Philos Stud (2009) 143:33–57
DOI 10.1007/s11098-008-9317-y

Models, measurement and computer simulation: the changing face of experimentation

Margaret Morrison

Published online: 24 January 2009

Abstract The paper presents an argument for treating certain types of computer simulation as having the same epistemic status as experimental measurement. While this may seem a rather counterintuitive view it becomes less so when one looks carefully at the role that models play in experimental activity, particularly measurement. I begin by discussing how models function as "measuring instruments" and go on to examine the ways in which simulation can be said to constitute an experimental activity. By focussing on the connections between models and their various functions, simulation and experiment one can begin to see similarities in the practices associated with each type of activity. Establishing the connections between simulation and particular types of modelling strategies and highlighting the ways in which those strategies are essential features of experimentation allows us to clarify the contexts in which we can legitimately call computer simulation a form of experimental measurement.

Keywords Simulation · Modelling · Experiment · Calculation · Material systems

1 Introduction

Lord Kelvin is well known for his emphasis on the role of models in producing scientific knowledge.

> I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it. As long as I cannot make a

M. Morrison (✉)
Philosophy Department, Trinity College, University of Toronto, 319 Larkin Building, Toronto, ON M5S 1H8, Canada
e-mail: mmorris@chass.utoronto.ca

> mechanical model all the way through I cannot understand; and that is why I cannot get the electromagnetic theory. I firmly believe in an electro-magnetic theory of light, and that when we understand electricity and magnetism and light we shall see them all together as parts of a whole (1884, pp. 6, 132, 270–271).

In this remark Kelvin is criticizing Maxwell's Lagrangian formulation of electrodynamics and the lack of a suitable mechanical model that would should how electromagnetic waves could be propagated through space. However, in addition to these famous remarks on models he had this to say about measurement:

> I often say that when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot express it in numbers your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science whatever the matter may be (1891, pp. 80–81).

On closer inspection these passages suggest a connection between models and measurement that is linked with scientific understanding. Kelvin claims he cannot understand a phenomenon unless he has a mechanical model that illustrates how it is constructed and how it can be integrated into a broader theoretical context. An important part of that task is the ability to quantify various aspects of the model in a way that is related to the measurement of different values and parameters. Kelvin does not elaborate, specifically, on this connection but it is an important one, not just for him but also for contemporary aspects of modelling and measurement.

As we shall see below, determining the possible connections between models and measurement is important for distinguishing between measurement and calculation and also, I want to claim, between experiment and simulation. Initially these distinctions might seem relatively straightforward. Measurement typically involves some type of causal interaction with the material world via instruments, something that is also characteristic of experiment. Calculation, on the other hand, is a mathematical activity involving abstract reasoning, pencil and paper or computers. Because simulations are frequently associated with computers, their status as experiments is linked to the notion of 'numerical experiments' that involve using numerical methods to find approximate solutions to mathematical models/equations that cannot be solved analytically. But, on reflection these distinctions are less than straightforward. In recent years computer simulations have evolved from the kinds of number crunching activities suggested by the label 'numerical experiments' to tools for investigative research in areas such as climate science and astrophysics.[1] The issue is complicated by, among other things, the links between simulation and experiment and whether the outputs and data produced by simulations can be characterised as measurements.

What, exactly, would it mean to characterize the product of a simulation as a measurement and the simulation itself an experiment? I should point out that my

[1] For an excellent paper on modelling in climate science see Norton and Suppe (2001).

concern here is not with theory of measurement discussed by Krantz et al. (1971, 1989) but rather with an analysis of the activity of measuring—what is involved in the *process* of measurement. The issues related to measurement in many sciences focus primarily on the technical features of instrumentation, specifically the development of the appropriate type of techniques to precisely and accurately measure a property or feature of a phenomenon as well as methods for the interpretation of results. Current accounts of computer simulation can easily be accommodated in this picture: the computer functions as the apparatus and the act of running the simulation is regarded as a numerical experiment. But this does not really tell us much about whether or why we can classify simulation outputs as measurements. In what follows I argue that we can answer that question by looking at the role that models play in the context of both traditional experiment and simulation. Not only do models allow us to interpret so-called measurement outputs, but I want to claim that the models themselves can function as measuring instruments and it is this latter role that figures importantly in the analysis of simulation as measurement/experiment.[2]

In some cases the idea of models as measuring instruments is rather straightforward, as in the case of physical models that can be manipulated in a manner similar to models used by nineteenth century British field theorists. These models were used to determine features or quantities associated with the electromagnetic field which were in turn used to modify Maxwell's original equations. In this case the model operated as device that was, at least in the hands of Kelvin, able to 'measure' certain quantities supposedly associated with the aether. Although the status of these measurements was highly questionable a less problematic example is the use the physical plane pendulum to measure local gravitational acceleration. Although this seems to involve a more direct notion of measurement, it is important to note that even in this case there is an extensive reliance on theoretical models and approximation techniques that are necessary to guarantee the accuracy of the measuring process/outcome. The connection between models and measurement in this case extends beyond the theoretical interpretation of outputs to an understanding of the apparatus itself and how the representation of physical phenomena 'embedded' in the model/instrument enables us to use it as a measuring device. Although the sophistication of the apparatus (the physical pendulum) determines the precision of the measurement of G, the accuracy of the measurement depends on the analysis of all the incidental factors that are part of the 'experimental' setup.[3] Indeed the ability of the physical pendulum to function as a measuring instrument is completely dependent on the presence of these models. That is the sense in which models themselves also play the role of measuring instruments.

[2] A model can be based on some theoretical belief about the world that is suggested by the data or it can sometimes be understood as simply a statistical summary of a data set. Data assimilation is an example that extends beyond straightforward issues of description; here models are used to fill in gaps in observational data and to translate noisy observations into a collection of initial model-states. They function as tools for generating data.

[3] Consequently we need very sophisticated models in order to represent the apparatus in the appropriate way. Because the knowledge associated with the measurement comes via the model the apparatus and the model function, in many respects, as a single system.

In what follows I want to explore some of these issues in an attempt to illustrate how models function both in measurement contexts and as measuring instruments, and discuss the implications of this for assessing the epistemic status of computer simulations.[4] Drawing attention to the mediating role of models in measurement contexts helps to clarify how models themselves are often the focus of investigation. They mediate between us and the world and between us and our theories by acting as objects of inquiry and the source of mediated knowledge of physical phenomena. This mediated knowledge is characteristic of the type of knowledge we acquire in measurement contexts.

I begin with a brief discussion of the relation between models, experiment and measurement in Kelvin's work. His emphasis on models and measurement provides a context for seeing how the debate over abstract methods versus materiality has developed historically. It also highlights the way models were seen as experimental tools for bridging the gap between measurement and calculation. From there I go on to explore questions concerning the status of simulation as an experimental activity and connect features of simulation with the characterization of models as measuring instruments. Ultimately, I want to claim that the connection between models and measurement is what provides the basis for treating certain types of simulation outputs as epistemically on a par with experimental measurements, or indeed as measurements themselves.

2 Models and manipulation

In Galileo's *Dialogue Concerning the Two Chief World Systems* Simplicio is urged to put forward arguments and demonstrations rather than just texts; his debate with Salviati must engage the sensible world and not just the world on paper. This distinction nicely captures many of our intuitions regarding the distinction between experiment and simulation. The former is thought to directly engage the physical world while the latter is seen as a more abstract type of activity. This distinction between abstract theorizing and more material-based investigation was also important in the development of nineteenth century electrodynamics. In formulating his field-theoretic account Maxwell relied heavily on mechanical models of the aether, models that consisted of mechanical depictions drawn on paper whose functions were described in detail using mathematical equations. Realizing that they

[4] What is less clear is the connection between models, measurement and representation when we are dealing with statistical models. Statistics, in particular structural equation modelling, makes use of measurement models which specify the relationships among latent variables and determines how these variables will be measured. This is sometimes referred to as model specification and is probably the most important and difficult step in the process since if the model is incorrectly specified or the wrong measured variables are chosen to reflect the latent ones, there is little hope the model will give any kind of useful information. Here the model takes centre stage in defining the object of study and it is the model itself that is investigated. Again, the question of whether this kind of activity is best described as measurement or calculation will determine, at least to some extent, the kind of results we take the model to provide. Space prevents me from addressing issues related to this type of modelling in the paper. I mention it simply to highlight the pervasiveness of the connection between models and measurement.

were merely fictional entities used as heuristic devices for deriving a set of field equations, Maxwell jettisoned them in later formulations of the theory in favour of the abstract analytical mechanics of Lagrange. This enabled him to ignore difficult questions about hidden mechanisms or material causes that might be responsible for the production of electromagnetic phenomena in the aether.[5]

Maxwell's contemporary Lord Kelvin was extremely critical of this approach to modelling. He saw models as tactile structures that could be manipulated manually, thereby allowing one to know the target system or phenomenon as a "real" thing. In fact, it was a common theme throughout his *Baltimore Lectures* that theoretical expressions were acceptable only to the degree that they could be observed or felt; and because mechanical models could deliver this kind of tangible knowledge they provided convincing evidence for the properties and objects they represented. For example, when demonstrating the model of a molecule interacting with the aether Kelvin used a ball in a bowl of jelly and produced vibrations by moving the ball back and forth. For him this provided a more realistic account of the relation between matter and the aether because one could see how vibrations were produced and traveled through the medium. Kelvin claimed these models were demonstrative in a way that Maxwell's "paper" models were not. In fact, he claimed that one could predict features of the molecule based on model manipulation alone, without recourse to traditional experimental methods.[6] This emphasis on the power of models was especially prominent in cases where the phenomena in question were difficult to access—like the structure of molecules and the aether. By contrast, the mathematical formulas and imaginary constructions characteristic of Maxwell's electromagnetic theory had no basis in reality and gave rise to what Kelvin called nihilism—a system comprised of only words and symbols. But what exactly was the basis for this link between manipulation and reality? In other words, why did Kelvin think that manipulating mechanical models was akin to experimental demonstrations?

Part of the answer is that, for Kelvin, the debate about models was a debate not only about the methodology of modelling but also the justification of certain theoretical features of the new electrodynamics—features at odds with his brand of model-based empiricism. According to Kelvin the elastic solid aether model, which was the basis for the wave theory of light, provided the only viable account of electromagnetic wave propagation. The foundation of Maxwell's field theory was the introduction of the displacement current which was responsible for the dissemination of electricity in the spaces surrounding magnets and conductors. The details of Maxwell's account of displacement are not important for our story; what

[5] Despite this "paper" approach to theorizing Maxwell claimed that his field equations were deduced from experimental facts with the aid of general dynamical principles about matter in motion. This allowed him to treat the field variables as generalized mechanical variables that contained terms corresponding only to observables. Once a physical phenomenon can be completely described as a change in the configuration and motion of a material system the dynamical explanation is complete. Until that time, however, we must rely on mechanical images, analogies and models that furnished visual conceptions of the phenomena, conceptions whose sole function was heuristic. For more on Maxwell's claims about deduction from phenomena see Morrison (1992).

[6] See *Baltimore Lectures*, 120, pp. 282–283.

is important though is the fact that once Maxwell abandoned the mechanical aether model there was no way to explain how charge or electricity in general was produced.[7] Instead it was the displacement current, as the source of the magnetic field, together with Faraday's law of electromagnetic induction that provided the so-called "physical basis" for the interaction between the **E** and **H** fields that constituted electromagnetic waves. This enabled Maxwell to derive a wave equation with solutions in the form of propagating waves with magnetic and electrical disturbances transverse to the direction of propagation. But here displacement was merely a term in the equations; so in what sense could it/they furnish a "physical" foundation for a theory, especially if there was no accompanying causal–mechanical explanation of how displacement was produced?

This was exactly the problem that vexed Kelvin. The derivation of these transverse waves involved no mechanical hypotheses, only electromagnetic equations.[8] According to his elastic solid model the aether had a structure of a jelly like substance with light waves described as bodily vibrations in the jelly. Neither it nor any modifications of it were capable of supporting transverse waves. Moreover, his elastic solid aether model formed the foundation for a good deal of his work on the transatlantic telegraph cable. The telegraph theory entailed the existence of longitudinal waves—they were simply telegraph signals in wires with their propagation beyond the wire supported by an elastic solid aether. For Maxwell, of course, not only were electromagnetic waves transverse but they traveled in the space surrounding the wire, not in the wire itself. In that sense there was complete conceptual dissonance between the two accounts. Any account of displacement could not be separated from matter, but matter and motion required a mechanical model in order to understand its nature and effects. Hence the displacement current was not only methodologically suspected but it defied the very physical foundations of electromagnetism as understood by Kelvin. Theory and method were deeply intertwined.

As I mentioned at the outset, Kelvin saw mechanical models as intimately connected to measurement and experiment. He considered numerical calculation measurement as long as it was performed in the context of model construction, testing and manipulation. All of these features enabled one to know an object "directly" rather than simply becoming acquainted with a mere representation. This combination of a mechanical model and its practical application in the success of the Atlantic cable stood as a bold challenge to the kind of paper models and abstract mathematics found in Maxwell's work. The cable required a number of precise measurements that simply were not possible without a mechanical model. Displacement, understood as a field process, could not be measured because he was unable to build a model that could replicate it. Because the manipulation of these models often took the place of direct experiment, the aether, although inaccessible in

7 For a more extensive account see Morrison (2008).

8 Kelvin complained that even the so-called measurement of c involved a purely electromagnetic procedure, the only use made of light was to see the instruments. c was determined by measuring the electromotive force used to charge a condenser of a known capacity which was then discharged through a galvanometer. This would give an electromagnetic measure of the quantity of electricity in the galvanometer.

many respects, was considered real by Kelvin insofar as he was able to experiment on it using models. As we can see it was largely the model that came to define and constrain reality and not vice versa. What was physically realizable was what could be mechanically constructed and what was mechanically constructed could be measured. This constituted the experimental basis for theoretical beliefs and ontological commitments. The model-based unity of theory, method and ontology created a web of belief that was virtually impenetrable.[9]

From the discussion above one can begin to see how the calculation versus measurement distinction is important here. Kelvin's liberal attitude toward what counted as measurement resulted in a notion of experiment that extended much further than what one would normally allow as the basis for inference.[10] Maxwell was clearly able to calculate the velocity of wave propagation and values associated with the displacement current but there was no physical basis for these calculations. However, the materiality Kelvin associated with mechanical models could not provide the appropriate foundation for ontological claims about measurement. Neither mechanical models nor the success of the telegraph was sufficient for grounding a methodology that simply *assumed* a connection between physical reality and the ability to manipulate mechanical models that supposedly represented it.

Yet, if we look to modern science we often find extensive reliance on models as the source of knowledge of physical systems, especially when these systems are largely inaccessible, like the aether was for Kelvin. In disciplines like astrophysics and cosmology modelling has given way to computer simulation enabling researchers to investigate the formation of stars, galaxies and the large scale structure of the universe. Even in biological contexts like population genetics, highly abstract mathematical models and computer simulations are often the focus of inquiry rather than the evolution of natural populations. Although models and modelling techniques play an important role in these types of exploration they are also crucial in more traditional forms of experimentation. The interesting question is whether the prevalence of models and simulations in these contexts indicates a narrowing of the gap between measurement and calculation, particularly in cases where the model/simulations function as the object of investigation—as a stand-in for the physical system. In order to answer that question we need to evaluate the epistemic role and limits of model-based inference in both experiment and computer simulation. A crucial part of the answer will involve an assessment of the status of simulation as experiment. Several philosophers have drawn attention to connections between modelling and simulation as well as the similarities and differences between simulation and experimental inquiry.[11] Although opinion seems divided

[9] Even after the experimental proof of the existence of electromagnetic waves by Hertz in 1888, Kelvin remained critical of Maxwell's theory and its postulation of the displacement current. For more on Kelvin's views about measurement and models see Smith and Wise (1989).

[10] To that extent it is more than a little ironic that the empiricism that motivated Kelvin's use of mechanical models evolved into a form of methodological and theoretical dogmatism that prevented him from understanding and accepting the fundamental components of the most successful theory of his time.

[11] Humphreys (1990) was one of the first to draw our attention to the importance of simulation in understanding features of mathematically oriented theorizing and their connection to mathematical

about whether simulation can be properly classified as experimental inquiry, many who argue in favour of that position typically cite the materiality of laboratory experiments as not epistemically important (cf. Parker 2008; Winsberg 2008).[12] While I agree with this conclusion, nagging questions about materiality arise when the object of investigation is to establish the *existence* of a specific phenomenon or effect. Although this seems to suggest an important role for materiality I argue below that the situation is less straightforward than might appear. The role of models in the experimental process calls into question appeals to materiality that are allegedly responsible for grounding ontological claims.

In what follows I want to show why I think there is often very little basis on which to epistemically differentiate the activities involved in some types of simulation and experimentation. What I want to claim is that in many cases this notion of 'being in contact with the system' is less than straightforward and that the materiality has no relevant epistemic or ontological implications for the way the outcomes are evaluated. A consequence of this claim is that any sharp distinction between the outputs of simulation and experimental measurement is called into question. Once again, the reasons for this can be traced to the roles that models play in both experiment and simulation. In order to argue my point and show how models and simulation are linked with measurement and experiment, let me first look at some of the philosophical issues associated with simulation.

3 Computer simulations and experiment: highlighting the differences

The straightforward notion of simulation is used in a variety of scientific contexts where one type of system is used to represent or mimic the behaviour of another. An example is the use of wind tunnels to simulate certain aerodynamic features of flight. Computer simulation builds on this notion not just through the calculational power of computers but in the actual processes that are involved in the simulation itself. While we can capture a good deal of the straightforward aspects of the wind tunnel simulation in the traditional framework of modelling, computer simulation extends modelling practices in a variety of ways. Its use in condensed matter physics provides a clear example.

Very briefly, one uses an algorithm or set of instructions which, when presented with some initial microscopic arrangement for the system concerned, generates a different microscopic arrangement. The earliest arrangements in the sequence generated by repeated applications of the algorithm reflect the specific choice of the initial arrangement. However, the algorithm is designed such that irrespective of the

Footnote 11 continued
models and modeling practices more generally. That discussion continues in his 2004. Hartmann (1995) has also emphasised several important features of simulation including their dynamical aspects. For a discussion of issues surrounding the materiality of simulations see excellent papers by Winsberg (2008) and Parker (2008) as well as Morgan (2003a, b). Hughes (1999) has an interesting discussion of computer simulation and the Ising model.

[12] The notion of experiment I am interested in here is one that goes beyond numerical experiment intended as a kind of mathematical number crunching.

initial arrangement, the arrangements appearing later in the sequence do so with a frequency defined by an equation from statistical physics which gives the probability that a physical system at a particular temperature will have a particular microscopic arrangement of energy. In other words, the simulation tracks the dynamical evolution of the system. The desired properties can then be determined by computer “measuring” the average value of the property in question over a suitably large number of arrangements. This type of simulation involves renormalization group methods and is used to investigate systems at critical point where the system is forced to choose, as it were, among a number of macroscopically different sets of microscopic arrangements. The computer simulation illustrates the way the ordering process takes place. The basic ingredients of the algorithm take the form of a set of probabilities that an *s*-variable in a given state with its immediate neighbours (with which it interacts) in a given arrangement will, in unit time, jump to a different state.

From this example we can see that the computer simulation not only enables us to represent the evolution of the system but also to “measure” the values of specific properties as it approaches critical point. The question of whether this so-called “measurement” is properly characterized as such, or whether it is better described as calculation, depends, to a great extent, on whether simulation can be labelled an experiment. As is typically the case in these kinds of debates, there are persuasive arguments on both sides. If simulations enjoy the same epistemic status as experiments then why spend money building large colliders designed to test the results produced by Monte Carlo simulations of particle interactions. On the other hand, there are a variety of contexts where computer simulations do take centre stage as the source of experimental knowledge simply because the systems we are interested in are inaccessible to us. An example is the evolution of spiral structure in galaxies. Because the typical time and distance scales are 10^8 years and 10^{23} m the only way experiments can be performed is by simulating the system on a computer and experimenting on the simulation. In the case of stellar evolution what this involves is solving partial differential equations for heat transfer within a star simultaneously with equations determining the burning of nuclear fuel. This type of study has been so successful that the life histories of most types of star, from birth to death and all the stages in between are thought to be well understood, with no suggestion that the knowledge we gain in these contexts is in any way inferior.

In the astrophysics case we may want to say that simulation is an acceptable source of experimental knowledge simply *because* we are unable to conduct materially based experiments the way we can with other types of systems. But, in cases where both simulations and material experiments are utilized are there good reasons to give priority to the latter? In other words, what *justifies* an appeal to materiality in motivating an epistemic distinction between simulation and experiment and is there something significant about the access to material systems that renders experiments epistemically privileged?

Winsberg (2008) cites Gilbert and Troitzsch (1999, p. 13) who discuss simulation in the social sciences and claim that the major difference between simulation and experiment is that in the latter case one is controlling the actual object of interest while in a simulation one is experimenting with a model rather than the phenomenon itself. In a similar vein Morgan (2002, 2003a, b) claims that what differentiates simulation from experiment is the materiality of the latter which is an important feature in establishing the external validity of the experiment. Because simulations bear only a formal or abstract similarity to the target system their validity is much more difficult to assess.

Both Winsberg (2003, 2008) and Parker (2008) have also drawn attention to ways of distinguishing experiment and simulation. Parker defines a computer simulation as a time ordered sequence of events that serves as a representation of some other time ordered sequence of events where the simulating system has certain properties that represents the properties of the target system. Experiment, on the other hand, involves intervening on a system and observing how certain properties change in light of the intervention. The former do not qualify as experiments but computer simulation studies do because they involve intervening on a system (the digital computer) to see how properties of interest change. In that sense it is the activity, the intervening, that renders something an experiment. And, as Winsberg (2008, p. 4) points out, if we fixate on the experimental qualities of simulations we can classify them as numerical experiments that produce certain kinds of data. As a way of methodologically distinguishing experiment and computer simulation I take this claim to be relatively unproblematic. However, if our goal is a comparison of the data produced by experiment and computer simulation then we need to move beyond a methodological distinction to an epistemologically analysis.

Winsberg attempts to address this issue by distinguishing two kinds of simulation, SR is a type of representational entity and SA is an activity on a methodological par with experiment but is distinct from ordinary experiments. He thinks there are epistemological differences between experiment and simulation even though the former may not be intrinsically more epistemically *powerful* than the latter. Although both types of activities face epistemic challenges, he claims that the kinds of challenges each faces are fundamentally different. Essentially what distinguishes the two is the character of the argument given for the legitimacy of the inference from object to target and the character of the background knowledge that grounds that argument. In a simulation the argument that the object can stand in for the target and that their behaviours can be counted on to be relevantly similar is grounded in certain aspects of model building practice. If we are studying the dynamics of some target system then the configuration of the object of investigation will be governed by the constraints used to model the target system. The reliability of those model building principles will provide the justification for the external validity of the simulation. Although this practice is also common in ordinary experimentation Winsberg claims that in experiment we are interested in modelling what he terms "the object" under study rather than the target system and hence the model building principles are used to justify the internal validity of the experiment.[13] In simulation we are interested in the target system and its relation to the outside world, hence external validity is what we are concerned with (2008, p. 24).

[13] I take it that "object" here means apparatus whereas in computer simulation the object is the simulation run on the digital computer.

In the end Winsberg concludes that none of these conditions means that simulation has less potential to make strong inferences back to the world. In fact, a simulation of the solar system that relies on Newton's laws is epistemically stronger than virtually any experimental setup simply because the ability to construct good models for this type of system is not in question. Hence, it is the quality not the kind of background knowledge and skill in manipulating it that is important for grounding inferential practices in both simulations and experiments.

While I certainly agree that simulation is equally able to make inferences back to the world, it is not clear to me that the distinction between experiment and simulation should be grounded in the difference between internal versus external validity. Moreover, emphasising such a difference has implications for both the kind and strength of the inferences we make about the world based on the results of experiment and simulations. When establishing a claim or hypothesis about a physical system the internal validity of the experiment is simply the first step in that process. Without that we have no basis for assuming the legitimacy of the result. But external validity is crucial in assuming that experimental findings are true of the physical world and not just the shielded environment of the experiment. Focusing only on internal validity when characterizing experiment leaves us without any methodology for making the type of inferences required to extend experimental results.

In addition, if one locates the justification for inferential practices with the models used in each context, then the types of inferences made in both experiments and computer simulations begin to appear very similar. This becomes especially apparent when one looks at the way that models are used in experiment where both the object and target systems are modelled. For example, in particle physics we typically do not directly observe the object being measured in an experiment, like the spin of an electron or the polarization of a photon. Instead extremely complex equipment is used together with a theoretical model that describes its behaviour by a few degrees of freedom interacting with those of the microscopic system being investigated/observed. It is by means of this model that the behaviour of the apparatus is related to the assumed (modelled) properties of the microscopic system. In other words, this type of comparison is what constitutes a measurement, with the model of the target system and the model of the apparatus playing an essential role. Moreover, in simulation a great deal of emphasis is placed on calibration where we test not only the precision of the machine (the computer) but the accuracy of the numerical procedure; both of which rely on calibration models to establish external and internal validity.

To illustrate this point in more detail let me introduce the notion of a "simulation system" and compare it with two different types of experiment, one which is rather straightforward and the other less so. My claim is that the way models function as the primary source of knowledge in each of the three contexts is not significantly different; hence we have no justifiable reason to assume that, in these types of cases, experiment and simulation are methodologically or epistemically different. As we shall see, the causal connections between measurement and the object/property being measured that are typically invoked to validate experimental knowledge are also present in simulations. Consequently the ability to detect and rule our error is comparable in each case.

4 Connecting simulation with experiment

The kind of "simulation system" I want to focus on is computer simulation using particles.[14] This method involves tracing the motion of tens of thousands of particles and is typically used for evolutionary calculations—for the temporal development of an initial value–boundary value problem. The initial state of the system at time $t = 0$ is specified in some finite region of space on the surface of which certain boundary conditions hold. The simulation involves following the temporal evolution of the configuration. What is central to computer simulation and something that both Winsberg and Parker stress is the importance of representing the target system in terms of an appropriate mathematical model. The particle method provides an appropriate way of discretizing the mathematical model for cases where we have a corpuscular system or the system is succinctly described by a Lagrangian formulation.[15] Some examples of physical systems appropriate for particle methods include semiconductor devices, ionic liquids, stellar and galaxy clusters and many others.

What, then, is the relation between this discretizing method, the target system and the simulation itself? We can think of the process as involving a bottom up framework that begins with the phenomenon or target system and a mathematical model of that system. We then use a specific approximation method to discretize the model in order to make it amenable to numerical solution. These discrete algebraic equations describe a simulation model, in this case a particle model, which can then be expressed as a sequence of computer instructions (an algorithm) that yield the simulation programme. In other words, the simulation model is the result of applying a particular kind of discretization to the mathematical/theoretical model. The representation provided by the simulation model is an important feature in developing the algorithm out of which the programme is constructed. A trade off between the quality of the representation and computational costs/power needs to be established in order for the programme to fulfill certain conditions like ease of use and modification. What I call the "simulation system" consists of the computer, the programme and the simulation model and allows us to follow and investigate the evolution of the model physical system (i.e. the mathematical model/representation of the target system).

There are, I think, several reasons for characterizing this type of investigation as an experiment, or more properly, a computer experiment. First of all, we can think of the simulation system as both the apparatus and the material under investigation. Considered as an apparatus it allows us to create the kind of

[14] For an extensive treatment of simulation using particle methods see Hockney and Eastwood (1994). A good deal of my discussion of this method borrows from their account. I introduce the notion of a "simulation system" as a way of isolating what I take to be the crucial features of the particle method.

[15] Other methods of discretization include finite difference approximations and finite element approximations.

controlled environment where one can vary initial conditions, values of parameters etc. In that sense it functions like a piece of laboratory equipment, used to measure and manipulate physical phenomena. Unlike traditional experiments the situation is not constrained by conditions such as linearity, symmetry or large numbers of degrees of freedom. The properties of the material, the physical system being modelled, are represented in accordance with the limitations of the discrete representation on the computer, the particle/simulation models. To clarify, the material is, strictly speaking, what the experimenter can represent in the simulation models which are then investigated using the computer and programme. One could refine the distinction even further by isolating the computer plus programme as the apparatus and the simulation model as the material under investigation.

It is tempting to object here that when speaking of the 'material under investigation' I am equivocating between the model-representation and the properties of the material system.[16] Strictly speaking the model is what is being investigated and manipulated in a simulation. However, that model functions as a representation of a physical system via its relationship to the mathematical model of the target, so to that extent the system itself is also the object of inquiry.[17] As we shall see below this type of model/system relation is not simply a peculiarity of computer simulations but is also present in more traditional forms of experimentation. Consequently it does not provide a basis for epistemic differences between the two. Moreover, in contexts like astrophysics and cosmology the object of investigation is also the model, yet that does not prevent us from claiming that the object of inquiry is also the physical system and that the knowledge we obtain is of that system.

This picture differs from other accounts of simulation and comparisons of simulation with experiment (e.g. Humphreys 2004; Parker 2008) because it locates the materiality not in the machine itself but rather in the simulation model. What justifies this assignment? In most applications the particles in the simulation model can be directly identified with physical objects (via the theoretical model of the system). Each computer particle has a set of attributes (mass, charge etc.) and the state of the physical system is defined by the attributes of a finite ensemble of particles, with the evolution determined by the laws governing the particles.[18] In molecular dynamics simulation each atom corresponds to a particle and the attributes of the particle are those of the atom while in simulation of stellar clusters stars correspond to particles. The relationship between the physical particles and the simulation particles is determined by the interplay of finite computer resources

[16] Unlike an experiment one is not directly manipulating a "material object" in a computer simulation, unless of course we think in terms of the materiality of the digital computer itself. Parker (2008) refers to computer simulation studies as material in this latter sense.

[17] There are issues here about the degree of departure between the discretized simulation model and the mathematical model, that is, whether the former is an accurate representation of the latter. To some extent these issues are practical problems resolved via calibration and testing and hence are not reasons for general scepticism regarding the representational capacity of the simulation model itself. .

[18] For more details see Hockney and Eastwood (1994, p. 3). As they point out, one advantage is that a number of particle attributes are conserved quantities so no need to update the system as the simulation evolves.

together with the length and timescales important for the physical systems. The trick is to devise a model that is sufficiently complex that is reproduces the relevant physical effects without making the calculation too onerous. There are different types of particle models, each of which is useful for particular types of problems.[19] The motivation for claiming the simulation model functions as the 'material' under study arises not only from the correspondences mentioned above but also from the fact that it is crucial that the results from the simulation model be physically meaningful. In that sense the "physical" success of the computer simulation rests, ultimately, with the simulation model.[20]

In their book on particle simulation models Hockney and Eastwood (1994) draw some comparisons between simulation, characterized as a "computer experiment", and more traditional types of experiment. For them the computer experiment occupies a middle ground where the gap between theoretically and experimentally realizable objectives is difficult to bridge. Theory requires a vast number of simplifications and approximation techniques, a small number of parameters etc., and laboratory experiments are forced to deal with enormous complexity making measurements difficult to both make and interpret. The computer experiment, in its ability to overcome these difficulties, becomes the "link between theory and experiment".[21]

They claim that computer experiments, like their material counterparts, can be divided in roughly three categories. The first involves simulations designed to predict the workings of complex devices, allowing different configurations to be evaluated before the apparatus is actually constructed. The second involves the collection/production of data that is not obtainable using ordinary laboratory techniques, as in the simulation of galaxies. The third involves what Hockney and Eastwood call the "theoretical experiment" which is intended as a guide for theoretical development in that it enables us to examine complex situations by dissecting them into smaller components with an eye to obtaining a clearer theoretical picture of the target system. The film output from the computer experiment provides a more precise picture of the dynamics than the mental images/thought experiments of the theoretician. They claim that the computer experiment in conjunction with theory and physical experiment provides a more effective way of obtaining results than either approach or pair of approaches together. Theory uses mathematical analysis and numerical evaluation, experiment—apparatus and data analysis, and computer experiment—simulation programme and computer. Taken together they constitute a sophisticated methodology for scientific investigation.

While Hockney and Eastwood want to situate simulation squarely in the experimental camp, their characterization of different types of computer experiment applies equally well to modelling, especially when we think of how

[19] The decision to use a particular method depends on a variety of factors including the size of the system, whether the forces are long range, non-zero interaction, smoothly varying etc.

[20] Provided of course that the other relevant calibrations on the machine have been performed.

[21] Winsberg (2003) has suggested a similar view in his claim that simulation, like experiment, has, to use Hacking's phrase, a life of its own.

models provide a link between theory and experiment.[22] But we can take the similarities much further. Just like simulations, models are often constructed before building an apparatus or physical structure, they are sometimes the focus of investigation when the target system is inaccessible; and, they often provide the stimulus for new theoretical hypotheses. There are, of course, the computational resources of simulation that make it different from modelling, but given the various functions of simulation outlined above one could certainly characterize it as a type of "enhanced" modelling.

There are many reasons why the picture of simulation as modelling may seem attractive. Above we discussed the various ways in which models, both mathematical and what I have called the 'simulation models', play a role in the computer simulation of the target system. Indeed many of those who have written on simulation have stressed the link with models. Winsberg (2003, p. 107), for example, claims that simulation "provides a clear window through which to view the model-making process". He uses the term simulation to refer to the process of building, running and inferring from computational models. Simulations themselves, he claims, are based on models insofar are they incorporate model assumptions and produce models of phenomena. And, as we saw above, he also sees the justification for simulations as directly related to the strength of the underlying model of the target phenomenon/system and the legitimacy of the modelling techniques themselves. Humphreys (2004) also emphasises the role of computational models and has provided a helpful characterization in terms of a computational template which provides the basic form of the model, other construction assumptions, output representation and other factors. Computational science is then defined in terms of the development, exploration and implementation of computational models.

While this characterization seems relatively unproblematic in one sense it raises the following concern: If we identify simulation with modelling then how can we account for the experimental features typically associated with simulation? Moreover, it we understand model manipulation as calculation then there is no basis whatsoever for linking simulation with measurement.[23] To successfully argue for the experimental status of computer simulations requires more than just an appeal to the fact that intervention with and manipulation of models and machines are involved. Similarly, to categorize computer simulations simply as 'numerical' experiments narrows their scope to cases where their primary function is calculation. More principled arguments are required to establish robust epistemic similarities between experimental results and those arrived at using computer simulation, arguments that turn on the similarity of methodology and evaluation of results.

Many of the arguments that favour a strong distinction between experiments and simulations rely on the epistemic priority of materiality that is characteristic of

[22] This, of course, is just the sense in which models function as mediators between theory and material systems. I will have more to say about the specifics of this link below.

[23] Consequently we would be unable to capture many of the uses of computer simulation in experimental contexts, specifically astrophysics and cosmology.

physical experiment. Locating the materiality of computer experiments in the machine itself, however, carries with it no epistemological significance. While the materiality argument has intuitive appeal I show below why it is extremely difficult to motivate philosophically. The reason for this is the pervasive role of models at all levels of experimental practice. Above I claimed that models are experimental tools capable of functioning as measuring instruments, but they are also objects of inquiry in cases where they, rather than physical systems, are being directly investigated and manipulated.[24] What this suggests is a close connection between models and experiment and as we shall see below it is this link that figures prominently in establishing the experimental character of simulations. In other words, arguments illustrating the experimental features of models can also establish the experimental nature of simulation. In what follows we see how models act as measuring devices and how this role is crucial in defining their experimental character. The association with these types of modelling practices also underwrites our ability to classify the outputs of computer simulation as measurements. Put slightly differently, the modelling features of simulation are co-extensive with its experimental character making any epistemically relevant differences between experiment and simulation very difficult to articulate.

5 Measurements, models and mediation

In discussing the "experimental" features of models it is important to distinguish between what is known as a model of the experiment and what I refer to as experimental models. The former are typically called models of experimental results and are discussed (albeit in a slightly different form) in some of the early work in statistics (Pearson 1911) and, among other places, in the literature on the semantic view of theories, particularly Suppes (2002, p. 7). He describes the relation of theory to data as a process that puts experience "through a conceptual grinder" and produces experimental data in canonical form. These data constitute a model of the experimental results which, given its discrete and finite character, is a different logical type than a model of a theory. Statistical methodology then relates the model of the experimental results to some designated model of the theory. In order to do this there needs to be a theory of experimentation in place that can not only adequately connect the two but also estimate theoretical parameters in the model of the theory from models of the experimental results (or data). This is the famous model selection problem in statistics and there are many different approaches one can take to these issues (AIC, Bayesian methods, Classical N–P statistics etc.). However, my concern here is to focus on how we should understand the more general relation between models and experiment; that is, how models enable us to extract information from our apparatus and how should we understand their role in experimental practice more generally.

[24] The use of the term 'direct' is important here. It is not that the target system is *not* being investigated in contexts where there is a strong reliance on models but rather there is no direct access to the target system making the model the immediate object of inquiry.

Below I look very briefly at two different types of experiments, the measurement of gravitational acceleration and spin measurement. Although each of these might appear to involve different types of measurement, one more direct than the other, we will see that this in itself is not significant. In both cases and indeed in most experimental measurements of physical quantities or effects, models occupy centre stage. The first example I consider is the measurement of gravitational acceleration using a simple plane pendulum (Morrison 1999). This is the kind of experiment one usually does in the lab as a student and although it may seem relatively straightforward a good deal of theoretical modelling is required in order to achieve an accurate result.

We begin with a *theoretical* object—the ideal point pendulum—supported by a massless, inextensible cord of length l. From Newton's laws we know the equation of motion for oscillations in a vacuum and for the period. If the cord length and period are known we can solve for the acceleration of gravity. So, the experimental problem is to measure l and T_0. However, once the measurements are made on the *physical* apparatus a vast number of corrections must be applied since the equation used to solve for g as a function of l and T_0 describes a pendulum that deviates significantly from the one used in the measurement.

The period of oscillation of the real pendulum T is related to its idealized counterpart T_0 using a series of corrections depending on the desired level of accuracy. They involve (1) finite amplitude corrections, (2) mass distribution corrections which take account of the bob, wire connections and its mass and flexibility, (3) effects of air which include buoyancy, damping, added mass, and theoretical damping constants (4) elastic corrections due to wire stretching and motion of the support. Some of these corrections, especially (1) and (2) are relatively straightforward but the others involve rather complex applications of modelling assumptions and approximation techniques.[25] The level of sophistication of the experimental apparatus determines the precision of the measurement but it is the analysis of correction factors that determines the accuracy. In other words, the way modelling assumptions are applied determines how accurate the measurement of g really is. This distinction between precision and accuracy is very important—an accurate set of measurements gives an estimate close to the true value of the quantity being measured and a precise measurement is one where the uncertainty in the estimated value is small. In order to make sure our measurement of g is accurate we need to rely extensively on information supplied by our modelling techniques/ assumptions.

The point I am making here is not just the well known one of requiring a model in order to apply the law of harmonic motion. Rather, it is that using the physical pendulum as an experimental apparatus requires us to have in place a very sophisticated model of that apparatus that renders it a measuring instrument. In other words, the corrections applied to the model pendulum as a description of the physical system are the source of the knowledge and reliability associated with the

[25] This is a different issue than the kinds of considerations required for measurement by, for example, National Bureau of Standards where it is necessary to take into account variation in temperature, change in length due to atmospheric pressure etc. Those kinds of considerations are over and above the correction factors described above.

measurement. While we tend to think about theory or models as informing our interpretation of experimental data or results, we often ignore the specific role that models play in informing the entire activity including how the apparatus is used.[26] This is not just as issue that concerns the pendulum. In exploration gravity surveys field observations usually do not yield measurements of the absolute value of gravitational acceleration. Instead, one only derives estimates of variations of gravitational acceleration. The primary reason for this is the difficulty in characterizing/modelling the recording instrument well enough to measure absolute values of gravity down to 1 part in 50 million. Simply put: Without models there is no measurement.

This brings me to a related issue which is the status of the measurement of g. As we saw above, from values of the cord length and period we can obtain a value for the acceleration of gravity. If we take the view that models are abstract symbolic entities then questions arise about where to draw the line between measurement and calculation. In other words, how do we determine when the manipulation of models counts as measurement and when it is simply calculation? The standard definition that comes from measurement theory would not help us here. It tells us that measurements are mappings from objects being studied to some numerical representation called a variable. Calculation, on the other hand, is usually defined as involving the transformation of inputs into outputs. Both seem to be satisfied in our pendulum example. One possible way of differentiating the two involves the distinction (which goes back to Carnap 1950) between fundamental and derived measurement. Fundamental measurements deal strictly with observing relations among physical objects, performing operations on those objects and using some type of counting procedure to quantify or assign numbers to those relations. Derived measurements incorporate these procedures but also involve abstract calculation of the sort used to measure, for example, the instantaneous acceleration of a body. In this latter context there is no observational counterpart for the mathematical operations. Although this gives us an intuitive way of differentiating measurement and calculation, where only the former is based on observation, virtually none of the measurements performed in modern physics can be classified in this way. Even the measurement of gravitational acceleration done with our plane pendulum does not qualify as an observation in the sense described above.[27] As Luce and Tukey (1964) point out (albeit for slightly different reasons) fundamental measurement is largely a theoretical property.

There is, however, a methodological counterpart of this distinction which appeals to direct vs. indirect measurement.[28] The latter are simply those that presuppose

[26] It is important to note here that this is a different issue from what we typically call the "theory of the apparatus" which involves the theoretical knowledge required to build a machine that will function in the desired way. However, I should also point out that in designing machines like the LHC, computer simulations of particle paths provide crucial data for the alignment of magnets etc.

[27] Batitsky (1998) argues that even things like length measurements cannot be justified solely on the basis of our perceptual interactions with the world. Hence, the notion of fundamental measurement is essentially an empiricist's myth.

[28] Sometimes this distinction is absorbed or collapses into the fundamental/derived distinction. Causey (1969) defines a derived measurement as one which has dimensions expressed in terms of previously

prior measurements while the former do not. In the pendulum case we solve for g if we know the length and period, which seems to satisfy the conditions for indirect measurement. Indeed many of the quantities we typically 'measure' are the products of indirect measurement or involve what is sometimes referred to as indirect variables. Velocity, for example, is defined as the distance covered per unit time while the notion of a force field is something that is only measurable in virtue of its effect. This distinction requires that there be fundamental quantities like mass, length and time, with other quantities being monomial functions of these. Although the representational theory of measurement is formulated on the basis of this distinction it is questionable whether there really are any truly direct measurements. Even if there were, it does not solve the problem referred to above, which is defining the boundary between indirect measurement and calculation.[29]

The issue becomes more complicated in the second example I want to consider—the measurement of spin, an intrinsic angular momentum associated with the electron and other elementary particles. Although most physics texts tell us that the Stern–Gerlach experiment was the first experimental observation of spin on a beam of neutral silver atoms it was not identified as such at the time because the spin hypothesis had not yet been put forward.[30] However, one of the most peculiar things about the Stern–Gerlach experiment was its failure to be associated with electron spin by its discoverers Goudsmit and Uhlenbeck (1926) and by Pauli himself, who first proposed the idea of a fourth quantum number. Moreover, both Pauli and Bohr claimed that it was impossible to observe/measure the spin of the electron separated fully from its orbital momentum by means of experiments based on the concept of classical particle trajectories (Pauli 1932, 1958). The implication of this remark is that question of whether the free electron really has a spin can never be decided on the basis of experimental measurement.

Renewed attempts to 'measure' electron spin took place in the 1970s with the results described as an improvement of three or four orders of magnitude in the accuracy attained in magnetic moment measurements. But this is not really a measurement of spin in the sense discussed by Bohr and Pauli. Moreover, since the experiments were different from classical quantum measurements, questions arose as to how to theoretically interpret them. Various approaches involving generalizations of quantum mechanics and non-linear stochastic Schrodinger equations have been suggested but none has proven satisfactory. That said, this is not just a problem of how to interpret the experiments themselves. As in most 'measurements' of microscopic systems what is involved is the comparison of different models. A theoretical model is constructed where the behaviour of the macroscopic equipment is described using a few degrees of freedom, interacting with those of the microscopic system under observation. By means of this model, the observed

Footnote 28 continued
given fundamental quantities. Fundamental measurements are those that those that can be made independently of the possibility of making any other measurements.

[29] I come back to this point below.

[30] The experiment did however measure the magnetic moment of the silver atoms.

behaviour of the apparatus is related to the observed properties of the microscopic system.[31]

My claim is not that this is inherently problematic, or that it should not qualify as measurement; rather what the example emphasizes is the importance of the point mentioned above—in order to have any understanding of what is happening in these experiments one requires highly sophisticated models that are themselves sometimes hypothetical. Spin measurement now involves electron spin resonance techniques and scanning tunnelling microscopy, extremely effective methods for detecting precession frequencies of spin. But even here it requires about 10^{10} electrons to get a measurable signal. MRI techniques are also useful but the spatial resolution of MRI is limited, partly because it takes around 10^{15} nuclei to generate a detectable signal. In none of these cases are we actually measuring the spin of a single electron.

Despite these considerations no one would claim that spin is not a measurable property of certain types of elementary particles. Nor would one say that the extensive use of modelling assumptions and corrections somehow undermines the claim that we can measure the acceleration of gravity to four significant figures of accuracy using the pendulum. So, what is going on here? Clearly the use of models in these examples is not sufficient to characterize the activity as simply calculation. With respect to spin the Dirac equation is thought to provide the requisite calculation of electron spin and as we saw above, Maxwell's equations furnished a calculated value for the propagation of electromagnetic waves. Neither of these could, under any interpretation, be called a measurement. To that extent we have a distinction between measurement and calculation that matches our intuitions—what we do in manipulating mathematical equations is calculation, measurement involves another type of activity. More specifically measurement has some causal connection to instruments that generate data from some physical source. In that sense a measurement, even those that are indirect, bears some type of causal connection to the property being measured. Although various calculational techniques are undoubtedly part of the measurement process the activity of measuring involves a different set of constraints. The accurate measurement of g required a model as the primary source of this knowledge. But it was the intervention with the model and its interplay with the physical pendulum that we use to characterize the measurement process.[32]

Given this depiction of the role of models in experiment, what, if anything, differentiates experiment from simulation? Can we claim that simulation also involves measurement? The cases of experimental measurement discussed above illustrate the centrality of models in acquiring knowledge of the relevant parameters; questions about 'materiality' or some type of direct comparison

[31] For a detailed discussion of these measurements see Morrison (2007).

[32] In the pendulum case we are interested in both precision and accuracy and hence concerned not only with the status of the physical pendulum as a measuring device (internal validity) but also with its ability to measure the true value of g (external validity). The correction factors applied to the idealized model are relevant for both. Hence, contrary to the distinction introduced by Winsberg, the modelling involved in experimental contexts relates to both internal and external validity and includes the physical system under investigation as well as the object being manipulated.

between object and target system based on 'materiality' does not play a justificatory role. And, in the case of spin measurements, what is allegedly measured is the value of spin via inferences based on modelling assumptions. Indeed, such is the case with most measurements on elementary particles like quark masses. Experimental measurement is a highly complex affair where appeals to materiality as a method of validation are outstripped by an intricate network of models and inference.

If we look at the practice of computer simulation we see that the situation is strikingly similar to experiment. We start with a mathematical model of the target physical system and apply the appropriate discretizing approximations which replace the continuous variables and differential equations with arrays of values and algebraic equations. This yields the simulation model which together with various numerical algorithms produces the programme which is run on the computer. This translation of the discrete simulation model into a computer programme involves devising a sequence of operations which constitute the numerical algorithm. When these operations are performed they solve for the unknowns of the discrete model equations. This sequence is expressed as a series of computer instructions. The quality of the algorithm is judged by the number of arithmetic operations required to solve a given set of algebraic equations. The programme itself is tested and calibrated to ensure that it reproduces the behaviour of the target system (as described in the mathematical model) in a reasonably accurate way before it is put to work in actual experiments. In that respect it functions like an apparatus or piece of machinery in a typical laboratory experiment. As in an experiment we can trace the linkages back to some material features of the target system that are modelled in the initial stages of investigation.

Simulations also serve as the basis for how to build sophisticated experimental equipment. For example, the computing project developed to handle the vast amounts of data generated by the LHC was started to support both the construction and calibration of the machine. It involves simulating how particles will travel in the tunnel, knowledge which determines how the magnets should be calibrated to gain the most stable "orbit" of the beams in the ring. In that sense the simulation acts as a piece of experimental data required for determining the proper operation of the collider.

Within this framework there is also a natural division between calculation and measurement, a distinction that be used to motivate the notion of a computer simulation as an experiment. The simulation involves the evolution of the configuration/system in temporal increments by what is called the timestep DT. DT is an important modelling parameter and should be as small as possible consistent with the requirements of the problem and computer time available. This timestep cycle constitutes the main part of what we can call the 'calculation' and typically a few thousand timesteps are required to get useful results from a computer experiment. The calculation generates a large amount of data which requires that they be appropriately modelled in order to render them interpretable. Only by doing that can we say that the computer experiment, like an ordinary experiment, has measured a particular quantity. In both cases models are crucial. And, just as in the pendulum example where we are interested in both the precision and accuracy, similar concerns arise for simulation where the precision of the machine and the accuracy of the programme must be tested. This can be done by using the apparatus (machine + programme) to reproduce known results, as in simulating a model that is analytically solvable.

There is, however, one last point to consider. While models may play an important role in the experimental context the goal in some simulations is still a type of comparison with laboratory experiments, as in the case where we want to determine the existence of a subatomic particle. So, what is the justification for classifying simulations as experiments if the ultimate test remains the more traditional type of experimental confirmation? There are of course the examples of simulations used to test theories in ways that laboratory experiments cannot. For example, molecular dynamics simulation has been widely used to study phase changes, particularly glass formation and melting in ionic systems. Here the simulation can test microscopic features of theory that were not possible in the laboratory, features like the hardness of the ions which could be varied independently of their radii. Because different theories of melting predict different dependencies of the melting point on hardness it is fair to say that the simulations provided new and independent experimental tests.

However, in cases where simulations are stand-ins until laboratory experiments can be performed any comparison of the two needs to take account of whether the structure of experimental knowledge differs in any way from that produced in a simulation. At this point our intuitions incline toward the 'materiality' of experiment and the notion that this somehow confers epistemic legitimacy and priority on the outcome. Morgan (2005), for example, argues for the epistemic priority of material experiments claiming that inferences about target systems are more justified when both the experimental and target system are made of the same stuff. Parker (2008) challenges this view and claims that when comparing computer simulations with experiments the emphasis on materiality is misplaced. The justification of inferences about target systems depends on having good reasons to think that the relevant similarities, whether material, formal or some combination of the two, between the experimental and target systems hold.

Although the emphasis on the materiality of laboratory experiments appeals to our intuitions about empirical verification, as we saw above, the sense in which experiment and the knowledge gained therein provides a causal connection with concrete physical systems is typically via models. Consequently the comparisons are not between a directly accessible physical system and a hypothesis but between various levels of modelling. Obviously one cannot determine the *existence* of physical phenomena using computer simulations, only material experiment is capable of that. But, the ability of those experiments to *establish* ontological and epistemic *claims* about those phenomena requires a complex network of modelling assumptions that are an inseparable part of the experimental practice. In that sense the appeal is not to materiality as the source of justification but to the model of that materiality that best accounts for the data. But this is also true of computer simulations in that we always begin with a material target system that we want to investigate. My point here is not to deny the role of materiality in establishing ontological claims but only to point out that in complex experimental contexts materiality is not an unequivocal standard that functions as the way to epistemically distinguish simulation outputs from experimental results.

Moreover, it is frequently the case that simulations function as the confirmation of experiment. For example, computer simulations of particle collisions provide data that enable researchers to deduce the nature of the debris produced in a collision. These results are then compared with actual experiments as a way of determining what exactly the experiment has shown. In that sense the 'measurements' produced by the simulation provide an important source of data that are crucial features of the discovery process. This comparison between a computer experiment and a laboratory experiment seems no different in kind than the comparison of a model of the apparatus with a model of the target system that occurs in a laboratory experiment. In each case we adjust the parameters and calibrate the instruments to ensure the legitimacy of the agreement. Given the dependence on computer simulation in these and other contexts like climate modelling and astrophysics there seems no justifiable reason to deny the products of simulation the same epistemic status as measurement, or indeed to identify them as measurements.

Finally, it is important to stress, once more, that this does not involve an abandonment of empiricism. The hierarchy of modelling that is characteristic of computer simulation experiments begins with a mathematical model of the physical target system that is then discretized to produce the simulation model. While the computer simulation may represent states of a system that have not yet occurred, our ability to determine how the system will evolve under certain conditions plays an important role in the way scientific knowledge is put to practical use. In that sense the outputs of computer simulations are frequently classified as knowledge claims in a manner similar to experimental results. And, unlike Kelvin's models of the aether, there is a similar type of causal link with the target system as occurs in a material experiment.

6 Conclusions

We have seen how models function as both the object of investigation and as a type of experimental tool in different kinds of investigative contexts. The physical apparatus, while an important part of experimental investigation, is only one feature of the overall measurement context, we need models in order to extract meaningful information about both the apparatus and the target system. We have also seen how the simulation system comprised of the simulation model, programme and computer functions as both the apparatus and material under investigation. Here too the computer plus simulation programme functions as the apparatus while the simulation model represents features of the target system that can be implemented in the programme. This enables us to represent the target system in a way that facilitates computer measurement. By focusing on how models function as measuring instruments in experimental inquiry I have tried to shed some light on the way simulations, as a form of modelling, can fulfill the same role. The conclusion, that simulation can attain an epistemic status comparable to laboratory experimentation, involved showing its connections with particular types of modelling strategies and highlighting the ways in which those strategies are also an essential feature of experimentation.[33] The role that models play in these contexts highlights the difficulties in justifying (epistemically) any intuitive appeal materiality might have in elevating traditional experiment over computer simulation.[34]

References

Batitsky, V. (1998). Empricism and the myth of fundamental measurement. *Synthese, 116*, 51–73.

Carnap, R. (1950). *Testability and meaning*. Graduate Philosophy Club, New Haven: Yale University Press.

Causey, R. (1969). Derived measurement, dimensions and dimensional analysis. *Philosophy of Science, 36*, 252–270.

Gilbert, N., & Troitzsch, K. (1999). *Simulation for the social scientist*. Philadelphia: Open University Press.

Goudsmit, S., & Uhlenbeck, G. (1926). Die Kopplungsmoglichkeiten der quantenvektoren im atom. *Zeitschrif fur Physik, 35*, 618–625.

Hartmann, S. (1995). The world as a process: Simulations in the natural and social sciences. In R. Hegselmann (Ed.), *Simulation and modelling in the social sciences from the philosophy of science point of view*. Theory and Decision Library, Dordrecht: Kluwer.

Hockney, R. W., & Eastwood, J. W. (1994). *Computer simulation using particles*. Bristol: Institute of Physics Publishing.

Hughes, R. I. G. (1999). The Ising model, computer simulation and universal physics. In M. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 97–145). Cambridge: Cambridge University Press.

Humphreys, P. (1990). Computer simulations. In A. Fine, M. Forbes, & L. Wessels (Eds.), *Philosophy of Science Association Proceedings* (Vol. 2, pp. 497–506). East Lansing, Michigan.

Humphreys, P. (2003). *Extending ourselves*. New York: Oxford University Press.

Humphreys, P. (2004). *Extending ourselves. Computer science, empiricism and scientific method*. Oxford: Oxford University Press.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement. Vol. 1*: *Additive and polynomial representations*. New York: Academic Press.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1989). *Foundations of measurement. Vol. 2*: *Geometrical, threshold, and probabilistic measurements*. New York: Academic Press.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1990). *Foundations of measurement. Vol. 3*: *Representation and axiomatisation*. New York: Academic Press.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement. *Journal of Mathematical Psychology, 1*, 1–27.

Morgan, M. (2002). Model experiments and models in experiments. In L. Magnani & N. J. Nercessian (Eds.), *Model based reasoning: Science, technology and values* (pp. 41–58). Dordrecht: Kluwer.

[33] A general epistemological issue that emerges from this discussion concerns the changing nature of both measurement and experiment. In the end it may be that the account of computer simulation I have argued for is one that is more appropriate to the natural rather than the social sciences. However, given the complex nature of measurement in the social sciences I am inclined to think that the two contexts may be remarkably similar. Justification of that intuition will have to wait for another time.

[34] Support of research by the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. I would also like to thank the participants of the Oberlin Colloquium, particularly Ron Giere and Wendy Parker for their comments and questions. Thanks also go to Eran Tal for many useful conversations, comments and challenges.

Morgan, M. (2003a). Model experiments and models in experiments. In L. Magnani & N. Nersessian (Eds.), *Model-based reasoning: Science, technology and values* (pp. 41–58). New York: Kluwer.

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

Morgan, M. (2005). Experiments vs. models: New phenomena, inference and surprise. *Journal of Economic Methodology, 12*(2), 317–329.

Morrison, M. (1992). Some complexities of experimental evidence. In D. Hull, M. Forbes, & K. Okruhlik (Eds.), *Philosophy of Science Association Proceedings* (Vol. 1, pp. 49–62). East Lansing, Michigan.

Morrison, M. (1999). Models as autonomous agents. In M. Morgan & M. Morrison (Eds.), Models as mediators (pp. 38–65). Cambridge: Cambridge University Press.

Morrison, M. (2007). Spin: All is not what it seems. *Studies in History and Philosophy of Modern Physics, 38*, 529–557.

Morrison, M. (2008). Fictions, representation and reality. In M. Suarez (Ed.), *Fictions in science: Philosophical essays on modelling and idealization* (pp. 110–138). London: Routledge.

Norton, S., & Suppe, F. (2001). Why atmospheric modelling is good science. In C. Miller & P. N. Edwards (Eds.), *Changing the atmosphere: Expert knowledge and environmental governance* (pp. 67–105). Cambridge, MA: MIT Press.

Parker, W. (2008). Does matter really matter: Computer simulations, experiments and materiality. *Synthese* (forthcoming).

Pauli, W. (1932). *Proceedings of the Sixth Solavy Conference 1930* (pp. 183–86; 217–20; 75–80). Brussels: Gauthier-Villars.

Pauli, W. (1958). In S. Flugge (Ed.), *Handbuch der physik*. Berlin: Springer.

Pearson, K. (1911). *The grammar of science* (3rd ed.). London: Black.

Smith, C., & Wise, N. (1989). *Energy and empire: A biographical study of Lord Kelvin*. Cambridge: Cambridge University Press.

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

Thomson, W. (1884). Baltimore lectures. Cambridge, MA: MIT Press. (Reprinted in Kelvin's Baltimore lectures and modern theoretical physics by P. Achinstein & R Kargon, Eds.).

Thomson, W. (1891). *Popular lectures and addresses* (Vol. 1). London: Macmillan.

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

Winsberg, E. (2008). A tale of two methods. *Synthese* (forthcoming).