Experiments, Simulations, and Epistemic Privilege

Emily C. Parke*†

Experiments are commonly thought to have epistemic privilege over simulations. Two ideas underpin this belief: first, experiments generate greater inferential power than simulations, and second, simulations cannot surprise us the way experiments can. In this article I argue that neither of these claims is true of experiments versus simulations in general. We should give up the common practice of resting in-principle judgments about the epistemic value of cases of scientific inquiry on whether we classify those cases as experiments or simulations, per se. To the extent that either methodology puts researchers in a privileged epistemic position, this is context sensitive.

1. Experiment versus Simulation. Scientific practice in the twenty-first century is increasingly blurring the lines between experiment and simulation. While it was once common for individual scientists, laboratories, or even entire subfields to focus on only one of these methodologies, experimental and computational methods are now increasingly combined. This has led to new ways to do science, as well as opportunities to reexamine the roles that experiment and simulation play in scientific inquiry and their changing natures in practice. These trends are reflected in increased attention from philosophers of

Received October 2013; revised May 2014.

* To contact the author, please write to: University of Pennsylvania, Department of Philosophy, 249 S. 36th St., Cohen Hall 433, Philadelphia, PA 19104; e-mail: eparke@sas.upenn.edu.

† I am grateful to Mark Bedau, Mark Colyvan, Karen Detlefsen, Zoltan Domotor, Mary Morgan, Margaret Morrison, Daniel Singer, Kyle Stanford, two anonymous reviewers, and especially Brett Calcott, Paul Sniegowski, and Michael Weisberg for their valuable feedback on drafts of this article. Thanks also to audiences at conferences where I presented earlier versions of this work for helpful comments and discussion: the Australasian Association of Philosophy 2012, Philosophy of Biology at Dolphin Beach 6, Philosophy of Scientific Experimentation 3, the International Society for the History, Philosophy, and Social Studies of Biology 2013, and a University of Pennsylvania Ecolunch Ecology and Evolution seminar. This work was supported by National Science Foundation grant DGE-0822.

Philosophy of Science, 81 (October 2014) pp. 516–536. 0031-8248/2014/8104-0007$10.00
Copyright 2014 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

science to experimentation (e.g., Hacking 1983; Franklin 1990; Radder 2003; Weber 2004), to simulation (e.g., Humphreys 2004; Winsberg 2010; Weisberg 2013), and to their methodological and epistemic points of convergence and contrast (e.g., Guala 2002; Peck 2004; Morgan 2005; Barberousse, Franceschelli, and Imbert 2008; Morrison 2009; Parker 2009; Winsberg 2009; Peschard 2012).

This article contributes to that discussion by challenging a widespread idea that the difference between experiment and simulation tells us something, in principle, about epistemic value. There is a pervasive view among philosophers and historians of science, and scientists themselves, that experiments have epistemic privilege over simulations. That is, they allow us to make better inferences about the natural world or generate more reliable scientific knowledge. A number of people have recently put this kind of idea in writing. For example: "[Simulation's] utility is debated and some ecologists and evolutionary biologists view it with suspicion and even contempt" (Peck 2004, 530); "simulations are supposed to be somehow less fertile than experiments for the production of scientific knowledge" (Guala 2002, 4); and "the intuition of [noneconomic] sciences and philosophy is that experiment is a more reliable guide to scientific knowledge" (Morgan 2005, 326).

All of scientific inquiry involves engaging with some object of study—a model, a physical system in the laboratory or field, or a combination of these—to learn about some target of inquiry. The methodological distinction between experiment and simulation is certainly important for making judgments about epistemic value. But I argue that this is the case only in a context-sensitive way, not as a generalization across science. The experiment/simulation distinction should not be used as a basis for in-principle judgments about epistemic value.

Before proceeding, some clarification of what is meant by 'simulation' is in order. Much of the literature on experiment versus simulation focuses on computer simulation: studies of computational models with some dynamic temporal element. But there is another broader understanding of 'simulation' in which the object of study in question could be any kind of model: mathematical, computational, or concrete (physical). Simulations in this broader sense are taken to include studies of computer models, model organisms in laboratories, and model airplanes in wind tunnels. Discussions of the relationship between experiment and simulation in the literature have focused sometimes on computer simulation (Humphreys 2004; Peck 2004; Morrison 2009; Parker 2009) and sometimes on simulation in the broader sense (Guala 2002; Morgan 2005, 320; Winsberg 2009, 2010). I highlight this contrast up front because being clear about which sense of 'simulation' is at stake matters. In particular, I do not believe that there are any clear methodological or epistemic distinctions to be drawn between experiment and simulation in the broader sense.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

Experiments are thought to have two particular virtues which give them a privileged status over simulations. First, they generate greater inferential power, or external validity: experimenters are in a better position to make valid claims about their targets of inquiry in the natural world. Second, experiments are a superior (or, the only) source of surprises or genuinely novel insights. I argue that both of these claims are mistaken as generalizations. To the extent that the difference between experiment and simulation carries epistemic weight, this weight is context sensitive; the mere fact that a case of scientific inquiry counts as an experiment or a simulation is no indication of its epistemic value. There is a lot of important work to be done in understanding what grounds and validates inferences from objects of study to targets of inquiry in the natural world. Focusing on whether to classify cases as simulations or experiments, per se, muddies the waters of that task.

In arguing that there is no in-principle difference in epistemic value between experiments and simulations, I am not challenging the idea that, within the context of a particular research area, there might be good reasons to think that experiments have epistemic privilege over simulations or vice versa. Furthermore, I am not challenging the idea that experiments have priority over computer simulations in the greater picture of what fundamentally grounds scientific knowledge. We know how to do good computer simulations precisely because we have gained knowledge about the world through observation and experiment. In the grand scheme of things, empirical data are fundamental for answering scientific questions about the natural world. But people often seem to blur the lines between this general empiricist claim and a separate issue, namely, which methodology, now, will generate better scientific knowledge. This is the target of my objection: the kind of thinking that goes into claims about experiments' general superiority, like those cited above, or claims that a case of research is less epistemically valuable because "it's just a simulation."

2. Inferential Power and Materiality. Belief in the epistemic privilege of experiments over simulations is often grounded in ideas about their relative inferential power. In particular, the idea is that experiments lead to better inferences about natural systems or phenomena than simulations do (this is sometimes referred to as the issue of external validity).[1] This difference has to

1. A number of people have also argued versions of the claim that (at least some kinds of) simulations are experiments (e.g., Peck 2004; Barberousse et al. 2008; Morrison 2009; Parker 2009). This often takes the form of arguing that simulations allow us to achieve methodological or epistemic goals regarded as characteristic of experiments, such as actively intervening in an object of study or testing causal claims, and thus that simulations deserve to be thought of as epistemically on a par with experiments. This is a different sort of point from the one I am addressing here, but it still subscribes to the idea that judgments about epistemic value hinge on categorizing cases as simulations or ex-

https://doi.org/10.1086/677956 Published online by Cambridge University Press

do with the relationship between objects of study and targets of inquiry. When a researcher studies one system intending to make inferences about another, the former is her *object* and the latter is her *target*. These terms gives us a useful neutral language for discussing views on experiment versus simulation.[2]

Judgments about experiments' privileged status are often driven by the intuition that experimental objects of study have a privileged link to targets in the natural world in virtue of their shared materiality. Winsberg nicely describes this as "the suspicion (or conviction) [that] the experimenter simply has more direct epistemic access to her target than the simulationist does" (2010, 55). Guala (2002) and Morgan (2005) have argued that material object–target correspondence is a defining feature of experiments and that this correspondence is responsible for experiments' advantage over simulations in terms of inferential power. I call this shared view of theirs the *materiality thesis*.[3] Guala puts it in terms of experiments' objects having "deep, material" correspondence to their targets, while simulations' objects have only "abstract, formal" correspondence to their targets. Morgan puts it in terms of experiments' objects "replicating" or "reproducing" parts of the world, while simulations' objects only "represent" parts of the world. These authors cite examples of experiments in economics and psychology to support their points: experimental objects of study (in those cases, humans) are the same kind of thing in the laboratory and outside the laboratory. Experiments are thus supposed to have greater inferential power in virtue of their objects of study being "made of the same stuff as the real world" (Morgan 2005, 322); as Morgan puts it, "The fact that the same materials are in the experiment and the world makes inferences to the world possible if not easy. . . . The shared ontology has epistemological implications. We are more justified in claiming to learn something about the world from the experiment because the world and experiment share the same stuff" (323).[4]

periments (with experiments getting the upper hand). That is precisely the core idea I am objecting to here.

2. I am using these terms in this way following Winsberg (2009) and others. It is important to clarify this usage because people have used the word 'target' in different ways: usually to refer to what I call the target here, but occasionally to refer to what I call the (experimental) object here (Winsberg 2010, 31, 35; Peschard 2012), which can get confusing.

3. Harré (2003) also argues for a version of this thesis, although in the context of discussing inferences from experiments, not comparing experiments to simulations.

4. Morgan's paper is on experiments versus models. I take simulations to be studies of models with a dynamic temporal element, and it is clear from the context of her paper that the models Morgan has in mind involve being studied in this way (cf. n. 11 below). So her point here about object-target relationships in models applies straightforwardly to simulations. Some people define simulation differently, e.g., not as a study of a model but rather as a methodology in between modeling and experimenting (Lenhard 2007), but this is not the view I am assuming here.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

This view lines up with the above-mentioned intuition: experimental inquiry gets us closer to the natural world because experimental objects are samples, instantiations, or reproductions of their targets; simulations, in contrast, have as their objects (only) models of their targets. As Morgan puts it, "on grounds of inference, experiment remains the preferable mode of inquiry because ontological equivalence provides epistemological power" (2005, 326).

There is certainly some truth in the intuition behind the materiality thesis. Experimenting on an actual sample or physical approximation of a target in the natural world is often the best way to get traction on understanding it when we know very little about the target in question or when the relevant theoretical background is minimal. For example, we have relatively low confidence (today) in our understanding of how new drug cocktails will work in the human body. It makes sense for us to place more confidence (now) in tests on physical proxies, like mice or, even better, human clinical trial volunteers, than in a computer simulation of the human body. But often is not the same as always; material object–target correspondence is not, and should not be thought of as, necessarily the best route to valid scientific inferences. I return to this point below.

In section 3 I discuss an example of a research program that could reasonably be thought of as an experiment or as a simulation in the broader sense. In section 4, through analyzing this example, I argue two points: first, we should reject the idea that material object–target correspondence is characteristic of experiments and confers greater inferential power on them; second, we should stop looking to the experiment/simulation distinction altogether as a basis for making wholesale judgments about inferential power.[5]

3. One Object, Multiple Targets: Examples from Experimental Evolution. Experimental evolution involves propagating populations of organisms in the laboratory as a means to study evolution in real time. This research area provides a good framework for thinking through issues regarding experiment versus simulation because, as I spell out further below, it blurs the

5. I agree with concerns that both Parker (2009) and Winsberg (2009, 2010) have raised about the materiality thesis; in particular, their points that material correspondence does not always entail greater inferential power and that it is difficult to even make sense of the distinction between material and formal object-target correspondence. I am objecting to the materiality thesis in a different context than theirs, namely, that of arguing that we should do away altogether with relying on the experiment/simulation distinction to tell us anything in principle about epistemic value. Winsberg (2010) seems sympathetic to my conclusion, but I am putting the point more strongly: it is not just that there are exceptions to the generalization that experiments have epistemic privilege, but rather, thinking in terms of such generalizations is the wrong way to approach judging the epistemic value of cases of scientific inquiry.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

lines between experiment and simulation in the broader sense. Experimental evolution is different from studying evolution in natural populations because the controlled laboratory setting allows for more direct, comprehensive analysis of the populations' genetic makeup and evolutionary history. It is also different from artificial selection experiments, because the populations are evolving via natural selection; the researcher does not choose which individuals will carry on to the next generation on the basis of any particular trait(s) they possess. Experimental evolution commonly uses bacteria as objects of study. They are ideal for a number of reasons: they have short generation times, large populations can be kept in very small spaces, and they can be frozen and revived, allowing for comparisons and competitions between evolved populations and their ancestors (Lenski et al. 1991; Gentile et al. 2011).[6]

A notable example is Richard Lenski's long-term evolution experiment. Lenski's research group used a single ancestral cell of *E. coli* to found 12 genetically identical populations in 12 identical environments (flasks of minimal liquid growth medium), let them evolve, and watched what happened (Lenski et al. 1991; Vasi, Travisano, and Lenski 1994; Travisano, Vasi, and Lenski 1995). They began this experiment in February 1988; the populations passed the 50,000 generation mark in 2010 and are still going strong (Lenski 2014).

The initial goal of the experiment was to study these 12 populations to learn about the long-term dynamics of adaptation and diversification. Lenski and colleagues have also used this same object of study to make inferences about a number of different targets in the natural world. I briefly describe two examples and then discuss how the differences between them put pressure on the materiality thesis.

3.1. High Mutation Rates. The Lenski populations have been used to study the evolution of high mutation rates over long evolutionary timescales. Most new mutations are deleterious, which intuitively makes sense: there are more ways to mess something up at random than there are ways to improve it. One might expect that in well-adapted populations in unchanging environments, genomic mutation rates would stay the same over time or perhaps even decrease. In just three of the 12 Lenski populations, the opposite happened: their genomic mutation rates increased by two orders of magnitude after 10,000 generations. The explanation for how this happened involves mutator alleles, which raise the genomic mutation rate by inhibiting mechanisms like DNA mismatch repair or proofreading.

6. See Garland and Rose (2009) and a 2013 special issue of *Biology Letters* (Bataillon, Joyce, and Sniegowski 2013 and others) for excellent overviews of experimental evolution.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

Sniegowski, Gerrish, and Lenski (1997) discuss how mutator alleles arise spontaneously, by mutation, and then hitchhike to high frequencies in the experimental populations, which is possible because those populations are completely asexual.[7]

From this point about high mutation rates evolving in three of the laboratory populations, the authors make an inference about populations outside of the laboratory. They conclude that high mutation rates might evolve via the same hitchhiking mechanisms in similar clonal populations of cells with high mutation rates in the natural world—in particular, pathogenic *E. coli* and *Salmonella* (Sniegowski et al. 1997, 704).

3.2. Punctuated Evolution. In this second example, a different sort of inference is made from the same object of study. Lenski and colleagues used results from their evolving *E. coli* populations to make claims about punctuated equilibrium: the notion that evolution occurs in long periods of relative stasis punctuated by short periods of rapid change (Elena, Cooper, and Lenski 1996). Punctuated equilibrium was a hot topic in contemporary discussions of macroevolutionary trends in the fossil record, notably championed by Gould and Eldredge in their discussion of events like the Cambrian explosion (1977).

Over the first few thousand generations of the Lenski experiment, the 12 populations of *E. coli* increased in both average fitness and average cell size (Lenski et al. 1991; Lenski and Travisano 1994). In addition to these overall trends, average cell size increased in a step-like pattern. In one population it remained stable for the first 300 generations, increased by over 25% in the following 100 generations, and then remained stable for another 300 generations before dramatically increasing again (see Elena et al. 1996, fig. 1). This led the researchers to claim that punctuated evolutionary trends could be observed on (relatively) very short timescales, associated with the rise of beneficial new mutations in the population that rapidly sweep to fixation.[8] At the end of the paper they conclude with the following claim about natural populations:

7. Mutator hitchhiking is a phenomenon whereby beneficial mutations arise in a genome and happen to be linked to mutator alleles. Strictly asexual populations do not undergo genetic recombination. So if a beneficial mutation is linked to a mutator allele, and natural selection favors that beneficial mutation, it will spread through the population in subsequent generations, and the mutator allele can "hitchhike" along with it, resulting in a population with a higher genomic mutation rate.

8. The authors are careful to use the term 'punctuated evolution' rather than 'punctuated equilibrium'. Still, the paper is written so as not to avoid being interpreted as saying something about the much more long-term dynamics of punctuated equilibrium involving speciation events.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

> The experimental population was strictly asexual, which may have increased our ability to resolve punctuated changes. However, any difference between sexual and asexual populations with respect to the dynamic of adaptive evolution breaks down when two conditions are met: (i) standing genetic variation for fitness is exhausted, as will eventually happen in any constant environment, and (ii) beneficial mutations are so rare that they occur as isolated events. To the extent that these conditions are fulfilled in nature, then the selective sweep of beneficial alleles through a population might explain cases of punctuated evolution in the fossil record. (Elena et al. 1996, 1804)

4. Against the Materiality Thesis. The materiality thesis implies that inferential power is proportional to the degree of material correspondence between object and target. To think through what this means exactly, we first need to clarify what this correspondence is supposed to consist in. Morgan and Guala refer to it in various ways: "material correspondence," "material analogy," "ontological equivalence," or being "made of the same stuff" (Guala 2002; Morgan 2005). They must of course mean "made of the same *kind of* stuff," in the sense that object and target are both instances of the same material type. But there are various ways to interpret what this means. One (uncharitable) way would be "made of the same material stuff at the most fundamental level." On this interpretation, almost all biology experiments could be said to achieve strict material correspondence because their objects and targets are made of carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur.

Here is a more reasonable interpretation of "material correspondence": the highest degree of material object–target correspondence would be an identity relation. Certain field experiments might achieve this, for example, when they involve studying every individual in a small, clearly delimited population in order to make inferences about particular features of exactly that population. But this sort of object-target identity is far from the norm in science.

The next level of material correspondence would be an object that is a token of the same relevant ontological category as the target, at a sufficiently fine-grained level for the purpose at hand. When chemists study samples of uranium to learn about the properties of uranium in general, they are achieving material object–target correspondence in exactly this sense. The object is a token of the target's type at the finest level of grain with which we classify the relevant kinds: chemical elements. Further from this extreme would be studying mice to learn about humans. Both are living organisms, but they are not of the same type at even close to the finest level of grain with which we classify organisms. They are both mammals but do not belong to the same species or even genus. Even further from this extreme would be studying plastic models of mice to learn about mice.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

I take this to be the most plausible way to make sense of the idea of degrees of material correspondence and will from here on use the term in this sense: in terms of variations in grain of correspondence at the relevant "material" (e.g., chemical, physical, biological) level of categorization. Adopting this scheme, were we to endorse the materiality thesis, we should hope to be able to at least roughly rank research programs in terms of their achieved degree of material object–target correspondence. However, this gets difficult if we try to think through the two examples of inferences from the Lenski system in these terms.

In the high mutation rates case, it is not so difficult. The inference is from experimental populations of *E. coli* to natural populations of *E. coli* and *Salmonella*: the mechanism posited for explaining the evolution of high mutation rates in the former might be the same mechanism responsible for the evolution of high mutation rates in the latter. There is straightforward correspondence, in the sense outlined above, between object and target: *E. coli* in the laboratory belong to the same type as *E. coli* in nature, at a fine-grained level of classification of biological types: species. *E. coli* and *Salmonella* belong to the same type at a level that is not as fine-grained but still, arguably, significantly fine-grained.[9] Thus, in the high mutation rates case, the object (experimental populations) and target (natural populations) in question differ in their exact genetic makeup and environments, but they correspond materially according to the scheme laid out above for making sense of the materiality thesis.

In the punctuated evolution case there is not such straightforward material correspondence. It is not even clear how to go about evaluating it. We know what the object is: the same set of 12 experimental populations of *E. coli* as before. But what exactly is the target? How is it classified materially? In this case, the claim was that rare, beneficial mutations sweeping to fixation explain the punctuated evolutionary dynamic in the laboratory populations and that this same process could explain punctuated evolutionary dynamics in nature. This claim rests on the experimental populations sharing properties of their evolutionary history with an arbitrary set of populations, traces of whose evolutionary history are left in the fossil record. Those properties, in particular, were (i) displaying a certain macroevolutionary trend, (ii) existing in a constant environment, and (iii) having rare beneficial mutations. Note that none of these properties has to do with what the population is "made of": its phylogenetic classification or any specifics of its phenotype. How are we to assess material object–target correspondence here? Both comprise or once comprised living organisms; that is a start. But the target is not iden-

9. While phylogenetics can get messy for bacteria, *E. coli* and *Salmonella* are in any case closely related; see, e.g., Fukushima, Kakinuma, and Kawaguchi (2002).

https://doi.org/10.1086/677956 Published online by Cambridge University Press

tified in such a way that its material correspondence to *E. coli* can be straightforwardly evaluated.

Material correspondence is not necessarily characteristic of experimental object–target relationships. This case illustrates that nicely because it can be said of the exact same experimental object that it corresponds materially to one target, but it is unclear whether it corresponds materially to another or how to even evaluate such correspondence.

A separate issue, however, is this: what conditions would need to hold for the inferences in question, from the laboratory population of *E. coli* to these different targets in the natural world, to be valid? In the high mutation rates case, the inference relies at least in part on the object corresponding materially to the target, in the sense outlined above. It relies on mutator hitchhiking actually being the mechanism responsible for the high mutation rates observed in the laboratory populations and on the reasonability of assuming that the same mechanism could explain the same observation in closely related natural populations. Here the researchers seem to rely on material correspondence in the way Morgan and Guala had in mind: identifying a mechanism in a physical object of study (experimental populations) allows you to make an inference about materially corresponding targets in the natural world (natural populations in the same biological/phylogenetic class: asexual pathogenic microbes).

In the punctuated evolution case, material correspondence is not playing such a role. There the focus is on evolutionary mechanisms and environmental particulars, not what the target populations are "made of" in anything like the material correspondence sense. For the punctuated evolution inference to be valid, the population-level mechanisms and environmental conditions responsible for the evolutionary dynamics in the laboratory populations would need to correspond to mechanisms and conditions responsible for evolutionary dynamics in the relevant natural populations. But physical, physiological, or phylogenetic particulars of the populations in question are not figuring in to the researchers' reasons to think that is the case.

I agree with Parker (2009) that material object–target correspondence does not necessarily entail greater epistemic value. The above considerations establish a further point: it does not even follow from the fact that we have a material system as our object of study (that is, we are doing an experiment) that material correspondence is doing, or is even meant to be doing, the work in validating an inference. If we want to assess inferential power in cases like the punctuated evolution one, we need to look somewhere other than success or failure at achieving material correspondence.

The picture is further complicated because proponents of the materiality thesis compare experiment to simulation in the broader sense (Guala 2002; Morgan 2005), while its opponents have sometimes followed suit (Winsberg 2010) and sometimes focused only on computer simulation (Parker

https://doi.org/10.1086/677956 Published online by Cambridge University Press

2009). I mentioned in section 1 that it is particularly difficult to establish a difference between experiment and simulation in the broader sense. Here is why. If we consider the difference between the two examples from section 3 from the perspective of the materiality thesis, we can generate some puzzling questions. Should we think of both cases as experiments because their objects are the same kind of physical experimental system in a laboratory (we call the research area "experimental evolution," after all)? Should we think of the high mutation rates case as an experiment and the punctuated evolution case as a simulation in the broader sense? The rationale for this might be that the former takes the object as an instance of some target and makes inferences grounded in material (and, in this case, phylogenetic) similarities, while the latter arguably takes the object as more of a representation or model and makes inferences grounded in more formal similarities. Or should we think of both cases as simulations in the broader sense because the object in question is a population of model organisms, a kind of concrete theoretical model?[10]

Plausible arguments could be made for affirmative answers to all three of those questions, and people argue about just these kinds of questions (see, e.g., discussion in Guala [2002] and n. 10). But if we are interested in accounting for how and why cases of scientific inquiry differ in inferential power, these sorts of questions are red herrings. This discussion of experimental evolution, and the different kinds of inferences that can be made from the same object of study, highlights the fuzziness of the distinction between experiment and simulation (at least in the broader sense). Whether we classify a case of scientific inquiry as an experiment, a simulation, a hybrid, or explicitly both at once, per se, should not make a difference to how we judge its inferential power. It is beyond the scope of this article to assess the validity of the particular inferences in the examples discussed above. But if one were more valid than another, the right way to think about this would be in terms of specific details of the case and how well the object of study captures features relevant to making a good inference about the target in question—not in terms of how we categorize it.

It does not follow from a case of inquiry being categorized as an experiment that any inference it makes is (actually or intended to be) grounded in its degree of material object–target correspondence. Furthermore, categorization as an experiment or a simulation alone does not tell us anything about inferential power. We should not look to the experiment/simulation distinction as a basis for in-principle judgments about inferential power. There are certainly particular contexts in which experiments put us in a

10. Some (Harré 2003; Frigg and Hartmann 2012; Weisberg 2013) say so of model organisms, in any case. Levy and Currie (2014) have recently argued that model organisms are not (theoretical) models.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

better position for making valid inferences than simulations. But it does not follow that experiments are always better generators of inferential power in principle.

5. Surprise. Even if we agree on the points I have made so far, people still want to say that there is a further difference between experiments and computer simulations that affects their epistemic value: simulations cannot surprise us the way experiments can (henceforth, I refer to this as *the surprise claim*). The thinking behind the surprise claim has to do, again, with the nature of experimental objects of study: while experimenters usually design at least some of their object's parts and properties, they never design all of them, and in some cases they design none of them, such as in some field experiments. A simulationist's object of study, however, is a model: she made or programmed it herself, so she knows all of the relevant facts about its parts and properties. It is thought that experiments, in virtue of the nature of their objects, can thus surprise us in ways that simulations cannot.

The strongest form of the surprise claim would be that simulations cannot genuinely surprise us at all. More commonly, proponents of the surprise claim have in mind the idea that simulations and experiments differ, quantitatively or qualitatively, in their capacity to surprise us. Paul Sniegowski puts it in the following terms: "Although surprises do emerge in simulations, in general what goes into a simulation is well known and surprises are not anticipated. In contrast, surprises and exceptions to anticipated results are fairly common in experimental systems" (personal communication; cited with permission). This is supporting the claim to a more quantitative difference; arguing for a more qualitative difference, Morgan (2005) says that while simulations may be able to surprise us, experiments can both surprise and confound. Focusing on examples from experimental economics, she writes:

> New behaviour patterns, ones that surprise and at first confound the profession, are only possible if experimental subjects are given the freedom to behave other than expected. . . . This potential for laboratory experiments to surprise and confound contrasts with the potential for mathematical model experiments only to surprise.[11] In mathematical model construction, the economist knows the resources that went into the model. Using the model may reveal some surprising, and perhaps unexpected, aspects of the model behaviour. Indeed, the point of using the model is to reveal its implications, test its limits and so forth. But in principle, the constraints on the model's behaviour are set, however opaque they may be, by the economist who

11. By "mathematical model experiment" Morgan means what I am calling a simulation: a study of a model (in this case, a mathematical or computational model) with some dynamic temporal element.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

> built the model so that however unexpected the model outcomes, they can be traced back to, and re-explained in terms of, the model. (324–25)

When we talk about whether and how different states of affairs can surprise us, we might mean two different things. We could be talking only about epistemic features *in us*: our reactions to those states of affairs in light of our background knowledge. Or we could be talking about features of how the world is independent of us: the properties of those states of affairs themselves, regardless of what any agent knows about them. The distinction between surprise and confoundment sounds like purely the former kind, about researchers' epistemic states regarding results borne out in their research. It is the distinction between reacting to (i) something going otherwise than expected, but in a way still consistent with one's relevant theoretical background or worldview, versus (ii) something emerging from one's object of study that is sufficiently puzzling so as to spark questioning or reevaluating one's theoretical background or worldview.

Differences in researchers' epistemic states, alone, seem like the wrong grounds for tracking a distinction between experiment and simulation. We do not want a research program's status as experiment or simulation to hinge only on facts about researchers' epistemic states; in the extreme, that would mean that the same instance of inquiry could count as an experiment for one researcher or research community and a simulation for another. I suggest shifting from talking about only our reactions, to talking as much as possible about *sources of surprise* in the object of study itself.[12]

There are (at least) two relevant kinds of sources of empirical surprise. The first is *unexpected behaviors*: surprising states or phenomena in one's object of study exhibited over the course of studying it. The second is *hidden mechanisms or causal factors*: sources of surprise that in some important sense were "there all along." These are features of the object of study itself, which one was genuinely unaware of before studying it. We could talk about hidden mechanisms existing in our objects of study at different levels: (i) the individual, molecular, or atomic level; (ii) the level of interactions among individuals; or (iii) the aggregate or population level. (I remain loose with the wording here because depending on the area of inquiry, what exactly we call these levels of organization in our object of study will change. For example, the relevant levels in chemistry might range from atomic to aggregate, in ecology from individual to community, etc.)

12. Just to be clear, it is impossible to talk about sources of surprise without talking at the same time about how and why they are surprising. When I say that we should focus on features of the objects of study themselves, I am not claiming that we can do so independently of researchers' epistemic states. The distinction I am making is just between focusing on those epistemic states alone versus focusing on features of the objects themselves.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

Our background knowledge plays a role in this distinction; it is impossible to talk about differences in sources of surprise without relying on some epistemic features of the surprised agents. But against the background of relevant knowledge in a given case, the key difference between unexpected behaviors and hidden mechanisms is a difference between potential sources of surprise which (i) emerge over the course of or at the end of a simulation or experiment versus (ii) were in the object of study itself to begin with. I discuss these two kinds of sources of surprise in turn.

Unexpected behaviors are found in experiments all the time. A classic sort of example is the 2007 discovery of aluminum-42 (Castelvecchi 2007). Physicists were using a particle accelerator to test how many neutrons can bind to an atomic nucleus and colliding atoms in attempts to create magnesium-40, a very heavy isotope of magnesium. They succeeded but ended up creating a significant amount of atomic debris in the process; when they sifted through this debris they were surprised to find aluminum-42, another very heavy isotope that nobody had thought would physically exist. The appearance of aluminum-42 was an unexpected behavior of their object of study.

Another example of unexpected behavior is in the high mutation rates case from section 3. Three of the 12 populations in the Lenski experiment developed high mutation rates; Lenski and colleagues did not expect this to happen, but when it did, they drew on their theoretical background to explain what was going on.

Unexpected behaviors also occur all the time in simulations. Examples abound in the area of agent-based modeling. In agent-based models (ABMs, also known as individual-based models), individual agents and their properties are represented and the consequences of their dynamics and interactions are studied via computational simulation. Common applications of ABMs include in ecology and the social sciences, in which agents represent organisms and their interactions, locations, behaviors, life history traits, and so forth. Behavior patterns can emerge from simple initial conditions comprising agents, their properties, and their interactions, such as complex cycles of fluctuation in population size or flocking behavior (Railsback and Grimm 2011; Grimm and Railsback 2013).

One example of unexpected behavior from ABMs, keeping with the theme of studying evolving populations, is evolved predator avoidance in Avida. Avida is an ABM in which self-replicating "digital organisms" compete for resources in the form of CPU memory (Ofria and Wilke 2004). Ofria and colleagues discuss a case in which they wanted to study a population that could not adapt but would accumulate deleterious or neutral mutations through genetic drift. Digital organisms are perfect for this: researchers can examine each new mutation as it occurs by running a copy of the mutant in a test environment and measuring its fitness. The test allowed them to identify agents in the primary population with beneficial mutations and kill them off, which

https://doi.org/10.1086/677956 Published online by Cambridge University Press

would in theory stop all future adaptation. Surprisingly, however, the population continued to evolve. It turned out that the agents had developed a method of detecting the inputs provided in the test environments; once they determined that they were in a test environment, they downgraded their performance, the authors say, so as to avoid being killed. As they put it, the organisms "evolved predator avoidance."[13]

It makes sense that simulations and experiments can both involve sources of surprise in the form of unexpected behaviors, if we consider their methodological points in common. An experiment starts with choosing or designing an object of study and specifying a protocol. A simulation starts with the object of study, a model, in some initial state with a set of transition rules specifying how it will update to future states. In both cases, a researcher sees what happens to her object of study over time. The examples of unexpected behaviors I just discussed were cases of subsequent states or properties of the object of study differing in unexpected ways from its initial states or properties. There are equal opportunities for this sort of thing to happen, in principle, in both experiments and simulations.

The extreme version of the surprise claim—that a researcher cannot be genuinely surprised by her simulations because she knows everything about them—is plainly false. A modeler will often, but not always, know everything about her model's initial conditions and transition rules. A straightforward reason why she might not know everything is when she did not write the model herself and is ignorant of aspects of how it was programmed or how it works. But there are more interesting reasons, too. For example, she might write the model in a high-level programming language and fail to understand all of its low-level details. Or she might program the model in a way that leads to its initial conditions having unintended features or its transition rules entailing unintended consequences. Furthermore, very complicated models are often written by teams, rather than single modelers (such as in climate modeling); in some such cases, no individual researcher running simulations might be said to understand everything about the model's initial conditions and transition rules.

In any case, knowing "everything" about a model's initial conditions and transition rules does not entail knowledge of its future states. Setting an initial state and deciding which rules will govern its change over time does not tell you what will happen—that is why you must run the simulation. Similarly, finding out as much as you can about an experimental object of study and sorting out all the details of your protocol does not tell you what

13. One could argue about whether this is the right way to describe what went on here, but that is beside the point; the point is that unexpected behaviors can occur in simulations just as they can in experiments.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

will happen in the experiment. Any study of a system with an initial state and subsequent states has at least the potential to surprise us because it contains potential sources of unexpected behaviors as it changes (or fails to change) over time.

I now turn to hidden mechanisms or causal factors. Unlike unexpected behaviors, these are features an object of study already had, which a researcher was genuinely unaware of when she embarked on a study. A perfect example is the discovery of transposable genetic elements. Barbara McClintock was studying the genetics of maize patterns, trying to figure out what gave rise to them, and discovered that the gene regulating the maize's mottled pattern also made its chromosomes break. In the process of examining this breakage, she eventually discovered that genes can move from one place to another on the chromosome, with a sort of cut-and-paste mechanism, refuting the earlier belief that genes' positions on the chromosome are fixed (McClintock 1951). This research earned her the Nobel Prize in 1983. McClintock discovered transposable elements over the course of her studies of maize plants, but in an important sense she was discovering a hidden feature of the genome that had been there all along, which she did not know was there—nobody knew it was there.

The transposable elements case is an example of a hidden mechanism in the form of a molecular-level feature of an object of study. Simulations can also contain hidden mechanisms, at least in one sense: at the higher, aggregate level. Here is an example: the ABM Sugarscape is a simple model consisting of cells in a grid; every cell can contain different amounts of a resource (sugar). The basic setup of the model is that agents populate the grid, and with each time step, they look around for the nearest cell in their neighborhood with the most sugar, move, and metabolize. The simple local rules governing the model can give rise to population-level features that look remarkably like the macrostructures we see in societies of living organisms: structured group-level movement, carrying capacities, distributions of wealth, and migration patterns. The model's creators say of these emergent structures: "Now, upon first exposure to these familiar social, or macroscopic structures . . . some people say, 'Yes, that looks familiar. But I've seen it before. What's the surprise?' The surprise consists precisely in the emergence of familiar macrostructures *from the bottom up*—from the simple local rules that outwardly appear quite remote from the social, or collective, phenomena they generate. In short, *it is not the emergent object per se that is surprising, but the generative sufficiency of the simple local rules*" (Epstein and Axtell 1996, 51–52). Now, one might think: that is not a hidden mechanism. You had to run the model to see the macrostructures; they were not just sitting there in the initial conditions. That is true, but there is something revealing in what Epstein and Axtell say here, in the italicized last sentence: it is not the phenomena themselves that are

https://doi.org/10.1086/677956 Published online by Cambridge University Press

surprising but the fact that these simple local rules are sufficient to generate them. This object of study that looks very simple has generative properties that one would have never known about until studying it. And the interesting lessons in this case come from studying that fact and how it works, not the "familiar macrostructures" themselves. This gives us reason to think that simulations can contain hidden mechanisms of a sort.

The difference between unexpected behaviors and hidden mechanisms is another way to articulate the sort of idea I believe Morgan had in mind regarding the difference between surprise and confoundment. Namely, there are plenty of situations in which we can be surprised by unexpected behaviors, but only in special circumstances does surprise cause us to dig down and question what is actually going on with the object of study itself or realize that we did not know as much about it as we thought we did going into the research program. However, unlike Morgan I am arguing that neither form of surprise is unique to experiments—although studies of material systems arguably put us in a much better position to uncover one form of hidden mechanism, namely, the molecular, atomic, or individual-level ones.

It might be true that experiments in the end can contain a particular form of source of surprise that simulations cannot. If so, this is an important point, but it is consistent with my overall argument that we should not use the experiment/simulation distinction to make in-principle judgments about epistemic value. Surprise can be a good thing or a bad thing; it depends on the kind of scientific question one is asking. Most of our favorite stories of great scientific discoveries involve some element of surprise. Many people are drawn to science because of the prospect of discovering unexpected behaviors or hidden mechanisms. But the value of surprising results in a given context depends on what we are after. In a strict hypothesis-testing setting, good inferences come from having shown that we have eliminated sources of surprise, in a sense. This is part of the point of having controls. Surprise is certainly valuable for scientific inquiry. It plays a very important role in exploratory research, in particular. But it does not have any in-principle justificatory power that would make experiments better than simulations, even if experiments contained the only sources of genuine surprise (which they do not).

6. Conclusion. The difference between experiment and simulation matters for pragmatic reasons. Most often, doing an experiment will be more costly than doing a simulation. The supplies, reagents, and person hours needed to run a laboratory experiment tend to cost significantly more than running a model on a computer. Simulations can allow one to observe an object of study's dynamics over time much more quickly than doing so in a real-time experimental system.

This pragmatic advantage can come with epistemic costs. Many people have the intuition that it always comes with epistemic costs; this is an im-

https://doi.org/10.1086/677956 Published online by Cambridge University Press

portant part of the intuition that the materiality thesis tries to explain.[14] The idea is that studying a model as opposed to a material system involves sacrificing realism, and sacrificing realism reduces epistemic value. Again, this is a good intuition in contexts in which we know relatively little about the features of our target of inquiry relevant to designing a good experimental system or model. But science is not always operating in such contexts. One example of a context in which there is no such epistemic cost associated with simulation is the study of molecular bond angles in chemistry. We know enough about chemical bonding to answer questions via mathematical modeling and computer simulation about how atomic substitutions will affect the bond angle in a given molecule; for example, swapping atoms in the molecular backbone of DNA (as in Denning and MacKerell 2011). For answering questions about straightforward atomic substitutions in familiar molecules, we would not be in a better epistemic situation were we to carry out the relevant experimental manipulation (and it would certainly be far more pragmatically costly). So again, the point about the epistemic costs of simulation is a point that holds in many contexts. But it is not an in-principle epistemic difference between experiment and simulation.

The methodological difference between experiment and simulation is not purely pragmatic. It matters for making judgments about epistemic value—but only in a context-sensitive way. All of science is about engaging with some object of study to learn about some target of inquiry, and very rarely are the object and target identical.[15] We should not look to the experimental/simulation distinction alone to tell us anything in principle about the epistemic value of cases of scientific inquiry. Developing a full account of where we should look instead is beyond the scope of this article. But I will say a few words in closing about which kinds of considerations should go into such an account—in particular, the context-sensitive role that the experiment/simulation distinction should play in thinking about it.

The experiment/simulation distinction is relevant to judgments about epistemic value. But we need more information before we can conclude that, in any given context, one methodology has epistemic privilege over the other. There are several kinds of considerations that play key roles in our ability to draw such context-specific conclusions: first, how much and what we already know about the object and target. When we have little background knowledge, a physical sample of our target or a close approximation is often the best starting point. In some contexts, we know enough to build reliable simulations precisely because we have enough information from the world already. The molecular bond example discussed earlier is such a context. Experiments

14. I am grateful to an anonymous reviewer for highlighting this point.

15. See Sugden (2000) for an excellent discussion of inferences from models to the world outside the model and parallel points about experiments.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

tend to have epistemic privilege when we know very little; this is not the case in contexts in which we know enough to build reliable simulations to answer certain sorts of questions. It is also not the case in situations in which studying a physical approximation of the target would be unfeasible (such as large-scale climate studies; see Parker 2009) or certain kinds of physics at the nanoscale (see Lenhard 2006).

Other considerations include the related issues of (i) the importance of realism versus control and (ii) what sort of object-target correspondence we have reason to believe is most relevant to validating an epistemically responsible inference about the target. For addressing certain kinds of questions, realism and material correspondence seem paramount. For other kinds of questions, they do not. If we are interested, for example, in specifics of the molecular mechanisms of mutation in populations of *E. coli* in a particular environment, this seems like the kind of case in which material object–target correspondence might matter a lot: we care about the physiological and phylogenetic particulars of the organisms themselves, and the physical particulars of their environment. Contrast this with cases like the following, in which the goal is to understand the relative importance of different evolutionary processes in populations with high mutation rates. In a recent paper, Keller, Wilke, and Bull (2012) discuss computer simulations aimed at understanding which processes are responsible when null models of mutation-selection balance fail to predict a population's fitness equilibrium. They explain their choice of object of study as follows: "we do not pretend that our model captures any biological system. The property that is most appealing is a fitness landscape in which many different biological properties can evolve" (2308). In a case like this, control of certain high-level features is paramount, and there are a number of reasons to think that studying a material system, like laboratory populations of organisms, would put researchers in an epistemically worse situation with respect to answering the particular question at hand, namely, When the null models fail to predict fitness equilibrium, what sorts of other evolutionary processes might be responsible? This is because they would be sacrificing much-needed control for arguably unneeded realism (for example, it would be hugely more difficult to identify and measure the fittest genotypes in the population). This is where the kind of intuition underlying the materiality thesis comes in: if we are asking a scientific question that relies particularly on physical, physiological, or phylogenetic object-target correspondence, experiments are the best route to valid inferences. It is the conditional that is key here: not all scientific questions rely on such correspondence to achieve valid inferences about the target in question; in fact, some explicitly have goals that conflict with such correspondence.

Much more needs to be said to account for how we should evaluate the success or failure of scientific objects of study at informing us about targets

https://doi.org/10.1086/677956 Published online by Cambridge University Press

of inquiry, and the validity of inferences from the former to the latter. But in closing, I hope to have given convincing reasons why we should not look to the experiment/simulation distinction to tell us anything in principle about such evaluations, without looking further. I have shown that two senses in which experiments are commonly thought to have epistemic privilege over simulations—inferential power and capacity to generate surprises—do not generalize across science. Studying a material system as opposed to a computer model does not automatically entail better inferences or more opportunities for uncovering surprises.

REFERENCES

Barberousse, Anouk, Sara Franceschelli, and Cyrille Imbert. 2008. "Computer Simulations as Experiments." *Synthese* 169 (3): 557–74.

Bataillon, Thomas, Paul Joyce, and Paul Sniegowski. 2013. "As It Happens: Current Directions in Experimental Evolution." *Biology Letters* 9 (1): 20120945.

Castelvecchi, Davide. 2007. "Let There Be Aluminum-42: Experiment Creates Surprise Isotope." *Science News* 172 (17): 260.

Denning, Elizabeth J., and Alexander D. MacKerell Jr. 2011. "Impact of Arsenic/Phosphorus Substitution on the Intrinsic Conformational Properties of the Phosphodiester Backbone of DNA Investigated Using Ab Initio Quantum Mechanical Calculations." *Journal of the American Chemical Society* 133 (15): 5770–72.

Elena, Santiago, Vaughn S. Cooper, and Richard E. Lenski. 1996. "Punctuated Evolution Caused by Selection of Rare Beneficial Mutations." *Science* 272 (5269): 1802–04.

Epstein, Joshua M., and Robert Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings.

Franklin, Allan. 1990. *Experiment, Right or Wrong*. Cambridge: Cambridge University Press.

Frigg, Roman, and Stephan Hartmann. 2012. "Models in Science." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. http://plato.stanford.edu/archives/fall2012/entries/models-science/.

Fukushima, Masao, Kenichi Kakinuma, and Ryuji Kawaguchi. 2002. "Phylogenetic Analysis of *Salmonella*, *Shigella*, and *Escherichia coli* Strains on the Basis of the gyrB Gene Sequence." *Journal of Clinical Microbiology* 40 (8): 2779–85.

Garland, Theodore, and Michael R. Rose, eds. 2009. *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*. Berkeley: University of California Press.

Gentile, Christopher F., Szi-Chieh Yu, Sebastian A. Serrano, Philip J. Gerrish, and Paul D. Sniegowski. 2011. "Competition between High- and Higher-Mutating Strains of *Escherichia coli*." *Biology Letters* 7 (3): 422–24.

Gould, Stephan J., and Niles Eldredge. 1977. "Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered." *Paleobiology* 3 (2): 115–51.

Grimm, Volker, and Steven F. Railsback. 2013. *Individual-Based Modeling and Ecology*. Princeton, NJ: Princeton University Press.

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.

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

Harré, Rom. 2003. "The Materiality of Instruments in a Metaphysics for Experiments." In Radder 2003, 19–38.

Humphreys, Paul. 2004. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press.

Keller, Thomas E., Claus O. Wilke, and James J. Bull. 2012. "Interactions between Evolutionary Processes at High Mutation Rates." *Evolution* 66 (7): 2303–14.

https://doi.org/10.1086/677956 Published online by Cambridge University Press

Lenhard, Johannes. 2006. "Surprised by a Nanowire: Simulation, Control, and Understanding." *Philosophy of Science* 73 (5): 605–16.
———. 2007. "Computer Simulation: The Cooperation between Experimenting and Modeling." *Philosophy of Science* 74 (2): 176–94.
Lenski, Richard E. 2014. The *E. coli* Long-Term Experimental Evolution Project Site. Michigan State University. http://myxo.css.msu.edu/ecoli.
Lenski, Richard E., Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. 1991. "Long-Term Experimental Evolution in *Escherichia coli*." Pt. 1, "Adaptation and Divergence during 2,000 Generations." *American Naturalist* 138 (6): 1315–41.
Lenski, Richard E., and Michael Travisano. 1994. "Dynamics of Adaptation and Diversification: A 10,000-Generation Experiment with Bacterial Populations." *Proceedings of the National Academy of Sciences of the USA* 91 (15): 6808–14.
Levy, Arnon, and Adrian Currie. 2014. "Model Organisms Aren't (Theoretical) Models." *British Journal for the Philosophy of Science*, forthcoming.
McClintock, Barbara. 1951. "Chromosome Organization and Genic Expression." *Cold Spring Harbor Symposia on Quantitative Biology* 16:13–47.
Morgan, Mary S. 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.
Ofria, Charles, and Claus O. Wilke. 2004. "Avida: A Software Platform for Research in Computational Evolutionary Biology." *Artificial Life* 10 (2): 191–229.
Parker, Wendy S. 2009. "Does Matter Really Matter? Computer Simulations, Experiments, and Materiality." *Synthese* 169 (3): 483–96.
Peck, Steven L. 2004. "Simulation as Experiment: A Philosophical Reassessment for Biological Modeling." *Trends in Ecology and Evolution* 19 (10): 530–34.
Peschard, Isabelle. 2012. "Is Simulation an Epistemic Substitute for Experimentation?" In *Simulations and Networks*, ed. S. Vaienti. Paris: Hermann.
Radder, Hans, ed. 2003. *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press.
Railsback, Steven F., and Volker Grimm. 2011. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton, NJ: Princeton University Press.
Sniegowski, Paul D., Philip J. Gerrish, and Richard E. Lenski. 1997. "Evolution of High Mutation Rates in Experimental Populations of *E. coli*." *Nature* 387 (6634): 703–5.
Sugden, Robert. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7 (1): 1–31.
Travisano, Michael, Farida Vasi, and Richard E. Lenski. 1995. "Long-Term Experimental Evolution in *Escherichia coli*." Pt. 3, "Variation among Replicate Populations in Correlated Responses to Novel Environments." *Evolution* 49 (1): 189–200.
Vasi, Farida, Michael Travisano, and Richard E. Lenski. 1994. "Long-Term Experimental Evolution in *Escherichia coli*." Pt. 2, "Changes in Life-History Traits during Adaptation to a Seasonal Environment." *American Naturalist* 144 (3): 432–56.
Weber, Marcel. 2004. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
Winsberg, Eric. 2009. "A Tale of Two Methods." *Synthese* 169 (3): 575–92.
———. 2010. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.

https://doi.org/10.1086/677956 Published online by Cambridge University Press