
A Plea for (Good) Simulations: Nudging Economics Toward an Experimental Science

Simulation & Gaming
42(2) 243–264
© 2011 SAGE Publications
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1046878110393941
<http://sg.sagepub.com>



Julian Reiss¹

Abstract

In this article, the author argues that simulation is an undervalued technique to draw conclusions about empirical phenomena in economics. If the aim is to learn about the behavior of socioeconomic systems of interest, simulations have a variety of advantages relative to alternatives such as mathematical (pen and paper) modeling and laboratory experimentation. Therefore, the author has a good *prima facie* reason to exploit this method more fully. The author proceeds by demonstrating that frequently heard arguments against simulations are wrong, and finally the author discusses a number of more specific empirical phenomena, criticisms of one type of simulation methodology used in economics.

Keywords

advantages, agent-based modeling, computer simulations, economics, empirical phenomena experiments, methodology, simulation, simulations as experiments, socioeconomic systems

Although the number of experiments performed by economists is increasing every year, and it may be the case that “experimentalism” has finally reached economic methodology (Reiss, 2008, chap. 5), it is still true that one cannot subject whole economies to experimental control and that many smaller scale economic phenomena are inaccessible experimentally for ethical, technological, and other practical reasons. It is also true that observational studies in economics tend to suffer from what has been

¹Erasmus University, Rotterdam, Netherlands

Corresponding Author:

Julian Reiss, Faculty of Philosophy, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, Netherlands
Email: reiss@fwb.eur.nl

called the “problem of confounders” (e.g., by Steel, 2004)—the researcher’s inability to distinguish causal hypotheses by empirical means. Does this mean that those parts of the economy that are characterized by such difficulties for traditional empirical methods are inscrutable to rational evidence-based investigation?

My answer is “no,” and in this article, I argue that computer simulations can help economics on its way to a full-fledged experimental science. The reason is simple: some computer simulations are experiments. Of course, this does not mean that drawing valid conclusions about simulation models or using these conclusions to predict the behavior of real economies is a trivial business, but it does provide a strong reason why economists should be willing to explore the potential of simulation techniques.

My plea proceeds as follows: After briefly discussing various attempts to define *computer simulation* and giving my own proposal, I motivate my worry by showing that simulations are an underemployed method in economics. At least *prima facie* this seems odd as simulations are used far more frequently in other domains that share many relevant aspects with the subject matter of economics, such as complexity and lack of experimental control. I therefore discuss an attempt that has been made to explain why economists eschew simulation, but it is one thing to explain why a group of researchers rejects a method and quite another to give reasons for holding that they are justified in doing so. I discuss some arguments that could be made to that effect and show that they are unfounded. Finally, I introduce some methodological challenges faced by the simulationist and discuss proposals to overcome them.

What is a (Computer) Simulation?

Two definitions of (computer) simulation¹ dominate the methodological and philosophical literature today:

- (a) Simulation refers to the use of a computer in solving an equation that is not or cannot be solved analytically (see Frigg & Reiss, 2009, “narrow sense”; Humphreys, 1991; Küppers, Lenhard, & Shinn, 2006; Pritzker, 1979; Troitzsch, 1997, §1.1.; Winsberg, 2001)
- (b) Simulation is the mimicking of one process by another (computer) process (see Hartmann, 1996; Humphreys, 2004; Korb & Mascaro, in press; Pritzker, 1984; Troitzsch, 1997, §1.2; Zeigler, 1976)

Definition (a) has the advantage that it describes the notion that was relevant when simulations were first introduced in the sciences. An important factor in the original development of simulation techniques was that methods were sought to circumvent problems with the analytical intractability of equations describing nonlinear phenomena in the context of research into thermonuclear weapons at the Los Alamos National Laboratories (Galison, 1996; Keller, 2003). Today, however, the notion is more broadly used and includes application to so-called cellular automata (see, for example,

Wainer, Liu, Dalle, & Zeigler, 2010) and agent- or individual-based models in which the solution of equations plays no or at best an attenuated role.

Definition (b) seems to be more topical in contemporary parlance. However, that definition applies exclusively to dynamical models, and simulations are often employed in the context of investigating static and abstract objects (Lehtinen & Kuorikoski, 2007). The most important use of simulations in economics, for example, is still the examination of properties of statistical distributions in econometrics.

For the purposes of this article, I want to propose an alternative definition:

- (c) Simulations in economics explore the properties of computer-implemented models; they are aimed at drawing inferences about properties of a socio-economic system or socioeconomic systems of interest

Definition (c) is neutral on the static versus dynamic issue. It points out that simulations are used for a purpose—to learn about relevant aspects of an economy—but, for now, does not make explicit what kinds of inferences are made on the basis of simulations. I will say something more about this issue later.

The Dearth of Simulations in Economics

Economics is, to a large extent, a model-based science. Indeed, models pervade virtually all areas of economic inquiry. Naturally, theoretical models are the main engines of theoretical progress in economics, but models are just as important in applied areas, such as econometrics and economic policy. Even economic experiments usually (although not necessarily, see Guala, 2005, chap. 10) require models at some point or another.

Simulations, by contrast, are relatively rare in economics. It is certainly true that the number of publications citing simulations have become much more numerous in the past decades. In order to give some quantitative bite to this claim, I conducted a search on EconLit, a database of economic publications that covers some 400 journals, collections, books, working papers, and PhD dissertations. Between 1969 and 2006, the number of publications that had “simulation” or “simulate” (and its declensions) in the title or abstract grew from less than 30 publications per year in the first decade to more than 900 publications per year in last decade covered, or by about 4.4% per year (see Chart 1).

However, this does not mean that simulations are also widely employed in the profession. Although the percentage of articles about simulations has also steadily risen (see Chart 2), the share of simulation articles among all articles is still below 3%.² This is exacerbated by the fact that on average more than four fifths of all papers publishing simulation results are in fact Monte Carlo experiments that study properties of statistical distributions (i.e., mathematical) and only indirectly economic phenomena (Fontana, 2006).³

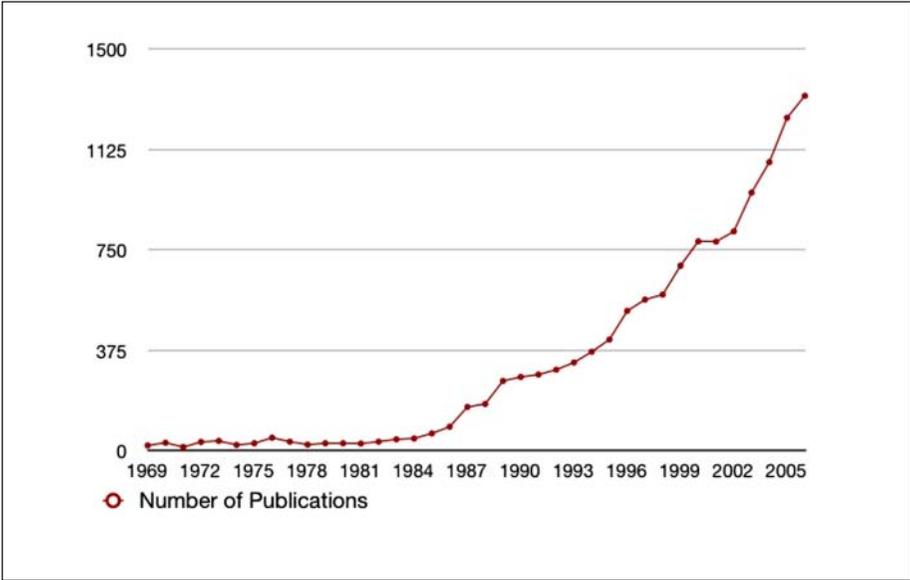


Chart 1. Number of economics publications that had “simulation” or “simulate” in title or abstract

Data obtained from EconLit—www.aeaweb.org/econlit/

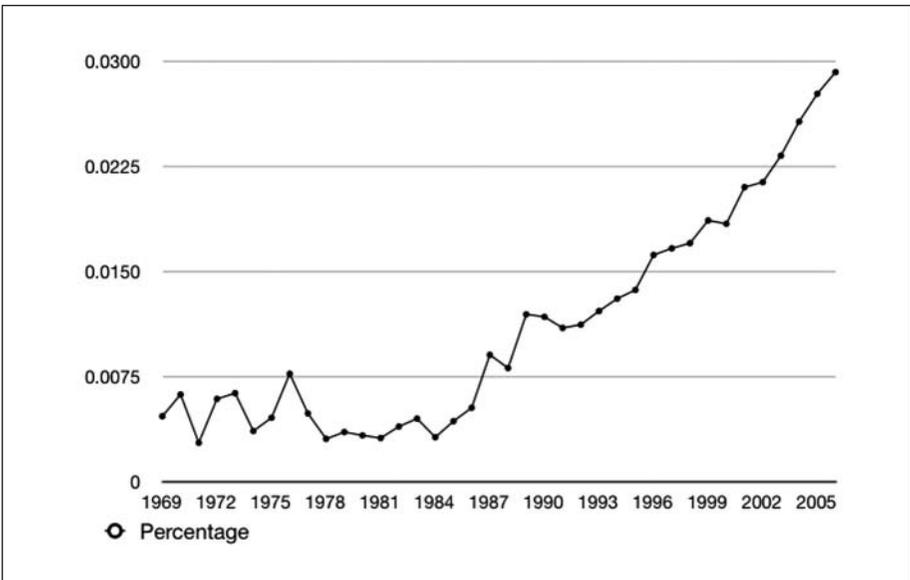


Chart 2. Percentage of economics articles about simulations

Data obtained from EconLit—www.aeaweb.org/econlit/

This is, to say the least, odd if we consider the many advantages simulations have. Paul Humphreys, for example, writes (Humphreys, 2004):

Because of their increased computational power, simulations usually allow a reduction in the degree of idealization needed in the underlying model, idealizations that often must be made in analytical modeling to achieve tractability. Because they are free from many of the practical limitations of real experiments, simulations and numerical experiments allow more flexibility and scope for changing boundary and initial conditions than do empirical experiments. (p. 115)

Simulations, thus, are said to have advantages vis-à-vis mathematical models as well as laboratory experiments. On one hand, simulations allow a larger class of empirical phenomena to be modeled because they require less stringent idealizations. On the other hand, simulations are not subject to many of the practical limitations of real experiments, which are always costly, often unethical, and sometimes impossible for technological reasons.

The potential usefulness of simulations because of the first advantage can be illustrated by a complaint that Nancy Cartwright makes about mathematical models in economics. She argues that an intended interpretation of many such models is a Galilean experiment, in which many causal factors are idealized away in order to focus on the operation of one factor all by itself (or a small number of factors by themselves). The problem is that (Cartwright, in press)

. . . most Galilean thought experiments have many more 'unrealistic' assumptions than those they should. Again, this would not be a problem if these assumptions did not play a role in deducing the final results. But of course generally they do - that is the point of including them in the first place. Just by inspection we can see that they are a necessary part of the deduction offered by the model.

In these cases I say that the results of the model are *overconstrained*. All the conditions sufficient to ensure that the model describes a Galilean experiment are met. So (pace mistakes in the driving principles) the results must be ones we would see in a real Galilean experiment. The problem is that the Galilean experiment takes place in a very special and unusual setting. What we see is indeed the result of the cause acting on its own without impediment but it is a very special result that we cannot expect in other Galilean experiments. We know we cannot expect it because we can see by inspection that the description of the special setting plays a necessary role in the derivation offered. So unrealistic assumptions that overconstrain the results are a problem for learning lessons that apply elsewhere even if the model does function as a Galilean thought experiment.

Simulations are vastly more flexible than mathematical models in that they require fewer assumptions to produce useful results (Mitchell, 2009). Simulation can thus be used in order to ameliorate the problem of overconstraining assumptions, thereby

allowing the modeling of systems that are more Galilean in nature (and, at the same time, reaping the epistemic fruit that comes with Galilean experiments).

Advantages that simulations have relative to real experiments include the following:

- the possibility of precise replication (but see Axelrod, 1997)
- the possibility of varying parameters that cannot be varied in nature
- dramatically reduced costs (financial and ethical)
- improved speed of implementation

For the sake of completeness, one might also mention that simulations have some advantages vis-à-vis thought experiments proper.⁴ About them, Humphreys says (2004),

Indeed, many simulations are examples of what would have been, in technologically more primitive times, thought experiments, as when one removes a planet from a planetary system to see how the orbits of the other planets are affected. The enormous flexibility and precision of simulation methods provide an opportunity to implement Gedankenexperiment in contexts providing much greater precision than is possible with traditional mental implementations, and they are free from the psychological biases that can affect even simple thought experiments. (pp. 115-116)

These various advantages relative to other methods provide at least a *prima facie* reason to expect simulations to play important roles on the economists' methodological landscape. Alas, this does not seem to be the case. The following section reports an attempt to explain this situation.

Why Do Economists Shun Simulations?

Simulations have various characteristics that appear to be desirable from the point of view of an economist who is interested in modeling real socioeconomics systems. However, simulations are, by and large, ignored by the community of economists. In a recent article, Aki Lehtinen and Jaako Kuorikoski (2007) have taken up the challenge to dissolve this seeming paradox. Formulating the paradox thus:

1. Simulation techniques are methods with desirable characteristics for economists
2. If methods have desirable characteristics, economists will employ them (with a high frequency)
3. Simulations are not used by economists (with a high frequency)

One can summarize Lehtinen and Kuorikoski's (2007) article by saying that they dissolve the paradox by denying the first premises: Economists do not find the characteristics of simulations desirable.

Essentially (and paraphrasing), Lehtinen and Kuorikoski (2007) argue that economists pursue two desiderata with their model building practices: certainty and (a specific kind of) understanding. Simulation delivers neither. Simulation results are not as certain as results reached by mathematical deduction—we cannot (usually) check them step by step, a large part of the work is done by what remains a black box to the economist; simulations may introduce more artifacts as they, unlike mathematical deductions, are never entirely general, but rather function by computing the results for specific combinations of parameter values.

Moreover, although techniques for verifying simulation results are available, these techniques are experimental in nature and therefore lack the conviction achieved by purely deductive methods. Nor does deriving a result by computer simulation bestow us with the kind of understanding of the model system that comes with deducing theorem?

[T]he cognitive process of solving a model [analytically!] constitutes the understanding of the model, and only by understanding the (perfect) model can “the economics” of a given social phenomenon be understood. (Lehtinen & Kuorikoski, 2007, p. 323)

Lehtinen and Kuorikoski (2007) then liken the type of understanding they argue economists seek to Philip Kitcher’s (and, one should add, Michael Friedman’s) account of understanding as unification by derivation from common argument patterns and point out that their project is descriptive, not normative:

However, we stress that this is strictly a descriptive claim, and that we in no way endorse Kitcher’s theory as a normatively cogent account of what good science *should* be like. Moreover, although Kitcher’s theory seems to be descriptive of economics in particular, we definitely do not wish to use it to defend mainstream economics. (p. 324, emphasis original)

The same is probably true of the other desideratum, certainty.

Although one may or may not endorse the claim that certainty and that particular sense of understanding are desiderata that economists actually pursue, many reasons exist to doubt their normative cogency. Increased certainty is always bought at the expense of reduced scope. An agent who requires all his beliefs to be certain before asserting or acting on them asserts and acts on little indeed. At any rate, the certainty achieved by building models of the kind that economists prefer is chimerical. Although we know that our results are true, given the assumptions, we never know where the results hold once we leave the model world (see Cartwright, 2007, chap. 15). In other words, the high degree of certainty achieved by mathematically deducing a result from a set of assumptions is paid for by the high degree of uncertainty in using the result outside the model.

Moreover, we questioned whether the understanding gained by going through the cognitive process of analytically solving a mathematical model is so useful and important. Economists certainly accept and use the results derived by other economists without always checking other researchers' proofs themselves. Why should simulations not provide understanding also? Already in the 1960s it was argued that (Buneman & Dunn, 1965)

One encounters, at times, a prejudice against computer experiments. Partly, such prejudice is based on mathematical snobbery (the formal description of the skin effect in Bessel-functions of complex argument enjoys higher prestige than a few graphs showing how it actually goes!). But often one hears the complaint that a computer can at best say "this is *how* it happens" and never "this is *why* it happens." The examples produced here should suffice to answer this complaint. The mere fact that the computer was able to produce the "how" has, many times, told us the "why." (p. 56; quoted from Keller, 2003, p. 208, emphasis original)

In the remainder of this article, I will therefore focus on defending simulations against those who accept that we can learn from experiments, but deny that simulations are experiments or that we learn from simulations in essentially the same way as we learn from other kinds of experiments rather than on defending simulations against those who seek certainty and understanding found in analytically solving mathematical models.

Simulations as Experiments

Instead of defining an experiment and then showing that simulations fall under the definition as a special case, I want to show that the epistemology of simulation is essentially an experimental epistemology. Naturally, the kinds of systems in which simulations are run—models implemented on digital computers—differ in more or less significant ways from the kinds of systems on which, say, physical experiments are performed. In a second step, I will therefore argue that this difference does not provide an in-principle reason against regarding simulation results against which to test theoretical claims.

Scientific experimentation is a kind of systematic observation aimed at drawing inferences about certain phenomena of interest. Now, at this point already an important distinction must be made. Sometimes these observations are of the phenomena of interest themselves. Observations, say, of our planetary system or of other galaxies are of this kind. Any inferences made about the systems responsible for these observations will, if true, automatically be true also of the systems responsible for the phenomena of interest because the two systems are the same. At other times, we observe one

system—a proximate or epistemically accessible system—in order to draw inferences about another system—the system we are ultimately interested in. For instance, we perform experiments on a sample to draw inferences about the underlying population; we observe the behavior of model animals to draw inferences about humans; we perform wind tunnel experiments to learn about flight behavior in the atmosphere. In these cases, we have two inferential steps: first, from the observation to the observed system and second, from the observed system to the system of interest.

A popular epistemology for the first kind of experimental inference is that of eliminating artifacts. According to this story, data—the observable outcomes of experiments, measurements, and so on—are not only the common product of numerous causal factors, including the phenomenon of interest, but also many idiosyncratic factors such as the specific details of the experimental system, the state of the observer, and so on. Inferring from data to phenomena consists in disentangling what is a property of the specific setup used to produce the observation and what is a property of the phenomenon of interest.

To give an example from social science, the difference in the test score achieved by a specific student before and after implantation of a new training program can be a data point, many of which are used to make inferences about the training program. The difference in test score may partly be due to the program, but partly also to factors of no interest to the experimenter, such as the student's ability to concentrate on the morning of the test, additional training received outside school, and the student's stage of intellectual development. The phenomenon of interest in this case is the efficacy of the training program. Information about it must be extracted from the data, for example, by averaging over a large number of students and controlling for possible confounders. Importantly, no matter how correct the results are about the test population studied in the experiment, a second step is always involved when inferences are made about other populations. When an experimental result is free of artifacts in inferring about the system on which the experiment is performed, it is said to be internally valid. When it is free of artifacts in inferring about some other system or target, it is said to be externally valid.

Especially, when observations are made using scientific instruments, epistemic strategies employed can be quite complex and involve calibration against known results, empirical examination of the equipment, triangulation (independent confirmation using different types of instrument), the elimination of or controlling for confounders, theories of the instrument, and various statistical techniques (see, for example, Bogen & Woodward, 1988; Franklin, 1986, 1990).

The literature on simulations distinguishes precisely the two mentioned inferential steps, and the strategies used to make the inferences reliable are analogous. Eliminating error in the first step is called *verification* and eliminating error in the second step is called *validation*. That is, a simulation result is said to be verified if it has been shown to be true of the model implemented on the computer. It is said to be validated if it is shown to be true of the target—the phenomenon or system of interest.

For verification, Nigel Gilbert proposes the following techniques among others (Gilbert, 2008, pp. 39f.):

- *Add assertions.* If you know that variables must take some values and not others, check for valid values as the simulation runs and display a warning if the value is out of range
- *Test with parameter values for known scenarios.* If any scenarios exist for which the parameters and the output are known with some degree of certainty, test that the model reproduces the expected behavior
- *Use corner testing.* Test the model with parameter values that are at the extremes of what is possible and ensure that the outputs are reasonable

They can all be regarded as kinds of calibration techniques. Further,

- *Observe the simulation, step by step.* Run the code one line or one function at a time, observing how the values of the variables, parameters, and attributes change and checking that they alter in the expected way
- *Use unit testing.* Unit testing is an increasingly popular software engineering technique for reducing bugs . . . It consists of writing some test code to exercise the program at the same time as you write the code itself. The idea is to develop the program in small, relatively self-contained pieces or units

These are part of the empirical investigation of the equipment. Grimm and Railsback (2005) also mention that “when software reliability is of utmost importance, it is common for two (or even more) teams to program the entire model independently and then compare both intermediate and final model results” (p. 291)—a way of triangulating results.

Testing whether the experimental system is a good model of the target is considerably less well understood. Consequently, fewer off-the-shelf techniques aimed at establishing validity are available—in the literature both on experiments and on simulations. One promising account is based on a method well known to social scientists, process tracing (Steel, 2008). The main idea is that the mechanisms responsible for the behavior of the experimental system are uncovered and then compared with the mechanisms responsible for the behavior of the target, but only at those points where the two are most likely to differ. It is then judged that greater the similarity of configuration and behavior of entities involved in the mechanism at these key stages, the stronger the basis for the inference from experiment to target (Steel, 2008).

This is precisely the methodology of validation too. Part and parcel of the process of validating a simulation is the analysis of the behavior of submodels and subsequent comparison with known facts about the target—before the various submodels are put together and aggregate patterns emerge. Other techniques known from experimentation that are also used for simulation validation are the independent prediction of results not known or used when the simulation model was built, the elimination of

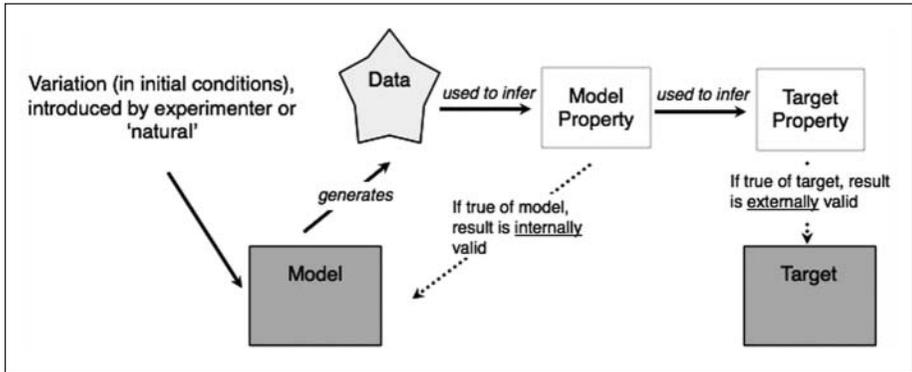


Figure 1. The joint epistemology of experiments and simulations

alternative hypotheses and sensitivity as well as robustness tests. The epistemology of simulation is the same as the epistemology of experiments (see Figure 1; the idea that experiments and simulations share an epistemology can also be found in Grimm & Railsback, 2005, chap. 9; Korb & Mascaro, in press; Peck, 2004).

Some Common Objections

Perhaps I reached my conclusion—that simulations can be used to empirically test theoretical hypotheses in the same way as experiments—a bit hastily. Extracting empirical knowledge from a computer-implemented model seems a bit like a hat trick. Instead, are simulations not tools for theory development, rather than empirical testing, and do we not require material experiments—experiments on the physical and social systems in which we are ultimately interested—to gain empirical knowledge? In this section, I will consider the following two objections:

- (a) Simulations as mere tools for theory development

A long⁵ tradition in the philosophical and methodological writing on simulation explicitly denies that simulation results constitute new empirical information.⁶ According to this literature, simulations do nothing, but explore the deductive consequences of theory. Hence, they are welcome tools for theory articulation and development, and they allow genuine learning—in so far as learning the deductive consequences of something constitutes genuine learning—but they do not teach us anything about the empirical world.

I believe that this literature has the notion of simulation wrong. Exploring deductive consequences of theories is indeed a function of some—especially Monte Carlo—simulations, but the concept is much broader.⁷ One kind of extension concerns equations that are not analytically solvable, but that require much additional input before a full simulation model can be built.

Eric Winsberg shows that between theory and final output—he calls it a model of the phenomena—a sequence of models exist with increasing degree of concreteness that use much besides theory in their construction: general physical modeling assumptions, parameters, boundary values, initial data, ad hoc modeling assumptions, approximations, degrees of freedom, discretization and coding, imaging techniques, data analysis, and interpretation (Winsberg, 1999). The simulation result or model of the phenomenon, then, does not simply represent a deductive consequence of the theory—without these additional ingredients no consequence would follow. Korb and Mascaro (in press) make a similar point about a specific example. They argue that it is in no way clear that the mechanism represented in Hinton and Nowlan's simulation result, which demonstrated the Baldwin effect, was in any sense implicit in Baldwin's original theory.

This may just be the difference between theory articulation on one hand and theory development on the other hand. Over 20 years of philosophical analysis of model building in the sciences have taught us that theories have little representational power⁸ by themselves, but they rather require idealizations, simplifications, approximations, and so forth—not much of which can be justified on the basis of theory alone.⁹ Simulation, then, does more than demonstrate the deductive consequences of theory by adding a variety of model building tools that help to expand and improve theory and make it applicable to concrete phenomena, but nevertheless the whole set of practices remains in the realm of theorizing. However, the lesson that we need to draw from the recent literature on modeling and experimenting is stronger. It is that there is no clear-cut distinction between theorizing and experimenting, speculating and observing, conjecturing and refuting—pick your favorite pair of opposites. As explained above, experiments are almost always¹⁰ done on models (i.e., on proximate, epistemically accessible systems that act as stand-ins for the systems of ultimate interest). However, if that is true, all experimental inference involves an inference from the observations on a model to the behavior of something outside the model. Experiments on theoretical models may differ from other experiments with respect to the substrate on which the experiment is performed (e.g., abstract theories, programming code, mental models, animal, mechanical, or other physical models), but not in the form of inference made when drawing conclusions about systems of interest.

Perhaps this, then, is where the crux of the matter lies: “Real” experiments are conducted on material systems, systems that display a great deal of resemblance with their targets in that they are made of the same kind of “stuff.” I will now argue that this is not so.

- (b) Simulations and real experiments have a distinct epistemology because they differ in their materiality

Mary Morgan (2003, 2004) has made some claims to this effect. Essentially, she argues that inferences from model to target are less problematic when both are made of the same kind of stuff and that the differences in materiality result in an inferential

gap, which results in reduced reliability. One way of supporting that claim would be to say that the reliability of inferences is grounded in the number of properties that model and target share. In the case of material models—such as wind tunnels and laboratory mice—a large number of properties are automatically shared. However, theoretical and computer models abstract from all material properties of the systems represented, which makes the number of shared properties necessarily smaller.

However, the reliability of inferences from model to target is not a linear function of the number of shared properties. Scientific practice is rife with counter examples to this idea. Here, I consider three types of counter example: laboratory reenactments, scale models, and animal models (see Frigg & Reiss, 2009):

Laboratory reenactments. Aerodynamic features of vehicles such as cars, ships, and planes are frequently tested on models in wind tunnels before the vehicles see the outside world for which they are designed. Although often suitable for the purposes of research and development, contexts exist where the very fact that wind tunnels are material simulacra of their targets is damaging—because walls introduce turbulence affecting airflow past models and causing systematic airflow variations, which can be ameliorated, but not eliminated by experimental design and calibration corrections. That is, although a wind tunnel is in a sense “more similar” to the target than a simulation model, we have no guarantee that inference conclusions drawn from wind-tunnel experiments will be more reliable than conclusions reached on the basis of simulations (Norton & Suppe, 2000): “Suitably done, enhanced computer modeling of data introduces vicarious control every bit as good as, sometimes superior to, traditional experimental control” (p. 72).

Scale models. Scale models are exact replicas of their targets, with the exception of size. They therefore share nearly all properties, but size matters. As Max Black has pointed out (Black, 1962),

Too small a model of a uranium bomb will fail to explode, too large a reproduction of a housefly will never get off the ground . . . Inferences from scale model to original are intrinsically precarious and in need of supplementary validation and correction. (p. 221)

Animal models. It is still standard to test drugs in animals for toxicity before the first human trials can be conducted. Animals are certainly not the same as humans, but depending on the species they share more or less genetic makeup and always materiality. However, animal models are notoriously bad predictors of toxicity in humans (see for instance Shanks, Greek, Nobis, & Greek, 2007), which causes much harm: adverse side effects, safe beneficial drugs that were never developed because of animal toxicity, loss of animal welfare, and so on. One reason for which the causal relationship between smoking and lung cancer was doubted for many years was that it does not have a counterpart in mice (Clemmensen & Hjalgrim-Jensen, 1980). With respect to the development of AIDS drug design, Greek and Greek (2000) write (<http://www.satyamag.com/march00/greek.html>, retrieved on April 23, 2009),

Computer models of HIV and AIDS have revolutionized AIDS drug design. Computer-designed nonnucleoside analogs have been shown to inhibit drug-resistant HIV. The computer model takes into account structural changes and residue characteristics resulting from mutations associated with clinical resistance to nonnucleoside inhibitors. Dr. Faith Uckun stated in *Bioorganic & Medical Chemistry Letters*, “We can predict how the AIDS virus will react to new agents that we develop with the computer model . . . we now have the opportunity to rationally design very effective drugs against the multidrug-resistant AIDS virus.”

In light of this evidence, it is clear that animal models of human disease have never been predictive or effective. Today, they are stealing money from areas of research that hold promise for a vaccine to prevent AIDS.

My argument here concerns only the strong claim that simulations must, because of their lack of materiality, be less reliable than material experiments. This is not to say that material experiments are more successful than simulations in some areas and not even that material experiments across all fields are on average more reliable. It is also true that simulation modeling requires a great deal of prior understanding of how the target works, and if we have no such understanding, then simulations are bound to fail. We should also realize, however, that inferring reliably from a material model also requires such an understanding. In particular, as mentioned earlier, we require some understanding of the mechanisms responsible for a phenomenon of interest to have a good reason to believe that a material model is a good model of a target of interest (see Steel, 2008). The question of reliable inference is therefore one that needs to be addressed case by case, on the basis of the detailed circumstances of the situation at hand and not on the basis of a priori argumentation.

Simulations Good and Bad

I have argued that the epistemology of simulation is the epistemology of the experiment, that there is no—a priori—reason to believe that simulations are less reliable than material experiments and that simulations have some advantages over mathematical models because they are more flexible. None of this means that useful simulation results are easy to have or that all simulations are equally valuable. After all, an explanation of economists’ skepticism toward simulations, besides the one mentioned by Lehtinen and Kuorikoski (2007), may be that current simulation practice in economics is problematic for reasons that have nothing to do with simulation methods as such.

In this section, I argue that this might indeed be the case. One area where simulations have become prominent is in the real business cycle literature that followed Finn Kydland and Edward Prescott’s seminal article “Time to Build and Aggregate Fluctuations” (Kydland & Prescott, 1982). Proponents of this so-called *calibration approach* use simulation techniques to analyze business cycles and draw policy implications from their analyses. Its basic methodology is as follows (see Hoover, 1995):

Firstly, a general-equilibrium dynamic model describing the time-evolution of the economy as an optimal growth path with stochastic shocks to technology is built. Concrete functional forms are chosen to capture some general features of business cycles. Then these concrete functional forms are parameterized using data from microeconomic studies or national-income accounting. When parameters cannot be specified from previous data, the model is calibrated such as to reproduce certain key variances and covariances of the data. Finally, the model is tested by generating a large number of realizations of technology shocks for the number of periods considered and variances and covariances of important variables such as output and consumption, investment, real interest rates and so on are computed and compared with actual data. (pp. 26ff.)

Kydland and Prescott's (1982) approach has been criticized as badly founded empirically on the following grounds, among others:

- The methodology does not allow distinguishing alternative causal hypotheses regarding the basic mechanism (in this case, regarding the source of business cycle fluctuations)
- There is no formal measure of *fit* or *success* of the model
- Neither estimation error nor model-specification error are part of the model's output
- No sensitivity analyses are conducted
- The use of the device of the *representative agent* leads to aggregation problems
- Independent—and false—predictions are ignored

I will now go through each of these points in slightly more detail and show—by means of a comparison with the so-called, pattern-oriented modeling (POM) approach of agent-based complex systems—that none of these criticisms concerns a feature that is essential to simulation modeling and that researchers working in this alternative paradigm have developed strategies precisely to counter criticisms of this kind.¹¹

Distinguishing Alternative Theoretical Hypotheses

Hansen and Heckman (1996) argue that the real business cycle modeler's practice of using only time-series averages as inputs, but not correlations (which are saved to test the models), makes it difficult to answer the main question, namely, how quantitatively important are alternative sources of business cycle fluctuations? They write,

Using intuition from factor analysis, it is impossible to answer this question from a single time series. From two time series, one can isolate a single common factor . . . Only using multiple time series is it possible to sort out multiple sources of business cycle shocks. The current emphasis in the literature on using

only a few “key correlations” to check a model’s implications makes single-factor explanations more likely to emerge from real business cycle analyses. (Hansen & Heckman, 1996, p. 96)

By contrast, distinguishing alternative theoretical hypotheses on the basis of the simulation results is a key feature of POM. POM is a type of agent-based modeling that uses simulations to replicate multiple “patterns”—systematic empirical features or phenomena such as a spatial mosaic pattern of successional stages of natural beech forests or the fact that the climax stage has closed canopy and little understory—for this purpose. Criticizing some studies that implement only one hypothesis—in this case, about agents’ decision-making behavior—Grimm et al. (2005) describe,

A more rigorous strategy for modeling agent decisions, or other bottom-up processes, is to use “strong inference” by contrasting alternative decision models, or “theories.” First, alternative theories of the agent’s decisions are formulated. Next, characteristic patterns at both the individual and higher levels are identified. The alternative theories are then implemented in a bottom-up model and tested by how well they reproduce the patterns. Decision models that fail to reproduce the characteristic patterns are rejected, and additional patterns with more falsifying power can be used to contrast successful alternatives. (p. 988)

No Formal Success Criterion

The success criterion employed by the Kydland-Prescott (1982) methodology is whether the second moments of selected simulated time series reproduce observed correlations. Explicitly, they reject the mimicking of the historical realization as irrelevant and do not use any statistical measure of goodness of fit. In the words of one observer (Fair, 1992),

The disturbing feature of the RBC [real business cycle] literature is there seems to be no interest in computing RMSEs [root mean square errors] and the like. People generally seem to realize that the RBC models do not fit well in this sense, but they proceed anyway. (p. 141)

Within the POM paradigm and agent-based modeling in general, a battery of statistical tests has been developed and is regularly employed to compare simulation output with observed patterns (see, for example, Grimm & Railsback, 2005, Section 9.5.4).

Parameter Value and Model Specification Uncertainty

Hansen and Heckman (1996) observe that data from the microstudies that are used as input for the business-cycle simulations are often incompatible with the general-equilibrium assumptions of the macromodel, which creates specification uncertainty. They argue that the uncertainty about model parameters should be incorporated into

the outputs of simulations. Alas, “Current practices in the field of calibration and simulation do not report either estimation error and/or model-specification error” (Hansen & Heckman, 1996, p. 98). POM researchers are fully aware of parameter value and model structure uncertainty and have developed various ways to cope with it. One way is to make models more structurally realistic, which often makes them less sensitive to parameter uncertainty (Mooij & DeAngelis, 2003). Another is to separate the analysis of parameter and structure uncertainty to some extent (Grimm & Railsback, 2005, Section 9.3.3).

Sensitivity and Robustness Analyses

If parameter values and model structure are subject to uncertainty, then sensitivity and robustness analyses help in model validation. Once more, what seems to be absent from real business cycle modeling is part and parcel of the POM methodology (e.g., Grimm & Railsback, 2005, Section 9.6).

Representative Agents

In their quest for microfoundations, Kydland and Prescott (1982) employ the fiction of the representative agent to model optimizing behavior. Hoover (1995) argues that this leads to aggregation problems:

To understand (verstehen) their [humans’] behavior, one must model the individual and his situation. [While t]his insight is clearly correct, it is not clear in the least that it is adequately captured in the heroic aggregation assumptions of the representative-agent model. The analogue for physics would be to model the behavior of gases at the macrophysical level, not as derived from the aggregation of molecules of randomly distributed momenta, but as a single molecule scaled up to observable volume—a thing corresponding to nothing ever known to nature. (p. 40)

The ability to model agent heterogeneity is, of course, one of the main advantages of agent-based modeling.

Independent Predictions

Real business cycle modelers explicitly ignore certain features that their models could be used to generate, in particular, facts about the nominal side of the economy. In private correspondence with Hansen and Heckman (1996), John Taylor has said about this practice:

I have found that the omission of aggregate price or inflation data in the Kydland-Prescott second moment exercise creates an artificial barrier between real business cycle models and monetary models. To me, the Granger causality

from inflation to output and vice versa are key facts to be explained. But Kydland and Prescott have ignored these facts because they do not fit into their models. (p. 96)

The correlation, Taylor describes, is one that could have been used to independently confirm the model on the basis of an observed fact not used in the construction of the model. The method of using additional implications of the model to confirm it is, once more, an important part of the POM approach: “Structurally realistic models can make independent and testable secondary predictions” (Grimm et al., 2005, p. 988; see also Grimm & Railsback, 2005, Section 9.9).

Conclusions

In a recent article defending an experimentalist stance toward simulations in evolution and ecology, Steven Peck (2004) argues,

Adopting the stance of simulation as experiment, currently being championed by philosophers and practitioners of science in the physical sciences, will help clarify the role that simulations can play in advancing ecology and evolutionary biology. (p. 533)

The same is true of economics. Simulations currently have bad press in economics because their characteristics are compared with theoretical models and of course simulations lack mathematical elegance, certainty of results, and analytical transparency, but simulations have a variety of advantages that are potentially highly useful for economics as science. Because they are far more flexible, they help with the problem of overconstraining assumptions that besets much mathematical modeling. They are also much more flexible, cheaper, and less ethically involved than laboratory experiments.

Of course, simulations are no panacea, although we currently witness a broad and growing literature on simulation modeling that is extremely conscious of the many methodological problems that the researcher faces when analyzing a model as a stand-in for some target system of interest, and, as I hope to have shown, initial plausibility can be gained in using simulations for economic analysis.

Acknowledgments

The author wishes to thank François Claveau, Till Grüne-Yanoff, an anonymous referee, and participants of the 2009 INEM conference in Xalapa, Mexico, for valuable comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interests with respect to their authorship or the publication of this article.

Funding

Research for this work has been supported by the research projects FFI2008-01580 and CONSOLIDER INGENIO CSD2009-0056 of the Spanish government.

Notes

1. In what follows, I will suppress the qualification *computer* and always mean by *simulation* “simulation using a digital computer.” Of course, analogue and many other kinds of simulations exist, even in economics. For instance, one could consider runs of the Phillips machine as simulations of an actual economy. However, the greatest part by far of simulations that are produced today are run on digital computers, so I will not distinguish between computer and other kinds of simulations. A potential exception is thought experiments. Arguably, thought experiments can be regarded as a type of simulation, and thought experiments are not rare in economics. These, however, demand a separate analysis, and I will exclude them here too. On thought experiments in economics, see Reiss (2008, chap. 6) and Schabas (2008).
2. As one commentator pointed out, these figures might well understate the extent to which simulations are used in economics as authors do not always emphasize the use of computers in deriving model results by putting the word in title or abstract. This is certainly true, and the data are to be understood as a rough indication rather than precise estimation of the use of the method in economics. Especially, in macroeconomics, for instance in the area of real business cycle modeling that I discuss below, simulations are certainly more frequent than these figures suggest. Fundamentally, the main point of this article is not about numerically solving equations—which is what computers do in these cases—but rather about using computers as experimental tools to draw inferences about socioeconomic systems. This latter activity remains rare in all areas of economics. See also Lehtinen and Kuorikoski (2007) who make the same observation.
3. Fontana’s figures may be somewhat biased because she classifies 44% of simulations as “unexplained.” However, even of those she does categorize, by far the greatest share comes from statistics and econometrics. For example, in 2004, nearly 300 publications using simulations were in that area, while each of the other categories (system dynamics, multilevel simulations, microsimulations, agent-based simulations, discrete events, and learning models) had fewer than 50 publications in that year.
4. Thought experiments proper—experiments on models implemented in the mind—are a relatively rare species in economics. For a discussion of these, see Schabas (2008).
5. “Long” is of course relative to philosophical writing on simulations in general, which, starting in the 1960s and properly taking off the ground only in the 1990s, is extremely young.
6. This point has been made most forcefully by Oreskes et al. (1994). See also Di Paolo et al. (2000) and, for economics, Axelrod (1997). Lehtinen and Kuorikoski (2007) do not explicitly argue in favor of this position, but they seem to presuppose it, which is made plain by the following quotations: “The dearth of simulation models is most conspicuous in the most widely respected journals *that publish papers on economic theory*” (p. 305, italics added); “Economists have historically considered physics a paradigm of sound scientific methodology . . . , but they are still reluctant to follow physicists in embracing computer simulation as an important tool in the search for *theoretical progress*”; “For some reason, true simulation is considered inferior to analytically solvable equilibrium models *in the construction of economic theory*” (p. 312).
7. Evelyn Fox Keller aptly calls these simulations *experiments in theory* (Keller, 2003). This is the kind of simulation that stood at the beginning of the development of this technique, and

they are still very important in many areas of science, but they are only one among various kinds of simulations.

8. What I mean by this term is simply the ability to represent concrete empirical phenomena.
9. This aspect of simulation is also emphasized by Humphreys (2004).
10. They are always done on models if we regard a particular—say, our planetary system at time t —as a stand-in for the corresponding type—say, our planetary system at all times and all systems that are exactly like it. I prefer to make a difference in the terminology because it is one thing to infer from the observation of a particular about the behavior of the corresponding token and a different thing to infer from the observation of a particular that belongs to Type X about the behaviour of a particular that belongs to Type Y, which differs from X or Type Y itself. Perhaps they are both problems of induction, but the first is mainly a philosophical problem, whereas the second has enormous methodological and practical repercussions.
11. One may argue that my comparison is unfair as the criticized approach was developed during the 1980s, and even the mentioned criticisms are all over 10 years old, whereas pattern-oriented modeling (POM) is very recent. However, here the point is less to criticize the calibration methodology or its proponents as such but rather to show that certain features of the calibration methodology that economists might take to be features of simulation methodology in general are in no way essential to the latter.

References

- Axelrod, R. (1997). Advancing the art of simulation in social science. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena* (pp. 21-40). Berlin, Germany: Springer.
- Black, M. (Ed.). (1962). Models and archetypes. In M. Black (Ed.), *Models and metaphors: Studies in language and philosophy* (pp. 219-243). Ithaca, NY: Cornell University Press.
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, 97, 302-352.
- Buneman, O., & Dunn, D. A. (1965). *Computer experiments in plasma physics* (Report No. SU-05254-2). Washington, DC: NASA.
- Cartwright, N. (2007). *Hunting causes and using them*. Cambridge, UK: Cambridge University Press.
- Cartwright, N. 2010, 'Models: Parables v Fables', in Frigg, R. and M. Hunter (eds), Beyond Mimesis and Convention: Representation in Art and Science, *Boston Studies in the Philosophy of Science* 262, 19-31
- Clemmensen, J., & Hjalgrim-Jensen, S. (1980). On the absence of carcinogenicity to man of phenobarbital. In F. Coulston & S. Shubick (Eds.), *Human epidemiology and animal laboratory correlations in chemical carcinogenesis* (pp. 251-65). New York, NY: Ablex.
- Di Paolo, E. A., J. Noble, and S. Bullock (2000). Simulation models as opaque thought experiments. In M. A. Bedau, J. S. McCaskill, N. H. Packard, and S. Rasmussen (Eds.), *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life*, Cambridge, MA, pp. 497 -506. MIT Press.
- Fair, R. (1992). The cowles commission approach, real business cycle theories, and new Keynesian economics. In M. T. Belongia & M. R. Garfinkel (Eds.), *The business cycle: Theories and evidence* (pp. 133-47). Boston, MA: Kluwer.

- Fontana, M. (2006). Simulation in economics: Evidence on diffusion and communication. *Journal of Artificial Societies and Social Simulation*, 9, 1-15.
- Franklin, A. (1986). *The neglect of experiment*. Cambridge, UK: Cambridge University Press.
- Franklin, A. (1990). *Experiment, right or wrong?* Cambridge, UK: Cambridge University Press.
- Frigg, R., & Reiss, J. (2009). The philosophy of simulation: Hot new issues or same old stew? *Synthese*, 169, 593-613.
- Galison, P. (1996). Computer simulation and the trading zone. In P. Galison & D. Stump (Eds.), *Disunity of science: Boundaries, contexts, and power* (pp. 118-157). Stanford, CA: Stanford University Press.
- Gilbert, N. (2008). *Agent-based models*. Los Angeles, CA: Sage.
- Greek, R. and J. Greek 2000: 'Effective or not? Animal Models of AIDS', Satya (March 2000): p. 18
- Grimm, V., & Railsback, S. (2005). *Individual-based modeling and ecology*. Princeton, NJ: Princeton University Press.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W., Railsback, S., . . . DeAngelis, D. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310, 987-991.
- Guala, F. (2005). *The methodology of experimental economics*. Cambridge, UK: Cambridge University Press.
- Hansen, L. P., & Heckman, J. (1996). The empirical foundations of calibration. *Journal of Economic Perspectives*, 10, 87-104.
- Hartmann, S. (1996). The world as a process: Simulation in the natural and social sciences. In R. Hegselmann, U. Müller, & K. Troitzsch (Eds.), *Modelling and simulation in the social sciences from the philosophy of science point of view* (pp. 77-100). Dordrecht, Netherlands: Kluwer.
- Hoover, K. (1995). Facts and artifacts: Calibration and the empirical assessment of real-business-cycle models. *Oxford Economic Papers*, 47, 24-44.
- Humphreys, P. (1991). Computer simulations. *Philosophy of Science, PSA 1990*, 497-506.
- Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford, NY: Oxford University Press.
- Keller, E. F. (2003). Models, simulation, and "computer experiments." In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 198-216). Pittsburgh, PA: University of Pittsburgh Press.
- Korb, K., & Mascaro, S. (in press). *The philosophy of computer simulation*.
- Küppers, G., Lenhard, J., & Shinn, T. (2006). Computer simulations: Practice, epistemology, and social dynamics. In G. Küppers, J. Lenhard, & T. Shinn (Eds.), *Simulation: Pragmatic construction of reality* (pp. 139-151). Dordrecht, Netherlands: Springer.
- Kydland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica*, 50, 1345-1370.
- Lehtinen, A., & Kuorikoski, J. (2007). Computing the perfect model: Why do economists shun simulation? *Philosophy of Science*, 74, 304-329.
- Mao C, Sudbeck EA, Venkatachalam TK, Uckun FM 1999, 'Rational design of N-[2-(2,5-dimethoxyphenylethyl)]-N'-[2-(5-bromopyridyl)]-thiourea (HI-236) as a potent non-nucleoside inhibitor of drug-resistant human immunodeficiency virus', *Bioorg Med Chem Lett*. 1999 Jun 7;9(11):1593-8.

- Mitchell, S. (2009). *Unsimple truths: Science, complexity, and policy*. Chicago, IL: University of Chicago Press.
- Mooij, W., & DeAngelis, D. (2003). Uncertainty in spatially explicit animal dispersal models. *Ecological Applications*, 13, 794-805.
- 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, PA: Pittsburgh University Press.
- Morgan, M. (2004). Simulation: The birth of a technology to create "evidence" in economics. *Revue d'Histoire des Sciences*, 57, 341-377.
- Norton, S., & Suppe, F. (2000). Why atmospheric modeling is good science. In C. Miller & P. Edwards (Eds.), *Changing the atmosphere: Expert knowledge and environmental governance* (p. 72). Cambridge, MA: MIT Press.
- Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). 'Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, *Science*, 263 (5147)(Feb. 4, 1994), pp. 641-646
- Peck, S. (2004). Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology & Evolution*, 19, 530-534.
- Pritzker, A. A. B. (1979). Compilation of definitions of simulation. *Simulation*, 33, 61-63.
- Pritzker, A. A. B. (1984). *Introduction to simulation and SLAM II*. New York, NY: John Wiley.
- Reiss, J. (2008). *Error in economics: Towards a more evidence-based methodology*. London, UK: Routledge.
- Schabas, Margaret 2008, 'Hume's monetary thought experiments', *Studies in History and Philosophy of Science* 39: 161-69
- Shanks, N., Greek, R., Nobis, N., & Greek, J. (2007). Animal and medicine: Do animal experiments predict human response? *Skeptic*, 13, 44-51.
- Steel, D. (2004). Social mechanisms and causal inference. *Philosophy of the Social Sciences*, 34, 55-78.
- Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford, UK: Oxford University Press.
- Troitzsch, K. (1997). Social science simulation—Origins, prospects, purposes. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena* (pp. 55-72). Berlin, Germany: Springer.
- Wainer, G., Liu, Q., Dalle, O., & Zeigler, B. P. (2010). Applying cellular automata and DEVS methodologies to DIGITAL GAMES: A survey. *Simulation & Gaming*. 41(6): 796-823
- Winsberg, E. (1999). Sanctioning models: The epistemology of simulation. *Science in Context*, 12, 275-292.
- Winsberg, E. (2001). Simulations, models, and theories: Complex physical systems and their representations. *Philosophy of Science*, 68(Proceedings), S442-S454.
- Zeigler, B. (1976). *Theory of modeling and simulation*. New York, NY: John Wiley.

Bio

Julian Reiss is an associate professor at the Faculty of Philosophy of Erasmus University, Rotterdam, and specializes in the philosophy of economics and general philosophy of science. In 2009, he was awarded the International Research Prize 2009 of his faculty among other things for the book *Error in Economics: Towards a More Evidence-Based Methodology* (Routledge, 2008).