) that serve as data and the hypotheses the data are taken to support, when these are categorically articulated. As examples, think of the difference in content between descriptions of patterns of tracks in cloud chambers and claims about the behavior of elementary particles, or patterns of hemoglobin oxygenation and deoxygenation

in brain tissue measured via magnetic resonance imaging and claims about specific brain/ mind activity or between recordings of seismic activity on the surface of the earth and claims about the deep structure of the planet. If we think not of a mythical end of investigation, but of the here and now, and represent the matter not as the possibility of multiple empirically equivalent theories, but as the problems of 1) fixing evidential relevance and 2) evaluating the background assumptions that facilitate such fixing, underdetermination presents a serious challenge for epistemologists of science. In contrast to the holist interpretation, we may call this the contextualist interpretation of underdetermination. It is hypothesis specific and does not presume holism of any kind. In my earlier books, I showed how background assumptions needed to establish evidential relevance could enable entry of social values into scientific reasoning and also how moving to a social account of knowledge helped to preserve the objectivity of science. Hitherto underappreciated developments in the sciences themselves both underscore the continued relevance of the underdetermination problem to contemporary philosophy and sciences and require extending the analysis to address these developments. I will discuss three of these. One derives from the complexity of some of the phenomena we seek to understand. Another derives from the character of statistical data and hypotheses. And a third arises because of the increasing use of computer modeling to understand and predict the behavior of complex systems.

Many complex phenomena are investigated by multiple disciplines. In my latest book, *Studying Human Behavior*, I identified a common feature of the investigative approaches whose evidential structure I was comparing. They each required what I call a parsing of the causal space. And, the parsings were different. That is, each approach considered only a portion of the possible causal factors that influence the establishment of dispositions to behave in particular ways. Researchers seek to understand variation in human behavior: why one individual displays one set of behavioral characteristics, while another displays a different set. The empirical research focuses on individual members of the sets of possible behavioral traits and dispositions: risk-taking versus fearfulness, shyness versus sociability, nurturant/caring behavior, aggressive behavior, sexual behavior, etc. The different research approaches used to study variation in such human behaviors, taken together, reveal a large number of different kinds of factor that are involved. One can display these factors in a horizontal grid:

Genotype 1 [allele pairs]	Genotype 2 [whole genome]	Intrauterine environment	Physiology [hormone secretory patterns; neuro-transmitter metabo-lism] Anatomy [brain structure]	Non-shared envi-ronment [birth order; differ-ential parental attention; peers]	Shared (intra-family) environment [parental attitudes re discipline; communica-tion styles; abusive/ nonabusive]	Socio- Economic Status [parental income; level of education; race/ethnicity]

Figure 1. From Longino (2013).

Each research approach is directed at understanding the effects of one or a subset of these factors on behavior. To do this they must hold other factors fixed. That is, in the particular studies they perform, they must assume that the other factors are not playing a role. So the causal universe that they study is not the whole set of possible causal factors, but only a subset. Correspondingly, the grid shrinks to include only those factors among which their methods can discriminate. For example genetic methods must shrink the causal universe to the following:

Genotype 1 [allele pairs]	Genotype 2 [whole genome]	

Figure 2.

Neurobiological research methods can explore elements within this causal universe:

Physiology [hormone secretory patterns; neurotransmitter metabolism] Anatomy	other

Figure 3 .

The “other” category in each grid represents whatever causal factors are responsible for the variation not accounted for by the elements under study. So, they simply represent that portion of the sample conditions not correlatable with the causal factor under investigation. The study is designed to minimize that quantity, by holding all factors other than the ones under study constant as much as possible, but no study can reduce it to zero. The correlation, therefore, that any particular study purports to find between genomic and behavioral data is only evidence for a hypothesis about the particular degree of influence of the genetic variation on the behavioral variation in light of an assumption that no other types of factor are operating, that is, in light of an assumption that the causal parsing is correct. Similarly for the other approaches. Because the causal factors each investigates are different, each assumes a different causal parsing. Because each of the approaches can produce results showing the relevance of the particular factor or set of factors they study to behavioral variation, however, and because there is not an empirical investigative method that can compare factors from different causal parsings, there can be no empirical argument to the empirical superiority of one of the approaches over the others.

Putting this in the terms introduced by Duhem, the background, or auxiliary, assumption brought to the phenomena by any one of the approaches reduces the space of causes and thereby determines the correlations that will be fed into statistical machinery to support hypotheses about the degree of difference any given factor makes to the behavior. The approach selects what is relevant from a complex set of interacting factors. Thus the background assumptions both determine just what data will be counted as evidence and establish the relevance of those data to the hypotheses under investigation. What the experimental methods each employs are best designed to do, however, is to discriminate between the relative influence of potential factors among those it investigates. In spite of their interpretation in the general press, they are not designed to discriminate between the relative influence of one type of factor, say, genetic factors, as against another type of factor, neurobiological or social factors. And to be even more precise, an approach can discriminate among factors in the causal spaces it investigates only presuming all (non-measured) others remain the same or do not interact with the factors under measurement.

This underdetermination situation is a function of the complexity of the phenomena being investigated. While I cannot demonstrate this, I venture to say that the evidential situation of any comparably complex phenomenon will share the same features. Thus, in order to appraise just what to conclude in a general way about such a phenomenon from its investigation, it will be important to take into account the range of approaches involved in its study. Discrimination among approaches will require appeal to supra-empirical considerations, which will themselves vary depending on the context of investigation. (These may include such considerations as simplicity, but also considerations such as applicability, coordination with current theory in related domains and so on.) What makes this underdetermination, rather than multiple independent investigations studying different phenomena, is that the different aspects of the phenomena are, in the actual world, inseparable. That a selected category of data serves as evidence for a selected category of hypothesis rests on assumptions about the non-interference of other factors, that is about the independence of the factor measured.. We may achieve this condition in the laboratory, or even in the design of an observational study, by placing constraints on the situation, but can only assume the condition holds as we transfer from the world of investigation to the world of action.

My second and third examples reflect areas that I have not yet fully investigated, but that seem to me to exhibit classic features of the problem identified in the contextual interpretation of underdetermination, as well as slightly distinct features. Even as the semantic gap persists in the way I've just described, most evidence now is statistical in character, and hypotheses, too, are frequently articulated as statistical or probabilistic relations. How does the change from categorical to probabilistic or statistical expressions affect our understanding of the structure of evidential relations? First of all, we might note that the original underdetermination problem is expressed as a question about the relation between a single data point, the result of a single experiment, and a categorical hypothesis, that is an hypothesis attributing a dependency relationship between one kind of thing and another kind of thing, the collision between a muon and a pion and the subsequent disintegration of one into even smaller particles, for example, or a particular genetic profile and a particular phenotype. But contemporary science, especially science directed at unraveling complex phenomena, does not generate support for categorical hypotheses, but for statistical hypotheses. And the evidence is not a single data point, but an array of data, a set of measurements. In those cases where a single data point does acquire evidential relevance, it is only against a background of statistical information. Different episodes of measurement will yield slightly or even grossly different sets of measurement, licensing different statistical hypotheses about the phenomena. Meta-analysis was introduced to address this confounding feature of statistical analysis. But meta-analyses are only as good as the studies that they include. The measurement data must be indexed to the system used to generate the data: the instruments, the study population, etc. Some commentators, for example, Ioannidis (2005), will conclude that "most published research findings are false." This conclusion assumes that there must be one correct frequency of the measured phenomenon, and so one true value for the hypothesis of relationship. All others are false. But we could as easily see the multiple results as evidence for random fluctuation in nature. Each study takes its data to support a particular hypothesis in light of assumptions about the representativeness of its measured data, about the reliability of the instruments (whether material or in the form of questionnaires/surveys), about the accuracy of identification of measured phenomena. And the variation in the set of studies can be interpreted as a sign that most of them are false or as a symptom of natural fluctuations, again depending on assumptions about fixity or fluctuation in nature.

Finally, one must acknowledge that computer modeling and simulation now play important roles in arguments about the actual or expected behavior of complex systems. How should evidence from computer modeling be evaluated in comparison with straight extrapolation from data? This is a question philosophers have raised about debates concerning the nature and future course of climate change. (Lenhard and Winsberg 2010, Lloyd 2010, Parker 2010). Forms of the question also recur in discussions about evidential reasoning in ecological sciences and other sciences of complex phenomena. Once again, it is an example from contemporary scientific practice that underscores the contextual character of evidence. There is no empirical way to evaluate the comparative evidential relevance of modeling versus straightforward extrapolation. Assumptions that go beyond the empirically determinable are required in order to establish the relative reliability of one method compared with another. These assumptions may have to do with the potential gravity of the issues, as well as with the prospects of additional evidence in the future. The point is that the evidential relevance of the different kinds of data available from extrapolation and from modeling depends on assumptions that themselves require quite different support than is available in the data available to the different methods.

Each of these examples from current practice in science develops a different dimension of Duhem's original problem. Together they show the continuing importance of Duhem's initial insight into the underdetermining character of evidential support for hypotheses. Far from being dismissable as holding in only an empty way, as we may be tempted to conclude when approaching Duhem's argument from a theory-centered or holist perspective, underdetermination is a reminder of the contextual nature of evidence and of the adjustments to scientific epistemology such contextuality requires.

In previous work (Longino 1990, 2002), I have argued that, from the perspective of traditional individualist epistemology, the underdetermination problem undermines the credibility of scientific claims and theories by making the background assumptions arbitrary. The social account of objectivity and by extension of knowledge proposes that traditional epistemology neglects the key ingredient of criticism. What makes background assumptions involved in the selection and classification of data and in mediating the relation between evidence and hypotheses not arbitrary is the subjection of background assumptions to criticism from multiple points of view. Because criticism can go unheeded, the social approach proposes norms that, when observed by a community, makes their criticism transformative, that is, objectivity or knowledge enhancing.

My (2002) takes the earlier analysis a bit further in its proposal of a new term of epistemic appraisal. What is wanted of scientific representation is some form of contact with the phenomena being represented. This contact I term semantic success. Truth, as classically understood, is too narrow a form of semantic success, and it leaves too many forms of scientific representation either false or beyond evaluation. I proposed instead an umbrella term, "conformation," intended to include multiple forms of success: truth, where appropriate, but also approximation, fit, similarity, isomorphism, homomorphism, calibration, and so on. Two important features of conformation are that it is a matter of degree and of respects. A representation can conform to a greater or lesser degree to the intended object of representation. And the respects in reference to which conformation is evaluated may vary. Any complex phenomenon has many aspects and we may only be interested in or able to access a subset of those aspects. A representation may conform to one or several aspects but not conform to others. This means that for any evaluation of conformation, the degree and respect in which conformation is sought must be specified. The degree and respect will be a function of the reasons for which the knowledge is sought, that is, of the goals of the inquiry. What those goals are must also be agreed upon. Furthermore, the specification of the kind of conformation sought in any particular context will often be the outcome of tradeoffs between additional desiderata, for example, precision and applicability. My analysis of the sciences of behavior shows that each approach has the capacity to be successful and unsuccessful in its own terms. Their representations of relations between the potential causal factors they each investigate and observed phenomenon may well conform to whatever the relations actually are even though there is not one single "true" and precise representation of the relation between the phenomenon and all its causal factors. The problem of multiple statistical generalizations about the relationship between or among the same set of factors can also be moderated by changing from the exactness of "truth and falsity" to the greater flexibility of "conformation." We might even say that this is what meta- analysis (the practice of pooling results of many varying statistical studies) achieves – not a true quantitative representation of the relationship, but one that is close enough to permit relying on it in future action or research. To say that it is close enough is to say that it conforms, without needing to say that it is strictly speaking true. A future study that gives close enough, but not exactly identical, results may or may not require adjustment of the precise quantitative estimate of the relationship. But it would be incorrect to say that the new study has shown the meta-analysis to be false. Whether adjustment is sought will depend on the degree of exactness of conformation that is sought, which will in turn depend on the goals of the research.

Determination of degree and respects must be determined socially, that is, through negotiation among those interested in obtaining an estimate of the relationship, that is, those who will be evaluating the strength and reliability of the estimate. Conformation is not a context independent relationship holding between representation and object of representation, but is relative to context and to what communities wish to accomplish with the information. Nevertheless, to be at all meaningful as a normative concept doing the same kind of work as "true/false", those degrees and respects must be specified antecedently to any test of conformation. And whether conformation is achieved will be determined in use, that is, in the success achieved by relying on the representation in cognitive or practical pursuits.

Why do these issues in the philosophy of science matter? While I believe that there are several kinds of public consequence relevant to the broader intellectually and socially engaged community, I will mention only one here. We in the industrially developed world are enveloped in a science based economy. Much contemporary technology has grown from developments in so-called basic science, while the direction of scientific inquiry is in turn science-dependent. And, many public decisions require scientific information. Efforts to develop informed policy are efforts to develop science-based policy. The difficulty is that science is treated has become something like one of Bacon's idols. Both scientists and philosophers have promulgated unrealistic expectations of science. As a consequence, many in the general public also have unrealistic expectations. In the United States, one can see scientific claims evaluated by public figures against a concept of evidence derived from the legal standard of "beyond a reasonable doubt," which suggests conclusive determination. On both categorical and statistical representations of the underdetermination problem, scientific evidence falls short. Yet, scientific inquiry is the best way we have of generating and evaluating representations of the empirical world, representations that will guide policy and interventions. Trying to persuade skeptics by arguing that scientific inquiry meets the legal standard is a losing strategy. Science is never beyond a reasonable doubt, it is all about doubt. The legal standard is inappropriate for the evaluation of scientific evidence. This evaluation always requires assessment of statistical and probabilistic judgments and of the background assumptions that form the context in which evidential relevance is assigned.

We can never achieve certainty, but must be satisfied with the best that empirical evidence can provide, always accepting the risk of fallibility. A more general appreciation of this feature of scientific reasoning would, I believe, make for a great improvement in the character of public debate about publicly relevant science.

To conclude, underdetermination is a genuine feature of scientific inquiry. It is not about the creative adjustment to theories or multiple interpretations of the same formalism, but about the character of our data relative to the claims about nature and society we want those data to support. Rather than treating underdetermination as a dirty little secret best kept from a public used to easy answers, it should be fully acknowledged. As philosophers of science our response ought to be to develop accounts of scientific knowledge that help the nonscientific public appreciate both the inherent uncertainties in our attempts to understand the natural world and that provide tools for holding scientific practice accountable to its ideals.

References

Duhem, Pierre. 1956 *Aim and Structure of Physical Theories*, (first published 1914) trs. Philip Weiner. Princeton. N: Princeton University Press.

Ioannidis, John, 2005. "Why Most Published Research Findings Are False." *PLoS Medicine*. 2,8: 696-701

Laudan, Larry, and Jarrett Leplin. 1991. "Empirical Equivalence and Underdetermination" *The Journal of Philosophy* 88, 9: 449-472.

Lenhard, Johannes, and Eric Winsberg. 2010. "Holism, Entrenchment, and the Future of Climate Model Pluralism." *Studies in the History and Philosophy of Modern Physics* 41: 253-62

Lloyd, Elisabeth. 2012. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77, 5: 971-984.

Longino, Helen. 1990. *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

Longino, Helen. 2002. *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

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

Maxwell, Nicholas. 1974. "The Rationality of Scientific Discovery." *Philosophy of Science*. 41, 2: 123-153.

Parker, Wendy. 2010. "Predicting Weather and Climate: Uncertainty, Ensembles, and Probability." *Studies in History and Philosophy of Modern Physics* 41: 263- 272.

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

Quine, W.V.O. and Joseph Ullian. 1970. *The Web of Belief*. New York: Random House.

Stanford, P. Kyle. 2006. *Exceeding Our Grasp*. New York: Oxford University Press.

STS Occasional Papers

1 **Joe Cain** No Ordinary Space: A Brief History of the Grant Museum's New Home at University College London. 2011.

2 **Simon Schaffer** Mutability, Mobility and Meteorites: on Some Material Culture of the Sciences. 2014.

3 **Hsiang-Fu Huang** Ouranologia: an Annotated Edition of a Lenten Lecture on Astronomy with Critical Introduction. 2015.

4 **Helen Longino** Underdetermination: a Dirty Little Secret? 2016.

occasional **papers**
Department of Science and Technology Studies