Synthese (2013) 190:2367–2390
DOI 10.1007/s11229-011-9976-7

How simulations fail

**Patrick Grim · Robert Rosenberger ·
Adam Rosenfeld · Brian Anderson ·
Robb E. Eason**

Received: 11 October 2009 / Accepted: 24 June 2011 / Published online: 12 August 2011

Abstract 'The problem with simulations is that they are doomed to succeed.' So runs a common criticism of simulations—that they can be used to 'prove' anything and are thus of little or no scientific value. While this particular objection represents a minority view, especially among those who work with simulations in a scientific context, it raises a difficult question: what standards should we use to differentiate a simulation that fails from one that succeeds? In this paper we build on a structural analysis of simulation developed in previous work to provide an evaluative account of the variety of ways in which simulations *do* fail. We expand the structural analysis in terms of the relationship between a simulation and its real-world target emphasizing the important role of aspects intended to correspond and also those specifically intended *not* to correspond to reality. The result is an outline both of the ways in which simulations can fail and the scientific importance of those various forms of failure.

Keywords Computer simulation · Game theory · Model · Simulation · Scientific methodology

1 Introduction

How should scientific simulations be evaluated? Our work here uses the tools of our tripartite structural analysis of simulation, keyed to the different scientific purposes

---

P. Grim · A. Rosenfeld · B. Anderson
Stony Brook University, Stony Brook, NY 11794, USA

R. Rosenberger (✉)
Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: estragon10@yahoo.com

R. E. Eason
Emerson College, Boston, MA 02116, USA

to which simulations are put (Eason et al. 2007). We start with a summary of main points from that analysis.

Simulations are employed for at least four scientific purposes: prediction, explanation, retrodiction, and what can be called ‘emergence explanation.’ A common example of *prediction* is the use of computer simulation for weather forecasting, in which data regarding current meteorological conditions, together with a programmed understanding of factors that can change those conditions, are used to predict tomorrow’s weather. An example of *explanation* is the use of a centrifugal simulation by the Army Corps of Engineers, reproducing conditions of a hurricane in small scale, in order to understand what factors led to the rupture of the 17th Street levee in New Orleans during Hurricane Katrina (Interagency Performance Evaluation Task Force 2006). Simulations are also used for purposes of *retrodiction*, in which we attempt to understand how the world was in some point in the past, supporting for example the plausibility of a specific asteroid impact fatal to the dinosaurs or a planetoidal impact that produced our Moon (e.g. Bottke et al. 2007; Canup 2004). We call a fourth purpose of simulation *emergence explanation*; the attempt to show that complex features of our world can emerge from the operation of very simple rules. Here examples include the many game-theoretic accounts of altruistic and cooperative human behavior (e.g. Axelrod and Hamilton 1981; Nowak and Sigmund 1992; Kitcher 1993; Grim 1995; Grim et al. 1998). [Links to animations and further resources regarding some of the simulations referred to can be found at www.pgrim.org/hsf.]

Simulations serving all of these purposes can be conceptualized in terms of a tripartite structure of input conditions, mechanism, and output conditions.

Input conditions refer to the starting configuration before a simulation is run. In computer simulations, input conditions include the particular values chosen for operant variables. In a computer simulation used for predicting tomorrow’s weather, for example, the input conditions include the values chosen to reflect the current wind patterns, temperature, and barometric pressure.

Mechanism refers to the means by which the simulation produces output from input conditions during the course of a run. In terms of a weather prediction simulation, the programmed differential or updating difference equations constitute the mechanism of the simulation.

Output conditions refer to the resulting configuration of the simulation after it has completed its run or at a series of specific points in the run. Output conditions represent the result of a simulation mechanism working from initial inputs, often taken to represent the causal consequence of natural forces in specific conditions. In a computer simulation of tomorrow’s weather, the output conditions are those features of the simulation that the simulators read off and use as a weather prediction.

This tripartite account emphasizes the fact that scientific simulations are structured in terms of a beginning and end point, exploiting a mathematical, computational, or physical mechanism as a means of transition. There are clearly differences between these: input and output are causally related in the case of a physical mechanism and typically (though not inevitably) succeed each other in time. Although often thought of metaphorically in similar terms, the relation between input and output in the case of mathematical functions remains purely abstract rather than temporal, logical rather than causal. Computational mechanisms occupy a middle ground, conceivable either in terms of the algorithm at issue—and thus on the model of mathematical functions—or in terms of the operation of the physical device and the temporal steps of programming. Within scientific simulation, the role of mechanism might be played by any of these. Our contention is that simulation operates in the same way, with the same scientific structure and purpose, in each case.

The tripartite account also emphasizes the ‘purpose-relativity’ of simulations; it is those operating the simulation, for the specific purposes of the simulation, who read aspects of a process as ‘input,’ ‘mechanism,’ and ‘output.’ If it is the progress of a storm between now and tomorrow that is of interest, rather than merely tomorrow’s weather at noon, it will be a series of states in the simulation that will be taken as outcome rather than merely a single state.

Purpose-relativity also extends to those aspects of the structure which are taken as points of new information—those points at which investigators look to see, relative to their purposes, what the simulation has to tell them. The tripartite structure is exploited in different ways by simulations put to the purposes of prediction, explanation, retrodiction, and emergence explanation. In all of these cases, certain aspects of the simulation are posited to correspond to ways that we know the world to be, with other aspects of the simulation taken to reveal new information about the world.

The four different uses are distinguished, however, by which parts of the structure are taken to correspond to what we already know about the world and which parts are scrutinized as sources of new information. In cases in which a simulation is used to make a *prediction*, it is the output conditions that are taken as the point of new information. Both input conditions and the mechanism of the simulation are taken to correspond to the way we know the world to be. The prediction of future states of the world is read off the output of running that mechanism with those input conditions.

Where a simulation is used in the search for *explanation* it is the mechanism that is taken to be the point of new information. To the extent that the input and output conditions accurately reflect conditions in the world, the success of the simulation’s mechanism in transforming one to the other offers a potential explanation of how natural forces in the world, targeted for explanation, may operate.

In cases in which simulation is used for *retrodiction*, it is the input conditions that are read for new information. If the output conditions correspond to the current state of the world, and if the simulation’s mechanism plausibly corresponds to ways in which we know the world to work, the input conditions indicate a possible previous state of the world.

In the case of *emergence explanations*, both input conditions and mechanism represent points of new information.

The structural forms of simulation appropriate to these different scientific purposes are outlined in Table 1, with X marking aspects of posited or proposed correspondence and O marking points of new information. Of course correspondence can be posited or proposed with various degrees of confidence, and the strength of our conviction in points of new information will depend in part on the conviction with which we posit the other points of correspondence.

Our aim in what follows is to use this account of the structures and purposes of simulation to respond to the critique that simulations cannot fail—that they are ‘doomed

Table 1 The relationships between the four purposes of simulations and the tripartite structure

	Prediction	Explanation	Retrodiction	Emergence explanation
Input	X	X	O	O
Mechanism	X	O	X	O
Output	O	X	X	X

X's refer to points of posited correspondence with the real world. O's refer to points of new information

to succeed.' In so doing, we hope to sketch in broad detail a means for distinguishing success from failure in simulations. We think it is clear that simulations *can* fail in scientifically important ways. They fail most crucially when a simulation fails to properly correspond with reality, but it is important to note that different uses of simulation are open to different forms of correspondence failure in different ways and with different implications. Our hope is that appreciation of those details will be a first step toward more systematic rubrics for the calibration, validation, and evaluation of scientific simulations.

Our strategy is bottom-up: starting from the perspective and terminology of those currently working in simulation, we attempt to abstract general methodological points regarding simulation in particular. A parallel might be the attempt to start with the current statistical practices of risk analysis, working from there toward better understanding and evaluation in terms of general principles in analyzing risk (see for example Mayo and Hollander 1994). Our approach contrasts with the top-down strategy more common in the philosophical literature, which starts with a general definition of 'model' applicable across theory and experiment generally, for example, deriving application in simulations as an extended case. For the top-down strategy the parallel would be the attempt to begin with axioms of probability, deriving statistics important for Biology as a special case. Both approaches can be of value, can operate in tandem, and should ultimately converge. Our attempt here is to emphasize points evident from a bottom-up approach that tend to be missed from an approach top-down.[1]

The crucial concept of correspondence is explored in Sect. 2. In Sect. 3 we develop the notion of 'intentional non-correspondence' in order to emphasize a neglected distinction between aspects of a simulation that are intended to correspond to the world and the aspects of the simulation that are intended *not* to correspond, with important consequences for evaluation. All of this allows us a discussion in Sect. 4 of the importantly different ways in which simulations can fail.

2 Correspondence

We have analyzed a variety of uses for simulation in terms of a tripartite structure. The general pattern is one in which those operating the simulation posit a correspondence

---

[1] We do diverge from common terminology among simulators by avoiding the use of the term 'model' in favor of the term 'simulation' wherever possible, simply because different technical uses of 'model' are already established in various aspects of philosophy of science. In everyday practice, those working in simulation use both the terms 'models' and 'simulations' to refer to what they build.

between some aspects of the simulation's structure and known aspects of reality, allowing them to read other aspects of this structure as corresponding to unknown aspects of reality. But precisely what *is* this crucial notion of correspondence?

Correspondence between a simulation and its target is frequently discussed in terms of isomorphism (van Fraassen 1980; Suppes 2002), partial isomorphism (Da Costa and French 2003), or similarity (Giere 2004; Teller 2001). As Frigg and Hartmann note, however, the correspondence between simulations and their targets may range from nearly-identical scale models to abstractions or idealizations (2006). We believe that, from the perspective of simulation design, correspondence is not a single relation but many—that what is really at issue is a wealth of relevant relations between representations and what is represented, and between simulations and what is simulated. Different conceptions of correspondence will be relevant to the different kinds of simulations that simulators make, and the different purposes to which their simulations are put. Different aspects of a single simulation may correspond to aspects of reality in different ways. On this approach, appreciation for the scientific function of simulations demands not that we find the one magic tie that is 'correspondence' but that we open our eyes to the many different ways in which simulations can simulate.

The notion that correspondence is a complex and plastic relationship is consistent with Mauricio Suarez's call for recognition of the intentional component in representation. Suarez argues that scientific representation, of which simulation is one species, cannot be adequately accounted for by merely appealing to some objective similarity or isomorphism between a representation and its representandum (2003). Simulation does not simulate its target all by itself. It is the designer's or user's intentions that determine what a simulation is a simulation of and what features are to be taken as corresponding with reality. The fact that intention is integral to simulation in this way, however, clearly does not mean that intention is sufficient for simulation success. A successful simulation—indeed any successful representation—must be capable of objectively sustaining the correspondence intended.

That correspondence can take many forms. What we want to emphasize is the inevitable balance between similarity and distortion in the relationship between simulations and their targets, always with an eye to relevance, in the many relations that are grouped together as 'correspondence.'

2.1 Spatial and temporal correspondence

One might simulate spatial relations by spatial relations. Precisely that type of isomorphism was intended in the Mississippi Basin Model, begun with German prisoners of war in World War II, which simulated on approximately 200 acres all crucial aspects of waterflow for an area of 1.25 million acres: rivers, tributaries, levees, dikes, floodwalls, and control reservoirs both actual and proposed (Fig. 1). Trees were simulated with folded screen wire at a height taken from aerial photographs, and the roughness of different river channels was simulated by scoring and brushing concrete channels in the simulation. The simulation was used to gauge the impact of changes in the watercourse, to estimate the necessary height of levees at particular points, to project the course of a flood and eventually, in 1972, to guide policy in response

Fig. 1 The Mississippi River Basin Model (courtesy of David Preziosi and Mississippi Heritage Trust)

to flooding on the lower Mississippi (Robinson 1992).[2] A physical simulation of San Francisco Bay constructed in the 1950s and attempts to simulate the structure of levies and watercourses in New Orleans under the impact of Hurricane Katrina offer further examples of precisely this kind of isomorphism (Huggins and Schultz 1967, 1973, Interagency Performance Evaluation Task Force 2006). [Links to animations and further resources regarding many of the simulations referred to can be found at www.pgrim.org/hsf.]

None of these spatial simulations, of course, map only space to space: they also map time to time. The Mississippi Basin Model was particularly useful because a full day of shifting water in the Mississippi could be simulated in 5 min with the model, allowing engineers to know not only where but when flooding impacts could be expected. In the other simulations it is equally true that longer intervals of time in the real targets

---

[2] Interestingly, horizontal scale in the simulation was 1:2,000 while vertical scale was 1:100. Robinson (1992) writes:

> "To be whimsical, if one were a Lilliputian resident in this model conforming to its laws, he or she would be about 3/4 inch tall, and because of the distorted scale, as thin as tissue paper."

There are two further aspects of the history of the Mississippi Basin simulation that tie it closely to our study. One is that the very idea of simulating major hydraulic systems met great initial resistance within the profession, the government, and the Army Corps of Engineers. Simulations of this sort were regarded "as mere toys for youngsters of the profession" (Robinson 1992), much as computer simulations are sometimes regarded today. The other interesting fact is that the simulation, after proving its use in simulating flood incidents and the impact of possible changes and breakdowns in the watercourse, was eventually superseded by something that is now used to do precisely the same work: a computer simulation.

map onto significantly shorter intervals of time in the physical simulation, just as large spatial expanses map onto the significantly smaller areas of the simulation.

In work just one step removed, a group affiliated with The Santa Fe Institute has constructed a simulation of population dynamics among the ancient Anasazi of the Long House Valley. These studies employ an abstract spatial simulation, imaged on a computer monitor, intended to correspond to real spatial dimensions and real geographical features on the ground (Dean et al. 1999; Axtell et al. 2002; Gumerman et al. 2003). Simulations designed to predict influenza pandemics in particular regions of the world (Ferguson et al. 2006; Committee on Modeling Community Containment for Pandemic Influenza 2006) construct an artificial space to correspond to a real space in much the same way, often with an emphasis on transportation routes and hubs.

It should be noted, however, that correspondence between a simulation and reality cannot be judged merely point by point or additively, variable by variable. There may be a correspondence in linear dimension between simulation and reality, a correspondence in weight, a correspondence in velocity, and a correspondence in surface area. If these correspondences are not appropriately coordinated with regard to the purposes of the simulation designer/user, however, correspondence overall may very well fail. It was a familiar and frustrating fact for those pursuing manned flight at the turn of the twentieth century that a range of successful flying toys failed when ‘scaled up.’ The precise character demanded for successful scaling in multiple variables became a focus area at the intersection of mathematics, physics, and engineering, with particularly valuable contributions on physically similar systems by John William Rayleigh, Riabouchinsky, and Edgar Buckingham (Sterrett 2005).

In these examples, one can begin to see how complex an affair even relatively straightforward relationships like isomorphism can be in the practice of simulation. Expansion or compression of scale (spatial, temporal, or otherwise) offer significant advantages for simulations over “real-world” target systems, but those advantages do not come without the risk of failure to adjust for the shift in scale with respect to all relevant variables. Occasionally, these variables may not be amenable to simple scaling up or down. Whether the effects of shifting scale are complicated by surface area to volume ratios, the particulate nature of the material medium of one’s simulation, or some other factor, seemingly innocuous and even apparently isomorphic distortions may demand additional distortions in order to simulate successfully.

2.2 Abstraction

Even where space in a simulation corresponds to space in reality, that correspondence may be handled at different degrees of abstraction. That is, simulations can differ in terms of whether the target to which they relate is more of a general and broad-ranging phenomenon or instead a more specific and concrete occurrence. For example, in the simulation for smallpox epidemics constructed by Cummings et al. (2004), there is no particular set of communities that is simulated, and no specific geometry to which the geometry of the simulation corresponds. What is crucial is merely propinquity and mobility of infectious agents—the changing spatial relations

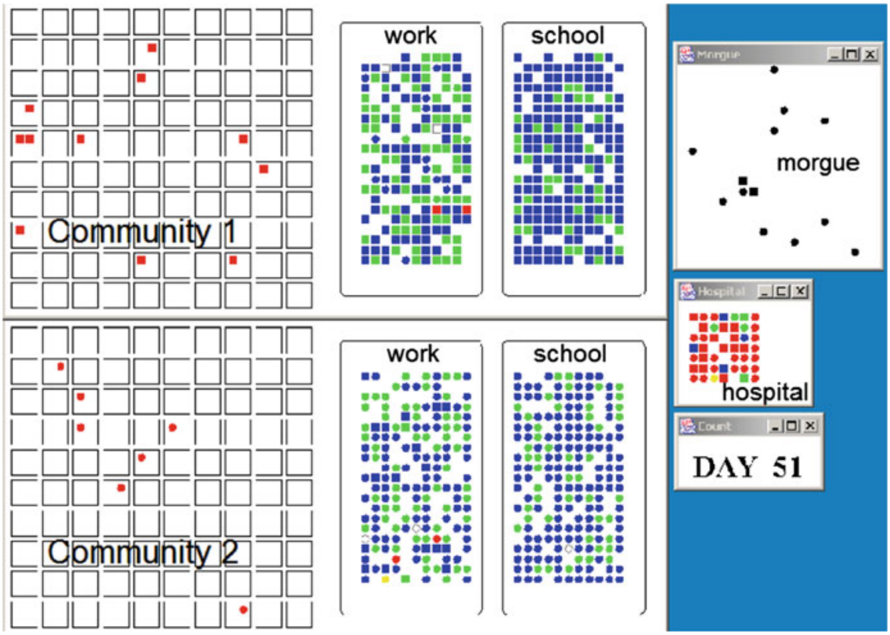

Fig. 2 Simulation of a smallpox epidemic (Cummings et al. 2004). Two communities are represented merely by spatial relation of agents, with contact merely at home, work, or school. (courtesy of J. Epstein)

of carriers and those who may become infected. Aspects of space other than ‘at home,’ ‘at work,’ ‘in the hospital,’ or ‘in the morgue’ are left out (Fig. 2). The abstraction at work in this example involves taking out many features of the world (i.e., abstracting them) so that what remains in the simulation are only general features posited to be crucial to the target phenomena. This strategy presents advantages and disadvantages. The fact that only certain aspects of space are simulated—the morgue, home, work, the hospital—weakens any claim of specific spatial correspondence with any specific community. At the same time, however, it increases the power of the simulation to explain and predict dynamics common across the different spatial specifics of many different communities.

The space of Thomas Schelling’s simulation of residential segregation is similarly abstract: pixels represent neighboring households, but with no outline of streets, parks, city centers or school districts (more on this below, Schelling 1978). The spatial dimension of our own simulation of social psychological theories of prejudice reduction may not simulate space at all: it uses propinquity in space to simulate interactive contact between people that carries both a potential for discriminatory action and for behavioral imitation. In the real human relations that are the target of the simulation, that interactive contact might be instantiated in any of various spatial relations, or in relations that are not primarily spatial at all (Grim et al. 2004, 2005).

Different purposes demand different degrees of abstraction. This is true for the examples above, in which space is used to simulate something like space, but becomes even more evident when we move to the status of agents in a simulation. In the simulations of the Anasazi of the Long House Valley, the relevant agents are envisaged not as people but as households, though the simulated households come with a complex range of plausible historical characteristics—location, marital characteristics, probability of dissolution or division by marriage of children, nutritional needs, and grain stocks, for example (Dean et al. 1999; Axtell et al. 2002; Gumerman et al. 2003). Often the agents that appear in computer simulations are envisaged as abstract individuals—'they could be anybody'—with characteristics and behaviors limited to the particular targets of the simulation. In some simulations, for some purposes, agents may be behaviorally identical; it may not be necessary to include behavioral differences from agent to agent to simulate real groups of real people. In agent-based game-theoretic simulations, for example, agents are standardly conceived merely as individual economic maximizers. The purpose of other simulations is to explore the impact of precisely the opposite assumption—here it is precisely heterogeneity of agents that is the correspondence at issue. Granovetter's simulations of collective action and Epstein's simulations of civil unrest both emphasize the impact of different patterns of heterogeneity in a population (Granovetter 1978; Epstein 2002).

2.3 Relevance

The purpose of simulation in all its forms is to capture the essentials of a real phenomenon within the constraints of a simpler structure. It is not correspondence of all elements but rather of essential elements that is crucial. In terms simply of some standard of optimized similarity over all, the only simulation fully similar to a reality would be that reality itself, which if taken as a 'simulation' would clearly defeat the practical utility of working with simulations, whether such utility involves simplifying complex processes, distorting phenomena for ease of interpretation, exemplification of important features, or simply finding more convenient approaches to costly, dangerous, or time consuming real phenomena.[3]

Whether a simulation captures the relevant aspects of a phenomenon is related both to (a) what the investigator wants to know and (b) the real mechanisms of the target phenomenon. A simulation of protein-folding designed to indicate the role of a particular amino acid, for example, will contain relevant features only if (1) it tracks a representation of the amino acid in question and (2) within representational constraints its operation parallels the reality of protein folding. Here (1) is that aspect of relevance intimately tied to an investigator's focus: a simulation which did not track the amino

---

[3] This theme appears in fiction in the wonderful fragment by Jorge Luis Borges, "On Exactitude in Science":

> In time…the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations …saw that that vast map was Useless…they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars… (Borges 1998, p. 325).

Dorathy Grim was educated in a one-room schoolhouse in rural South Dakota. When she first crossed state lines, so the family story goes, she was surprised that the ground color did not change as it did on the schoolhouse map. That was a failure on her part to recognize the conventional status of certain features of a representation. Like maps, simulations also similarly represent reality in specific ways, and thus some features of a simulation can also maintain a conventional status.

acid in question would fail to represent features relevant to the investigation at hand. (2) is that aspect of relevance that is largely independent of investigator intent. A simulation that ignores either of those factors will be a failure that omits relevant features—a failure of relevance above and beyond what the investigator may either know or intend.

Evaluating correspondence requires an articulation of which parts of a simulation's structure are intended to correspond to which parts of reality, always with an eye to relevance. What makes that particularly tricky is the fact that relevance is itself double-sided, tied both to the intentionality of representation and to how reality actually functions, independent of intent, within the constraints of such a representation.

3 Intentional non-correspondence

A primary question in evaluating both simulations and experiments is whether aspects intended to correspond to reality really do. An equally important question in evaluation, often overlooked, concerns distinguishing between those aspects intended to correspond to reality and those intended to *not* correspond in both simulations and experiments.

We use a highly refined and calibrated microscope to peer into the microscopic world, investigating experimentally the reaction of microscopic organisms to chemical changes in their environment. In both experimental and control conditions, the microscope, the lab table, and the fact that the organisms have been moved from a pond to a slide are all aspects of the experiment implicitly or explicitly understood *not* to factor into the experiment's result. We posit that the organisms will continue to swim in the drop on the slide as if they remained in the pond, unmagnified, and unstudied. Lab table, microscope, slide and observer are all aspects of the experiment that are understood to *not* correspond to the reality of pond life with which we are ultimately concerned, but this non-correspondence is taken not to impact the experimental outcome.

In any of the various senses of correspondence considered above, there are aspects of a computer simulation that are intended to correspond to aspects of reality. In simulating ecological pressures and the history of settlement in the Long House Valley, the artificial topography programmed into the simulation and displayed on the screen is intended to correspond to the real contours of the valley (Gumerman et al. 2003). The agents of the simulation are intended to correspond to individual households. The representation of maize productivity and caloric intake needs of individuals are matched as closely as possible to known data. Other aspects of the simulation are *intentionally* non-corresponding features. Settlement patterns are represented in terms of small square regions, but no-one thinks households or people are square. Rainfall and temperature are invariant across the terrain at any one time, though no-one thinks they ever were. The simulation evolves over the course of minutes, although the whole purpose of the simulation is to understand events that took place over centuries. These are aspects of the simulation that are understood to not correspond to reality but whose lack of correspondence is understood to make no relevant difference. The purposes of simulation, like the purposes of theory generally, demand simplification. Square people, perfectly distributed weather events, and mismatches of temporal scale are merely part of that simplification.[4]

In the logic of argumentation based on either experiment or simulation, intentionally non-corresponding features should not influence results. The chemistry of glass slides is understood not to make any difference to the chemistry of pond water held on a glass slide. If that postulation is false, experimental results will be unreliable. The precise shape of agents in many spatialized game-theoretic simulations, simultaneous updating, the number of neighbors in a lattice, and even precise payoffs (within the Prisoner's Dilemma, for example) have often been posited to not to make a difference to simulation results (Nowak and Sigmund 1992; Nowak and May 1993; Grim 1995, 1996). For cases in which those postulations are false, simulation results will not be what they pretend to be, reflecting instead mere artifacts of the simulation process.[5]

One way to demonstrate that specific intentionally non-corresponding features of a simulation do not factor into the results is by proving the simulation to be robust across variations, with no relevant differences despite deliberate variation over repeated runs. Spatialized game-theoretic simulations, for instance, can and should be run with different lattice structures, differences in updating schedules (synchronic and asynchronic), differences in numbers of neighbors, and differences in payoff matrices (Grim et al. 2006, 2008). Sensitivity analysis of this type is growing in use and sophistication (Chattoe et al. 2000). In evaluating simulations, we need more than assurance that aspects claimed to correspond to reality really do. We also need assurance that aspects purported not to correspond are as innocuous as they are posited to be.

Whether non-corresponding features are genuinely innocuous can be a matter of heated scientific debate. In the 1990s, simulators in meteorology coupled atmospheric and ocean simulations for the first time, resulting in a more accurate simulation of climate. A crucial element of that coupled simulation, however, was a Flux Adjustment—an artificial mechanism introduced so that the newly-coupled simulation would not produce 'unrealistic' data. The Flux Adjustment was a mechanism operating solely within the simulation with no purported correspondence to reality other than to ensure that the simulation would indeed simulate reality. Reaction was mixed. Some saw the Flux Adjustment as nothing more than a kludge—a convenient way to skew the simulation to output desired results. Others defended the Flux Adjustment on the grounds that artificial mechanisms belong to the very methodology of work in simulation (Küppers and Lenhard 2006).

---

[4] This combination of arbitrary designation coupled with other features that are not so freely assigned is reminiscent of Kendall Walton's (1990) discussion of the relationship between "props" and "principles of generation" in his *Mimesis as Make Believe*. By Walton's account, while any prop can be designated by fiat as standing for something, there are still more or less definite rules which determine what he refers to as "fictional truths." Walton's discussion of mimetics is explicitly concerned with artistic representation and fictions, and while we are open to the relationship between scientific simulation and fiction, it is beyond the purview of this essay. For more on the link between fiction and science, see *Fictions in Science* edited by Mauricio Suárez (2009).

[5] In criticism of Nowak and May's assumption of simultaneous updating, for example, see Huberman and Glance (1993).

The virtues of intentional non-correspondence, usually addressed under terms like ‘idealization,’ ‘abstraction,’ ‘exemplification,’ or even ‘distortion’ have been fairly well rehearsed in philosophy of science literature (e.g. Cartwright 1983, 1999; Elgin 2009; van Fraassen 2008). Many of the accounts of the relationship between the truth of scientific theories and their explanatory power are often couched in anti-realist commitments. However, the general point that an effective explanation involves representations which employ, at the very least, some sort of idealization, abstraction, or *ceteris paribus* qualifications is one which is compelling for scientific realists and anti-realists alike. Thus, no matter what our ontological commitments to unobservable entities are, a representation of a real-world entity qualifies as a representation not despite the fact that it lacks certain features of that entity, but precisely *because* it is lacking those features. Experiments also exemplify the salient features of a phenomenon under investigation by intentionally leaving out putatively non-essential features. Whether one is inclined to think of a simulation as a representation of a theory, or as more like an experiment, the point holds equally well; a certain degree of intentional non-correspondence is not only something that may be tolerated in scientific representation, but is an essential virtue of it.

4 How simulations fail

Failure is crucial to scientific procedure and scientific progress. It is crucial to the scientific status of theories, hypotheses, conjectures, methodologies and experimental tests that they can fail. If there is a similar place for simulation in science, simulation must be capable of failure as well.

Rodney Brooks reflects the views of many in saying that ‘the problem with simulations is that they are doomed to succeed.’ The purpose of the quip is to question the possibility of simulation failure and thereby the scientific value of simulation.[6] Where simulational techniques are impugned, it is often on precisely these grounds: on the grounds that ‘you can prove anything with simulation’, or ‘you can produce any result you want by tweaking parameters.’ If simulations were genuinely immune from failure, they would be scientifically useless. Simulators face these concerns differently in different fields, and in the fields in which simulational techniques are regarded with suspicion, this kind of argument is encountered.

The truth is that simulations can fail miserably. They can fail in all the ways that theories can fail, we want to argue, and in all the ways that experiments can fail as well. Properly handled, prospects for simulational failure are as great or greater than prospects for other aspects of science. An understanding of the many ways in which

---

[6] A related criticism is that simulation amounts to ‘fact-free science.’ John Maynard Smith, for example, writes:

> …I have a general feeling of unease when contemplating complex systems dynamics. Its devotees are practicing fact-free science. A fact for them is, at best, the output of a computer simulation: it is rarely a fact about the world (Smith 1995, p. 30).

Much of what we say about correspondence and failure of correspondence applies to this form of the critique as well.

simulations can fail offers both a better grasp of appropriate evaluation regarding simulation and a deeper understanding of how simulations fit within the scientific enterprise as a whole.

There is of course a consolation-prize sense in which attempts at simulation, at theory, and at experimentation are all ‘doomed to succeed.’ Incorrect and even wrongheaded theories, like disappointing and even badly designed experiments, can form historically important parts of a line of scientific research. Even the most disastrous venture in theory, experiment, or simulation can convince us that a particular approach does not work, thereby ‘succeeding’ in furthering the more general purposes of research by eliminating an unfruitful line of inquiry.

In speaking of the success or failure of a theory, an experiment, or a simulation, we generally have something much narrower and more specific in mind. In the broadest and most naïve terms, science is an attempt at grasping an external reality. Various aspects of science can be judged in terms of success or failure in grasp. A theory is designed to capture a general truth about some phenomenon. Indications that it does not—that reality is not as advertised in theory—are indications that the theory fails in its purpose. An experiment is designed to reveal an aspect of reality to observation. Indications that the experiment does not genuinely reveal, or does not reveal what it purports to, are indications of its failure as an experiment. Aspects of a scientific simulation are also intended to represent aspects of a larger reality beyond the simulation. Computer simulations are designed as simulations of a reality beyond and independent of their own computational algorithms. Failure to represent that independent reality in relevant respects constitutes simulation failure.

In each case—that of theory, experiment, and simulation—it should be noted that there is a crucial element of intentionality or epistemic purpose. A theory is proposed as targeting a particular phenomenon, and it is that intended phenomenon, not some other, that is used to measure its success or failure. It is experimenters who propose an experiment as revealing a specific targeted phenomenon, and it is in terms of that intentional target that success or failure is measured. Simulations have the targets that they do by virtue of the intentions of their users, and success or failure occurs in terms of how well a simulation represents the target given its intended purpose.

4.1 Failure at the point of new information

The tripartite structure outlined for simulation in the introduction, applied with an eye to the variety of purposes to which simulations can be put, allows a first outline of ways in which simulations can fail.

In the three main uses of simulation (prediction, retrodiction, and explanation), it is different points of the tripartite structure—input, mechanism, and output—that are the focus of attention as points of potential new information. If what a simulation offers at that crucial point clashes with reality rather than corresponds to it, there is a clear sense in which the simulation is a failure.

Predictive simulations can fail in precisely the way predictive hypotheses can fail: the prediction generated as output can fail to fit the facts. How close a prediction has to be to count as a ‘fit’ will undoubtedly vary with context, but a weather simulation that predicts a snowstorm in Hawaii for a month in which the temperature never dips below ninety degrees will, for almost all purposes, count as a failure. If the focus of prediction is a point far in the future—a prediction as to whether the universe will end in heat death, a big crunch or a big bounce, for example—any verdict of this sort regarding the simulation will have to be as indirect as other evidence regarding the universe’s fate.

In the absence of the final verdict of a conclusive test, we may judge the plausibility of a predictive simulation in terms of the plausibility of its prediction. We have suffered no widespread contemporary influenza pandemic on the order of the Spanish flu epidemic of 1918. Because we have no contemporary case for comparison, the output of predictive computer simulations regarding the impact of closing airports, hospitals, or schools in such a case must necessarily be judged in terms of plausibility (Ferguson et al. 2006; Committee on Modeling Community Containment for Pandemic Influenza 2006). We have been wrong in plausibility judgments before, however, and a predictive simulation rejected as implausible at one point might nonetheless be vindicated by later events.

A retrodictive simulation is open to failure in parallel ways. A simulation of the formation of the Moon which uses a mechanism employing standard laws of physics, builds in known paths of planets and comets, shows as output the current size and position of the Earth’s Moon, and depends crucially on a major asteroid impact at a particular point in the past as input offers a clear retrodiction that a major asteroid did hit Earth at a particular time. If that historical event did occur, the simulation is a retrodictive success. If it did not—if the Moon was formed in a very different way—the simulation fails. If we have independent evidence for an asteroid-produced moon at a particular point, we have independent evidence of simulation success. If we have independent evidence against such a hypothesis, we have evidence against the simulation that accords with it. Here too plausibility judgments can play a role: short of final historical confirmation, we may tentatively judge a simulation a success or failure in terms of independent indications of plausibility or implausibility regarding the history it offers.

Retrodictive simulations are rarely presented as giving strong conclusions of this type, however—a fact that calls for further explanation. Weather predictions are standardly put forward as predictions of what tomorrow’s weather *will* be like, not merely as indications of what the weather tomorrow *might* be like. Retrodictions of the Moon’s formation or the extinction of the dinosaurs, in contrast, are generally presented in weaker terms: as outlines of what *might* have produced the Moon or *may* have killed the dinosaurs. Retrodictive simulations are often presented merely as showing that the dinosaurs *could* have been driven extinct by climate changes, or that the size and orbit of the Moon are consistent with an asteroid impact in the remote past. Modal qualification of this type makes evaluation more difficult. If presented with modal qualifications in place, a retrodictive simulation is not necessarily a failure if the input events of its point of new information did not in fact occur. There is, we will see, a range of ways in which such a simulation can fail nonetheless.

Explanatory simulations fail if the explanations they offer are incorrect. The fact that the simulation mechanism offers a causal path from known input to known output gives reason to believe that a corresponding path in reality may explain the course of real events. If that mechanism does not in fact correspond with reality, there is a strong sense in which the simulation has failed in its purpose, though the attempt to determine such a failure may be difficult (see discussion below). It is well known that empirical data standardly underdetermine causal structure: a variety of causal structures may be consistent with a data set (Glymour et al. 1987; Spirtes et al. 2000; Pearl 2000; Woodward 2003). Despite correspondence with available empirical data regarding input and output, the causal mechanism embodied in a simulation may therefore fail to match the genuine structure of causes in the world. If so, there is something wrong with the simulation despite correspondence at input and output—a form of simulation failure that may be revealed when the same simulation structure fails to generate accurate predictions on other data, for example.

As in the case of retrodictive simulations, however, explanatory simulations are often (though not always) offered with modal qualification and in a softer spirit. Explanatory simulations are often presented not as giving us *the* mechanism of a phenomenon but as offering a *possible* mechanism, not as giving us *the* explanation for a phenomenon but as offering *an* explanation of the phenomenon.[7]

Why are retrodictive and explanatory simulations, in contrast to predictive simulations, so often articulated in qualified modal terms, as offering a possible explanation or a possible course of events rather than as giving us the explanation and the actual course of events *tout court*? That contrast is explained by the fact that simulations are structured in such a way that input and mechanism give us merely sufficient conditions for the output. Those who construct the simulation are able to show that input conditions of a particular type together with a core mechanism of a particular type produce output of a particular type. That structure does not show that *only* input conditions of that type or *only* a mechanism of that type will produce that output, and therefore explanatory simulations generally do not offer an indication that proposed inputs and mechanisms are either necessary to produce their outcome or are guaranteed to track the actual historical processes.

In the case of prediction, if we have confidence that our input conditions and our simulational mechanism correspond to reality, the sufficient-condition character of simulations assures us that our output predictions must also correspond to reality. If the input and mechanism are sufficient conditions for the output, then given the input and mechanism, the output is necessarily entailed. Hence the confidence with which we offer predictions not merely of what the weather *might* be tomorrow but of what it *will* be.

The story is different in the case of explanatory and retrodictive simulations. In explanatory simulations, even if we have confidence that the input and output of such a simulation correspond to reality, the fact that the simulated input and mechanism are sufficient for the output tells us only that the proposed mechanism *may* be the missing piece. It does not tell us that it must be, and it does not rule out alternative mechanisms. In retrodictive simulations, even if we have confidence that our mechanism and output conditions correspond to reality, a simulation will at best tell us only that a particular

---

[7] This is true of our work on prejudice, for example. See Grim et al. (2004, 2005).

input is *sufficient* to produce observed results. It will not show that such an input is a necessary condition, nor will it tell us that it had to be such an input that actually produced the observed results. The modal qualification characteristic in presentations of retrodictive and explanatory simulations therefore follows quite directly from the fact that simulations characteristically offer merely sufficient conditions from input and mechanism to output.

Occasionally even predictive simulations are qualified modally: as telling us not how things will be, but merely how they might be.[8] Qualification in that case, we suggest, cannot be justified in terms of the inherent structure of simulation; it can only be justified in terms of lack of confidence in the simulation or its components. It is when we lack confidence in our ability to accurately gauge or set relevant input conditions, or in our ability to offer a mechanism adequate to the complexities of the situation, that we weaken claims from predictions as to how things will be to mere forecasts of how they might be. Modal qualification of this form makes evaluations of prediction failure more difficult: a claim that the future might turn out a particular way is not shown to be strictly incorrect when the future does not in fact turn out that way. Modal qualification does not, however, make failure impossible: in some cases independent evidence may show not only that things did not turn out a certain way, but that they could not have turned out in the way it was predicted that they 'might.'

There is a common use of simulations that may seem removed or exempt from the failures of correspondence outlined here. As noted above, the standard structure of simulations involves input and mechanism as sufficient conditions for an effect. That structure can be exploited in offering simulations—even very simple simulations—as evidence that reality is *less* complicated than we might take it to be: in particular, that fairly minimal, widespread, or low-level conditions are sufficient for what we might otherwise take to be complex, specialized, or high-level effects.

Thomas Schelling's simple simulations of patterns of residential segregation offer a clear example. Schelling (1978) attempts to show that even low levels of ethnic preference (a wish to have 30% of one's neighbors of one's own ethnic background, for example) will predictably result in patterns of segregated housing that one might otherwise have thought to be evidence of strong racial prejudice. Our own simulations of prejudice reduction under conditions of contact attempt to show that a well-established phenomenon in social psychology may be explainable in terms of low-level game-theoretic advantage rather than the highly cognitive mechanisms of stereotype-reconstruction and friendship previously proposed (Grim et al. 2004, 2005).

In offering sufficiency simulations such as these, aspects of reality—including aspects taken to be relevant in other approaches—are deliberately left out. Do sufficiency simulations then fail in virtue of a failure of correspondence—in particular, in virtue of omitting relevant aspects of reality? The answer is that they do not *automatically* fail. The whole point of sufficiency simulations is to offer evidence that alternative explanations for real phenomena are unnecessarily complex: that features taken as crucial in those alternatives may be secondary, derivative, or unimportant. To the extent that they offer evidence that other aspects of reality are *not* relevant to

---

[8] Nicholas Rescher has termed these 'scenario projections,' sharply distinguishing them from predictions proper. See Rescher (1998, p. 40 ff).

the processes at issue, sufficiency simulations also offer evidence that they *do* capture the relevant factors. Their success at demonstrating (a) that the conditions at issue are genuinely sufficient for the effect, and (b) at defending themselves against a charge of relevant omission, are linked and simultaneous forms of success. This does not, however, make sufficiency simulations in any way immune from failure; they can fail in all the ways outlined for simulation failure above.

The recent criticisms of Schelling's simulation by Elizabeth Bruch and Robert Mare (2001, 2006) are a case in point. One criticism regards correspondence at the point of new information. Bruch and Mare argue that Schelling's results have rarely been gauged against real data regarding residential segregation, and may not fare well when they are. That is, the output of the simulation fails to correspond to reality. A second ground for criticism is artificiality: that the results of the simulation depend crucially on the fact that preference for neighbors of one's own type has a sharp cut-off, but that this does not match the reality of questionnaire results or other behavioral data. When the assumption of a sharp preference threshold is replaced with stepwise or continuum preferences, the simulation does not produce the same results. In the case of this second ground for criticism, it is the correspondence of the mechanism that is at issue. This criticism points out that it does not seem plausible that relevant real-world preference corresponds to that yielded by the simulation.

4.2 Wider failures of relevant correspondence

In all the forms we have outlined, correspondence at the point of new information is crucial if a simulation is to achieve its purpose. It is for that reason that failure of correspondence at the point of new information constitutes clear and dramatic simulation failure. But options for failure exist beyond merely failing to correspond at the point of new information. Failure of relevant correspondence at *any* point in the simulation structure is a mark of simulation failure. In a predictive simulation, for example, it is not merely the output that must correspond to reality: if it is to be a success, the input and mechanism of the simulation must correspond to reality as well, with realistic output the product of realistic input conditions and mechanism. Retrodictive and explanatory simulations are similarly open to question if relevant correspondence fails not merely at the point of new information but at any point in the structure.

Here as elsewhere, however, it is *relevant* correspondence that is crucial. The whole purpose of simulation, like theory-construction, is to simplify a phenomenon so that its reality can be comprehended. The whole idea is to leave things out, and all simplification entails loss of correspondence. For that reason it is not enough to point out that 'the reality is much more complicated' or that the simulation 'leaves things out.' Lack of correspondence constitutes failure only if the simulation leaves out relevant variables, and only if reality is more complicated in ways relevant to the goals of the simulation.

Here failures can be sketched in line with the discussion of correspondence and intentional non-correspondence in the previous sections. Whether a lapse in correspondence is due to sins of commission (including features in a simulation which play a significant role in the production of output but which are not present in the target phenomenon) or sins of omission (failing to include features in a simulation which correspond to real features which play a significant role in the target phenomenon), in order for this to bear on the success or failure of the simulation, the features in question must be demonstrably relevant. In the case of either a sin of commission or a sin of omission, failure occurs when a simulation produces its results based on features that do not correspond to the features of the phenomenon being simulated. Results in such cases will be artifactual. If arbitrary simulation choices are crucial to a simulation's results, there is reason for concern that predictions, retrodictions, or explanations offered on the basis of that simulation will be skewed or incorrect.

What matters in simulation success and failure is not correspondence per se, we have noted, but relevant correspondence. The relevance of correspondence or the lack of it may remain an open question even where the fact of correspondence or the lack of it is not. Establishing whether a demonstrable omission in a simulation is an omission of a relevant factor in reality depends on knowing which features in reality are important parts of a dynamic process. That may be far from obvious, and may in fact be one of the things being explored in the process of simulation itself.

The hint of circularity evident here should be familiar. Simulation is a form of scientific procedure, and scientific procedures quite generally have this character. In order to establish whether a causal hypothesis is true we would need to know what the causal structure of the world really is—precisely the issue we are exploring in the spinning of causal hypotheses. An experiment is inadequate when experimental conditions are not relevantly like those characteristics of the target phenomenon 'in the wild.' Establishing that they are relevantly alike depends on knowing which features in the wild operate in what ways, which may be one of the things we are attempting to explore in the process of experimentation itself. Simulation failure is as real as experimental or theoretical failure. It should not be too surprising that detection of simulation failure may sometimes be as complex as detection of experimental or theoretical failure.

There is another minor but noteworthy way in which simulations may fall short of their goal: excessive correspondence. Though different than correspondence lapse, we consider this too a form of failure. There will be features of reality that are not important for the process under investigation. If a simulation includes features that simulate these, and if those simulation features are as idle in the simulation as their correlates in reality, the simulation has included irrelevant aspects of reality. The simulation is baroque in ways it does not need to be, and perhaps harder to understand than it needs to be. This echoes our earlier comment on the virtues of intentional non-correspondence and is an issue that is not limited to simulation, but representation in general. Catherine Elgin (2009, p. 86) describes this failure of excessive correspondence with the example of the space shuttle Challenger explosion, in which engineers warned of the vulnerability of the rockets' O-rings in a faxed report. The data suggesting potential problems with the O-rings were obscured by other less-relevant data, and were consequently overlooked. This sort of failure to represent is especially noteworthy for simulations, which are often employed precisely because of their ability to highlight the relevant features of their target by abstracting them from irrelevant ones. Thus, the same potential for failure holds for simulations that do not effectively draw our attention to the relevant aspects of their target phenomena because they are needlessly complex. In extreme cases, such simulations will fail to simplify the reality they represent, and hence fail as simulations and as scientific tools. The art of simulation involves balancing the desire for intelligibility with the need for relevant correspondence.

4.3 Success and failure: theory, experiment, and simulation

Simulation is sometimes spoken of as occupying a realm somewhere between theory and experiment—like theory in its remoteness from direct contact with reality, perhaps, but like experiment in its potential for manipulation and new information. We are suspicious of any portrait of the wide repertoire of scientific exploration that attempts to divide the enterprise into two clean categories of theory and experimentation. It is nonetheless possible, and in fact helpful, to distinguish theoretical from experimental components in scientific praxis.[9]

Theories can fail if their claims are not true—if the claims they make do not correctly represent reality. Theories can also fail if the terms in which they are written fail to refer, or if their categories fail to cut nature at its joints. Phlogiston theories failed because the postulated substance released in combustion did not exist. ‘Consumption’ was a term applied to what are now recognized as a range of distinct diseases with distinct causes; the term failed to cut nature at its joints.[10]

Simulations have a clear theoretical character in that, like a theory, a simulation makes a claim about reality: the claim that target input conditions will produce certain output conditions by means of a specific mechanism. Those claims can be false, and the simulation can therefore fail as an adequate representation of its target. It can also be that the categories crucial to the simulation fail to correspond to anything in reality, fail to fully represent their target, or fail to cut the target reality at its crucial joints. Phlogiston and consumption simulations will fail for precisely the reasons that the corresponding theories fail. By virtue of their theoretical character, simulations can fail in all the ways that theories fail.

Yet simulations also have a certain experimental character, and they can also fail in all the ways that experiments fail. Experiments fail if their samples are not representative of the target population—if the small ore sample tested is not genuinely representative of the quality of the vein, if the DNA samples are not representative of the variation in the gene pool, or if those who filled out the questionnaire are not a random cross section of the voting population. They can also fail if experimental conditions are not relevantly similar to the natural conditions at issue—if the difference in pressures at the surface and underground are important, if DNA decay conditions are different in the biology lab from what they are in the open air, or if the context for viewing pornography is importantly different in the marketplace from its testing

---

[9] Examples of recent work on the distinction between simulation and experimentation include (Guala 2002; Morgan 2003; Winsberg 2003, 2009; Barberousse et al. 2009; Parker 2009). While determining the relation between simulation and experimentation and theory is certainly important, here we limit our comments to the simulation’s potential for failure with respect to the ways both experiments and theories can fail.

[10] To use a contemporary example, we do not yet know whether Alzheimer’s is one disease or many; in this case we do not yet know how to cut nature at its joints.

in college labs. These are problems of correspondence—precisely the problems we have emphasized for simulations throughout.

Additionally, experiments can fail if technique is shoddy, if human error is introduced in measurement, if analytic equipment is calibrated improperly, and if statistics are mishandled. All of these can appear in physical simulations, with parallels at various levels in the programming of computer simulations.[11] Just as any experimental apparatus must undergo a long and careful process of tuning before it is ready to perform reliable experiments, so too must a simulation be subjected to test runs in order to ensure that it is running properly. Computer programs are notoriously buggy, and simulations often crash and burn. A computer simulation can simply fail to run, producing no output at all. Just as an experiment may produce bizarre or uninterpretable results—usually attributed to something gone wrong—a computer simulation can produce gibberish, gobbledygook, or nonsense. Both experiments and simulations can go wrong by producing results that are investigator artifacts rather than genuine insights into the target reality.

Experiments can also fail in the sense that they fail to produce the results that a favored theory would lead us to expect. Since that failure may indicate a flaw in the guiding theory rather than in experimental procedure, an experimental 'failure' of this sort can also constitute a point of scientific progress, successfully disproving the theory at issue by failing to produce the theoretically predicted results. Here as elsewhere the difference between failure and success may lie in the intentions of the experimenter. The same experiment may be considered a failure or success depending on whether the experimenter's intention is to confirm a theory or challenge it. It should also be noted that it is only by virtue of our confidence in experimental techniques that we can take the failure to produce expected results as an invalidation of theory. Simulations that fail to produce expected theoretical results can function in precisely the same way, with similar ties to intention and with similar provisos regarding confidence in simulational techniques.

In their roles as the proving grounds for theories, both experiments and simulations can fail by failing to adequately instantiate the conditions required by the theory at issue. When this happens, the results of an experiment or simulation, whether successful in obtaining the predicted results or not, fail to generate any reliable knowledge regarding our theories and the world.

5 Conclusion

It is not true, then, that simulations are 'doomed to succeed.' A crucial fact about simulations is that they are intended as representations of reality, much as theories

---

[11] The identification of sources of error in experimental execution is its own line of inquiry, in which one attempts to determine whether a certain potential source of error would be consistent with observed results. Wendy Parker, following Deborah Mayo's error-statistical approach to scientific practice, points out the utility of computer simulations for determining what observable effects may result from possible sources of error in experimentation. This enables experimenters to evaluate the likelihood that these sources of error are actually present and influencing experimental results. If an experiment is simulated to include a potential source of error and the simulated output does not match observed experimental results, then the simulation provides grounds for the dismissal of this error. See Parker (2008).

are representations of reality. They can fail as any representations can fail. But it is also a crucial fact about simulations that they are intended to offer conclusions on the basis of variable manipulation under simplified conditions. In that regard, they are like experiments, and can fail in all the ways that experiments can fail.

This account of simulation failure suggests a number of general normative claims regarding the evaluation of simulations. The recognition of the tripartite structure of simulation, the specific ways that the various purposes of simulation relate to this structure, and the fact that correspondence itself takes many forms, puts a general onus on both simulators and their critics. In sharp contrast to current practice, which often merely encourages an inference of correspondence, simulators should attempt to make claims of correspondence as explicit as possible. At the same time, however, critics of a simulation must specify how lapses in correspondence constitute *relevant* failures.

Current procedure is often to leave it to the reader to simply ‘see’ the relevant correspondences. That informality may be adequate for demonstrations intended merely to be suggestive—which also have their place—but it will not be satisfactory as simulations are progressively put to serious work in attempts at convincingly complete explanation and convincingly reliable retrodiction and prediction. As experiments have become more sophisticated over the last 100 years, experimental procedures and their reporting have become explicit. As simulations have become more sophisticated over the last 10 years, the statistical tools used in their analysis have become more sophisticated as well. It is important that a sophistication similar to the statistical, and analogous to the experimental, be demanded of simulation design in terms of an explicit outline of elements of intended correspondence and the character of the correspondence intended.

Along with the general onus on simulators and their critics, the tripartite account identifies specific places where the evaluation of simulation should focus. A clear focus will be on those aspects of the simulation taken as the point of new information, dependent on the intentions of the investigators. Does new information regarding output, input, or mechanism at that point actually correspond to reality? From there the correspondence of the further structural features that produce the simulation’s results should be evaluated. It is important to be sure the simulation neither commits a sin of commission, in which the simulation’s operation involves a feature intended to correspond to reality but does not, nor a sin of omission, in which a crucial aspect of the target reality is left out of the simulation. Intentionally non-corresponding features of the simulation should be evaluated to be sure that they are not playing critical and therefore artifactual roles in the simulation’s results.

There is no simple litmus test for evaluating simulations apart from structure, context, and purpose any more than there is a simple litmus test for evaluating experiments independent of structure, context, and purpose. But sophisticated standards, rubrics, and heuristics have evolved for statistical evaluation and experimental design, and it is to be expected that similarly sophisticated tools will evolve for the design and evaluation of simulations. Simulations serve a complex variety of scientific purposes in a complex variety of ways; we have argued they do so in terms of a central structure. The development of better evaluative tools regarding simulation demands an appreciation of both the central structure and the complexities of its scientific applications.

Acknowledgments We are grateful to John Norton, Koffi Magli, Michael Baumgartner, Delphine Chapuis-Schmitz, and Mehment Elgin, Fellows with Patrick Grim at the Center for Philosophy of Science at the University of Pittsburgh, for helpful comments on an earlier draft. We are grateful to Susan Sterrett for discussion and sources regarding the Mississippi River Basin model and to Michael Weisberg for calling our attention to the San Francisco bay model.

References

Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedland, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., & Parker, M. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. In B. J. L.Berry, L. D. Kiel, & E. Eliott (Eds.), *Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling,* Proceedings of the National Academy of Sciences of the USA, (Vol. 99, Suppl. 3, pp. 7275–7279). Washington, DC: National Academy of Sciences.

Barberousse, A., Franceschelli, S., & Imbert, C. (2009). Computer simulations as expriments. *Synthese, 169*, 557–574.

Borges, J. L. (1998). *Collected fictions* (A. Hurley, Trans.). New York, NY: Penguin Books.

Bottke, W., Vokrouhlický, D., & Nesvorný, D. (2007). An asteroid breakup 160 myr ago as the probable source of the K/T impactor. *Nature, 449*, 48–53.

Bruch, E. E., & Mare, R. D. (2001). *Spatial inequality*, *neighborhood mobility*, *and residential segregation*. Los Angeles, CA: California Center for Population Research On-Line Working Paper Series.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology, 112*, 667–709.

Canup, R. M. (2004). Simulations of a late lunar forming impact. *Icarus, 168*, 433–456.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Cartwright, N. (1999). Aristotelian natures and modern experimental method. In *The dappled world*. Cambridge: Cambridge University Press.

Chattoe, E., Saam, N. J., & Möhring, M. (2000). Sensitivity analysis in the social sciences: Problems and prospects. In R. Suleiman, K. G. Troitsch, & N. Gilbert (Eds.), *Tools and techniques for social science simulation*. pp 243–273 Heidelberg: Physica-Verlag.

Committee on Modeling Community Containment for Pandemic Influenza. (2006). Modeling community containment for pandemic influenza: A letter report. Institute of Medicine of the National Academies. http://www.nap.edu/catalog/11800.html

Cummings, D. A. T., Chakravarty, S., Singha, R. M., Burke, D. S., & Epstein, J. M. (2004). *Toward a containment strategy for smallpox bioterror: An individual-based computational approach*. Washington, DC: Brookings Institute Press.

Da Costa, N., & French, S. (2003). *Science and partial truth: A unitary approach to models and scientific reasoning*. Oxford: Oxford University Press.

Dean, J. S., Gumerman, G. J., Epstein, J., Axtell, R. L., Swedland, A. C., Parker, M. T., & McCarrol, S. (1999). Understanding Anasazi culture change through agent based modeling. In T. A. Kohler & G. J. Gumerman (Eds.), *Dynamics in human and primate societies: Agent based modeling of social and spatial processes.* (pp. 179–206). New York, NY: Oxford University Press.

Eason, R., Rosenberger, R., Kokalis, T., Selinger, E., & Grim, P. (2007). What kind of science is simulation? *Journal of Experimental and Theoretical Artificial Intelligence, 19*(1), 19–28.

Elgin, C. (2009). Exemplification, idealization, and scientific understanding. In M. Suárez (Ed.), *Fictions in science: Philosophical essays on modeling and idealization*. New York, NY: Routledge.

Epstein, J. M. (2002). Modeling civil violence: An agent-based computational approach. *Proceedings of the National Academy of Sciences, USA, 99*, 7243–7250.

Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., & Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature, 442*, 448–451.

Frigg, R., & Hartmann, S. (2006). Models in science. *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/entries/models-science.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science, 71*(Supplement), S742–752.

Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (1987). *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*. San Diego, CA: Academic Press.

Granovetter, M. (1978). Threshold models of collective behavior. *American Sociological Review, 83*, 1420–1442.

Grim, P. (1995). Greater generosity in the spatialized Prisoner's Dilemma. *Journal of Theoretical Biology, 173*, 353–359.

Grim, P. (1996). Spatialization and greater generosity in the stochastic Prisoner's Dilemma. *Biosystems, 37*, 3–17.

Grim, P., Mar, G., & St. Denis, P. (1998). *The philosophical computer: Exploratory essays in philosophical computer modeling*. Cambridge, MA: MIT Press.

Grim, P., Au, R., Louie, N., Rosenberger, R., Braynen, W., Selinger, E., & Eason, R. E. (2006). Game-theoretic robustness in cooperation and prejudice reduction: A graphic measure. In L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone, & A. Vespignani (Eds.), *Artificial life X* (pp. 445–451). Cambridge, MA: MIT Press.

Grim, P., Au, R., Louie, N., Rosenberger, R., Braynen, W., Selinger, E., & Eason, R. E. (2008). A graphic measure for game theoretic robustness. *Synthese, 163*(2), 273–297.

Grim, P., Selinger, E., Braynen, W., Rosenberger, R., Au, R., Louie, N., & Connolly, J. (2004). Reducing prejudice: A spatialized game-theoretic model for the contact hypothesis. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. A. Watson (Eds.), *Artificial life IX* (pp. 244–249). Cambridge, MA: MIT Press.

Grim, P., Selinger, E., Braynen, W., Rosenberger, R., Au, R., Louie, N., & Connolly, J. (2005). Modeling prejudice reduction: Spatialized game theory and the contact hypothesis. *Public Affairs Quarterly, 19*, 95–125.

Guala, F. (2002). Models, simulations, and experiments. In L. Magnani & N. Nersessian (Eds.), *Model-based reasoning: Science, technology, values* (pp. 59–74). New York, NY: Kluwer.

Gumerman, G. J., Swedland, A. C., Dean, J. S., & Epstein, J. M. (2003). The evolution of social behavior in the prehistoric American Southwest. *Artificial Life, 9*, 435–444.

Huberman, B., & Glance, N. (1993). Evolutionary games and computer simulations. *Proceedings of the National Academy of Science, USA, 90*, 7716–7718.

Huggins, E. M., & Schultz, E. A. (1967). San Francisco bay in a warehouse. *Journal of the Institute of Environmental Sciences and Technology, 10*(5), 9–16.

Huggins, E. M., & Schultz, E. A. (1973). The San Francisco bay and delta model. *California Engineer, 51*(3), 11–23. http://www.spn.usace.army.mil/bmvc/bmjourney/the_model/history.html.

Interagency Performance Evaluation Task Force. (2006). *Performance evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System*: *Draft final report of the Interagency Performance Evaluation Task Force* (Vol. 1). www.asce.org/files/pdf/executivesummary_v20i.pdf.

Kitcher, P. (1993). The evolution of human altruism. *Journal of Philosophy, 90*, 497–516.

Küppers, G., & Lenhard, J. (2006). From hierarchical to network-like integration: A revolution of modeling style in computer-simulation. In G. Küppers, J. Lenhard, & T. Shinn (Eds.), *Simulation: Pragmatic construction of reality* (pp. 89–106). Dordrecht: Springer.

Mayo, D. & Hollander, R. D. (Eds.). (1994). *Acceptable evidence: Science and values in risk management*. New York, NY: University Press.

Morgan, M. (2003). Experiments without material intervention: Model experiments, virtual experiments and virtually experiments. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 216–235). Pittsburgh: University of Pittsburgh Press.

Ngenkaew, W., Ono, S., & Nakayama, S. (2007). Multiple pheromone deposition in ant-based clustering as an ant foraging concept. In S. Sahni (Ed.), *Proceedings of the 3rd IASTED International Conference*, *Advances in computer science and technology* (pp. 432–436). Anaheim, CA: Acta Press.)

Nowak, M., & May, R. (1993). The spatial dimensions of evolution. *International Journal of Bifurcation and Chaos, 3*, 35–78.

Nowak, M., & Sigmund, K. (1992). Tit For Tat in heterogeneous populations. *Nature, 355*, 250–252.

Parker, W. S. (2008). Computer simulation through an error-statistical lens. *Synthese, 163*(3), 371–384.

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

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, MA: Cambridge University Press.

Rescher, N. (1998). *Predicting the future*. Albany, NY: SUNY Press.

Resnick, M. (1997). *Turtles, termites, and traffic jams: Explorations in massively parallel micro-worlds*. Cambridge, MA: MIT Press.

Robinson, M. C. (1992). Rivers in miniature: The Mississippi Basin Model. In B. W. Fowle (Ed.), *Builders and fighters*: *U.S. Army Engineers in World War II*. Fort Belvoir, VA: Office of History, United States Army Corps of Engineers.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. New York, NY: Norton.

Smith, J. M. (1995). Life at the edge of chaos? *New York Review of Books, 42*(4), 28–30.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.

Sterrett, S. G. (2005). *Wittgenstein flies a kite*. New York, NY: Pi Press.

Suarez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science, 7*, 225–244.

Suarez, M. (Ed.). (2009). *Fictions in science: Philosophical essays on modeling and idealization*. New York, NY: Routledge.

Suppes, P. (2002). *Representation and invariance of scientific structures*. Stanford, CA: CSLI Publications.

Teller, P. (2001). Twilight of the perfect model. *Erkenntnis, 55*, 393–415.

van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective*. New York, NY: Oxford University Press.

Walton, K. (1990). *Mimesis as make-believe*. Cambridge, MA: Harvard University Press.

Winsberg, E. (2003). Simulated experiments: Methodology for a virtual world. *Philosophy of Science, 70*, 105–125.

Winsberg, E. (2009). A tale of two methods. *Synthese, 169*, 575–592.

Woodward, J. B. (2003). *Making things happen: A theory of causal explanation*. New York, NY: Oxford University Press.