Why Experiments Matter[1]

Adrian Currie, Sociology, Philosophy and Anthropology, University of Exeter

Arnon Levy, Department of Philosophy, Hebrew University of Jerusalem

Penultimate Version, Forthcoming in *Inquiry*

Abstract

Experimentation is traditionally considered a privileged means of confirmation. However, how experiments are a better confirmatory source than other strategies is unclear, and recent discussions have identified experiments with various modeling strategies on the one hand, and with 'natural' experiments on the other hand. We argue that experiments aiming to test theories are best understood as *controlled* investigations of *specimens*. 'Control' involves repeated, fine-grained causal manipulation of focal properties. This capacity generates rich knowledge of the object investigated. 'Specimenhood' involves possessing relevant properties given the investigative target and the hypothesis in question. Specimens are thus representative members of a class of systems, to which a hypothesis refers. It is in virtue of both control and specimenhood that experiments provide powerful confirmatory evidence. This explains the distinctive power of experiments: although modellers exert extensive control, they do not exert this control over specimens; although natural experiments utilize specimens, control is diminished.

1. Introduction

Traditionally, experiments are a privileged method of bringing the empirical and theoretical into contact. In particular, experiments are taken as an especially powerful way of confirming theoretical hypotheses. When experimentation is infeasible—in cosmology, geology, much of evolutionary biology and other areas—this is seen as a barrier to progress,

---

[1] For comments and assistance with previous versions of the article we're grateful to: Megan Delehanty, Alkistis Elliott-Graves, Emily Parke, Angela Potochnick, Kirsten Walsh, Eric Winsberg, an anonymous referee for this journal and audiences at IHSPSSB 2015 (Montreal) and the University of Geneva.

because it hinders confirmation. There is a large literature on confirmation, but surprisingly little of it deals with experimentation per se[2]. Further, several recent authors, noting similarities between experiments and other forms of confirmation – observations of natural phenomena (so-called ‘natural experiments’) and theoretical modeling (especially computer simulation) – contest the distinctive character of experiments, blurring the lines between experimentation, observation and modeling. Thus, a discussion of the role of experiments in confirmation is apropos.

We argue that the traditional view is right: experiments are a privileged means of confirmation; attempts to collapse experiments and other epistemic methods fail. We do so by presenting an account of experimentation and by diffusing competing arguments. On our view, experiments are: (1) controlled manipulations; that (2) pertain to typical instances of natural phenomena – or, as we designate them, specimens. Performing controlled manipulations generates fine-grained, discriminating information. Examining specimens underwrites moving from such information to more general conclusions. It is the *combination* of these two features which uniquely endows experiments with distinctive confirmatory power. These general contours should not be surprising: our view is a natural way of understanding traditional takes on experimentation. However, the details matter. Among other things, these details show why recent attempts to identify experiments with modeling and with ‘natural’ experiments fail. This, in turn, sheds further light on the place of experimentation in science.

After presenting an illustrative case, section 2 expands on ‘control’ and ‘specimenhood’. We then utilize this account in criticizing arguments for identifying experimentation with modeling, (section 3), and then with natural observations (section 4). This critical discussion is not merely negative in character; it grounds our claim about the superiority of experiment as a means of confirmation. Our criticisms will illuminate just what is so important about the combination of control and specimenhood for confirmation.

Our discussion focuses on hypothesis-driven experiments and their role in confirmation: studies aiming to test theoretical hypotheses against the world. Many empirical investigations, including experiments, have other structures and aims. Some are ‘exploratory’, seeking new phenomena, (Franklin, 2005). Others act as ‘possibility proofs’

---

[2] Significant exceptions are Hacking (1983), Brandon (1997), Cleland (2001), Okasha (2011) and Roush (2015). We rely on Okasha's ideas below. But we but cannot offer detailed discussions of the other authors here.

(Wimsatt, 2015), or provide novel results which drive further research (Morgan, 2005; Parke 2014; Currie 2017a). Still others verify the proper functioning of complex instruments (Galison, 1987). These types of experiments are important and interesting, but we won't discuss them here and remain neutral about whether our claims extend to them.

Furthermore, we are concerned with a certain aspect of confirmation. Issues surrounding confirmation can be divided into (at least) two types. The first concerns evidential *sources*: where and how evidence is generated, the relative significance of various sources, the relation between evidence-based and other forms of knowledge, and so on. The second type concerns the structure of evidential *relations*: are they probabilistic or logical? Discrete or graded? etc. These are connected, but can often be analyzed separately. We tackle the former issue: the significance of experiments relative to other means of obtaining evidence. We do not presuppose a specific view on the latter, structural issue. So far as we can see, our claims do not depend on answering such questions.

Finally, our discussion is normative, in the sense that we describe the evidential status of *successful* experiments. It is possible that many experiments fall short of the standards embodied in our account. We think real-life successes are not uncommon, but we shall not assess this matter explicitly. We outline an ideal: how experiments, when done right (and when luck agrees) provide particularly powerful confirmatory oomph. As an ideal, our account is fruitfully considered as a set of dimensions along which various investigations might fall. Those which approach the ideal—'bone-fide' experiments—have impressive confirmatory powers, while those departing from it are diminished in this regard. Crucially, this does not conflict with modeling or more passive observation figuring importantly in experimentation or confirmation. Indeed we accept, as explained below, that generating experimental data often requires the use of models. But we deny that models or natural experiments should be identified with experiments, or that they are epistemically similar.

2. Experiment as controlled investigation of specimens

Our account involves two central concepts - *control* and *specimen* – and highlights the importance of their combination. In this section we articulate these notions, providing a first pass analysis of experimentation. Subsequent sections further elaborate the concepts and clarify the importance of combining them, via comparisons between experimentation and other investigative methods.

To fix ideas, we begin with a simple, yet historically important, example: Isaac Newton's *experimentum crucis* (Walsh, 2012, 2017). Newton carried out the experiment sometime between 1665-1667 at his home in Woolsthorpe. Initially presented to the royal society in 1672, it was foundational for his *Opticks*, in 1704. The experiment dealt with the composition of white light (or at least sunlight). On the prevailing view of the time, such light is homogenous, i.e. composed of rays with identical properties. Different colours come about via modifications of white light. A potential competing hypothesis held that white light is heterogeneous, composed of rays of different colours.

Newton used an aperture in a window-shutter to isolate a beam of sunlight, transmitted through a prism. The prism projected the beam onto a screen, producing a full colour spectrum. A further aperture in this screen isolated a small part of the spectrum (another beam), transmitted through to a second prism. The second prism projected the light onto another screen. Shifting the angle of the first prism, and noticing the position of the light on the final screen, made it possible to record the light's 'refrangibility'—its disposition to refract.

Newton reasoned that if sunlight is homogenous, then isolating different parts of the spectrum will not change the position of the light projected onto the far screen: the prism's angle will not make a difference. Alternatively, if light is heterogeneous, then the position of the light will change, depending on which part of the spectrum is isolated (this is on the assumption of a one-to-one correspondence between refrangibility and spectral colour). He performed the experiment to find the latter option holds: varying the angle of the first prism varied the position of the light on the final screen. Newton concluded that light was heterogeneous:

> And so the true cause of the length of that Image was detected to be no other, then that Light consists of Rays differently refrangible, which, without any respect to a difference in their incidence, were, according to their degrees of refrangibility, transmitted towards divers parts of the wall. (Newton 1672, 3086)

The *Experimentum crucis* is a paradigmatic experiment, and illustrates our view. To begin, let us distinguish the object of an experiment and its target (Winsberg, 2009; Parke, 2014). The *object* is the concrete system which is manipulated; the immediate source of experimental data. Newton examined isolated beams of light, coming through his chamber window, under contrived conditions. That was his object. *Targets* of investigation, in contrast, are the phenomena which we ultimately seek to understand. Newton was interested in the properties of sunlight in general. That was his target. The two aspects of experiment which we identify track, first, our epistemic access to the object; and second, our capacity to extrapolate from object to target. These are discussed in turn.

2.1 Control

Like other paradigm experiments, Newton's study was *controlled* and this allowed him to draw rich, stable information about his object of study. Specifically, it involved three key features:

First, the experiment *isolated* an object of study, namely a ray of "fully coloured" white light. Light is abundant. It interacts with a wide variety of objects and forces under a wide variety of conditions. Newton's experiment depended on isolating rays of sunlight via the aperture: he was able to separate some rays of light from their regular environment and restrict their interactions to those relevant to his hypothesis. More generally, isolation concerns the ability to examine a system separately from the "raw" empirical world. Isolating an object entails severing, as far as possible, its causal connections to its environment, leaving the object's focal properties (those central to the experiment's aim) undisturbed.

Second, Newton *manipulated* his object of study by passing rays of light through a pair of prisms at varying angles. This altered the light's behaviour in subtle ways, allowing him to observe a telling effect: the change in the light's position on the screen, indicative of its 'refrangability'. Manipulation involves causally interacting with the relevant properties of the object, while holding other factors fixed. Thus, while isolation involves severing ties to the environment, manipulation consists in making changes to specified properties of the object, and only to those properties. With isolation, manipulation allows the generation of effects which tell for or against the hypothesis under consideration[3].

Third, Newton's experiment was *repeatable*. It could be carried out over and over again, allowing multiple data to be obtained. Moreover, the experimental procedure could be mimicked by anyone with the required equipment and knowhow. It did not depend on transient features nor on knowledge that could not, in principle, be passed along. When the

---

[3] This includes non-lab-based interventions, as in randomized controlled trials.

potential to be replicated is realized, experiments have strong confirmatory powers. A well-isolated, controlled but unrepeated experiment is less powerful than one which has been repeated. Confirmation requires bodies of data, and experimental validation (which we discuss below in light of specimens) demands that experimental results not be fluky.[4]

Further, combined with isolation and manipulation, repeatability allows an experimenter to conduct rich explorations of her object by finely varying the experimental circumstances while generating many significant data. This provides thorough information about the object.

Thus, an object is subjected to control when isolated from its natural environment and intervened upon in a replicable way. However, the power of experimentation does not rely on control alone. It also depends on the kind of objects over which control is exerted, namely specimens.

2.2 Specimens

Control concerns the interaction between an experimenter and her object. The object's being a *specimen* underwrites the move from knowledge about the object to conclusions about the investigation's target.

Hypothesis-driven experiments aim to ascertain empirical facts pertinent to the confirmation of a specified hypothesis, a hypothesis about some portion of the natural world. Paradigm experiments, like Newton's, can confirm (or disconfirm) hypotheses because their object of study is, in some sense, informative about the target of investigation. What must the object be like for this to be the case? We propose that the object should be a specimen, i.e. a *typical instance* of the target. By working with a specimen an investigator brings part of the world into the lab, so to speak. Roughly, we hold that typicality and the potential for confirmation are positively linked: all else equal, the more typical a specimen, the greater its confirmatory potential.

Being a specimen, then, depends on an object's being 'typical'. But what is typicality? In Newton's case, the answer is fairly straightforward: he was interested in the composition of sunlight. To this end, he examined instances of sunlight. These instances were typical

---

[4] In some experimental contexts a single "result" is suitable; we have in mind successful randomized control trials (RCT). But note that a single RCT in fact consists of a large number of repeated experiments, where results are summaries or aggregates. So we think an RCT is best seen as a series of (repeated) experiments.

inasmuch as they shared the composition of sunlight in general. Here, composition was a *focal property*: a property of the target which the hypothesis directly concerns came under experimental test. Sharing of focal properties is what made Newton's experimental object – sunlight from his window – informative about his target – sunlight in general.

This can be generalized: experimental objects inform us about targets insofar as the former are typical instances of the latter. Typicality can be understood in terms of similarity – in the limit, sameness – of focal properties. A specimen, we suggest, is typical inasmuch as it shares focal properties with the target. Can this notion be further specified? Some philosophers restrict attention to material similarity between objects and targets. In this view, what makes something an experimentally relevant object (a specimen, in our terms) is its being made of the 'same stuff' as the target. This has been cashed out by contrasting 'deep', material continuity, with mere formal or organizational continuity (Guala 2002), and by an experiment materially 'replicating' a part of the world (Morgan 2005). We grant that material similarity matters in some contexts, but we doubt this line of thought's generality. For one thing, it is often not obvious what counts as similar material composition (Parke, 2014). For another, material continuity (however understood) is neither necessary nor sufficient for capturing experimentation's power. Consider properties like being a parasite, a predator, or a voltaic cell; or think of the Stroop effect, Compton scattering, or the founder effect. The definitive features of these phenomena are not material composition, and experimental investigation relevant to them needn't be performed on materially similar objects (Parker 2009).

There are diverse kinds of object-target similarities that may undergird experimental relevance. And there do not seem to be in-principle limits on which kinds of properties good experimental objects share with their targets. Material composition can be central. But sometimes structural, dynamical and/or dispositional features are key. For this reason, we do not think it possible to specify, in general terms, the properties that an experimental object must share with its target. An object's being a specimen depends on its sharing focal properties with the target. But what these relevant focal properties are depends on the experimental context and the hypothesis under consideration.

If material composition cannot constrain specimenhood, what does? On our view constraint comes from the procedures underwriting how specimen are obtained. This is critical from an epistemic standpoint, since the epistemic value of a specimen depends on not its merely *being* a specimen, but on its specimenhood being *known* (or justifiably held). The key, we suggest, is the employment of an *unbiased drawing procedure* – i.e. a procedure for selecting an object, 'collecting it from the world', that preserves typicality. The underlying intuition is that in making an unbiased drawing we are "bringing the world into the lab" without introducing interference and distortion, preserving those aspects of the world that we are interested in understanding.

"Clean" unbiased drawing involves statistical sampling. A subset is selected from a large population in order to estimate some property. Statistics provides methods for making such selections, enhancing the sample's representativeness, i.e. ensuring that the distribution of values of the relevant feature in the sample resembles the distribution in the general population. In the simplest cases, random selection suffices. Randomness is important because it reduces (in the limit, eliminates) bias: each member of the population is as likely as any other to be sampled. Therefore, irrelevant factors are unlikely (in the limit, guaranteed not) to affect the estimate.

In some cases, an experimental object can be the product of this kind of random sampling, and its typicality is thereby assured. But we can extend the notion of unbiased selection and understand an unbiased selection process as one that reduces the risk of selecting an unusual object, and which preserves typicality by avoiding problematic alterations of the object. All this is, of course, relative to the aims of the experiment. Newton's experiment probed ordinary light. He obtained his experimental object by letting light through his window, presuming there was nothing special about the sunlight hitting his windowpane, in terms of its color composition. Furthermore, he presumed that the passage of the light through the window did not alter it. These assumptions constitute grounds for taking his object of study as a specimen, a typical instance of white light. These kinds of procedures allow us to ensure our objects of study are specimens – by proper statistical sampling when possible or by unbiased collection, given the experiment's aims. Of course, the exact nature of the collection process will depend on the subject matter and the object sought. But the general idea is to bring a part of the world, "as is", into the lab.[5]

---

[5] Readers may be reminded of Catharine Elgin's notion of exemplification – "the relation of a sample, example, or other exemplar to the features or properties it is a sample or example of." (2013, 400). An exemplar, explains Elgin, instantiates properties of the thing exemplified. Like specimens, exemplars can instantiate both concrete, material properties and abstract or structural ones. Moreover, exemplification, like specimenhood, can be selective, such that only some properties are exemplified.

Below, in sections 3 and 4, we argue that it is the combination of control and specimenhood which underwrite experimentation's privileged status *vis-à-vis* confirmation. But first, we should discuss the possibility of tension between these two attributes.

2.3 Do control and specimenhood conflict?

Objects and targets differ, sometimes dramatically. Objects must be *procured*—they must be drawn from the world and isolated in an experimental setting, and they must be *prepared*— made amenable to control. Procurement and preparation potentially involve departures from typicality, when they "tamper" with focal properties. This situation underlies some skeptical stances towards experimentation (e.g. Cartwright 1999). Briefly put, the worry is that the data we generate in our experiment is an artifact of the prepared nature of our object— or at any rate, that data pertains to the procured and prepared object, but not necessarily to the natural system of interest.

This is an important issue; but it does not present insurmountable scientific challenges, nor does it generate special worries for our account. Typicality, as we have delineated it, holds relative to the features specified by the hypothesis under consideration, and to the drawing procedures involve in obtainment. And so, only processes of procurement and preparation which affect focal properties and biased drawing procedures are potentially problematic. Newton's light was in an unusual circumstance: light beams do not often spend time squeezing through apertures, bending in prisms, then bouncing off walls. However, *vis-à-vis* his hypotheses, these departures from the usual are not departures from typicality, as they do not affect the composition of white light—the focal property in question.

However, many experimental systems depend on more extreme procurement and preparation. Model organisms, extensively used in biology, are cases in point: they are raised for generations under artificial conditions and bred to have unusual characteristics like genetic homogeneity, making them more suitable for lab-based study (Levy & Currie, 2015). Surely

---

However, there are important differences. First, in contrast to Elgin, we place weight on *how* a specimen was obtained. We think this matters for the specimen's epistemic role: a specimen is an object drawn (in an unbiased way) from the world. Assumptions about how it was obtained, i.e. whether the drawing was indeed unbiased and why, matter for justifying conclusions drawn from the specimen regarding the object of study. A further difference, perhaps pursuant to this one, is that Elgin employs the notion of exemplification to account for how models relate to the world (2010). We argue against such an application of specimenhood.

under such circumstances the needs of control clash with specimen procurement? Not necessarily.

The relationship between object and target—in particular the object's typicality—is an empirically tractable matter. As mentioned in §2.2, there are better and worse ways of procuring and preparing depending on context, and scientists often have highly developed and sophisticated methods for doing so. For instance, biologists working on model organisms standardly compare their results against 'wild-types'. Wild-type organisms are lab-reared, but retain some similarities with specimens prior to experimental trials. They thus provide a baseline from which mutants diverge. Wild-type organisms, among other things, provide a kind of middle ground between truly wild specimens and highly prepared organisms, thus bridging the divide between lab and world. So, scientists can experimentally probe the distance between their objects and targets. If there are departures from typicality, such studies can clarify these. Consequences, regarding how informative the object is *vis-à-vis* the target, can then be assessed. Control in fact aids us in solving such "external validity" issues.

Furthermore, cases where control and specimenhood do conflict are not paradigm experiments, and just in the way predicted by our account. If, in order to conduct a controlled investigation, I must change my object such that it is not (or is less) relevantly representative of my target, then my capacity to confirm hypotheses is diminished. Vice-versa, when control is sacrificed for specimenhood, the confirmatory significance of the results is similarly weakened. It is the combination of specimenhood and control that marks off experiments: where they conflict, the investigation at hand diverges from the experimental. How these two properties underwrite the confirmatory power of experiment is the subject of the next two sections, in which we compare experiments to other epistemic practices.

3. Experimentation versus modeling

The distinction between modeling and experimentation is a special case of the broader distinction between theoretical and empirical. Models, and theoretical devices more generally, are representations of the world – attempts to say something about some range of phenomena. In representing some natural system, a theory or model tells us to expect, or to entertain the possibility, that the system is a certain way. In contrast, empirical work, like experiment and observation, is a means of making causal contact with the world. We will not offer an account of representation (or of causal contact, for that matter). But we do assume that theories and models carry content in an intentional sense – they are *about something*, typically something in the actual empirical world. And what they say about the world may be correct or incorrect, accurate or inaccurate.

Models are a kind of theoretical device, typically geared towards a relatively specific target system and usually containing deliberate deviations from accuracy and completeness such as idealizations, abstractions, and approximations. Views differ as to how, exactly, models represent the world. We remain ecumenical, relying on no particular account. We only suppose, first, that models are theoretical representations, in the minimal sense discussed above. And, second, the widely shared idea that models can be, and often are, studied independently of the world, as self-contained simplifications of a complex reality. This is often expressed by stating that models have "autonomy" (Morgan and Morrison, 1999) or that models serve as surrogates for complex systems in the world.[6]

Modeling and experimentation have similarities. Models are often constructed on the basis of broader theoretical assumptions, before being subjected to analysis, which often involves tweaking various variables and ascertaining the consequences. Nowadays this often involves computer simulation, a form of numerical analysis, wherein the model is implemented via computer software, values are assigned to variables, parameters, and (when needed) initial conditions, and the software is run, often repeatedly. One then (usually) takes the results and forms a claim about the target. Thus in both modeling and experimentation there is a setup phase – of the experimental apparatus or the model – that may be grounded in background theory. This is followed by an investigation of the model/object – a run of the experiment or of the simulation – in its own terms, i.e. in the lab or on the computer. And finally, a move is made from the result-in-the-object to the target-in-the-world.

However, these rough similarities do not undermine, and should not obscure, the key difference we identified above: Models are representations; they are *about* the world, while experiments are not. Modeling results can be correct or incorrect, or more or less accurate, depending on the soundness of assumptions and the adequacy of analysis. Experiment, by contrast, are causal interactions with instances of targets. The experimental object is drawn from the world, and if the drawing is done appropriately, preserving typicality, then one can

---

[6] However, we remain neutral on how to understand such surrogation. Specifically, we do not take a stand, here at any rate, as to whether modelling involves indirect representation (Godfrey-Smith, 2006; Weisberg, 2013; Currie 2017b, Levy, 2015).

infer from the experimental results to conclusions about the target. That is to say: experiments are not representations and do not carry semantic content. They are not the sort of thing that can be right or wrong, accurate or inaccurate. Instead, they are concrete physical setups which allow scientists to engage a specimen of their target. The light shining through Newton's window, refracting through prisms as it did, was not a representation of the phenomenon of white light. It was *an instance* of an actual ray of light.

Crucially, this is reflected in the manner in which one justifies object-to-target inferences in modeling on the one hand, and experimentation on the other.[7] When inferring from the results of a simulation to a claim about the world, an investigator must have confidence in what her model says about the world. She must be sure of her background theory, and the assumptions and techniques used to construct the model, program the simulation and so on. She must know, in other words, that she has adequately represented the world in her model and in the way she has implemented and analyzed it. The grounds underlying an inference from experimental objects to their targets are different. While both must worry about external validity, the modeller is concerned with the adequacy of her representation, the experimentalist is concerned with the typicality of her specimen. As we've seen, this consists in an experimenter's confidence that her object was appropriately drawn from the world, and that experimental manipulation—preparation and the exertion of control—has not undermined the object's typicality. What is required isn't, primarily, sound theoretical knowledge (at least, not knowledge of the target), but confidence that one is experimenting on the relevant object, that one's object is a specimen, taken "as is" from the world. Put differently: whereas a successful experiment confronts theory with the world, a model fleshes out existing theoretical knowledge and/or assumptions. To be sure, modeling may show us that a certain hypothesis is impossible or unlikely, and thus provide new information about the world. But this differs from confirmation, in the sense of a direct "meeting" between theory and reality. A model may inform us about the world, but when it does, it is not by testing a theoretical hypothesis against the world.

Having outlined the distinction between modeling and experimentation in positive terms, we now discuss – and reject – recent arguments that seek to identify the two.[8]

---

[7] For a related view, focusing on simulation, see Winsberg (2009).

[8] These discussions tend to ask whether models and experiments are on epistemic par, simpliciter. We, however, restrict our claims to confirmation.

3.1 Experiments as representations?

Uskali Mäki (2005) explicitly identifies models with experiments. One of his arguments concerns control: he argues that both models and experiments are means of isolating a system for the purpose of studying a specific feature or process. One may question whether the interaction between modeller and model (or simulation) is causal and therefore whether a model is genuinely subject to control. But we will not pursue this. We accept that something akin to control occurs in modeling: A model represents a given system or phenomenon, often ignoring other systems entirely (i.e. isolation) and making idealized and approximate simplifying assumptions (i.e. manipulation). Runs of a simulation can and usually are repeated, often many times (i.e. repetition). Indeed, in exerting control, modellers, like experimenters, produce rich information about their objects. But Mäki also claims that experiments involve representation, and here we disagree.

Mäki holds that an object represents if, first, it is used by an agent as a stand-in, i.e. as an object that substitutes for something in the world; second, the object resembles the target, in respects relevant for the task at hand. Given this definition, he says, it follows that experiments are models: "[E]xperimental systems are representatives of some real non-experimental systems: they are material models of aspects of the rest of the world." (2005, 306). We believe this argument equivocates: it conflates something's being an intentional, semantic representation, in the sense discussed above, with the rather different notion of *representativeness*.

Recall that specimens are typical members of the kind under investigation, obtained by drawing from the world. We could say: a specimen is *a representative instance* of the target. But this sense of 'representation' is critically different from that applicable to models. Experimental systems, when suitably chosen, represent similarly to how statistical samples do – by being not-unusual subsets of the larger class, members of the kind in question. This is not representation in an intentional sense, and the difference is reflected in the epistemology: in a successful experiment the object is a specimen, and confirmation is possible because it has been procured via an unbiased procedure (or the bias is known and accounted for). To be sure, experimental specimens have been brought into the lab and manipulated. But as argued above, if properly chosen and handled, typicality often survives nonetheless. In contrast, a model represents the world by being *about* it. Often models are sets of equations or computer simulations: not in any relevant sense drawn from the world, and thus a "sampling"-like justification for being representative is irrelevant. Instead, what justifies thinking that the model represents is the theoretical considerations that led to the model's construction and the reasons for thinking that this way of representing the world is appropriate and accurate for the purpose at hand. Modellers examine their objects in great detail to ensure that they capture the intended dynamics.

This difference is sufficiently dramatic that regarding models as experimental objects is misleading, at best. Moreover, it illustrates our point that mere control is insufficient for experimental confirmatory power: because modellers do not exert control *over specimens*, they lack something distinctive of experimenters.

Let’s address a potential worry: if structural properties are amongst those that can be shared across specimens and targets, would that not open the door for models to count as specimens? After all, models often instantiate structures that are present in the target, or at least structures that resemble those of the target. The worry stems from a correct observation: specimens can share structural properties with targets and in this sense models do exhibit some of the features of specimens. But merely sharing properties with the target is not sufficient for specimens to play the epistemic roles they do; it matters how this fact is ascertained (§2.2). Specimens are drawn from nature, in the sense explained above. Models are not. To know that a model shares structural (or other) properties with a target we need confidence in our modeling assumptions and techniques, as just outlined—and these do not concern preserving typicality through procurement and preparation. Thus, while a model may share structural properties (as well as other properties) with a target and in that respect fulfil some of the conditions for specimenhood, it falls short of specimenhood insofar as our justification for believing property sharing holds is significantly different in character.

3.2 The model-ladeness of experiments

We now turn to a different attempt to blur the line between models and experiments, presented by Margaret Morrison (2009). Morrison examines the role played by models in ascertaining unknown physical quantities. She argues that modeling can function much like measurement, a type of experimentation.

Morrison highlights circumstances where simple measurement is infeasible unless scaffolded with models, suggesting therefore that measurement inherently involves modeling and simulation. We consider this an argument from the model-ladeness of measurement. She examines measuring the gravitational acceleration of the earth by using a simple plane pendulum. This involves a Newtonian equation that relates three variables: the pendulum's length l, the period of oscillation T and the acceleration due to gravity g. It might seem that ascertaining g is simple: measure l and T, and solve for g. However, the equation is idealized, neglecting various features of real-world pendulums. To apply it, we must employ a range of corrections—accounting for the non-uniform mass distribution of the bob, the effects of the surrounding air and so on. If our measurement is to be accurate, this requires delicate modeling of the apparatus we are working with. "Simply put" Morrison says "without models there is no measurement." (2009, 50).

Although Morisson's discussion is about measurement, her argument is readily adapted to hypothesis-driven experimentation (Massimi & Bhinji 2015 also emphasize how simulations are often intertwined with experimental practice).[9] Hypotheses are often idealized, necessitating local modeling to facilitate contact with relevant empirical systems. Indeed, if one wants to test whether the ideal pendulum equation accurately describes any real pendulum, similar modeling is required. Morrison could just as well have stated, then, "without models there is no confirmation".

We agree that, often enough, there is no measurement, nor confirmation, without models. But the fact that models are often necessary for obtaining data does not blur the model/experiment distinction, nor does it imply that models can serve in a confirmatory capacity. To see why, it will help to use a distinction between data and phenomena (Bogen and Woodward, 1988). *Data* are the immediate records of a causal interaction between a scientist and her object of study. The readings of one's instruments, the notes in one's lab book, photographs of tracks in a bubble chamber, etc. *Phenomena* are the stable features of the object which give rise to data, and that our theories aim to explain or predict. Observations and experimental "runs" produce instances of phenomena, which we collect data from. These data then serve as evidence for phenomena: tracks in the bubble chamber

---

[9] Massimi & Bhinji take arguments for the epistemic power of experimentation to turn on a distinctive causal interaction between scientists and the objects of their investigations which simulations lack. They argue that in experimental physics, intervening and representing are inextricable, and that no particular interaction between scientists and their experimental objects used in experimentation is unique, distinguishing it from simulations. In our terms, they deny that there is a special kind of experimental control. This argument, whatever its merits, is irrelevant for our position because we do not rely on a special kind of control. Rather, we rely on the notion of specimenhood, which pertains to the relationship between the object of an investigation and its target, namely, typicality.

are evidence of the movement of a charged particle. Similarly, Newton recorded data pertaining to the position of the light on the far wall of his room. His phenomenon was the refraction of white light when passed through a prism at varying angles.

Data are evidence for phenomena and moving from data to phenomena often requires a non-trivial inference. In the present context, it is perhaps best to speak of ‘experimental phenomena’, or ‘experimental effects’ – these are the stable behaviors of the object of study under experimental manipulation. It is these behaviors that the experimentalist aims to ascertain, and she does so by collecting data in the course of the experiment.

The records of l and T produced by the pendulum apparatus in Morrison’s example are data. The experimental phenomenon is the actual, true length of the bob and its period of oscillation. The data, i.e. what is recorded from the pendulum, is highly sensitive to interfering factors, as Morrison describes. Therefore, substantial modeling is necessary to arrive at an accurate estimate of the phenomenon – the value of g, for instance. However, this does not entail that the model is an experiment, nor that it is doing confirmatory work. For the model is still a means of representing the phenomenon, as well as, in this case, various interfering factors, which must be dealt with to infer the phenomenon from the data. Moreover, the results are relevant to confirming hypotheses about the pendulum in virtue of being obtained via engagement with a specimen. It is because of this relationship that the particular pendulum is informative about pendulums in general, even if model-based corrections are needed to make the move from data to phenomenon. Although models facilitate the inference from data to phenomena, the phenomenon’s being produced by a specimen is fundamentally responsible for the confirmatory (or dis-confirmatory) relationship between experimental results and hypotheses.

Generally speaking, experimentation is a complex epistemic process, and the confirmatory business of generating experimental data, inferring phenomena, and testing hypotheses is frequently underwritten by sophisticated theory. Often this includes “… the use of computer-aided techniques and theoretical assumptions to control background noise, potential sources of errors, and model data to extract meaningful signals…” (72, Massimi & Bhinji 2015). Much in science is theory-mediated, and often theoretical representations are useful—perhaps even necessary—for such mediation to occur. However, recognizing the role of theory in data production does not undermine our contention that the basis of an experimental result's confirmatory power is grounded in a controlled manipulation of a specimen.

Thus, while Morrison and others rightly highlight the importance of modeling in empirical work, it is best understood via the data/phenomena distinction, applied to the experimental object (see also Parker 2015). This is a subtle and interesting matter, but it does not undermine the distinctiveness of experiments.

3.3 Borderline cases

A third strategy for collapsing models and experiments proceeds by highlighting borderline cases. Several authors have commented on such cases – e.g. Wimsatt (2015), Parker (2008) and Franklin (1989) – but the most sustained discussion occurs in Parke (2014). Parke's central examples are 'bottle experiments', especially those performed in Richard Lenski's lab at Michigan State University. Lenski and his collaborators grew multiple lines of *E. coli*, over several decades, under various selection regimes, and investigated their long-term evolutionary dynamics. As Parke points out, often Lenski brings the same object, i.e. information about the same *E. coli* lineages, to bear on multiple targets. In some cases he targets features specific to *E. coli*, while in other cases the hypothesis concerns large-scale evolutionary patterns. The former pertain to relatively circumscribed phenomena in a particular species. The latter deal with abstract features of evolutionary processes, as many models do. Parke asks: are Lenski's investigations an example of experimentation or modeling? Her answer: it depends on context. We do not necessarily dispute this, but it is important to explain how it relates to our account and why it does not conflict with it.

There are two ways of understanding the kind of situation Parke describes. First, we could have a single experiment, relevant to hypotheses about different targets. This poses no challenge to our account, since a single object can serve as a specimen of more than one kind. On the (highly plausible) assumption that kinds are nested – one kind may be a sub-kind of another, like dogs and mammals or copper and metal – it is not surprising that an experiment can test multiple hypotheses by testing a claim about features that the sub-kind shares with its parent kind.

Alternatively, we could take the case as an experiment with respect to *E. coli*, and a model with respect to macroevolution writ large (this is Parke's reading). This interpretation too is fine by our lights. A single object of study can play multiple roles, so long as these roles can be reasonably well distinguished and the relevant inferences can be kept reasonably distinct. When the inference is from lab *E. coli* to the truth or falsity of a hypothesis about natural *E. coli* or related bacteria, we have experimental confirmation. Lenski and company are conducting a controlled investigation of specimens, and the inference from object to target operates along the lines we have discussed. When the targets are macroevolutionary (or ecological) events, *E. coli* are a kind of model, i.e. they are taken as a representation of large-scale trends over evolutionary time (see Levy & Currie 2015, section 6.3, also Currie forthcoming). Such a model is a material theoretical device, but that does not make it unusual (Weisberg, 2013). On this understanding, Lenski-type work consists of both experimentation and modeling, depending on its target.

These multiple roles are a prevalent feature of scientific practice. When a simulationist is constructing, calibrating and testing their object – 'verifying', in simulation-talk – they are treating the simulation as a specimen. This is because they are interested in the object's properties. In verification, the model is not treated as a representation (although the model's intended representational role matters for determining which properties are examined). Further, an object of investigation can act, as we have seen, as both a material model and as an experiment because it is the underlying justification of inferences from object to target that matters. Manipulation of the same object can tell us about some part or aspect of nature because it was drawn from it, and hence it is a specimen with respect to that part or aspect. But it can also serve as a representational construct, informing about nature due to theoretical assumptions about its resemblance to the relevant part of the natural world. In the latter case it would *function as* a model. Thus, on our account, whether an object is a model or being used experimentally depends in part on epistemic context: in particular, the focal properties under investigation and the manner in which their resemblance to the target's properties is justified. As focal properties and the underlying justification of object-to-target inferences shift, so too do crucial epistemic features of the investigation.

Our account, then, accommodates important borderline cases—and moreover, it captures just what makes them borderline. Surely, the inferences drawn by Lenski and his collaborators are firmer when the extrapolation is from experimental *E. coli* to wild *E.coli*, relative to when they move from evolutionary dynamics observed in a test-tube to dynamics on the macro-scale. One is less paradigmatically experimental than the other and this is reflected in their confirmatory properties.

Generally speaking, comparing experiments to models and other theoretical devices demonstrates the central importance of specimenhood for generating confirming evidence. Models are subject to something like control, indeed our capacity to repeatedly manipulate models often outstrips what we can do with experiments. This control makes models ideal tools for understanding how various parts of our knowledge interact: they can bring background knowledge together with observations to generate data and phenomena, they can test for coherence, and draw out the empirical consequences of our theories. And all of this can matter for how well supported we think a theory is. However, models are not specimens like the objects of experiments are. As such, they do not involve a confrontation between our theoretical ideas and the world. This underlies a central difference between the validation of models and experiments: whereas experimental validation depends upon the typicality of the specimen in question, considerations of typicality do not enter into model-target inferences. This difference in validation also highlights a difference in confirmatory power.

4. Experiments versus natural occurrences

Our discussion in the previous section pointed to a key difference between experimentation and modelling. The former engages with specimens, and may therefore test hypotheses *vis-à-vis* the kind of system they pertain to. In contrast, models are representations rather than specimens. To the extent that they play a part in confirmatory studies, there is a different underlying logic, and their role is less direct. We now turn to observations, a different sort of empirical work. Here, we argue for the mirror image of the previous contrast: observations resemble experiments with regards to specimenhood but lack crucial control-related features. We will discuss both ‘ordinary’ field-work and so-called ‘natural experiments’. In doing so, we fill in our account of control’s epistemic features, and complete our argument that experimentation’s distinctive status stems from the *combination* of control and specimenhood.

4.1 Observational Fieldwork

In observation one goes out to the field (or looks through a telescope, sends probes to the ocean floor etc.) and collects data about some portion of the world. In ‘ordinary’ observation this is relatively indiscriminate: the scientists record instances of the type of system in question, determining whether they have the relevant properties pertinent to the hypothesis. To fix ideas, suppose that the hypothesis takes the form of a quantified conditional such as 'all (or some, or such-and-such a proportion of) Fs are G'. In observation, 'ordinary' or otherwise, one may well be gathering data from specimens. What distinguishes 'ordinary' observation from experiment is that the specimens are not under the observer's control: she just happens upon them, as it were, recording the prevalence of Fs that are (and are not) G. Such information is relevant to confirming the hypothesis, on the assumption that one's data pertains to specimens. However, such data is not as powerful, confirmationally speaking, as that generated in an analogous experiment – i.e. one where an object is subjected to a controlled manipulation, bringing about *Fness*, and determining whether it leads to *Gness*. In elaborating this point we draw on work by Samir Okasha (2011).

Okasha relies on the 'Hypothetico-Deductive' principle, i.e. the idea that scientific hypotheses are confirmed when their predictions, in particular their direct 'positive instances', are borne out.[10] Suppose we test a hypothesis of the form "all Fs are G".[11] As Okasha points out, the most relevant observations would take the form $Fa \rightarrow Ga$ (i.e. positive instances of the generalization). In particular, these observations would be preferable to those of the form $Fa \,\&\, Ga$. After all, the former is a direct consequence of the hypothesis, and hence of direct confirmatory relevance; while the latter, conjunctive claim, needn't be an instance of the generalization in question (it may, e.g., be an instance of 'all Gs are F'). Hence it has less confirmatory relevance.

When exerting control, an object of study *a* is isolated, *Fa* is bought about, and we check whether *Ga* (co-)occurs. In contrast, in performing a non-experimental (i.e. non-controlled) observation one cannot, as it were, assume that *Fa* and test whether *Ga*. Instead one merely observes an *a* that is an *F* as well as a *G*, obtaining the information that these facts coincide, which is of lesser confirmatory power with regards to the generalization at issue. So an experimental test supplies direct confirmation of 'all Fs are Gs' in a way that the mere coincidence of *Fness* and *G*ness does not. Generally, in an experiment, one obtains

---

[10] Okasha relies on Bayesian ideas to argue for the Hypothetico-Deductive (HD) principle. In line with our general attitude of separating, as far as possible, the discussion of the means of confirmation from a discussion of the confirmatory relation, we here assume the HD principle, but we are *not* presupposing a Bayesian (nor any other specific) justification for it.

[11] The argument applies beyond universal generalizations, to any claim that has a conditional form, as noted above. We suspect that the argument can be extended (perhaps in a modified form) to other sorts of hypotheses, but since conditional claims are a central and common cases, we will not look beyond them explicitly.

information going beyond the fact that certain conjunctions obtain. Specifically, one obtains information about what happens when certain antecedent conditions have been fulfilled.

Thus, controlling a specimen provides a qualitatively stronger confirmatory position *vis-à-vis* 'passive' observation. So far so good; and, we think, relatively uncontroversial. But what about naturally occurring circumstances which appear to mimic experimental control – so-called natural experiments?

4.2 Natural 'experiments'

'Natural experiments' have been compared to traditional experiments. Woodward, for example, takes them to have the same basic form, differing only in the source of control:

> 'Natural experiments' typically involve the occurrence of processes in nature that have the characteristics of an intervention [… However these] do not involve human action or at least are not bought about by deliberate human design (2003, 94).

Although Woodward is not in the business of problematizing the experiment/observation distinction here, others have (Morgan 2013; Bromham 2016; O'Malley 2016).[12] The term 'natural experiment' is heard most often in contexts where experimental access is limited: studies of the deep past, or of large-scale, complex systems. By our account of experiments, what it would take for a natural observation to count as 'experimental' is clear: isolation, repeatability and manipulation must be not only possible, but often realized.

We do not deny that there are significant analogies between experiments simpliciter and natural experiments. For instance, Bromham (2016) reports a study investigating the apparently high rates of molecular evolution in parasitic plants. Noting the multiple origins of parasitism in flowering plants, she adds:

> This is like an experiment that has been run again and again using different kinds of plants… Parasitic plants are highly variable, but if you have enough of these independently derived parasitic lineages you can start to ask whether, despite the great

---

[12] Sometimes, such claims are made within science, with respect to a particular area or set of questions. For example, Stephen J. Gould explicitly takes events in life's history to play experiment-like roles (e.g. in his 1989. See also Beatty, 2006).

> differences between them, there are also any commonalities. Each of these parasitic lineages is like a single experimental treatment line... (Bromham 2016, 48).

Indeed, under such conditions ‘natural experiments’ may very well deliver confirmatory goods. However, 'natural experiments' rarely repeat and are, virtually by definition, not replicable. Repetition underwrites both inferences from data to phenomena and establishing a result’s external validity *vis-à-vis* a target.[13] Generating such knowledge often requires an enormous number of variations, test runs, and so forth. Although natural experiments can play a confirmatory role, their power is limited by a lack of finely varied repetitions—in seeing this, we also expand on the confirmatory significance of control.

Repetition is required for establishing reliability, and sifting data from noise. Moreover, repeated manipulation allows experimental scientists to explore and characterize an effect in some detail. Newton systematically varied the angle of the prism and observed the angle of the light on the far screen. Thus, he established not only that light passing through a prism will refract, but also *how* it will refract: there is a regular mathematical relationship between changes in the prism’s angle and the position of light on the far wall. Repetition, then, allows an experimentalist to reliably examine fine grained dependencies between variables – how one variable changes across a range of alterations to another variable. Such information is inaccessible from a small set of ‘natural’ experiments.

Paradigm experiments produce multitudinous data of this sort. In this they are unlike the methods Bromham discusses. Newton’s account of the *experimentum crucis* does not describe a single event. It is, rather, a formula or recipe for an experimental setup and one that was carried out, by Newton and others, many times. Even in cases like Bromham’s, it is extremely unlikely that natural experiments could provide the same level of detail—and repeated detail—provisioned by successful experiments.

Summing up, there are important similarities between some natural observations and experiments. In some circumstances naturally occurring events provide something like control: in principle, they may provide test cases of the sort Okasha highlights. However, experimentation’s special confirmatory power is due to its capacity to allow us to ‘make our

---

[13] ‘Repetition’ involves running the same experiment over and again. This capacity in combination with isolation and manipulability allows the running of several, slightly different interventions. Both of these are lacking in natural experiments.

own luck', via repeated experimental runs. It is, of course, possible for circumstances to conspire; nature *may* fortuitously work as an extremely busy scientist. Then we will obtain a rich set of cases, with enough repetition. In such situations we have something approaching a bone-fide experiment and this may supply substantial confirmation for a hypothesis. However, this is vanishingly rare; many observations that receive the label 'natural experiment' do not, in fact, have the epistemic power of experiments properly so-called. As such, specimenhood alone is not sufficient for the confirmatory prowess of experiments. Control allows the generation of bountiful fine-grained, relevant, data and evidence. Further, control provides flexibility in experiment design, allowing scientists to respond to the happenstances of empirical fortune.

5. Conclusion

Experiments have special confirmatory power in virtue of being controlled investigations of specimens. We have elaborated on these key notions – control and specimenhood – and have buttressed our account via a critical discussion of attempts to conflate experiments with models, on the one hand; and with certain kinds of observations, on the other hand. As we've seen, control endows experimentalists and modelers with the capacity to generate finely detailed, rich information about their object of study. Experimentalists and 'passive' observers often have specimens as objects of study, and this underwrites their evidential relevance, through the object and target's sharing focal properties. But only in experiments are control and specimenhood bought together, combining to generate the confirmatory power of that method—the traditional view is thus vindicated.

Our account shows why the experimenter is in a better situation *vis-à-vis confirmation* than other scientists. But it does not show, nor do we hold, that non-experimental scientists – modellers, observers, natural experimenters, and so forth – are somehow less 'scientific' than their experimental brethren. Confirmation is but one aspect of science among many. Further, our account explicitly targets an ideal: we're interested in *successful* experiments. As we've discussed, in some contexts control and specimenhood might conflict, and indeed the range of theories which are amenable to experimental testing might be limited. That said, insofar as confirmation is concerned, experiments are uniquely powerful in virtue of being controlled investigations of specimens.

Bibliography

Beatty, J. (2006). Replaying life's tape. *The Journal of philosophy*, 336-362.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, 303-352.

Brandon, R. N. (1997). Does biology have laws? The experimental evidence. *Philosophy of Science*, S444-S457.

Bromham, L. (2016). Testing hypotheses in macroevolution. Studies in History and Philosophy of Science Part A, 55, 47-59.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge University Press.

Cleland, C. E. (2001). Historical science, experimental science, and scientific method. *Geology*, 29: 987-990.

Currie, A. (forthcoming). Bottled Understanding: the role of lab-work in ecology. The British Journal for the Philosophy of Science.

Currie, A. (2017a). The argument from surprise. Canadian Journal of Philosophy, 1-23.

Currie, A. (2017b). From Models-as-fictions to Models-as-tools. Ergo.

Elgin, C. (2010), Telling Instances. In Frigg R. and M. Hunter (Eds.), *Beyond Mimesis and Convention: Representation in Art and Science*. Berlin: Springer.

-- (2011), Making Manifest: The role of Exemplification in the Science and the Arts. *Principia* 15: 399-413.

Franklin, A. (1989). The Epistemology of Experiment. IN: *The uses of experiment*. Gooding, Pinch & Schaffer (eds). Cambridge University Press.

Franklin, L.R., (2005), Exploratory Experiments, *Philosophy of Science* 72: 888-899

Galison, P. (1987), *How Experiments End*, Chicago, IL: University of Chicago Press.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy* 21 (5):725-740.

Guala, Francesco. 2002. “Models, Simulations, and Experiments.” In *Model-based Reasoning: Science, Technology, Values*, ed. Lorenzo Magnani and Nancy J. Nersessian, 59–74. New York: Kluwer Academic Publishers.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.

Levy, A. (2015), Modeling without Models. *Philosophical Studies* 172(3): 781-798.

Levy, A & Currie, A (2015). Model Organisms are not (theoretical) Models. *The British Journal for the Philosophy of Science* 66 (2), 327-348

Mäki, U. (2005), Models are Experiment, Experiments are Models, *Journal of Economic Methodology*, 12(2): 303-315.

Massimi, M. & Bhimji, W. (2015), Computer simulations and experiments: The case of the Higgs boson, *Studies in History and Philosophy of Science Part B,* 51:71-81.

Morgan, M. S. (2013). Nature's Experiments and Natural Experiments in the Social Sciences. *Philosophy of the Social Sciences* 43 (3):341-357.

Morgan, M. 2005. “Experiments versus Models: New Phenomena, Inference and Surprise.” *Journal of Economic Methodology* 12 (2): 317–29.

Morrison, M. 2009. “Models, Measurement and Computer Simulation: The Changing Face of Experimentation.” *Philosophical Studies* 143 (1): 33–57.

Newton, I. (1672), 'A Letter of Mr. Isaac Newton, Professor of the Mathematicks in the University of Cambridge; Containing His New Theory about Light and Colors: Sent by the Author to the Publisher from Cambridge, Febr. 6. 1671/72; In Order to be Communicated to the R. Society ', *Philosophical Transactions, 6*: 3075-3087.

Okasha, S. (2011). Experiment, observation and the confirmation of laws. *Analysis* 71: 222-232.

O'Malley, M. A. (2016). Histories of molecules: reconciling the past. Studies in History and Philosophy of Science Part A, 55, 69-83.

Parker, W. S. (2015). Computer Simulation, Measurement, and Data Assimilation. *The British Journal for the Philosophy of Science*, axv037.

Parker, W. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese* 169 (3):483 – 496.

-- (2008). Franklin, Holmes, and the epistemology of computer simulation. *International Studies in the Philosophy of Science* 22 (2):165 – 183.

Parke, E. C. (2014). Experiments, Simulations, and Epistemic Privilege. *Philosophy of Science*, 81(4), 516–536.

Roush, S. (2015). The epistemic superiority of experiment to simulation. (unpublished, philsci archive).

Walsh, K. (2017). How Many Colours?. In How Colours Matter to Philosophy (pp. 47-71). Springer, Cham.

Walsh, K. (2012). Did Newton Feign the Corpuscular Hypothesis?. In *Rationis Defensor* (pp. 97-110). Springer Netherlands.

Weisberg, M (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

Wimsatt, W. (2015). Models and Experiments? An exploration. *Biology and Philosophy symposium on Simulation and Similarity: Using Models to Understand the World*. 30 (2) pp 293-298.

Winsberg, E. (2010). *Science in the Age of Computer Simulation*. The University of Chicago Press.

-- 2009. “A Tale of Two Methods.” *Synthese* 169 (3): 575–92.

Woodward, J. (2003). “Experimentation, Causal Inference, and Instrumental Realism.” In *The Philosophy of Scientific Experimentation*, edited by Hans Radder, 87–118. Pittsburgh: University of Pittsburgh Press.