4-2004

Error as Means to Discovery

Kevin C. Elliott
University of South Carolina - Columbia, elliotkc@mailbox.sc.edu

4-1-2004

Error as Means to Discovery

Kevin Elliott
University of South Carolina - Columbia, elliotkc@mailbox.sc.edu

Error as Means to Discovery*

Kevin Elliott†‡

This paper argues, first, that recent studies of experimentation, most notably by Deborah Mayo, provide the conceptual resources to describe scientific discovery's early stages as error-probing processes. Second, it shows that this description yields greater understanding of those early stages, including the challenges that they pose, the research strategies associated with them, and their influence on the rest of the discovery process. Throughout, the paper examines the phenomenon of "chemical hormesis" (i.e., anomalous low-dose effects from toxic chemicals) as a case study that is important not only for the biological sciences but also for contemporary public policy. The resulting analysis is significant for at least two reasons. First, by elucidating the importance of discovery's *earliest* stages, it expands previous accounts by philosophers such as William Wimsatt and Lindley Darden. Second, it identifies the discovery process as yet another philosophical topic on which the detailed studies of the "new experimentalists" can shed new light.

1. Introduction. Scientific anomaly plays a crucial role in classic accounts of scientific reasoning, theory change, and discovery (e.g., Hanson 1958, 1961; Popper 1959; Kuhn 1970; Lakatos 1970; Laudan 1977; Shapere 1977). Several philosophers (e.g., Hon 1989; Mayo 1996; Allchin 2001) have recently begun to extend these traditional studies by exploring how scientists isolate the specific *errors* associated with anomalies.[1] This concern with scientific error flows, at least in part, from a recent emphasis

*Received June 2003; revised October 2003.

†To contact the author write to Department of Philosophy and Religious Studies, 106 Coates Hall, Louisiana State University, Baton Rogue, LA 70803, e-mail: kelliot1@lsu.edu.

‡I would like to thank Kristin Shrader-Frechette, Kelly Smith, Douglas Allchin, and two anonymous referees for helpful discussions and comments that strengthened this paper. I also thank Douglas Allchin and Lindley Darden for encouraging me to develop the paper.

1. The paper follows Allchin (2001, 38) in using 'error' to refer to "any mistaken conclusion or unintended outcome in science or technology." Thus, anomalies plausibly qualify as *instances* of error (i.e., as unintended outcomes), but anomalies also indicate that *further* errors (i.e., mistaken conclusions) must be present in order to produce the inconsistency or inexplicability between "fact" and theory that constitute the anomalies.

Philosophy of Science, 71 (April 2004) pp. 174–197. 0031-8248/2004/7102-03$10.00

in science studies on the details of scientists' experimental practices (e.g., Hacking 1983; Franklin 1986, 1997; Galison 1987; Mayo 1996). According to Deborah Mayo, for example, scientific experimentation is fundamentally a matter of testing for error and eliminating it. With regard to scientific anomaly in particular, Mayo claims that scientists perform experimental tests in order to probe for the specific errors associated with each anomaly (1996, 55). Allchin (2001, 53) also emphasizes that scientific anomalies reveal the presence of errors, and he provides a typology of error types that can help researchers to probe for the specific error that is present in each instance of anomaly.

This paper's thesis is that these recent studies of scientific error provide the conceptual resources to describe the earliest stages of scientific discovery as error-probing processes and that this description yields greater understanding of the entire course of discovery. Previous influential accounts of discovery (e.g., Wimsatt 1987; Darden 1991; Schaffner 1993) suggest that scientific inquiry often occurs via what Nickles (1996, 1997) calls a "multi-pass" progression, in which researchers gradually refine models and theories, frequently in response to anomalies. Nevertheless, these accounts have not focused much attention on the earliest stages of this process, in which scientists initially attempt to characterize the anomalies that they later try to eliminate. The current paper attempts to shed new light on these early stages by describing them as processes by which researchers probe anomalies for error. Specifically, it examines how researchers are probing for the errors associated with a contemporary biological anomaly called "chemical hormesis." It argues that the case study highlights the challenges posed by these early error-probing stages, the strategies with which researchers approach them, and their influence on the rest of the discovery process.

Section 2 discusses how the early stages of scientific discovery could be described as processes of probing anomalies for error. The next section illustrates this suggestion using the case study of chemical hormesis. Section 4 then argues, on the basis of the case study, that the entire discovery process can be illuminated by describing its earliest stages as error-probing processes. The paper is important for at least two reasons. First, it provides a more detailed description of scientific discovery's *earliest* stages, which have not been emphasized in recent descriptions of discovery (e.g., Wimsatt 1987; Darden 1991; Bechtel and Richardson 1993; Schaffner 1993; Nickles 1997). Second, it shows how the study of error can enrich the philosophical literature on scientific discovery. Both Mayo (1996, 51) and Allchin (2001) suggest that their work should be relevant for understanding the discovery process, but they have not explored this suggestion in detail. Thus, the paper constitutes an example of fruitfully integrating recent approaches to science studies (namely,

detailed studies of scientific experimentation) with a more traditional topic in the philosophy of science (i.e., scientific discovery).

2. Error Probing and the Early Stages of Discovery. This section of the paper argues, first, that recent studies of scientific discovery do not describe its earliest stages in detail. It then shows how Mayo's examination of scientific error provides the conceptual resources to describe those early stages as error-probing processes. As an example of recent philosophical work on discovery, the paper employs the studies of William Wimsatt (1987) and Lindley Darden (1991, 1992). Their work is chosen for at least four reasons: (1) it is influential, (2) it elucidates the role of anomaly in discovery, (3) it is broadly representative of other recent studies of discovery (e.g., Bechtel and Richardson 1993; Schaffner 1993; Nickles 1997), and (4) Darden's work in particular describes discovery's earliest stages in as much or more detail than any other previous account.

In Wimsatt's article "False Models as Means to Truer Theories" (1987), he argues that false models can play an important role in the development of scientific knowledge. He begins by noting that a model can be false in a number of ways: (1) by having only local applicability, (2) by being an idealization, (3) by being incomplete, (4) by misdescribing the interactions of variables, (5) by giving a totally "wrong-headed" picture of nature, (6) by giving purely phenomenological descriptions, or (7) by simply failing to describe or predict data (Wimsatt 1987, 28–29). Scientists obviously would prefer to have true models rather than false ones, but they don't usually have a choice in the matter; therefore, Wimsatt emphasizes that false models (especially ones that involve idealization or incompleteness) can frequently be of great value. He provides an extensive list of functions that such models might play in the search for better ones, including: (1) serving as a starting point for a series of more complex models, (2) suggesting new tests or refinements of established models, (3) serving as templates that account for large-scale effects and that make smaller effects noticeable, (4) serving as limiting cases that are true under certain conditions, (5) defining extreme cases between which all other cases lie, and (6) providing a simple arena for determining some properties of a system (1987, 30–31).

Wimsatt illustrates many of these functions with specific cases from the development of chromosomal mechanics by the Morgan School during the early part of the twentieth century. These examples clarify that, although Wimsatt does not *explicitly* state that anomalies are central to the early stages of discovery, they play a critical role in his account. For example, because Gregor Mendel's model for inheritance predicted that all genetic factors should be inherited either always independently or (as in the case of pleiotropic factors) always together, the existence of

"linked" factors that were *sometimes* inherited together "stood out" to researchers as a significant anomaly. This anomaly (i.e., linked factors) provided crucial inspiration for Sturtevant and Morgan to develop their more complex "linear linkage" model of genes on chromosomes (1987, 32). Then, when anomalous recombination frequencies systematically violated predictions of the linear-linkage model (in particular, the prediction that recombination frequencies would be additive), this violation indicated that the model was failing to take account of another causal factor (namely, double crossovers) (1987, 34–35). Finally, the less-than-predicted frequency of double crossovers between factors in close proximity helped to suggest particular hypotheses to Muller and Haldane about the mechanical features (such as rigidity and torsional strength) of the chromosomes upon which the genetic factors were located (1987, 36). Thus, according to Wimsatt, scientific discovery is largely the result of false models' giving rise to anomalous experimental results that stimulate the development of more sophisticated models and theories.

Lindley Darden's book *Theory Change in Science* (1991) provides a more comprehensive account of discovery, but it independently develops a very similar conclusion to that provided by Wimsatt. She argues that the growth of scientific knowledge is largely the product of gradual adjustments to previous theories in response to anomalies. Based on her analysis of the history of Mendelian genetics, she suggests five steps by which researchers change theories in response to anomalies: (1) they confirm the existence of the anomaly; (2) they localize the anomaly; (3) they change the theory by altering (e.g., deleting, generalizing, or complicating) a theoretical component or adding a new component; (4) they assess the new version of the theory (using criteria from the second set of strategies); and (5) (if the preceding four steps fail to resolve the anomaly) they consider whether the anomaly is so significant that it requires abandonment of the entire current theory or whether the anomaly can be set aside for a time (1991, 269).

Neither Wimsatt nor Darden provide much elaboration of the first two steps of this discovery process that Darden describes, however. With regard to the first step (i.e., "confirming" the anomaly), Darden claims:

> Before efforts are made to resolve an anomaly, the correctness of the anomalous data needs to be confirmed. Repeating experiments is one way to verify the existence of an anomaly. . . . The gene case [discussed throughout the book] provided few instances of a purported anomaly being explained away as based on faulty data. Presumably many such instances were never published in the scientific literature; scientists corrected them before publishing. (1991, 271)

Besides this suggestion that scientists can *repeat experiments,* Darden does not offer many further suggestions for the "confirmation" of anomalous data (see e.g., 1991, 271; 1992, 257–258). As she intimates, her case study did not lend itself to an investigation of the reasoning strategies by which scientists confirm that an anomaly is occurring.

Darden provides somewhat more detail concerning her second stage of discovery, the attempt to *localize* the source of an anomaly within one component of a theory (1991, 271). Specifically, she distinguishes *monster* anomalies from *model* anomalies. She claims that monster anomalies are those that can be localized in such a way that they do not pose a problem for *any* component of a particular theory, either because the cause of the anomaly is outside the scope of the theory's domain or because the anomaly is a rare, atypical phenomenon (1992, 258). In contrast, she describes anomalies that cannot be localized outside the domain of the theory as *model* anomalies, which require changes in some theory component. Nevertheless, her analysis of anomaly localization still leaves room for much further elaboration, because her account starts with researchers analyzing the relationship between an anomaly and *one particular theory.* Because some anomalies could be problematic with respect to a number of different theories or auxiliary hypotheses, a complete account of anomaly localization would ultimately need to describe how researchers identify *one particular theory* as problematic. For example, anomalous measurements of neutrino flux in the 1960's could have signaled a problem with theories from a wide variety of disciplines, including nuclear physics, astrophysics, neutrino physics, or radiochemistry (Collins and Pinch 1993).

Wimsatt's paper also fails to examine in detail these first two stages of scientific discovery (i.e., anomaly confirmation and anomaly localization). His goal was not to examine the *initial* steps by which scientists characterize an anomaly. Instead, he focuses on Darden's third step (i.e., theory revision), in which scientists change false models *in response to anomalies that scientists have already "confirmed" and "localized."* Nevertheless, careful attention to Wimsatt's own examples supports the importance of these initial steps. For example, a series of experiments and debates were necessary before the errors associated with the anomaly of non-independent assortment were clarified sufficiently so that Morgan and Sturtevant could postulate their chromosomal linkage model as a response to the anomaly. Initially, researchers had to consider the possibility that their anomalous data were invalid results of questionable experimentation. Thus, they had to "confirm" that the anomaly was a replicable phenomenon before they considered alterations to current theory. Furthermore, even if the anomaly represented a generalizable phenomenon, it could have been either (1) an example of typical patterns of inheritance

(thus indicating that Mendelian genetic theory might have to be abandoned or radically altered), (2) a mysterious phenomenon that would have to be temporarily ignored, or (3) (as Morgan and Sturtevant were able to view it) a specific, localizable error in the former theory. The anomaly served as a valuable instigation for theory alteration only because Morgan and Sturtevant ''localized'' the anomaly as a specific error (namely, overgeneralizing the phenomenon of independent assortment). Thus, the initial stages of characterizing an anomaly (which Darden describes as anomaly ''confirmation'' and ''localization''), although not discussed by Wimsatt and not fully explored by Darden, appear to play an important preliminary role in the discovery process.

This paper argues that Deborah Mayo's study of experimental error in *Error and the Growth of Experimental Knowledge* (1996) provides conceptual resources for extending Wimsatt's and Darden's preliminary descriptions of discovery's early stages. In particular, a central thesis of her book is that scientific experimentation is designed to ''probe'' for potential errors in scientific hypotheses. For example, she identifies four standard, or ''canonical,'' errors that pervade most areas of scientific practice: (1) mistaking chance effects or spurious correlations for genuine regularities, (2) mistakes about the quantity or value of a parameter, (3) mistakes about a causal factor, and (4) mistakes about experimental assumptions (1996, 449). She claims that scientific experimentation is designed to provide ''severe tests'' that identify these canonical errors (and other errors that are specific to particular domains of inquiry) when they are present.

With regard to anomalies in particular, Mayo claims that experimentation serves as a probe for the specific error that is responsible for anomalous results (1996, 147–148). This sort of probing is important because, as Mayo emphasizes, an anomalous result can signal not only a problem with a high-level theory but also a wide variety of more mundane difficulties. For example, because experimental data are often the result of extensive interpretation, the *anomalous data themselves* might be the result of experimental error (Mayo 1996, 128 ff.; see also Ackermann 1985; Hacking 1988). In other words, anomalous data might turn out to be the artifact of particular experimental procedures, or they might be the result of misguided interpretations of low-level experimental results. Therefore, researchers who encounter an anomalous result may not be sure precisely what error is responsible for the anomaly; it could be a fluke result, an experimental error, an instance of flawed auxiliary hypotheses, or an error in one or more current reigning theories. In order to isolate the specific locus of difficulty, Mayo advocates ''error probing,'' including the design of separate severe tests for the multiple auxiliary hypotheses associated with anomalous experiments (see also Allchin 2001).

Mayo's notion of error probing arguably facilitates a richer description of what occurs in the early stages of discovery (i.e., anomaly "confirmation" and anomaly "localization"). Darden characterized anomaly "confirmation" very briefly, claiming that it consists in showing anomalous data to be correct. Based on Mayo's "error-based" account of scientific practice, one might further characterize anomaly confirmation as the process of eliminating a variety of errors. These errors might include: (1) accepting an experimental artifact as a genuine phenomenon, (2) making an experimental mistake (e.g., with instrumentation or calculations), (3) accepting a fluke result as a generalizable phenomenon, (4) making an unwarranted statistical interpretation of experimental data, and (5) producing experimental data through fraudulent techniques (see, e.g., Star and Gerson 1986).

Similarly, one might develop Darden's description of anomaly localization more extensively by characterizing it as the process of probing for the locus of error in one or more experiments. Darden described anomaly localization as the process of either shifting an anomaly outside a primary theory's domain or identifying a particular theory component as problematic. Based on the notion of error probing, one might include the following activities as elements of anomaly localization: (1) identifying particular theory components (as opposed to other components of a theory) as either *plausible* or *unlikely* loci of error, (2) identifying particular theories (as opposed to other theories) as either *plausible* or *unlikely* loci of error, and (3) identifying particular auxiliary hypotheses as either *plausible* or *unlikely* loci of error. Describing anomaly characterization and localization in this way (i.e., as processes of eliminating error and probing for the locus of error) does not conflict with Darden's earlier descriptions of these processes. It merely provides added detail concerning what they involve, and it highlights studies of error as a source of insights concerning the confirmation and localization processes. Thus, by exploring in greater detail both the *sorts* of errors that may be associated with scientific anomalies and the *process* of probing these anomalies for error, one can plausibly enrich previous accounts of scientific discovery's early stages. The rest of this paper illustrates this error-probing process in a case study that is of great significance for contemporary biological science and public policy.

3. Chemical Hormesis: A Case Study of Anomaly and Error. This third section describes the error-probing process in the case of a contemporary biological anomaly called "chemical hormesis." For over 50 years, toxicologists in the United States have been attempting to establish standards for safe human exposure to hazardous substances. For most

substances, scientists have assumed the existence of threshold doses. In other words, they claim that some dose level exists below which a substance ceases to be toxic. This assumption fits well with the notion that organisms have adaptive capabilities that enable them to respond effectively to low levels of environmental stressors. According to some models of carcinogenesis, however, even one molecule of a carcinogen increases cancer risk. If such models were correct, this would imply that *no* threshold dose level exists, at least for carcinogens. Therefore, it is not intuitively clear whether scientists should predict the effects of toxins using models with a threshold or models without a threshold. Unfortunately, it is not easy to resolve *experimentally* this issue of the low-dose effects of toxins, because low-dose effects are usually quite small, they are difficult to measure with statistical significance, and they require large sample sizes. At present, federal agencies such as the Environmental Protection Agency (EPA), the Occupational Safety and Health Administration (OSHA), and the Food and Drug Administration (FDA) employ *threshold* models for estimating the low-dose effects of most toxins, but they employ *linear, no-threshold* models for carcinogens (National Research Council 1994, 31).

Because of the difficulties involved in measuring low-dose chemical effects, the extrapolation of the effects of toxic and carcinogenic chemicals at high doses down to their effects at very low doses continues to be a matter of dispute between the medical, environmental, industrial, and military communities. On the one hand, recent research has associated very low doses of some chemicals, especially chemicals that ''mimic'' hormones such as estrogen, with phenomena such as reproductive cancer, low sperm counts in male organisms, alteration of immune function, and decline in species populations. (See e.g., Birnbaum 1994; Colborn, Dumanoski, and Myers 1996; Krimsky 2000, NRC 2000.) This supports the notion that some toxic chemicals may have *extremely low* thresholds, or *no thresholds* at all. On the other hand, Edward Calabrese, an influential toxicologist working at the University of Massachusetts, Amherst, claims to have found extensive evidence for a phenomenon called ''hormesis,'' in which very low doses of toxic chemicals or radiation may produce an effect that is the *opposite* of the effect that the chemicals or the radiation produce at higher doses (Calabrese and Baldwin 1997, 1998, 1999, 2003).[2] This phenomenon results in a U-shaped dose-response

2. Calabrese is the chairman of the Biological Effects of Low Level Exposures advisory board, a group of scientists organized to develop a better understanding of biological responses to low doses of chemical and physical agents. He is editor of the journal *Biological Effects of Low Level Exposures,* he has organized several conferences related to the hypothesis of hormesis, and he recently published a summary of the hormesis hypothesis in the journal *Nature* (Calabrese and Baldwin 2003).

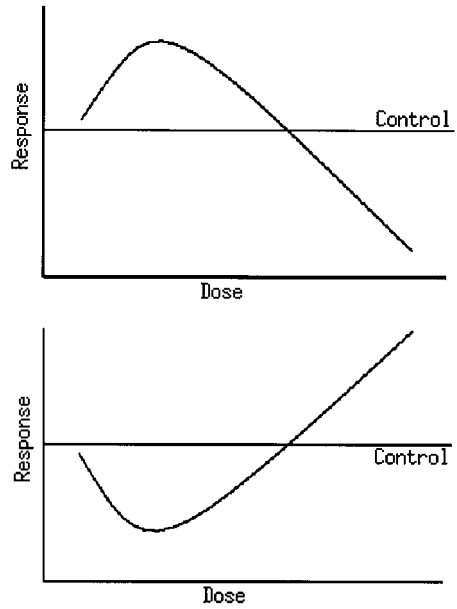

Figure 1. Examples of the general form of hormetic dose-response relationships. The bottom curve could represent the relationship between alcohol intake and human mortality, whereas the top curve could represent a hormetic relationship between dose of a growth inhibitor and plant growth.)

curve that displays a threshold below which a toxic chemical has either *no* effects or perhaps even *beneficial* effects (see Figure 1).

The dose-response curve for hormetic chemicals is U-shaped, because their biological effects *equal* control levels at dosages close to zero, their effects drop *below* control levels as their dosages increase, and then their effects appear to reverse and rise *above* control levels as the dosages increase further. These U-shaped curves can also be inverted, thus representing low-dose effects that rise above the level of controls and high-dose effects that reverse and drop below controls (see the top curve in Figure 1). A familiar example of this sort of dose-response is the effect of alcohol on human mortality. Low levels of alcohol consumption *decrease* human mortality rates *below* control levels. Nevertheless, high levels of alcohol consumption *increase* human mortality rates *above* control levels. Many other examples of chemical hormesis are more counter-intuitive than the case of alcohol. For example, some evidence exists that plant growth inhibitors may actually increase plant growth when the inhibiting chemicals are present at extremely low levels (Calabrese 1999). Similarly, very low levels of insecticides that are supposed to *inhibit* reproduction may actually *stimulate* reproduction (Calabrese and Baldwin 1998)!

Calabrese and Baldwin recently completed two of the most thorough examinations of the evidence for chemical hormesis to date (1997, 2001). They carried out extensive literature searches with quantitative methodologies designed to uncover evidence for chemical hormesis in previous toxicology studies. After finding numerous studies containing U-shaped dose-response curves, they concluded that chemical hormesis is a widely generalizable phenomenon and that "the concept of hormesis (i.e., low-dose stimulation/high-dose inhibition) is counter to the cancer risk assessment practices by U.S. regulatory agencies" (1998, 4 and VIII-1).[3] They note that government risk assessors use linear, no-threshold models that predict *harmful* effects from carcinogens at *all* dose levels, whereas Calabrese and Baldwin's data suggests that carcinogens may be likely to produce small *beneficial* effects at *very low* dose levels. In other words, Calabrese and Baldwin considered the phenomenon of chemical hormesis to be a serious *anomaly* that runs counter to the linear, no-threshold scientific models currently used for carcinogen risk assessment by U.S. governmental agencies.

The hormesis case is ideal for examining the reasoning by which scientists probe the potential errors related to anomalous data, because researchers have pointed out a variety of specific errors that might have given rise to the anomalous relationship between Calabrese and Baldwin's data and current carcinogen dose-response models. (See, e.g., Davis and Svendsgaard 1990; Davis and Farland 1998; Elliot 2000a, 2000b; Menzie 2001.) These errors can be grouped under at least three categories: (1) errors associated with the anomalous data itself, (2) errors associated with Calabrese and Baldwin's auxiliary hypotheses, and (3) errors associated with current carcinogen dose-response models. The elimination of errors in category (1) arguably constitutes what Darden called anomaly "confirmation." In other words, scientists "confirm" the hormesis anomaly by showing that the anomalous data appear to be justifiable. By probing for errors in categories (2) and (3), researchers are "localizing" the anomaly (i.e., identifying the particular auxiliary hypotheses, theories, or theory components that appear to be in error).

First, the anomalous data that Calabrese and Baldwin obtained in their literature search might themselves be in error. For example, the U-shaped dose-response curves that Calabrese and Baldwin have been studying might be statistical artifacts; if toxicologists perform a very large number of experiments, random statistical fluctuations in the data will result in a few studies with U-shaped dose-response curves, *even if all the dose-*

3. For further support of the hormesis hypothesis, see Gerber, Williams, and Gray (1999), who claim, "There can be no doubt about the reality of hormesis" (1999, 278).

response relationships explored by the experiments are actually linear (Davis and Svendsgaard 1990; Crump 2001). Or, these U-shaped dose-response curves might be produced by a wide variety of confounding factors, including interactive effects of different chemicals, multiple effects of the same chemical on different endpoints, adaptive or compensatory mechanisms, and atypical responses to unique laboratory conditions (Davis and Svendsgaard 1990, 75–77). U-shaped dose-response curves produced by these confounding factors would be experimental artifacts or, at best, the effects of unique conditions that could not be easily generalized to other circumstances. Finally, Elliot (2000b) has suggested that hormesis researchers are employing a confused array of different concepts of chemical hormesis. Therefore, there is a danger that researchers such as Calabrese and Baldwin will inadvertently make the error of concluding that experimental evidence supports the existence of phenomena associated with *many* or *all* concepts of chemical hormesis, even if experimental evidence supports only the existence of phenomena associated with *one particular concept* of chemical hormesis, if any.

Second, even if the data that Calabrese and Baldwin have been observing may be correct, they may be assuming erroneous auxiliary hypotheses that falsely make this data *appear* to conflict with the linear, no-threshold models used by government risk assessors. For example, Calabrese and Baldwin have not observed any studies in which a carcinogenic chemical produces hormetic effects on an endpoint associated with the *entire process* of carcinogenesis, such as cancer-related mortality. Instead, they have observed studies in which carcinogenic chemicals produce hormetic effects on endpoints associated with *some aspect* of carcinogenesis, such as DNA repair enzyme activity or cell division. Therefore, they must assume the auxiliary hypothesis that hormetic effects on endpoints associated with *some aspect* of carcinogenesis will result in hormetic effects on endpoints associated with the *entire process* of carcinogenesis. Calabrese and Baldwin must also accept the auxiliary hypotheses that hormetic effects observed during short-term studies will remain over extended periods of time, that hormetic effects are not eliminated by the interactive effects of multiple toxic chemicals, that a dose range exists at which both sensitive and non-sensitive members of the population experience hormetic effects, and that humans are not already exposed to dose levels that surpass the chemicals' hormetic dose ranges (Davis and Farland 1998, 380; Elliott 2000a, 185–192). If any of these auxiliary hypotheses are false, then government risk assessors may be justified in using no-threshold dose-response models even if hormetic effects occur under some circumstances. Therefore, the anomalous relationship that Calabrese and Baldwin claim to exist between their hormetic

data and the models used by U.S. government risk assessors may indicate merely that Calabrese and Baldwin are employing erroneous auxiliary hypotheses.

Finally, the anomalous data uncovered by Calabrese and Baldwin may actually indicate that the no-threshold models used by government risk assessors are in error. Even in this case, however, the linear, no-threshold models might be afflicted by several different sorts of error, just as Wimsatt (1987) points out that models can be false in a variety of ways. The no-threshold models might fail under some conditions (e.g., on particular endpoints or particular organisms or particular environmental conditions) but provide accurate descriptions of cancer risk under other conditions. Or, the models might simply fail to take account of some important variables. For example, no-threshold models might provide accurate predictions of cancer risk over *long* periods of time but provide inaccurate predictions over *short* periods of time.

Because the hormesis case study illustrates such a wide variety of errors that might be contributing to the hormesis anomaly, it supports the claim in Section 2 that the initial stages of scientific discovery can be fruitfully described as error-probing processes. ''Confirmation'' of the anomalous hormesis data requires the elimination of numerous potential errors (e.g., statistical errors, experimental errors, and poorly described experimental conditions) that might produce hormetic results as experimental artifacts. Furthermore, ''localization'' of the anomaly within a particular theory requires the identification and elimination of still other potential errors in order to elucidate the precise theory that is at fault. For example, Calabrese and Baldwin can ''localize'' the hormesis anomaly as a problem for theories of carcinogenesis only by showing that it is not erroneous to accept particular auxiliary hypotheses (e.g., that hormetic effects on carcinogenic *endpoints* provide evidence for hormetic effects on the entire *process* of carcinogenesis). Thus, only after the hormesis anomaly has been extensively characterized (by probing for potential errors) are scientists ready to begin altering former theories in response to it. Section 4 argues that this description of discovery's early stages as error-probing processes yields a better understanding of the entire course of scientific discovery.

4. Three Lessons for Scientific Discovery. This section argues, based on the hormesis case, that analyzing scientific discovery's earliest stages as error-probing processes yields a deeper understanding of the entire course of discovery. First, it provides a better understanding of the *challenges* that researchers face during the early stages of discovery. Second, it offers a richer understanding of the *reasoning strategies* that one can employ during those early stages. Third, it provides novel insights concerning the ways in which those early stages can *influence* the ultimate course of discovery.

4.1. The Challenges of Error Probing. First, the hormesis case illustrates the difficulties that scientists face as they probe anomalies for error during the early stages of scientific discovery. For example, Darden and Mayo have both emphasized performing *further experimental tests* as a reasoning strategy for these early stages. Darden suggests that researchers should repeat experiments in order to gain a better understanding of anomalous results (1991, 271), and Mayo argues that experimentalists should use error-statistical tests to evaluate the auxiliary hypotheses associated with anomalous data (1996, 147–148). The hormesis case illustrates at least three problems with this strategy of performing experimental tests in order to probe for the errors associated with anomalous results. These problems may not always be insuperable for the strategy of experimental testing, but they do limit its effectiveness.

The first problem is that even though auxiliary hypotheses play a crucial role in the experiments that yield anomalous data, *many of these hypotheses are very difficult to evaluate using experimental tests.* (See, e.g., Laudan 1997; Carrier 2001.) Therefore, the repetition of an experiment may help researchers to conclude that they did not make a one-time experimental error, but this repetition may not help them to determine whether all the auxiliary hypotheses associated with their experiment are legitimate. Mayo suggests that experimentalists can subject these auxiliary hypotheses to separate severe tests, but the hormesis case illustrates how difficult it can be to perform those experiments. To take one example, researchers such as Calabrese and Baldwin need to evaluate the auxiliary hypothesis that the U-shaped dose-response curves associated with particular experiments are not the result of unknown conditions that are unique to those particular experimental setups. Unfortunately, researchers may initially have very little idea what conditions might be involved in the production of a U-shaped dose-response curve, which makes this auxiliary hypothesis very difficult to evaluate. For example, some research suggests that hormetic effects are more likely to occur when organisms are exposed to adverse environmental conditions (Calabrese and Baldwin 1998, IV-3-4; Vichi and Tritton 1989). Similarly, hormetic effects may be much more likely to occur on some endpoints as opposed to other endpoints. Stebbling (1982, 1995) suggests that the endpoint of *growth* in particular may be governed by a number of feedback loops that are likely to produce overcompensation for the presence of small amounts of substances that normally inhibit growth. Therefore, researchers might find that a U-shaped, no-threshold dose-response curve consistently occurs on the endpoint of growth, but it might occur much more sporadically, if at all, on an endpoint such as longevity. Still other conditions to consider in the case of chemical hormesis include the organism type

(e.g., plant, fungus, bacteria, mammal, human), the chemical type (e.g., metal salts, growth inhibitors, carcinogens), the length of exposure to the chemical (e.g., a short period of time vs. an extended period of time), and exposure to combinations of toxic chemicals. Scientists might not even consider many of the other experimental conditions that could affect the production of the hormetic effect. Therefore, researchers must develop some understanding of the relevant variables that affect the production of U-shaped dose-response curves before they can design effective experiments to evaluate the auxiliary hypothesis that the U-shaped dose-response curve in a particular experiment is not the result of unique experimental conditions.

The second problem with experimental testing as an error-probing strategy is that researchers may not have a clear conception of what phenomenon they are trying to test. Elliott (2000b) claims that hormesis researchers employ at least seven distinct concepts of chemical hormesis, so an adequate test for one concept of hormesis may not be an adequate test for a different concept. For example, the hormesis dispute has been plagued by questions about whether the phenomenon of "hormesis" should be characterized so that it includes any case in which low-dose effects are the opposite of high-dose effects, whether it should include only cases of low-dose *stimulation* and high-dose *inhibition,* or whether it must be associated with a particular mechanism, such as overcompensation. If researchers cannot agree about precisely what chemical hormesis is, it is difficult for them to design an adequate severe test of the hypothesis that hormesis occurs in a regular fashion.

Yet a third problem is practical. Because the positive hormetic effects produced at low-dose levels are often quite small (130%-160% of the control), it is very difficult to construct a severe test for the hormesis hypothesis that also has sufficient statistical power to be likely to confirm the presence of hormetic effects. Therefore, if one designs a test with a small sample size, hormesis might occur, but the small effects characteristic of chemical hormesis would not be sufficiently different from controls to be statistically significant (and thus to pass the test). The upshot is that researchers must design a test with a very large number of organisms in order to be likely to confirm the existence of chemical hormesis. Because such tests are costly, researchers cannot obtain funding to perform them unless they can provide preliminary evidence that hormesis is likely to occur. The combination of these three problems makes it difficult for researchers to settle the hormesis issue using experimental tests alone. They cannot easily obtain funding for experimental tests until they have a sufficiently clear conception of the phenomenon that they are trying to test and until they know the conditions under which that phenomenon is likely to occur. Unfortunately, this is exactly the sort of information that

researchers who study anomalies may not initially have and that they are often trying to obtain by performing the experimental tests (see also Ashford and Miller 1998; Krimsky 2000).

4.2. Strategies for Probing Error. The primary value of the hormesis case study is not, however, the fact that it points out difficulties in others' strategies for probing the errors associated with anomaly. This case study is particularly valuable because it suggests other strategies that are useful for probing error, despite the difficulties just mentioned. Although these strategies do appear to *describe* the approaches actually taken by researchers in the hormesis case, and although they seem to have some *normative* force as good approaches for other researchers to follow, the strategies are probably most convincing when they are interpreted as *hypothetical* rather than as normative or descriptive. Following Darden, they are "hypotheses about strategies that could have produced the historical changes that did occur" or that may be producing the changes that are presently occurring (1991, 15). Consider the following seven strategies: (a) reexamining previous studies of the anomalous phenomenon; (b) looking for new variables that might be associated with the anomalous results; (c) suggesting plausible mechanisms that might produce the anomalous results; (d) relating the anomalous results to established theories; (e) explaining why earlier experiments failed to produce evidence for the anomalous phenomenon; (f) looking for evidence of the anomalous phenomenon under diverse conditions;[4] and (g) using informal error-statistical arguments in support of the anomalous phenomenon or the auxiliary hypotheses used to test the phenomenon.

First, the hormesis case study illustrates the value of strategy (a), which involves reexamining previous experimental studies of an anomalous phenomenon. This strategy is, admittedly, closely related to Mayo's and Darden's strategy of performing further experimental studies. Nevertheless, the strength of strategy (a) is that it involves examining *multiple* previous studies, whereas one or two new studies might not provide adequate information to probe successfully for error. Although Calabrese and Baldwin have not performed further experiments of their own, their fundamental approach to confirming the hormesis hypothesis involved the collection and analysis of hundreds of previous studies. In many of these previous experiments, researchers produced statistically significant evidence that low-level exposure of an organism to a chemical produced the

4. The conditions under which it is reasonable to look for an anomaly may be more or less diverse, depending on how general a claim the anomaly challenges. Nevertheless, strategy (f) advocates looking for evidence of the anomalous phenomenon under as broad a range of conditions as one could reasonably expect to find it.

opposite of the effects produced by the chemical at higher dose levels (Calabrese and Baldwin 1998). Most of the scientists who performed these studies placed little significance on their results, perhaps because they assumed that some experimental errors were vitiating their results or causing them to obtain data that would not be generalizable to many other cases. However, Calabrese and Baldwin's identification of a *large number* of such studies suggests that the theoretical models that conflict with these results may *themselves* be in error rather than the experiments that produced the results. These data have enabled Calabrese and Baldwin to support the anomalous hormesis hypothesis, despite the contrary scientific consensus against the occurrence of opposite effects of toxins at low doses.

An example of strategy (b) (i.e., adding new variables) is that hormesis researchers have argued that hormetic effects may make more sense when they are examined *over time*. Starting in the early 1980's, A. R. D. Stebbing (1982, 1998) suggested that the hormesis anomaly might reveal that previous researchers made the error of underestimating the importance of *time* as a variable in dose-response relationships. He suggested that chemical hormesis might involve feedback processes that result in *temporary overcompensation* for environmental stressors. In other words, Stebbing suggested that, after a toxin is administered, a linear or threshold model is likely to capture the relationship between the dose of the toxin and its effect. He argued, however, that organisms' defense mechanisms might later "kick in" and *overcompensate* for the stressor (at low doses), thus gradually changing the linear curve to a U-shaped dose-response curve for a period of time. Finally, he suggested that the overcompensation response might wear off after an extended period of time, thus returning the U-shaped dose-response curve to a linear or threshold curve (see Figure 2).

Stebbing's suggested time-dependence of hormetic effects is particularly significant, because it suggests a plausible set of *mechanisms* (namely, adaptive processes of overcompensation) that could explain hormetic phenomena. This is an example of strategy (c). Calabrese and Baldwin have suggested numerous other mechanisms that might also result in hormetic effects, including the removal or inactivation of enzyme inhibitors, the simultaneous actions of inhibitors as substrates in other reactions, dissociation of enzymes into active subunits, and alteration in cell permeability (1998, VII-3-7). Furthermore, in accord with strategy (d), Calabrese and Baldwin have attempted to associate hormetic effects with the theory of evolution by natural selection. They once again cite the work of Stebbing (1982), who argued that natural selection might have favored organisms that developed generalized adaptive strategies of mildly overcompensating for a variety of stressors (Calabrese and Baldwin

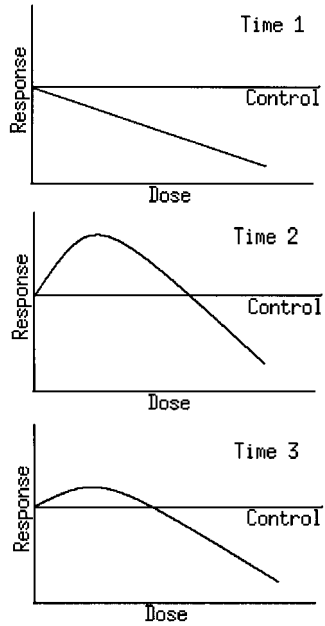

Figure 2. Example of the temporal dependence of the dose-response relationships characteristic of ''overcompensation hormesis.'' Time 1 represents an initial period of inhibition at all dose levels. Time 2 illustrates the organism's overcompensation to the toxin at low doses. At time 3, the overcompensation effect has begun to subside).

1998, VII-15). All these strategies assist in ''confirming'' the anomaly (by supporting the legitimacy of hormesis data) and ''localizing'' the anomaly (by suggesting particular theories that may need to be changed in response to the anomaly).

In accordance with strategy (e) (i.e., explaining why previous experiments failed to produce anomalous results), Calabrese and Baldwin claimed that few previous studies of the effects of toxic chemicals examined the *very low-dose* effects of toxins. Because the toxicologists designing the experiments were primarily concerned to find statistically significant evidence that a toxic chemical produced harmful effects at *high* doses, they did not bother to study low-dose effects (Calabrese and Baldwin 1999). Therefore, Calabrese and Baldwin argued, in effect, that most of these studies would not have been likely to display hormetic effects even if hormesis were a generalizable phenomenon. Their argument contributes to ''confirming'' the hormesis anomaly, because it shows that the lack of evidence for hormesis in the past should not be taken as evidence that hormesis is a fluke result or experimental artifact.

Calabrese and Baldwin also looked for evidence of hormesis under diverse circumstances (strategy f). A point that they emphasized in their initial literature search and in subsequent publications was the diversity of conditions under which hormetic effects have been observed:

> Hormetic responses are observed in numerous species from a broad range of taxonomic groups including microbes, plants, and animals (including humans). . . . Chemicals shown to induce hormetic effects represent a broad range of chemical classes: the most studied agents were metals, followed by alcohols, antibiotics, auxin related compounds, and numerous biocidal agents. . . . The types of hormetic responses indicate that the principal endpoint studied has been growth, followed by metabolic/physiological changes (e.g., enzyme activity), longevity, and various reproductive endpoints. (Calabrese and Baldwin 1998, 2)

This strategy makes perfect sense from an "error-probing" point of view. In toxicological experiments, a U-shaped dose-response curve might occasionally show up as an experimental artifact or as a fluke result. However, it is very *unlikely* that an experimental artifact or fluke result would show up in many different experiments, across a broad range of taxonomic groups, with a broad array of chemicals, on a number of biological endpoints. Thus, this example of strategy (f) is also an example of strategy (g); Calabrese and Baldwin are making an *informal* argument that their literature search "confirms" the hormesis anomaly. They claim that it would be exceedingly unlikely that toxicologists would have observed U-shaped dose-response curves under so many conditions if hormesis were not a generalizable phenomenon.

4.3. How Error Probing Influences the Course of Discovery. Another significant lesson to be drawn from the hormesis case is that the process of error probing in discovery's early stages can impact the ultimate course of scientific discovery in a complex variety of ways. As a preliminary attempt to begin understanding these impacts, this section identifies three specific sorts of influence (to be called "persistence," "identification," and "strategy" influences) by which error probing seems to be influencing the course of discovery in the hormesis case. By identifying these three impacts, the section extends the earlier work of Wimsatt and Darden by clarifying the role of the early stages of scientific discovery in the overall discovery process. As discussed in Section 2 of this paper, Wimsatt did not explore the influences of the early stages of scientific discovery (i.e., the stages of anomaly "confirmation" and "localization"). Darden claimed that the early stages of scientific discovery affect the later stages in at least two ways: (1) researchers do not alter former theories until an anomaly is "confirmed" and "localized," and (2)

anomaly localization identifies particular components within a theory as candidates for alteration. The persistence, identification, and strategy influences constitute three further ways in which the early stages of scientific discovery influence the later ones. The persistence influence is a further impact associated with the stage that Darden described as anomaly confirmation, whereas the identification and strategy influences are further aspects of anomaly localization. In order to illustrate these influences in the hormesis case, let us consider the time-scale error that Stebbing proposed (i.e., that previous dose-response models erroneously assumed that toxic effects do not change significantly over time).

The first way in which researchers' probing for this "time-scale error" seems to have influenced scientific discovery might be called a "persistence" influence. This influence is that the plausibility of the time-scale error seems to have made researchers more likely to "persist" in exploring ways to alter former models or theories in response to the hormesis anomaly. Throughout the early decades of the twentieth century, scientists frequently dismissed hormesis, claiming that it was probably the result of errors associated with researchers' experimental results (Calabrese and Baldwin 2000a, 2000b). Thus, the hormesis anomaly did not encourage them to explore alterations to their former theories. In contrast, Stebbing's proposed time-scale error is associated with former dose-response *models* rather than with experimental *results.* By proposing a way to explain the anomaly in terms of an error that is not related to researchers' experimental results, Stebbing has encouraged other researchers to take hormetic results seriously. Therefore, they are now more likely to explore ways to alter former theories or models. In sum, by shifting researchers' focus to a new locus of error (namely, dose-response models and physiological theories of compensation to toxins), Stebbing has affected the discovery process in the hormesis case by contributing to a much greater *persistence* in scientists' efforts to alter former models and theories in response to the hormesis anomaly.

The second, "identification," influence of the time-scale error is that it has contributed to *identifying* plausible theories for researchers to modify in response to the hormesis anomaly. Scientists are currently unsure to which theory the hormesis anomaly relates. Thus, the hormesis anomaly provides further evidence for the claim in Section 2 that Darden's account of anomaly localization could be extended to consider how researchers isolate *one particular theory* to which an anomaly relates. For example, hormesis might not be a problem for any general theory, if it is merely the result of experimental errors. Alternatively, it could reflect an error associated with physiological theories about how organisms compensate for exposure to toxins and carcinogens. Or, the anomaly could indicate that toxicologists have erroneously underestimated the diversity of physiological effects

produced by toxins. Because of this variety of theories that could be in error, the present paper has treated hormesis as an anomaly not for a particular *theory* but rather for the dose-response *models* that predict harmful or neutral effects from toxins and carcinogens at low doses. (Because these models are based on a wide variety of theories from physiology, carcinogenesis, and experimental design, the paper has been able to remain agnostic about the particular theory that might ultimately need to be changed in response to the hormesis anomaly.)[5] The time-scale error contributes to scientific discovery by *identifying* physiological theories about organisms' responses to toxins as the ''culprits'' for the hormesis anomaly. In other words, if the time-scale error is correct, current physiological theories have underestimated the frequency and extent to which organisms overcompensate for environmental stressors. Thus, Stebbing's proposal of the time-scale error directs the future course of discovery by providing researchers with plausible hypotheses for altering particular theories in response to the anomaly.

Finally, the time-scale error in the hormesis case provides a third, ''projects,'' influence on scientific discovery. In other words, Stebbing's proposal of that particular error provides researchers with *research projects* for gaining further insights about the hormesis phenomenon. In response to the suggestion that hormesis might be the result of a time-scale error, Calabrese established a research project of examining previous biological studies that measured toxic effects both at low doses *and at multiple points of time* (Calabrese 2001). Based on the study, he hoped to determine whether Stebbing's proposed error was a plausible explanation for most instances of the hormesis anomaly. Although Calabrese found some evidence that hormesis is an overcompensation phenomenon, he also found evidence that the toxins in other studies produced hormetic effects by affecting different receptor subtypes at different concentrations. Furthermore, he acknowledged that the mechanism of affecting different receptor subtypes might sometimes produce hormetic effects *immediately* (thus reflecting a different error than the time-scale error).[6] As a result,

5. In the present paper, dose-response models are distinguished from theories on the basis of the models' lack of explanatory detail. In other words, toxicologists' dose-response models (such as the linear-no-threshold model described in Section 3) provide rough predictions of toxins' effects while abstracting from the theoretical details of the mechanisms by which the chemicals produce those effects.Therefore, researchers could conclude that a dose-response model is probably erroneous while being quite unsure about the specific mechanisms (and thus the theories) that are problematic.

6. The *immediacy* of hormetic effects is a somewhat relative notion. Although there is bound to be some ''lag time'' between the administration of a toxin and the occurrence of its effects, Calabrese found that the stimulatory effects in some studies could be labelled ''immediate'' in the sense that they did not follow an initial period of inhibition produced by the toxin.

Calabrese seems to have altered his characterization of the hormesis anomaly. Whereas he suggested earlier that it might be possible to provide a unified characterization of the anomaly as an overcompensation phenomenon (1999), he has suggested in some of his most recent work that the anomaly may involve different types of mechanisms that produce similar phenomena but that occur in different physiological systems in response to different toxins (Calabrese and Baldwin 2002). Thus, even though the time-scale error may not be the only error associated with the hormesis phenomenon, it suggested a promising research project (i.e., examining the temporal effects of toxins) that has driven the discovery process forward.

5. Conclusions. This paper has argued that scientific discovery's early stages can be described as error-probing processes and that this description yields greater understanding of the entire discovery process. The paper's second section argued that one could provide an improved description of scientific discovery's early stages by characterizing them as processes for probing the specific errors that might be associated with scientific anomalies. The paper then supported this suggestion by showing how researchers are characterizing the contemporary biological anomaly of chemical hormesis by probing for the errors associated with it. Finally, the paper used the case study to show that this description of scientific discovery's early stages elucidates the challenges, research strategies, and influences associated with those stages.

Although the paper has focused on a case study from the biological sciences (i.e., chemical hormesis), it is plausible that its analysis could apply to a range of other scientific disciplines. For example, many of the errors that Section 2 identified as important ones for researchers to investigate in the early stages of the discovery process appear to be at least as relevant in the physical sciences as in the biological sciences. For example, Hacking (1983), Galison (1987), Collins and Pinch (1993), and Mayo (1996) have provided numerous illustrations of how researchers in the physical sciences must eliminate potential errors in their statistical interpretations of data, test experimental instrumentation, eliminate experimental artifacts, and distinguish fluke results from reliable ones. The authors' studies also illustrate many ways in which researchers in the physical sciences "localize" anomalies by probing auxiliary hypotheses for error before they assign blame to primary theories. To provide just one example, both Galison (1987) and Mayo (1996) describe how physicists investigating the anomalous phenomenon of neutral currents developed computer simulation models and performed detailed statistical analyses in order to determine whether auxiliary hypotheses concerning their experimental setup were free from significant errors. Therefore, it is plausible

that philosophers could gain increased understanding of the discovery process throughout a wide range of sciences by examining the role of error probing in the early stages of discovery.

By using recent studies of error to expand and elucidate previous philosophical investigations of scientific discovery, this paper makes at least three contributions to the philosophy of science. First, it extends previous work on error (e.g., Hon 1989; Mayo 1996; Allchin 2001). It clarifies weaknesses of experimental testing as a strategy for probing error, and it identifies a variety of alternative strategies for error probing. Second, it strengthens the literature on scientific discovery. It argues that the early stages of discovery can be described as an extended error-probing process by which researchers characterize anomalies. It shows that researchers can employ a variety of strategies for investigating anomalies during these stages. And, it elucidates ''persistence,'' ''identification,'' and ''project'' influences by which the early stages of discovery impact the rest of the discovery process. Third, the paper contributes to a unification of current science studies by bringing together traditional philosophical work on scientific discovery with the work of the ''new experimentalists.'' Thus, the paper illustrates that the literature on scientific discovery constitutes yet another locus at which recent studies of scientific experimentation can impact the philosophy of science.

REFERENCES

Ackermann, Robert (1985), *Data, Instruments, and Theory: A Dialectical Approach to Understanding Science.* Princeton: Princeton University Press.

Allchin, Douglas (2001), ''Error Types'', *Perspectives on Science* 9: 38–58.

Ashford, Nicholas, and Claudia Miller (1998), *Chemical Exposures: Low Levels and High Stakes,* 2d ed. New York: Wiley.

Bechtel, William, and Robert Richardson (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research.* Princeton, NJ: Princeton University Press.

Birnbaum, Linda (1994), ''Endocrine Effects of Prenatal Exposure to PCBs, Dioxins, and Other Xenobiotics: Implications for Policy and Future Research'', *Environmental Health Perspectives* 102: 676–679.

Calabrese, Edward (1999), ''Evidence that Hormesis Represents an 'Overcompensation' Response to a Disruption in Homeostasis'', *Ecotoxicology and Environmental Safety* 42: 135–137.

——— (2001), ''Overcompensation Stimulation: A Mechanism for Hormetic Effects'', *Critical Reviews in Toxicology* 31: 425–470.

Calabrese, Edward and Linda Baldwin (1997), ''The Dose Determines the Stimulation (and Poison): Development of a Chemical Hormesis Database'', *International Journal of Toxicology* 16: 545–559.

——— (1998), *Chemical Hormesis: Scientific Foundations.* College Station, TX: Institute for the Advancement of Chemical Technology.

——— (1999), ''Reevaluation of the Fundamental Dose-Response Relationship'', *BioScience* 49: 725–732.

——— (2000a), ''The Marginalization of Hormesis'', *Human and Experimental Toxicology* 19: 32–40.

——— (2000b), ''Tales of Two Similar Hypotheses: The Rise and Fall of Chemical and Radiation Hormesis'', *Human and Experimental Toxicology* 19: 85–97.
——— (2001), ''The Frequency of U-Shaped Dose Responses in the Toxicological Literature'', *Toxicological Sciences* 62: 330–338.
——— (2002), ''Defining Hormesis'', *Human and Experimental Toxicology* 21: 91–97.
——— (2003), ''Toxicology Rethinks Its Central Belief'', *Nature* 421: 691–692.
Carrier, Martin (2001), ''Critical Notice: Error and the Growth of Experimental Knowledge'', *International Studies in the Philosophy of Science* 15: 93–98.
Colborn, Theo, Diane Dumanoski, and J. P. Myers (1996), *Our Stolen Future.* New York: Dutton.
Collins, Harry M., and Trevor Pinch (1993), *The Golem.* Cambridge: Cambridge University Press.
Crump, Kenny S. (2001), ''Evaluating the Evidence for Hormesis: A Statistical Perspective'', *Critical Reviews in Toxicology* 31: 669–679.
Darden, Lindley (1991), *Theory Change in Science.* New York: Oxford University Press.
——— (1992), ''Strategies for Anomaly Resolution'', in Ronald Giere (ed.), *Cognitive Models of Science,* Minnesota Studies in the Philosophy of Science, vol. 15. Minneapolis: University of Minnesota Press, pp. 251–273.
Davis, J. Michael, and William Farland (1998), ''Biological Effects of Low-Level Exposures: A Perspective from U.S. EPA Scientists'', *Environmental Health Perspectives* 106:380.
Davis, J. Michael, and David Svendsgaard (1990), ''U-Shaped Dose-Response Curves: Their Occurrence and Implications for Risk Assessment'', *Journal of Toxicology and Environmental Health* 30:71–83.
Elliott, Kevin (2000a), ''A Case for Caution: An Evaluation of Calabrese and Baldwin's Studies of Chemical Hormesis'', *Risk: Health, Safety, and Environment* 11:177–196.
——— (2000b), ''Conceptual Clarification and Policy-Related Science: The Case of Chemical Hormesis'', *Perspectives on Science* 8: 346–366.
Franklin, Allan (1986), *The Neglect of Experiment.* Cambridge: Cambridge University Press.
——— (1997), ''Calibration'', *Perspectives on Science* 5: 31–80.
Galison, Peter (1987), *How Experiments End.* Chicago: University of Chicago Press.
Gerber, Linda M., George C. Williams, and Sandra J. Gray (1999), ''The Nutrient-Toxin Dosage Continuum in Human Evolution and Modern Health'', *The Quarterly Review of Biology* 74: 273–289.
Hacking, Ian (1983), *Representing and Intervening.* Cambridge: Cambridge University Press.
——— (1988), ''On the Stability of the Laboratory Sciences'', *Journal of Philosophy* 85: 507–514.
Hanson, Norwood Russell (1958), *Patterns of Discovery.* Cambridge: Cambridge University Press.
——— (1961), ''Is There a Logic of Scientific Discovery?'' in Herbert Feigl and Grover Maxwell (eds.), *Current Issues in the Philosophy of Science.* New York: Holt, Rinehart and Winston, 21–42.
Hon, Giora (1989), ''Towards a Typology of Experimental Errors: An Epistemological View'', *Studies in History and Philosophy of Science* 20: 469–504.
Krimsky, Sheldon (2000), *Hormonal Chaos: The Scientific and Social Origins of the Environmental Endocrine Hypothesis.* Baltimore: Johns Hopkins University Press.
Kuhn, Thomas (1970), *The Structure of Scientific Revolutions,* 2d ed. Chicago: University of Chicago Press.
Lakatos, Imre (1970), ''Falsification and the Methodology of Scientific Research Programmes'', in Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Scientific Knowledge.* Cambridge: Cambridge University Press.
Laudan, Larry (1977), *Progress and Its Problems.* Berkeley: University of California Press.
——— (1997), ''How About Bust? Factoring Explanatory Power Back into Theory Evaluation'', *Philosophy of Science* 64: 306–316.
Mayo, Deborah (1996), *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.

Menzie, C. (2001), "Hormesis in Ecological Risk Assessment: A Useful Concept, a Confusing Term, and/or a Distraction?", *Human and Experimental Toxicology* 20: 521–523.

National Research Council (1994), *Science and Judgment in Risk Assessment.* Washington, DC: National Academy Press.

——— (2000), *Hormonally Active Agents in the Environment.* Washington, DC: National Academy Press.

Nickles, Thomas (1996), "Deflationary Methodology and Rationality of Science", *Philosophica* 58: 9–50.

——— (1997), "A Multi-Pass Conception of Scientific Inquiry", *Danish Yearbook of Philosophy* 32: 11–44.

Popper, Karl (1959), *The Logic of Scientific Discovery.* New York: Basic Books.

Schaffner, Kenneth (1993), *Discovery and Explanation in Biology and Medicine.* Chicago: University of Chicago Press.

Shapere, Dudley (1977), "Scientific Theories and Their Domains", in Frederick Suppe (ed.), *The Structure of Scientific Theories,* 2d ed. Urbana: University of Illinois Press.

Star, Susan Leigh, and Elihu M. Gerson (1986), "The Management and Dynamics of Anomalies in Scientific Work", *The Sociological Quarterly* 28: 147–169.

Stebbing, A. R. D. (1982), "Hormesis—The Stimulation of Growth by Low Levels of Inhibitors", *The Science of the Total Environment* 22:213

——— (1998), "A Theory for Growth Hormesis", *Mutation Research* 403: 249–258.

Vichi, P., and T. Tritton (1989), "Stimulation of Growth in Human and Murine Cells by Adriamycin", *Cancer Research* 49: 2679–2682.

Wimsatt, William (1987), "False Models as Means to Truer Theories", in Matthew Nitecki and Antoni Hoffman (eds.), *Neutral Models in Biology.* New York: Oxford University Press, 23–55.