

Inquiry, Evidence, and Experiment: The “Experimenter’s Regress” Dissolved

Matthew J. Brown

September 10, 2008

Abstract

Contemporary ways of understanding of science, especially in the philosophy of science, are beset by overly abstract and formal models of evidence. In such models, the only interesting feature of evidence is that it has a one-way “support” relation to hypotheses, theories, causal claims, etc. These models create a variety of practical and philosophical problems, one prominent example being the *experimenter’s regress*. According to the experimenter’s regress, good evidence is produced by good techniques, but which techniques are good is only determined by whether they produce the evidence we expect. The best answer to this problem within the traditional approach relies on the concept of *robust* evidence, but this answer ultimately falls flat because it creates impossible requirements on good evidence. The problem can more easily be solved by rejecting abstract, formalistic models of evidence in favor of a model of inquiry which pays attention to the temporal complexity of the process of inquiry and the distinction between *observational* and *experimental* evidence.

1 Introduction

Several problems in the contemporary discussions of evidence only arise because of inattention to the complexity of the process of inquiry and the uses of evidence. While inattention to these matters seems endemic amongst philosophers, the problem has also come to infect certain of the social and medical sciences as well as policy-making. The source of the problem is the

use of idealized models of inquiry and evidence, in which evidence is characterized by a formal relation of support it has to a hypothesis or theory. In such models, evidence is useful for this one task, and the task can be characterized without recourse to factors of the particular context or temporal development of the inquiry in which evidence arises and is used.

This essay will show how a model of the temporal dynamics of inquiry and the functional role of evidence better addresses the problem known as the “experimenter’s regress,” popularized by H.M. Collins (1975, [1985] 1992). According to this model, there are different kinds of evidence, playing different roles in an inquiry, and one major distinction is that between *observational* and *experimental* evidence. Along the way, I will argue that the appeals to the robustness of evidence (e.g., Culp 1995) in attempt to solve the experimenter’s regress are bound to fail, and that the value of robustness itself is better understood according to the inquiry-model than by traditional approaches.

2 Evidence and the Model of Inquiry

The temporal dynamics of inquiry have received scant attention in philosophy of science. There are only two major models which develop the temporal complexity of inquiry. One has been developed by Kuhn and his followers and critics. Kuhn’s model discusses the career of large-scale theories or research paradigms that govern entire areas over a large span of time. These models are sufficiently large-scale and long-term that they are not useful for addressing current concerns about the nature of evidence. Problems like the experimenter’s regress deal not with the evolution of theories over the long run, nor the revolutionary replacement of theories or paradigms. The questions at issue are far more local.

The other major model of inquiry on offer is one introduced by C.S. Peirce and articulated by John Dewey. This model works best at the more local level of particular scientific inquiries, though it has some applications at the larger scale. I will not discuss potential conflicts or compatibilities between the two models, nor will I attempt a full defense of this model here,¹ though I will attempt to make it as plausible as possible and provide an illuminating example.

¹ I have done so elsewhere (Brown 2009, in preparation).

In the main outlines, the pragmatist model of the dynamics of inquiry goes has the following interlocking phases:

1. *Inquiry Begins with a Felt Perplexity.* There are many types of perplexity, but they are not in general a mere state of ignorance on the part of the inquirer. Rather, the state of the science, including theories and techniques, is discoordinated or indeterminate, and this is reflected in the feeling of perplexity on the part of the inquirer(s). There are conflicting tendencies within the situation of the field at the present time, requiring investigation. (Contrast this with the smooth application of a theory to a situation with immediate success.)
2. *The Institution of a Problem.* The situation must be assessed in order to *formulate a problem-statement* that adequately captures the felt perplexity. *Operations of observation* must take place in order to arrive at a statement of the problem, which evolves as the inquiry develops.
3. *The Determination of a Problem-Solution* The first pass at observation and problem-formulation *suggest hypotheses for solving the problem.* The hypothesis is usually but not always generated from a larger field of background theory, which may require special development in the specific context.
4. *The Coordination of Observations and Hypotheses.* A reciprocal process of coordination and improvement of observed facts and theoretical-hypothetical ideas is undertaken, in which hypotheses are developed or eliminated, new more refined observations are made, and the problem-statement is refined.
5. *The Necessity of Experiment.* A series of tentative, experimental applications of the hypotheses are made in order to estimate their efficacy as problem-solutions. Earlier experiments can suggest more refined experiments, or the necessity of further articulating data and hypothesis, or the need to “go back to the drawing board.” Solving any problem requires operations of making and doing. Ideas cannot be tested by a purely passive collection of facts through observation. Active interventions of an experimental nature provide necessary evidence about the prospective effectiveness of solutions.

6. *Judgment.* Inquiry continues until a hypothesis is adjudged to resolve the problem, while the alternatives have been ruled out, and the conclusion can be used as a reliable means to further inquiries.

This is obviously an idealized picture of the conduct of inquiry, but it is informed by Peirce's and Dewey's participation in and studies of the actual practice of science. It is a *normative-explanatory* model, attempting to capture and explain the lessons of successful inquiries past. It will be helpful to use a concrete illustration to clarify this abstract model. Consider the work of John Snow on the transmission of the disease cholera.²

3 Snow on Cholera

The basic outlines of the problematic situation are clear: cholera is a terrible disease, fatal in nearly all cases in Snow's time. It is tempting to say that the *problem* itself is clear from the beginning: how is cholera communicated, and how can its transmission be prevented or contained? While the idea of contagious diseases was not new in the middle of the nineteenth century, when Snow was at work on cholera, it was neither fully accepted nor clearly distinguished from views identifying disease as a punishment for sin. To regard some diseases as communicable, and to identify cholera in this way, is already to be well on into the inquiry. Cholera tended to be concentrated amongst the poor, and almost never infected the doctors who tended to the sick. This was taken as evidence that the disease was "a just punishment for the undeserving and vicious classes of society" (26). To regard the problem as one fixed prior to inquiry would be to take as fixed from the beginning many things that were at first unsettled.

Snow begins by collecting a variety of general facts, such as the beginnings of the disease in India and the spread via human interaction (29). He then moves to more specific cases (30–1). From the start, the evidence clearly suggests that *the disease is communicable*. But it doesn't fit well with the more popular "effluvia" theory of transmission of disease through the air, since spending time in the company of the sick doesn't necessarily lead to infection (31). Rather, a particular pattern of behavior (tending to the patient in

² My discussion here is taken from Goldstein and Goldstein (1978, pp. 25–62) who draw heavily on Snow's own manuscripts. Parenthetical references are to their discussion.

intimate fashion) and pathology (beginning with intestinal symptoms) suggest another hypothesis: *The disease spreads by some infected matter from a cholera patient being accidentally ingested in sufficient quantity* (33).

This hypothesis suggests some further observations. If it is valid, you'll find that certain people who come near to the patient do not get cholera (as we've seen), and that they avoided it by way of habits of cleanliness that would prevent them from accidentally ingesting any choleraic evacuations. Indeed, this is clearly the case with doctors, who do not generally contract cholera from their patients (33). Reasoning through the implications of the hypothesis, we can see that there are several reasons that people of different social classes would have different levels of risk of contracting the disease based on different living conditions and behavior around the sick (33–34).

One observation raises a puzzle, however. Cholera does sometimes spread to the rich despite the absence of the vectors of direct communication present in the case of the poor. Snow did not take this to invalidate the hypothesis, however. Rather, he supposed a further specification of the hypothesis in these cases that would provide the appropriate kind of transmission vector: *cholera can spread through the water supply* (35). Further cases support this hypothesis.

Having worked out the implications of the hypothesis and found corresponding facts is not where Snow stopped. The next phase requires experimental application of the hypothesis to real situations in order to test its adequacy. Experimental application is *not* just a special way of generating further observations. Certainly, techniques of observation are part of the experiment, but the function is nonetheless very different. The functions of observation are to fix the conditions of the problematic situation and the terms of the problem, as well as to suggest and refine hypotheses. The function on an experiment is to put the hypothesis *into practice*, in a limited and controlled fashion, in order to determine its efficacy in solving the problem.

Snow engaged in at least two experiments, neither of which is entirely satisfactory from the point of view of our model, though the details here are not so important to the main point. The final part of Snow's monograph on cholera is the most crucial, from the point of view of our model. In the last section, Snow provides a list of twelve recommendations for how to prevent the spread of cholera, based on his two hypotheses, plus some further reasoning about possible cases. For example:

- 1st. The strictest cleanliness should be observed by those about

the sick. . .

3rd. Care should be taken that the water employed for drinking and preparing food. . . is not contaminated with the contents of cesspools, house-drains, or sewers. . .

11th. To inculcate habits of personal and domestic cleanliness among the people everywhere. . .

Such recommendations are crucial to the eventual acceptance of Snow's explanation. No amount of convincing argument provided by a scientific manuscript can be the ultimate measure of a scientific judgment's warranted assertibility. What matters is that others take the results to be so settled as to provide a steady resource for further inquiry and that future applications, such as the ones suggested by Snow in this final section, are successful; these are the "decisive experiments" in favor of Snow's view.

4 The Experimenter's Regress

Many people regard the impact of theory on evidence as having problematic consequences. For example, Robert Hudson (2000) believes that if we cannot make room in our epistemology for direct perception, unmediated by theory or concepts, then we can never escape the "hermeneutic circle" and find some independent ground for our knowledge-claims unsullied by the question at issue. Sylvia Culp (1995), in a similar vein, worries about and attempts to solve the problem of the "experimenter's regress" raised by H.M. Collins (1975, [1985] 1992). According to Collins, good data is regarded as the product of a good experimental technique, but the test of an experimental technique is just whether it produces the expected data. The same worry can be put about the need to interpret "raw data" before it can become "data" or "facts" (Culp 1995, p. 439).

Something like the following picture, suggested by Culp, is surely right: what happens in the lab prior to interpretation is merely a brute happening, and brute happenings are not themselves evidence. We must then interpret those happenings, take them up as a certain item of fact, and, metaphorically speaking, teach them to speak the language of the theory, in order to see how they bear on the theory. This interpretation is never independent of theory, neither the theory of how the apparatus works nor the theory in question. All of this presents a problem, according to Collins and Culp, because we are left

wondering how interpretations of experiments that *themselves* presuppose controversial theories, including parts of the theory in question, can serve as solid ground to support those theories.

From the point of view of the inquiry-model, several crucial parts of the story have been left out. For one, it mentions only one direction on the two-way street of the coordination of factual and conceptual materials. Contra Culp's supposition, we don't only teach evidence to "speak the language" of theory. We also teach the theory to speak language of observation; that is, we must develop our hypotheses so that they have operational consequences, that they may *direct* activities of observation and so that experiments may be created that apply the hypothesis as a solution to the problem. Collins' and Culp's shared way of setting up the problem presupposes that theory is inert, and experiment must be constructed or interpreted in a way that meets it. But theory and experiment must meet in the middle.

Further, both parties to the debate construe the function of evidence extremely narrowly, collapsing the distinction between observational and experimental evidence. Evidence is taken to be exhausted by its function of supporting a theory, but this is a relatively minor function of evidence within the course of inquiry. Observation serves to help fix the problem, it suggests hypotheses for solution, helps improve hypotheses. Experiments put hypotheses to work in tentative application, trying them out as solutions to a problem. It is undeniable that in some sense, theories "produce" their own evidence, but this is only a problem if evidence only serves to justify theory, and theory is only justified by that body of evidence it produces. To the contrary, producing (not predicting) some events is the *point* of a theory; it is the adequacy of the consequences produced to solving the problem at hand, along with its usefulness in attacking new problems and supplementing new inquiries that are the ultimate test of the theory.

A key to the problem of the experimenter's regress is the issue of *calibration*.³ Early attempts to detect or measure some previously unobserved or unquantified phenomenon are faced with a problem of how to calibrate the technique, lacking any other techniques to check against. We have only theoretical expectations about what the phenomenon should be like to guide us.⁴ Later attempts are faced with the problem that their calibration depends on

³ See section I.B.1 of Franklin 2007.

⁴ Hasok Chang's work on temperature (2004) explicitly addresses the way that basic expectations guide this process.

previous measurements which themselves were not calibrated in a standard way. In both cases there is a troublesome regress; in the later case, the circle of data and technique is simply pushed back to earlier stages.

But the question we should ask is, “What is this experimental evidence *for*?” Under the impoverished model of theory-evidence relationships that regards the sole role for evidence to be either adding or removing support from a hypothesis (in context-free fashion), the experimenter’s regress is a serious concern. If evidence lacks independent plausibility, it cannot stand as support in the way this simple model would hope. Godin and Gingras (2002) have suggested that the “experimenter’s regress” amounts to just the classical problems of skepticism, and thus that we should get around it in the same way that we get around skeptical worries in epistemology generally. This answer will not do, however, as the problem is *internal* to the traditional model, once the facts of theory-dependence in evidence are accepted. One must either elaborate or replace this model to avoid them.

Sylvia Culp’s (1995) alternative solution to the problem posed by the experimenter’s regress is to appeal to the *robustness* of evidence. We need not have full independence of evidence from our expectations. Rather, what we need is evidence from a *variety* of different kinds of sources that are independent from each other, whose interpretation relies on the theory in question in quite different ways (if at all), and that all support the same conclusion. Evidence from a single source that seems to support the conclusion but only does so due to being calibrated that way would be problematically circular. A variety of different types of evidence, developed independently from each other, which all seem to support the conclusion but in fact are just the product of our expectations, so the argument goes, would be a miracle. A far better explanation is just the truth of the hypothesis.

The strategy is an appealing one. While no single thread can do the job, a rope woven in the right way can be strong enough. In Culp’s argument, she fully admits that no particular bit of evidence can be theory-free, that it doesn’t even make sense to talk of uninterpreted, bare “happenings” as evidence. Nonetheless, since she is committed to the metaphor of support, she attempts to find an arrangement of evidence that can be strong support our hypotheses. According to her argument, a set of evidence can be a foundation for theoretical knowledge if it is *robust*—if it comes from a variety of sources that are theoretically independent of each other.

This argument fails to meet the challenge posed by the experimenter’s regress, however. At least three difficulties arise, one empirical and two epis-

temological.⁵ The first is the difficulty of finding *really* independent sources of evidence. The history of the development of experimental techniques is replete with a variety of cross-calibration techniques. Chang’s (2004) discussion of the development of the modern thermometer shows the complex interdependencies of various new techniques for measuring temperature. Early errors propagate into later techniques and take a long time to disappear entirely, as in the case of measurements of the charge of the electron (Feynman [1974] 1999), because of the preponderance of cross-calibration. True independence may be difficult to determine.

The second problem, which springs off from the first, is that robustness doesn’t really solve the problem of calibration. For any particular measurement technique, there are two cases: either it is calibrated according to existing techniques, or it isn’t. In the former case, the possibility of independent techniques of measurement is seriously endangered. Furthermore, the question of how those pre-existing techniques were themselves calibrated must be examined. In the latter case, it would appear that all we have to go on to judge the results provided by the technique is the very expectations we hope to support. A variety of different types of evidence, all calibrated by reference to the same set of expectations also lack the independence required by the argument.

On the other hand, it may be that the different types of measurement, though originally calibrated in a suspect way, are calibrated with respect to different, independent sets of expectations.⁶ While problematic in those original circumstances, in a *present* case, they may be sufficiently independent from one another to provide robust, adequate evidence in the case at hand. Even supposing that this case passes the empirical test of independence discussed above, a larger question about whether we ought to rely on the evidence remains. Perhaps we ought to regard it as a miracle that a variety of such evidence purportedly supports a single conclusion, but why should we think that the truth of that conclusion explains the apparent miracle, given the story of evidence now on offer? A variety of methods, calibrated under highly suspicious circumstances, apparently providing no real support in the case of their original development, now all happen to agree on one conclusion. Do we have any reason to believe that this coincidence has anything

⁵ Compare to Jacob Stegenga’s “three easy problems” for robustness (unpublished). The first (empirical) problem is especially close to Stegenga’s discussion.

⁶ Though this seems unlikely in the light of Chang’s discussion of the underlying expectations that inform the development of measurement techniques.

to do with the truth of the conclusion? Not without some prior reason to think that the methods, taken individually, track the truth in even a modestly reliable fashion. But it is *precisely* the lack of such a reason in the case of individual techniques that leads to the demand for robustness in the first place.

A final problem arises for the attempt to solve this problem through the appeal to robustness. As mentioned before, in order to have truly independent sources of evidence, it is crucial that the measurement techniques not be calibrated to one another, lest the bias in one creep in to the other. The sources must be multi-modal, and they must be incommensurable, in the sense of not having any inter-modal standard of comparison (otherwise, they are probably calibrated to one another). If they are incommensurable in this way, however, we're left with a major worry: if we have no standard of comparison between the types of evidence, how can we say determinately that they support the *same* conclusion? If the interpretive framework at hand is the theory in question, of course, then it is easy to see how different pieces of evidence support the same conclusion. But if all the evidence can be interpreted by the theory in such a way as to allow cross-modal comparisons, it isn't really independent in the way that Culp demands. Suppose, then, that the sources of evidence are all independent from one another in the strong sense. How do you determine the relevance of each to your hypothesis?⁷ Evidence that meets the requirements of robustness, understood in the way it must be in order to solve the problem at hand, may be sufficiently incongruous that it would be difficult to make even qualitative comparisons. And in the common case where there is some discordance between different types of evidence, the necessary lack of an inter-modal standard of comparison prevents us from knowing how to resolve the conflict.⁸

The inquiry-model of evidence provides a very different answer to the question of the purpose of evidence. Evidence has a variety of functional roles within an inquiry, the main goal of which is the resolution of the perplexity which spurred the inquiry. In general, then, the experimenter's regress will not present any difficulty, since what matters is that the evidence fulfill its role well enough for the purposes of solving whatever problem presents itself. So long as we find a way to combat the disease and increase the life and

⁷ The problem of relevance is raised by Nancy Cartwright in her discussions of evidence-based policy (2009) and discussed by Stegenga (unpublished).

⁸ The terms "incongruity" and "discordance" and the associated problems are raised in the context of robustness by Stegenga (unpublished).

vitality of people, it doesn't matter than the experimental techniques have a variety of dependencies on the experimenter's expectations. Since experiment is *not* merely a procedure for producing neutral evidence, but *rather* a way of *making* and *doing* that puts the hypothesis *into practice*, there is a test of the experimental evidence, together with the hypothesis, that is independent of expectations *per se*. Expectation cannot prevent a bridge from falling down, nor can it cure disease, nor can it even make quantum mechanics compatible with general relativity.

5 The Value of Robustness

In attempting to respond to the experimenter's regress and related problems, the defenders of the value of robustness have created an insoluble dilemma. Robustness must on such accounts achieve independence from foreground expectations and background theory by being so independent from each other that the potentially infecting theories, assumptions, and expectations "cancel out," that it would be a "miracle" if such diverse techniques all point to the same conclusion. But in order to achieve independence, the members of the set of evidence must end up being mutually incommensurable, because commensurability requires the shared background and mutual calibration that endangers the needed independence. However, the incommensurability of evidence brings with it the problems of incongruity and discordance, which threaten the very possibility of determining how the evidence bears on a hypothesis.

Not only is robustness unnecessary to solve the problem of the experimenter's regress, which disappears when we move from an impoverished model of evidence to the inquiry-model, but "robustness" as we've been forced to defined it is actually an impossible requirement. If evidence cannot be integrated, then inquiry cannot move towards resolution. As a result, it may seem that robustness has no place as a scientific norm. This is an unacceptable conclusion, given the apparent obviousness of its epistemic value and its unanimous support amongst scientists. But it isn't the value of robustness *per se* that has been challenged in this chapter. Rather, it is the particular way of understanding robustness that Culp and others are forced into. Robustness, as it figures in the methodological platitudes which the defenders of robustness cite, is merely the recommendation to seek evidence of several type from different sources. The further requirement of complete

independence is forced by the purposes that Culp puts robustness to. If we relax these impossible restrictions on robust evidence, the value of robustness becomes more clear. A set of evidence that includes many different kinds of *physical* processes, and one that does not depend on controversial hypothesis that are unnecessary to the hypothesis in question or the materials need to integrate the evidence has the obvious value that we would expect. When robustness is not asked to do an impossible job, it is no longer plagued by irresolvable difficulties.

6 Conclusion

In closing, I would like to emphasize the *variety* of roles that evidence plays in the course of an inquiry. In many accounts, evidence is mono-functional: all evidence serves as a test of a theory/hypothesis, and it confirms or disconfirms, and there is no interesting difference between evidence garnered by observation versus that gotten by experimentation. In the model of inquiry I've been discussing, however, evidence serves many purposes. Observational evidence helps locate the problem; it provides information about fixed conditions; it guides speculation and hypothesis-formation; it helps us eliminate or improve our original hypotheses. Experimental evidence also serves as a tentative application of a developed hypothesis to check its consequences for future action and inference. In every case, it is not some abstract or formal relation between the evidence and the hypothesis by which the evidence serves to justify the hypothesis. It is rather a very concrete process of transforming a perplexity into a resolution that evidence serves, and which ultimately justifies any final judgment of the inquiry. The formal or symbolic features of evidence are only one small part. This complex, contextual, and pluralistic model of the role of evidence in inquiry can serve to resolve a variety of problems, as we've seen in the case of the experimenter's regress.

References

- [1] Brown, M.J., "Models and Perspectives on Stage" forthcoming in *Studies in the History and Philosophy of Science A*, Spring 2009.
- [2] Brown, M.J., "Scientific Significance and Genuine Problems," in preparation

- [3] Brown, M.J., *Science and Experience: John Dewey's Philosophy of Science*, dissertation manuscript in preparation
- [4] Cartwright, N.D., 'Evidence-Based Policy: What's To Be Done About Relevance', forthcoming in *Philosophical Models, Methods, and Evidence: Topics in the Philosophy of Science*. Proceedings of the Thirty-Eighth Oberlin Colloquium in Philosophy. Special issue of *Philosophical Studies*, Spring 2009.
- [5] Chang, H., *Inventing Temperature: Measurement and Scientific Progress*. New York: Oxford University Press, 2004.
- [6] Collins, H. M., (1975) 'The Seven Sexes: A Study in the Sociology of a Phenomenon, or The Replication of Experiments in Physics', *Sociology*, 9, 2, 205–224.
- [7] Collins, H. M., ([1985] 1992) *Changing Order: Replication and Induction in Scientific Practice*, Chicago: University of Chicago Press.
- [8] Culp, S., "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles" *Philosophy of Science*, Vol. 62, No. 3. (Sep., 1995), pp. 438–458.
- [9] Feynman, R.P., "Cargo cult science: some remarks on science, pseudoscience, and learning how to not fool yourself." In: Feynman RP, Robbins J. *The pleasure of finding things out*. Cambridge, Mass.: Perseus Books; 1999:205-16.
- [10] Franklin, A., "Experiment in Physics", *The Stanford Encyclopedia of Philosophy* (Fall 2007 Edition), Edward N. Zalta (ed.), URL = [http://plato.stanford.edu/archives/fall2007/entries/physics-experiment/.](http://plato.stanford.edu/archives/fall2007/entries/physics-experiment/)
- [11] Godin, B. and Y. Gingras, The experimenters' regress: from skepticism to argumentation, *Studies In History and Philosophy of Science A*, Volume 33, Issue 1, , March 2002, Pages 133–148.
- [12] Goldstein, M. and I.F. Goldstein (1980), *How We Know: An Exploration of the Scientific Process*, Da Capo Press.

- [13] Hudson, R.G., (2000), "Perceiving Empirical Objects Directly," *Erkenntnis*, Volume 52, Number 3 / May, 2000
- [14] Stegenga, J., "Robustness, Discordance, and Relevance" Hadden Prize Essay, Canadian Society for the History and Philosophy of Science, currently unpublished.