The Structure and Function of Experimental Control in the Life Sciences

Jutta Schickore*†

This article presents a new framework for the analysis of experimental control. The framework highlights different functions for experimental controls in the realization of an experiment: experimental controls that serve as tests and experimental controls that serve as probes. The approach to experimental control proposed here can illuminate the constitutive role of controls in knowledge production, and it sheds new light on the notion of exploratory experimentation. It also clarifies what can and what cannot be expected from reviewers of scientific journal articles giving feedback on experimental controls.

Introduction. Control experiments are a key element of study design in the life sciences.[1] As a recent editorial states: "Experimental design is the practice of planning a study to reach specified objectives. . . . Formal experimental design focuses on the *designation of appropriate controls*, random assignment of subjects that receive an experimental treatment, replication, and blocking to remove the potential effect of confounding variables" (*Nature Method* 2017,

Received February 2018; revised April 2018.

*To contact the author, please write to: Department of History and Philosophy of Science and Medicine, Morrison Hall 305, Indiana University, Bloomington, IN 47405; e-mail: jschicko@indiana.edu.

†I wrote this article while I was a member of the School of Historical Studies at the Institute for Advanced Study (Princeton, NJ). I am most grateful for the institute's support of my work. Discussions with Ann-Sophie Barwich, Stuart Firestein, Stephan Güttinger, Miriam Solomon, and students and researchers at the Center for Science and Society at Columbia University (New York) and at the Consortium for the History of Science/Greater Philadelphia Philosophy Consortium (Philadelphia) helped me to develop my ideas on experimental control. I also wish to thank two anonymous referees for *Philosophy of Science* for their helpful comments on the penultimate version of this article.

1. Control experiments are of course important in the physical sciences as well, but here I am going to limit my discussion to the life sciences. I will also not discuss the type of control experiment that is prevalent in clinical research, the randomized controlled trial (RCT). On the history and formal structure of RCTs, see, e.g., Hacking (1988), Cartwright (2010), and Keating and Cambrosio (2012).

Philosophy of Science, 86 (April 2019) pp. 203–218. 0031-8248/2019/8602-0001$10.00
Copyright 2019 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

929, emphasis added). Scientific articles routinely mention controls,[2] and handbooks and instruction manuals on methods and protocols in the life sciences regularly call for the inclusion or setup of control experiments.[3] Science journals commonly ask their reviewers to evaluate the appropriateness of controls; indeed this task is regarded as one of the core elements of successful reviewing: "Importantly, reviewers should identify critical weaknesses, suggest experiments to bolster support of the central claims, and comment on the robustness, validity and reliability of data. Feedback on the adequate description of methodology, the appropriateness of experimental design and *required controls*, the validity of data analyses and statistics, and the quality and reproducibility of data is central in assessing the strength of a manuscript" (*Nature Cell Biology* 2017, 1005, emphasis added). An experiment that is not controlled in the right way is not acceptable. Moreover, in informal conversations, scientists go even further and ascribe epistemic significance to controls, stating that finding a control is equivalent with "figuring out what the experiment is doing." Not only does control play justificatory roles for working scientists, but it has a constitutive function as well.

The notion of control is ubiquitous in scientific practice and scientific publications; it is taken for granted and rarely defined.[4] One of the

2. A couple of examples from the recent scientific literature may suffice to illustrate the point. The first is from an experimental study of the effects of different chemical forms of soil phosphorus on plant growth. The authors note: "We first supplied a controlled nutrient solution to eight plant species that had been sown together in greenhouse mesocosms mimicking phosphorus-limited European grassland communities" (Ceulemans et al. 2017, 1). The second example stems from a recent paper on the relation between soil moisture and CO2 and CH4 production, the authors report that CO2 and CH4 production varied significantly over time, "with initial suppression followed by subsequent enhancement in both the intermediate and saturated/drained treatments relative to the field capacity treatment, hereafter termed the control" (Huang and Hall 2017, 2).

3. Here are just two random examples from standard protocols in microbiology and molecular biology. Austin et al. (2015) describe methods for characterizing enzymes and pathways of anaerobic degradation of different natural and anthropogenic hydrocarbons. The chapter "Protocols for Purifying and Characterizing Integral Membrane AlkB Enzymes" deals with a metalloenzyme (AlkB) that catalyzes the oxidation of most medium-chain-length alkanes in the environment. The authors describe step by step how to investigate the mechanism of alkane hydroxylation with a diagnostic substance, noting: "A control experiment should be included, using all reaction components except AlkB" (143). Maia and Moura (2016) cover various detection methods for nitric oxide (NO), a signaling molecule in plants. The authors state: "Appropriate controls should be done to guarantee that, e.g., the absence (or weak) of EPR signal is not due to inhibition of the NO generating system; or the opposite, that an intense EPR signal is not the result of the inactivation of the NO target or scavenger" (90).

4. There are not many historical analyses of the development of control experiments, either, apart from studies of RCTs such as the ones referenced in n. 1. Existing studies typically present whatever the author identifies as "the first" instance of control exper-

https://doi.org/10.1086/701952 Published online by Cambridge University Press

few existing systematic studies is Boring (1954). In this article, psychologist and historian of psychology Edwin Boring distinguished several meanings of control and situated control experiments in the contexts of nineteenth- and twentieth-century science and philosophy of science. Apart from Boring's contribution, there is not much discussion in philosophy of science and indeed in the sciences themselves about what an appropriate control is, whether it is possible to give a general characterization of a control, or how the appropriateness of controls could be properly evaluated.

In everyday contexts, the term "control" is associated with oversight, management, and constraint. We seek to control riots, passports, pests, or temperatures. In this broad understanding of "control," experimental control means the check that an experiment is methodologically sound.[5] Here I am interested in experimental controls as particular kinds of checks—different from, say, replications or error statistics or validations of data analyses. In this article, I am taking my cue from Boring's work, but I will argue that Boring's explication of "control" is ambiguous, awkwardly straddling conceptual, normative, and practical issues. I propose a revised set of distinctions.[6] Along the way, I also show how the careful study of past scientific terminology can be philosophically fruitful.[7]

In a sense, my discussion of control experiments parallels the discussions in philosophy of science about different roles of experiments in scien-

iments in the history of science or medicine (see, e.g. McCartney 1942; Dunn 1997; Amici 2001; Chalmers 2001).

5. In the broad sense, the question of how best to control (validate) experiments has of course received philosophical attention. See, among many others, Franklin (1989), Mayo (1996), Staley (2004), Weber (2005), Bechtel (2006), Franklin and Perovic (2016), as well as the discussions of the role of values in experimental inquiry, e.g., Douglas (2000).

6. A note on terminology: as far as I can see, there is no consistent usage of the terms "controls," "control experiments," and "scientific controls," and even the meaning of the technical terms "positive control" and "negative control" varies in the scientific and philosophical literature. In this article, I use the term "experimental control" as an undefined umbrella term, as Boring does in his paper. The terms "perfectly controlled experimental design," "control-experiment," and "controlling experiment" take on specific meanings in the context of my argument, as will become clear later in the text.

7. There has been much discussion recently about how to combine historical and philosophical scholarship, especially about the proper role of historical case studies for philosophical analysis (see, e.g., Chang 2011; Kinzel 2015; Sauer and Scholl 2016; Schickore 2018). In this article, I am not presenting a historical case study. Instead, I examine how a problem of inquiry that we still have today—making sure that our experiments are sound and informative—was understood and conceptualized in the past. Recovering forgotten meanings can challenge the current understanding of our present-day concepts, enrich our interpretations, and thus clarify those concepts.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

tific practice. Philosophers of experiment have shown that experiments have a life of their own (Hacking 1983), that they can play exploratory as well as evaluative roles (Steinle 2002), and indeed that they have many different roles, including articulating an existing theory, measuring quantities that are of physical interest, demonstrating a successful new experimental technique, or correcting previous incorrect or misinterpreted results (Franklin 2016). My discussion highlights different functions for experimental controls in the realization of an experiment: experimental controls may serve as checks, and they may serve as probes.

Boring's approach does not account for an important dimension of experimental control, namely, its constitutive role in knowledge production ("figuring out what the experiment is doing"). The approach to experimental control that I propose can illuminate this role: I show that experimental controls as probes have a constitutive function in knowledge generation. My account of experimental controls also sheds new light on the conception of exploratory experimentation, and it clarifies what can and what cannot be expected from reviewers of scientific journal articles giving feedback on experimental controls.

Concepts of Control. Boring (1954) advanced two conceptual claims and one historical thesis. The first conceptual claim is that three meanings of "control" must be distinguished in the context of experimentation: "control" in the sense of restraint (keeping conditions constant), "control" in the sense of guided manipulation (causing an independent variable to vary in a specific manner), and "control" in the general sense of check or comparison. The third, general notion of control is the oldest. It stems from the French word *contre-rolle* and is etymologically tied to keeping a duplicate or a roll of accounts (573).

Boring's second conceptual claim is that control "in the sense of a check or comparison . . . appears in all experimentation because a discoverable fact is a difference or a relation, and a discovered datum has significance only as it is related to a frame of reference, to a relatum" (1954, 589). The historical thesis is that our philosophical understanding of experimental control originates in Mill (1843/2006), although Mill himself did not use the term "control" in this context.[8]

Boring's claim that check or comparison appears in all experimentation lends itself to different readings, none of which is entirely convincing. We could read it as an empirical claim about comparative experimentation. In a comparative trial, two entities (as similar as possible) are put in similar ex-

8. On Mill's method of difference as the first conceptualization of control experiments, see also Schaffner (1983), 202, and Gigerenzer (1989), 53.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

perimental situations, but only one receives treatment.[9] In this interpretation of Boring's, claim it would be incorrect that control appears in "all experimentation." Perhaps he meant to say that every proper experiment involves (or should involve) some kind of comparative act. Then we would be left with the task of defining what kind of comparison would be a proper experimental control. Or perhaps he meant to advance an interpretation of causality, along the following lines: we understand an experimental outcome (a claim about independent variable I and dependent variable D) against the background of an (implicit) assumption about what would have happened had the variable I not been changed. This claim is as plausible as counterfactual theories of causality are, but it is an interpretation of the meaning of the concept of cause and is not useful as a guide for good experimental practice.

Boring's historical claim does not help to clarify the issue. As evidence for the view that "the methodological status of control as a check or comparison is to be understood by reference to J. S. Mill's Method of Difference," Boring (1954, 589) cites the first edition of the *Oxford English Dictionary*, Murray (1893). This dictionary defines (the "modern scientific use of") control as "a standard of comparison used to check the inferences deduced from an experiment by application of the Method of Difference." In the same passage, the control experiment is defined as "a test experiment with this end in view" (Murray 1893, 927). Boring assumes, quite plausibly, that the dictionary definition alludes to one of Mill's methods of experimental inquiry.[10] However, it is important to keep in mind that Mill's method of difference captures an inference pattern to attain a conclusion, whereas Murray's definition invokes pragmatic strategies to "check the inferences": "control" in the sense of carrying out a set of complementary trials.

Mill described the method of difference as the only type of inference by which one can ever "arrive with certainty" at causes (Mill 1843/2006, 394). The secure arrival at a cause requires an ideal, the perfectly controlled experimental design, whereby an independent variable is caused to vary in a specific manner (it is manipulated in the first trial but not in the second), and the background is restrained (held constant across both trials).[11] Needless to say, in a concrete research context, the perfectly controlled experimental de-

9. Boring's explanation suggests that he meant to refer to an actual experimental strategy. He explained: "The control observation is often only implicit, but the control series or the control experiment is generally explicit, being taken into consideration in the design of an experimental investigation" (Boring 1954, 589).

10. In my view, Boring's thesis about Mill's role for the conceptualization of the methodology of experimental control is misleading, but this point is not central for my analysis in this article, and I will not discuss it here.

11. For a detailed reconstruction of Mill's method of difference in terms of a "perfectly controlled experimental design," see Guala (2005), 65–69. On the notion of background or causal field, see also Mackie (1974), chap. 3.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

sign will never be realized, as it is impossible to keep all background conditions constant across different trials and as there is an element of arbitrariness and context dependence in the delineation of the causal field that is relevant for the context in question.[12]

It is, however, possible to bring the philosophical conceptualization of experimental control closer to the realm of the working scientist by distinguishing different strategies for attaining experimental control. For this purpose, it is instructive to take a closer look at the scientific sources that served Murray as the sources for his dictionary definition of "control." Just like the current edition of the *Oxford English Dictionary* would, the first edition provides illustrative examples for the usage of the term "control," citing phrases from Charles Darwin's work on insectivorous plants (1875), Charles Darwin and Francis Darwin's work on the movement of plants (1880), and a journal article on immunology (Hankin 1890). Careful reading of these examples reveals that the scientific uses of control are structurally different. I contend that these differences matter epistemically and are relevant to the work of the scientific reviewers today.

Murray's sample phrase "Four bladders were first tried as a control experiment" is a quote from Darwin (1875, 4313). The phrase "Radicles without any attached squares, which served as standards of comparison or controls" is a quote from Darwin and Darwin (1880, 162). The phrase "Control mice died of tetanus within 48 hours" is a quote from a scientific article on immunology that was published in *Nature* (1890, 122; the original source reads "forty-eight hours"). All phrases refer to scientific experiments. All three experimental designs could be understood as real-world approximations of perfectly controlled experiments, but the important thing is that they exemplify different pragmatic strategies of comparison and checking. In the second and third examples, comparisons are being made between an entity that receives a certain treatment and another similar entity in a similar situation that does not. The Darwins compared the natural growth of radicles (primary plant roots) to the growth of radicles of the same plant with little cardboard bits attached. The article published in *Nature*, which announces a cure for tetanus and diphtheria, describes an inoculation experiment. Six mice received each 2 cc of a tetanus antiserum and withstood the subsequent inoculation with tetanus, while the "control-mice" that had not received the serum "died within forty-eight hours" (Hankin 1890, 122). In both cases, the assumption is that all relevant background conditions remain constant across the trials. This is a tacit assumption; neither the Darwins nor Hankin explicitly addressed the issue of background conditions.

12. In fact, Mill himself pointed out that his "method of difference" required an ideal experimental situation that could rarely (if ever) be attained in actual experimental contexts. The chapter following the formal presentation of the method of difference in Mill (1843/2006) illustrates this point with a number of examples from early nineteenth-century science.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

The first example is different. To learn about the process of digestion in insectivorous plants, Darwin (1875) investigated the "bladders" of certain plants (he was working with *Utricularia neglecta*, or bladderwort). The bladders are those parts of the carnivorous plant that ingest the prey. Darwin had observed that the absorption of decaying animal matter had generated protoplasm in the processes inside the bladders. He conducted experiments to establish whether the decaying animal substance was essential to the generation of protoplasm or whether the plants could also generate protoplasm by absorbing carbonate and nitrate of ammonium. To find out, he needed to make sure that the bladders he used in his experiments did not contain any rotting animal matter, or, in his words, to ascertain "positively that the bladder, though translucent, does not contain some minute animal in the last stage of decay" (413). These experiments were his control experiments. The best way to proceed was to cut bladders "longitudinally into two; the quadrifids were examined with No. 8 of Hartnack, then irrigated, whilst under the covering glass, with a few drops of the fluid under trial, kept in a damp chamber, and reexamined after stated intervals of time with the same power as before. *Four bladders were first tried as a control experiment*, in the manner just described, in a solution of one part of gum arabic to 218 of water, *and two bladders* in a solution of one part of sugar to 437 of water" (413, emphasis added). After these experiments, Darwin proceeded to test a number of bladders with various fluids, including a solution of one part of carbonate of ammonia to 437 of water, a solution of gum (one part to 218 of water), a solution of carbonate of ammonia, a solution of one part of urea to 218 of water, a "putrid and very offensive infusion of raw meat," and many more (413–15). In this series of trials, the control experiments were carried out before—in preparation of—the actual investigation. Their purpose was to determine the background conditions that might have an impact on the digestive processes within the bladders.

This strategy emerges even more clearly from the following passage in a later section of the book. Here, Darwin discussed the digestive function of *Drosera* (sundew). The question was whether the secretions of the plant contained a digestive ferment like pepsin (see Darwin 1875, 94–96). In this experiment, Darwin neutralized the acid of the secretion with an alkali and examined whether digestion would stop and would begin again when more acid was added. However, "it was necessary *first* to try *two control experiments* namely, whether the addition of minute drops of water of the same size as those of the dissolved alkalies to be used would stop the process of digestion; and, secondly, whether minute drops of weak hydrochloric acid, of the same strength and size as those to be used, would injure the leaves" (94, emphasis added). Again, the control experiment is the preparatory trial that was carried out to gather information about the background conditions for the actual experiment on the digestive function of sundew.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

Just a few years after the publication of Murray's (1893) *Dictionary*, another dictionary conceptualized the distinction that one can extract from Darwin's experimental reports. Whitney's (1897) *Cyclopedia* distinguished between "controlling experiment" and "control-experiment." The explication of "controlling experiment" is as follows: "Controlling experiment, in chem., a corroborating or confirmatory experiment" (1237). An example from Bowditch (1867, 149) illustrates the meaning of the term: "For a controlling experiment, the gas may be passed for a short time through the alcoholic ammonia alone." The phrase is taken from the description of a procedure to detect sulfide of carbon in the gas and is, in fact, itself a quote from a journal article published in 1861. In the experiment, the gas passes through a solution of alcoholic ammonia fortified with a few drops of a saturated solution of lead. The solution takes an orange color, which indicates the presence of sulfide of carbon in the gas. The dictionary also offers a definition of "control-experiment," which reads: "control-experiment . . . An experiment made to establish the conditions under which another experiment is made" (Whitney 1897, 1237). This definition is not illustrated with an example.

The two terms "controlling experiment" and "control-experiment" conceptualize two different strategies of experimentation in actual research contexts. In comparative trials, two similar experimental entities are placed in as nearly as possible the same situation and are subjected to as nearly as possible the same treatment except for the intervention of interest (the change of the independent variable).[13] "Controlling experiment" refers to one of these maneuvers. The controlling experiment either does not involve an intervention or it involves an intervention with a known effect. In the example from immunology, the controlling experiment is the experiment on the mice that did not receive antiserum. In Bowditch's case, the gas is passed through a chemical solution that does not contain an indicator.

The other strategy—termed "control-experiment" in the *Cyclopedia*—is the strategy of investigating those additional factors that may have an impact on (disrupt, change) the relation between the independent and the dependent variable in an experimental situation. This notion of experimental control is not covered in the account that Boring advanced. The two specific meanings of control that he distinguished both had to do with managing the known variables in an experiment: "control" in the sense of keeping known background conditions constant and "control" in the sense of causing the independent variable to vary in a specific manner. The control-experiment, as the *Cyclopedia* characterizes it, is preparatory. It is done to investigate the conditions before making the actual experiment and the controlling experiment.

13. In the philosophical literature on causation, "intervention" has become a technical term to designate the manipulations that are acceptable for manipulability theories of causation. For details, see Woodward (2003, 98; 2009, sec. 5).

https://doi.org/10.1086/701952 Published online by Cambridge University Press

We can now say that one of the examples from Murray's (1893) *Dictionary*, the example drawn from Darwin's book on carnivorous plants, is an example for a "control-experiment," that is, an experiment to establish the conditions under which another experiment is made. The experiments on infected mice and growing plants are examples for "controlling experiments" that are done to corroborate or confirm the original experiment. In all definitions and examples for scientific strategies of controlled experimentation, it is taken for granted that the experimenters use only known, stable investigative tools and operate in stable, known environments.[14] In the next section, we will see that the distinction between "control-experiment" and "controlling experiment" is epistemologically significant.

Functions of Control: Experimental Control as Probe and Test. So far, the nineteenth-century dictionaries have served me to expose two notions of experimental control in addition to the everyday notion of oversight and management and the ideal of the perfectly controlled experimental design (which enables "with certainty" the establishment of causes according to the method of difference). On the one hand, there is the strategy of conducting a controlling experiment for another experiment. In my interpretation of the notion, "controlling experiment" encompasses all three practical meanings of control from Boring's (1954) study. A controlling experiment is informative in the following circumstances: it subjects an entity that is similar to the entity in the original experiment to as nearly as possible the same treatment as in the original experiment, except for the intervention. The background conditions remain as stable as possible across both trials.

On the other hand, there is the strategy of conducting control-experiments. In my interpretation of the notion, a "control-experiment" establishes the conditions under which another experiment is made.

Using this interpretation emerging from the nineteenth-century dictionary definitions, it is now possible to distinguish different functions of experimental control. "Controlling experiments" function as checks, as attempts at confirmation of the outcomes of the original experiments. "Control-experiments," by contrast, function as probes.

We could say that the purpose of control-experiments is to build up a confounder repertoire, that is, a repertoire of factors that muddle (disrupt, distort) the relation between independent and dependent variables in an

14. Compare the two present-day examples I quoted in n. 2: in the first example, the effects of soil phosphorus on plant growth are investigated. The plants being studied are first treated with a "controlled nutrient solution." This treatment serves to create a known ("controlled") environment for the trial. The second example, by contrast, indicates that controlling experiments were performed. Saturated and drained fields are compared to "the field capacity treatment, hereafter termed the control."

https://doi.org/10.1086/701952 Published online by Cambridge University Press

experimental situation.[15] The term "confounder repertoire" resonates with Deborah Mayo's term "error repertoire," but the underlying idea is not the same. Mayo has argued that experimental inquiries can be understood as "piecemeal learning from errors," whereby complex experimental problems are broken down into manageable piecemeal questions. Each of these questions can be understood as a question about whether a certain type of error is absent. These types of errors include mistaking artifacts for real effects, mistakes about causal factors, or mistakes about the value of a parameter (Mayo 1996, 17–18). I prefer the notion of confounder to the notion of error because the presence of confounders does not necessarily lead to erroneous results; confounders may confuse, and no result can be obtained. More importantly, Mayo's "error repertoire" is a transcontextual catalog of "canonical errors"; it includes common types of experimental mistakes. According to Mayo, these types of errors can be found through the synoptic analysis of mistakes made in the past (452). In my account, the confounder repertoire is context specific; it encompasses the kinds of factors that are likely to interfere with an actual experimental inquiry.

An error repertoire as Mayo envisages it is built with the benefit of hindsight. Looking back and drawing on our current knowledge, we can characterize certain past experimental results as mistaken; we can point to the sources of these errors and can say what type of error was committed. The experimenters seeking to build specific confounder repertoires for their experiments do not have the benefit of hindsight. They may not even know where to begin their search.[16]

Consider again the strategy of conducting control-experiments. I extracted this interpretation of experimental control from a close reading of the *Cyclopedia* definition (Whitney 1897) quoted above. On yet closer reading, the characterization of "control-experiment" in the original *Cyclopedia* formulation lends itself to two different interpretations. "Establish[ing] the conditions under which another experiment is made" could mean figuring out what factors may have an impact on the experimental situation (identifying the repertoire of confounders). It could also mean establishing what the specific conditions are in a given situation. Both are important pretrial tasks in the practice of experimentation, but they are not the same. The *Cyclopedia* does not offer a quotation from a scientific source to illustrate this

15. I am using the notion of confounder informally. For formal definitions of "confounder," see Fisher (1942) and VanderWeele and Shpitser (2013); for the history of confounders and confounding, see Morabia (2011).

16. Boring hinted at this when he noted—albeit in passing—that controls require hypotheses: "When there is no acceptable hypothesis as to the independent variable in an experiment, it may be impossible to have a control, which may become available only when a good hypothesis comes along" (1954, 582).

https://doi.org/10.1086/701952 Published online by Cambridge University Press

definition, but if we consider again the example from Darwin's book on insectivorous plants, the distinction between two probing strategies becomes clearer. Diagnosing and determining the kinds of confounders that must be taken into account are two different tasks. In other words, there are two functions for probing strategies: diagnostic and determinative.

The diagnostic probing task is to figure out where to look for confounders. This step is constitutive for the knowledge that the experiment generates. In Darwin's case, it is the task of figuring out what factors might possibly affect the production of protoplasm and the digestive processes in carnivorous plants. Because Darwin published at a time when he had already accomplished this task, in his text it does not even appear as a task; the diagnostic steps are not described.

The determinative task for experimental controls is to determine what confounders are indeed present in a given experimental situation. In other words, even if we have our confounder repertoire, we still have to figure out how we could best control for each of the confounders. Once it was clear to Darwin that only empty bladders should be used to study the process of protoplasm generation in plants, he still had to figure out how the bladders could best be cleaned. Control-experiments as determinative probes pose logistic problems, which are not always easy to solve. On the contrary, the logistics of determining confounders may pose considerable, and at times insurmountable, engineering challenges.

My discussion of the twofold probing function of control-experiments—diagnosing and determining—illuminates the notion of exploratory experimentation that goes back to Friedrich Steinle.[17] Drawing on examples from the history of electricity, Steinle has characterized a kind of experimentation that is performed in situations where no well-developed theories are available.[18]

One important characteristic of exploratory experimentation, according to Steinle, is that it is a mode of knowledge *generation*. According to Steinle, exploratory experimentation generates a specific form of knowledge, namely, new concepts. Steinle also shows that past episodes of exploratory experimentation often encompass a range of experimental strategies, such as parameter variation, simplifying and purifying an experimental arrangement,

17. For a more detailed discussion of Steinle's conception of exploratory experimentation, see Schickore (2016).

18. This function of experimentation is contrasted with the justificatory function of experiments as tests. This view is usually associated with Popper (1934/2002), who famously stated that "the theoretician puts certain definite questions to the experimenter, and the latter, by his experiments, tries to elicit a decisive answer to these questions, and to no others. All other questions he tries hard to exclude" (89). The position that it is the role of experiments to test scientific hypotheses was the "standard view" in philosophy of science until philosophers of experiment began to distinguish different roles of experiments.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

and finding appropriate concepts to express the empirical rules governing the experimental project.

What exactly does it mean that exploratory experimentation has an epistemic function? After all, Steinle agrees that applying these strategies does not guarantee that new knowledge can be generated. Steinle's idea is as follows: in hindsight, we understand that in many experimental contexts we now regard as generative of new knowledge, exploration was involved. Charles Dufay's experiments on attracting and repulsing substances were exploratory in the sense that no theory of electricity guided these experiments. Retrospectively, we can characterize the products of such generative exploration as new concepts—Dufay's conception of positive and negative electricity in this case (see Steinle 2006). For Steinle, exploratory experimentation has an epistemic function in the sense that exploration typically contributes to the generation of new concepts, although these strategies of exploratory experimentation are not a recipe for cooking up new ideas and they do not guarantee that the new ideas thus generated will withstand further scrutiny.

Thinking about the structure and function of experimental control reveals that there is a deeper sense in which exploratory experimentation has an epistemic function, and thinking about experimental control can help to pinpoint this function. We have seen that the umbrella term "experimental control" covers different kinds of experimental strategies, some probing, others testing. The strategies lumped together under the umbrella "exploratory experimentation" are also of different kinds. Some of them—simplifying and purifying an experimental arrangement—are determinative. They presuppose that a confounder repertoire for this experimental arrangement is in place. Other strategies—finding a variety of parameters—are constitutive for an exploratory project in that they generate the confounder repertoire for the experiment.

There are no general recipes for finding confounder repertoires. Moreover, and more importantly, our knowledge of relevant confounders will always be limited, and we need to assume that there are unconceived alternatives to the confounders we identified.[19] The confounder repertoire enables our experimental learning, yet there is a degree of freedom in it. Mayo mentions in passing the "problem of choosing an appropriate test procedure for a given inquiry," but her emphasis is on what to do once one has "specified

19. For the problem of unconceived alternatives in theory choice, see Stanford (2006). Rowbottom (2016) has recently suggested extending the argument to observations, models, and experiments. Both Rowbottom and Stanford discuss the problem in the context of the debates about scientific realism. I propose an even further extension of the issue to the confounder repertoire for experiments and the challenges arising for scientific reviewing.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

the kind of argument from error one wishes to sustain." Once this is accomplished, the background knowledge "directs one to appropriate choices of test statistics and corresponding error characteristics" (Mayo 1996, 298–99). My distinction between two probing functions highlights the freedom in making these choices and clarifies what this freedom means for the knowledge gained through experiments. Not every choice is a plausible choice, yet there may be reasonable disagreement about the composition and comprehensiveness of confounder repertoires. Drawing up such a repertoire of confounders plays a constitutive role for our knowledge as it is produced in our experimental practice.

Conclusion. I have proposed an account of experimental control that is relevant to actual scientific practice. My framework comprises the following readings of "experimental control":

I. The ideal of the perfectly controlled experimental design (which enables "with certainty" the establishment of causes according to the method of difference).
II. Pragmatic strategies of experimental control

    1. Checks

        a) The task is to create and maintain a stable, known experimental environment.
        b) The task is to conduct a "controlling experiment" for another experiment.

    2. Probes

        a) Determinative: the task is to determine what confounders are indeed present in a given experimental situation ("control-experiment").
        b) Diagnostic: the task is to figure out where to look for confounders.

This framework for thinking about experimental control can do justice to the significance that scientists attach to controls and helps to better understand the tasks of reviewers for scientific journals who are expected to assess the appropriateness of required controls.

What matters for reviewers seeking to give feedback on controls are the pragmatic strategies (II). My framework makes clear that evaluating controls involves quite different tasks and different sets of evaluation criteria. Perhaps different kinds of reviewers are needed to accomplish these tasks

https://doi.org/10.1086/701952 Published online by Cambridge University Press

and to apply these criteria. Reviewers face several challenges, depending on whether they are assessing experiments functioning as checks (1*b*) or experiments functioning as probes for confounders (2) and whether they are concerned with the logistic problem of eliminating or accounting for known confounders (2*a*) and the diagnostic problem of devising confounder repertoires (2*b*).

Assessing experimental controls as checks requires assessing whether the experimental setting is stable enough to allow for any conclusions to be drawn from it (1*a*). It also means assessing whether these experiments yielded results that are significantly different from the original experiments (1*b*). Assessing logistic solutions for the problem of how to account for known confounders means assessing whether particular means can accomplish given ends, whether these means are satisfactory or maybe even constitute improvements in accuracy or precision, speed, or practicality (2*a*). Assessing the plausibility of confounder repertoires is a fundamentally different task because it raises the question of whether all relevant confounders have been taken into account (2*b*). The problem is that our knowledge of confounders is always incomplete. Here reviewers face the problem of unconceived alternatives to the proposed confounder repertoires.

It would thus be useful to remind authors and reviewers that controlling an experiment and evaluating controls is as complex and as demanding as making and evaluating the experiments themselves. It would be useful to address the different pragmatic dimensions of experimental control in instructions for authors and reviewers and to ask authors and reviewers for justifications and discussions of the choices of confounder repertoires and controlling experiments.

This account of experimental control explains why scientists refer to experimental controls as extremely important parts of their experiments, noting that finding a control is equivalent with knowing what they are doing in the experiment. My interpretation of control-experiments as diagnostic and determinative probes captures scientists' intuitions about the fundamental epistemic role of experimental control, the intuition that finding a control is equivalent with figuring out what the experiment is doing.

REFERENCES

Amici, Raffaele Roncalli. 2001. "The History of Italian Parasitology." *Veterinary Parasitology* 98:3–30.

Austin, R. N., D. Born, T. J. Lawton, and G. E. Hamilton. 2015. "Protocols for Purifying and Characterizing Integral Membrane AlkB Enzymes." In *Hydrocarbon and Lipid Microbiology Protocols*, ed. T. McGenity, K. Timmis, and B. Nogales, 133–47. Berlin: Springer.

Bechtel, Bill. 2006. *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge: Cambridge University Press.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

Boring, Edwin Garrigues. 1954. "The Nature and History of Experimental Control." *American Journal of Psychology* 67:573–89.
Bowditch, William Renwick. 1867. *The Analysis, Technical Valuation, Purification, and Use of Coal Gas*. London: Spon.
Cartwright, Nancy. 2010. "What Are Randomised Controlled Trials Good For?" *Philosophical Studies* 147:59–70.
Ceulemans, Tobias, Samuel Bodé, Jessica Bollyn, Stanley Harpole, Kristin Coorevits, Gerrit Peeters, Kasper Van Acker, Erik Smolders, Pascal Boeckx, and Olivier Honnay. 2017. "Phosphorus Resource Partitioning Shapes Phosphorus Acquisition and Plant Species Abundance in Grasslands." *Nature Plants* 3 (16224): 1–7.
Chalmers, Iain. 2001. "Comparing Like with Like: Some Historical Milestones in the Evolution of Methods to Create Unbiased Comparison Groups in Therapeutic Experiments." *International Journal of Epidemiology* 30:1156–64.
Chang, Hasok. 2011. "Beyond Case Studies: History as Philosophy." In *Integrating History and Philosophy of Science: Problems and Prospects*, ed. Seymour Mauskopf and Ted Schmaltz, 109–24. Dordrecht: Springer.
Darwin, Charles. 1875. *Insectivorous Plants*. London: Murray.
Darwin, Charles, and Francis Darwin. 1880. *The Power of Movement in Plants*. London: Murray.
Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67:559–79.
Dunn, Peter M. 1997. "James Lind (1716–94) of Edinburgh and the Treatment of Scurvy." *Archives of Disease in Childhood* 76: F64–F65.
Fisher, Ronald Aylmer. 1942. "The Theory of Confounding in Factorial Experiments in Relation to the Theory of Groups." *Annals of Eugenics* 11:341–53.
Franklin, Allan. 1989. "The Epistemology of Experiment." In *The Uses of Experiment: Studies in the Natural Sciences*, ed. David Gooding, Trevor Pinch, and Simon Schaffer, 437–60. Cambridge: Cambridge University Press.
———. 2016. *What Makes a Good Experiment? Reasons and Roles in Science*. Pittsburgh: University of Pittsburgh Press.
Franklin, Allan, and Slobodan Perovic. 2016. "Experiment in Physics." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/win2016/entries/physics-experiment/.
Gigerenzer, Gerd. 1989. "Survival of the Fittest Probabilist: Brunswik, Thurstone, and the Two Disciplines of Psychology." In *The Probabilistic Revolution: Ideas in Science*, ed. Lorenz Krüger, Gerd Gigerenzer, and Mary S. Morgan, 49–72. Cambridge, MA: MIT Press.
Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.
Hacking, Ian. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.
———. 1988. "Telepathy: Origins of Randomization in Experimental Design." *Isis* 79:427–51.
Hankin, E. H. 1890. "A Cure for Tetanus and Diphtheria." *Nature* 43:121–23.
Huang, Wenjuan, and Steven J. Hall. 2017. "Elevated Moisture Stimulates Carbon Loss from Mineral Soils by Releasing Protected Organic Matter." *Nature Communications* 8 (1774): 1–10.
Keating, Peter, and Alberto Cambrosio. 2012. *Cancer on Trial*. Chicago: University of Chicago Press.
Kinzel, Katharina. 2015. "Narrative and Evidence: How Can Case Studies from the History of Science Support Claims in the Philosophy of Science?" *Studies in History and Philosophy of Science* 49:48–57.
Mackie, John Leslie. 1974. *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon.
Maia, Luisa B., and José J. G. Moura. 2016. "Detection of Nitric Oxide by Electron Paramagnetic Resonance Spectroscopy: Spin-Trapping with Iron-Dithiocarbamates." In *Plant Nitric Oxide: Methods and Protocols*, ed. Kapuganti Jagadis Gupta, 81–102. New York: Springer.
Mayo, Deborah G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
McCartney, Eugene S. 1942. "A Control Experiment in Antiquity." *Classical Weekly* 36:5–6.
Mill, John Stuart. 1843/2006. *Collected Works*. Vol. 8, *A System of Logic*. Indianapolis: Liberty Fund.

https://doi.org/10.1086/701952 Published online by Cambridge University Press

Morabia, Alfredo. 2011. "History of the Modern Epidemiological Concept of Confounding." *Journal of Epidemiology and Community Health* 65:297–300.
Murray, James. 1893. *A New English Dictionary on Historical Principles*. Vol. 2, *C*. Oxford: Clarendon.
Nature Cell Biology. 2017. "Principles of Refereeing." *Nature Cell Biology* 19:1005.
Nature Methods. 2017. "Pen before Pipette." *Nature Methods* 14:929.
Popper, Karl. 1934/2002. *The Logic of Scientific Discovery*. London: Routledge.
Rowbottom, Darrell P. 2016. "Extending the Argument from Unconceived Alternatives: Observations, Models, Predictions, Explanations, Methods, Instruments, Experiments, and Values." *Synthese* 193. doi:10.1007/s11229-016-1132-y.
Sauer, Tilmann, and Raphael Scholl, eds. 2016. *The Philosophy of Historical Case Studies*. Boston: Springer.
Schaffner, Kenneth. 1983. "Clinical Trials: The Validation of Theory and Therapy." In *Physics, Philosophy, and Psychoanalysis : Essays in Honor of Adolf Grünbaum*, ed. R. S. Cohen and Larry Laudan, 191–208. Boston: Reidel.
Schickore, Jutta. 2016. "'Exploratory Experimentation' as a Probe into the Relation between Historiography and Philosophy of Science." *Studies in History and Philosophy of Science* A 55:20–26.
———. 2018. "Explication Work for the Sciences." *Journal of the Philosophy of History* 12:191–211.
Staley, Kent W. 2004. "Robust Evidence and Secure Evidence Claims." *Philosophy of Science* 71:467–88.
Stanford, Kyle. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.
Steinle, Friedrich. 2002. "Experiments in History and Philosophy of Science." *Perspectives on Science* 10:408–32.
———. 2006. "Concept Formation and the Limits of Justification: 'Discovering' the Two Electricities." In *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*, ed. Jutta Schickore and Friedrich Steinle, 183–95. Dordrecht: Springer.
VanderWeele, Tyler J., and Ilya Shpitser. 2013. "On the Definition of a Confounder." *Annals of Statistics* 41:196–220.
Weber, Marcel. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
Whitney, William Dwight. 1897. *The Century Dictionary and Cyclopedia, a Work of Universal Reference in All Departments of Knowledge with a New Atlas of the World*. Vol. 2. New York: Century.
Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
———. 2009. "Agency and Interventionist Theories." In *The Oxford Handbook of Causation*, ed. Helen Beebee, Christopher Hitchcock, and Peter Menzies. Oxford: Oxford University Press.

https://doi.org/10.1086/701952 Published online by Cambridge University Press