Studies in History and Philosophy of Science 56 (2016) 145–152

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Computer simulation and the features of novel empirical data

Greg Lusk [1]

University of Toronto, Canada

ARTICLE INFO

Article history:
Available online 7 November 2015

Keywords:
Computer simulation;
Measurement;
Data;
Evidence;
Empirical knowledge

ABSTRACT

In an attempt to determine the epistemic status of computer simulation results, philosophers of science have recently explored the similarities and differences between computer simulations and experiments. One question that arises is whether and, if so, when, simulation results constitute novel empirical data. It is often supposed that computer simulation results could never be empirical or novel because simulations never interact with their targets, and cannot go beyond their programming. This paper argues against this position by examining whether, and under what conditions, the features of empiricality and novelty could be displayed by computer simulation data. I show that, to the extent that certain familiar measurement results have these features, so can some computer simulation results.

1. Introduction

Computer simulation is now firmly entrenched in the methodology of science, so much so that simulations are widely designated as a third pillar of investigation along with experimentation and theorizing (see Reed et al., 2005, pp. 12-15). The increasing use of computer simulation to complement, and sometimes replace, instances of the other two pillars has recently led philosophers to focus on simulation's relationship with experimentation and theory (Beisbart & Norton, 2012; Guala, 2002; Humphreys, 2004; Morgan, 2003; Morrison, 2009; Parker, 2009; Winsberg, 2009, 2010). This literature analyzes theory's role in simulation design, the numerical techniques for executing the simulation, and the process of validating results, in order to stress the knowledge-making role that computer simulations have come to play in science.

By contrast, little emphasis has been placed on computer simulation's data-making capability. Simulations quite obviously produce large amounts of "data", but how this data should be characterized and treated is less than clear. Should computer simulation data be treated as novel and empirical, and allowed to play a role in the evaluation of theory? Or should it be treated as data of a lesser or different sort? And what grounds do we have for demarcating between novel empirical data, and data of other sorts? The answers to these questions bear on how and what scientists learn from computer simulation results: if simulations can produce novel empirical data, then they can be used to argue for the existence of phenomena and to provide support for other hypotheses about phenomena; if not, simulation results, without additional support, have little-to-no bearing on the likelihood of scientific claims.

It is often supposed that computer simulations could never produce novel empirical data for one of two reasons: they do not interact with the systems they are taken to produce data about and they cannot go beyond their programming to produce new knowledge of the systems they represent. I argue against this position. I claim that, insofar as certain common forms of measurement interact with their target and return new knowledge of their target system, simulations, under certain conditions, can as well. By analyzing common forms of measurement, I demonstrate how the features of empiricality and novelty are bestowed upon data. I then argue that if such features are bestowed upon data in these forms of measurement, then simulations, under certain conditions, can produce data that displays these features as well. My analysis reveals that we cannot deny the empiricality

E-mail address: glusk@uchicago.edu.
[1] University of Chicago, Department of Philosophy, 1115 E 58th St, Chicago, IL 60637, United States.

http://dx.doi.org/10.1016/j.shpsa.2015.10.005
0039-3681/

or novelty of simulation data for the above two reasons without simultaneously denying the empiricality or novelty of many measurement results.

I have purposely formulated the argument to come in terms of features of the data rather than features of the investigations that produce such data (though, the latter will be relevant to the former) to emphasize that it is the epistemic character of the results that we are investigating. Examining the character of the data also distinguishes my argument from other positions within debates regarding computer simulation, and avoids some of their problems. A significant portion of the existing literature has focused on whether computer simulations constitute a form of experiment or measurement. For example, both Marcel Boumans (2005) and Margaret Morrison (2009) have argued that models are (in some situations) measurement instruments, even if they never make contact with their targets. This position can be used to demonstrate that measurements and simulations are epistemically on par: if we understand simulations as a form of model, and models are measurement devices, then simulations are measurements. However, this position does not succeed in showing that models are measurement instruments as traditionally understood: traditional measurement typically requires causal interaction with the system being measured. In contrast, Giere (2009) and Beisbart and Norton (2012) all claim that since computer simulations never make contact with their target, they should not be considered measurement instruments. But this view will not do either: no one would deny that a computer could be programmed to simulate the workings of a stopwatch, and then used to measure the duration of an event. Whatever the kind of interaction a stopwatch has with an event, the simulation of the stopwatch run on the computer has the same interaction, and its results should be considered measurements.

The above debate focuses on the possibility of drawing a conceptual distinction between simulation and measurement (or experiment), and assumes that the epistemic character of the results will align with those conceptual distinctions. But drawing a conceptual distinction between two practices does not establish that their results could never share the same important epistemic features. I avoid this entrenched dialectic by focusing on the epistemic properties of the data itself. An examination of the production and handling of data illuminates how data gains and maintains its important epistemic character, and allows comparisons between the results of activities on this basis. Hence, in what follows, I do not argue for a conceptual distinction that allows simulations to serve as measurement devices; rather, I argue that in certain situations, simulations produce data with the same important epistemic characteristics as measurement. The argumentative strategy I employ is to show how measurement data comes to have two epistemically important features, and then to show that computer simulation data can obtain them in the same way. The significance of this argument is that it represents the first step in demonstrating that the features of data that make measurement epistemically significant can be extended to some simulation results as well.

2. The first feature: empiricality

To articulate what it means for data to be empirical and novel, it is useful to first specify what we mean by data. Intuitively, it may be popular to think of data as "elements of information that are taken for granted" in an investigation (Barberousse & Vorms, 2013, p. 31). However, this understanding of the term will not do. As, Humphreys (2013) notes, data's role in science is often to serve as evidence for some claim, but sometimes at least part of this evidence is not taken for granted, and instead is ignored or excluded from the investigation. Following Humphreys (2013, p. 13), I propose thinking of data as values of variables. This definition restricts the notion of data to quantitative values, but is advantageous because it recognizes statistical outliers or excluded information as data. A pitfall of this definition is that some scientific objects—for example cloud chamber photographs or the flushed face of a sick patient—do not count as data. This should not trouble us much here, because as we will see, the problem cases we are looking at are quantitative in nature. Furthermore, it is interesting to focus on quantitative data, because it is often this kind of data that undergoes character-changing transformations within a scientific activity.

The feature of empiricality grounds our belief that data conveys information about the investigative system that gave rise to it. It is often supposed that in order for an investigation to produce empirical data, the data must somehow be produced via an interaction with the system that the data is taken to represent. I will call this interaction a causal connection. There are at least two ways in which this causal connection can be established. One way is by physically interacting with the system, and another is through coextension with the system.[2] An example of the former is a pH meter that involves a physical interaction between the meter's probe and the substance whose pH is being tested. We interpret the results of the test to be about the substance because of the existence of this physical interaction between the substance and the data-producing device. Coextension can also establish a causal connection; for example, a scientist starts a timer when a phenomenon is observed and stops it when the phenomenon ceases. The data that results contains information about the duration of the phenomenon because the existence of the phenomenon and the operation of the timer coincided. A causal connection with the investigative system is necessary for producing data that has the feature of empiricality.

However, a causal connection with the investigative system is not sufficient for bestowing empiricality upon the data. There are, after all, many instances where causal connection does not produce data, and instances where the data produced using a causal connection might not be considered empirical. The former point is made simply by considering coextension: any two objects existing in time are coextensive, but such objects are not typically producing data about one another. To see the latter point, imagine setting up an apparatus that generated random numbers whenever it came into contact with water. Such a device could be placed in a lake, or put outside during a thunderstorm, but the numeric values that resulted would not display empiricality: there would be no reason to think that the data provided information about the lake's depth or the amount of rainfall. In what follows, I will use "causal connection" to indicate physical interaction, unless otherwise noted. Physical interaction is not only the more interesting case, but it is the kind of interaction relevant to our question; as mentioned earlier, no one would doubt that we could simulate a stopwatch and measure time.

Needing clarification are the conditions under which causal connections result in data that displays the feature of empiricality. To accomplish this, I turn to an exemplar of empirical data production: measurement. No one doubts that measurement data is empirical. Hence, data produced under the same conditions as measurement data should display the same feature of empiricality.

[2] It may be the case that these two forms of causal connection are subspecies of some more fundamental form. This possibility does not affect the argument to come.

2.1. Measurement data and empiricality

What conditions are sufficient to bestow the feature of empiricality upon data? Here I will explore three different types of measurement: fundamental, associative, and derived.[3] Important for our investigation is how each type of measurement is differentiated from the next by the inferential apparatus required to produce the resulting data. Investigating each of these kinds of measurement demonstrates how data comes to have empiricality, and how it can be maintained throughout an investigation.

Fundamental measurement is any particular form of measurement that does not depend upon prior measurements (Ellis, 1968, 56). This form of measurement estimates one quantity by comparing it with only other instances of that quantity. Put somewhat more formally, fundamental measurement enables the ordering of objects by quantity p without the need to appeal to quantities other than p. To see how this is accomplished, consider a fundamental measuring procedure. Typically a fundamental measurement procedure will consist of some measuring device, a set of operations for using that device, and assumptions about how that device works within the operation. Length is the most commonly cited instance of fundamental measurement. A ruler, for example, serves as a measurement device, and let us assume that it also serves as a measurement standard where one ruler equals one unit. To measure the length of an object, the ruler is placed flush with the object to be measured, initiating a causal connection.[4] If the object is longer than one unit, the ruler is moved to a new position, aligned with where it previously ended. The operation is repeated until the end of the object, and the number of times the ruler needed to be positioned is the result of the operation in ruler units. In this way, the ruler is used to map the length of the object into a quantification of ruler-units. However, in order for this quantification to be adequate, a number of assumptions need to hold.[5] For example, that the ruler remains rigid under transport (and does not change lengths) and that ruler units are strictly additive (i.e., there are no gaps when placed end-to-end, the ruler was not placed at an angle). Under such assumptions, the operation will provide a means for discerning when an object is longer, shorter, or the same length as another object. It will also allow the comparison of measurement outcomes between different objects whenever those assumptions hold and the procedure can be carried out. A method for evaluating whether these assumptions hold will be explored later on. What should be emphasized here is that fundamental measurement does not require prior measurements, it only requires some instrument to causally connect with the investigative object, that an operation to be performed, and that the operation accords with a set of assumptions.

The second type of measurement considered here is associative measurement, which uses one or more previously measured quantities to infer the value of a quantity of interest. Scientists' knowledge about the relationship between the two quantities is what permits the inference from the first quantity to the second. This knowledge often takes the form of a function that uses the first quantity as input and performs a transformation to produce values relating to the quantity of interest. Put more precisely, associative measurement involves there being some quantity p associated with the quantity q to be measured, such that when things are arranged in the order of p, under certain conditions, they are also arranged in the order of q (Ellis, 1968, p. 90). The prior quantity p can be measured fundamentally, or by another associative measurement. An example of associative measurement is counting the rings visible in the trunk of a tree to determine the tree's age. An operation for counting the tree rings is defined, and the function, often a simple one-to-one correspondence, is used to infer the growing years from the rings counted. Other associative measurements are less easy to spot. For example, to successfully measure temperature with a mercury thermometer, one must have a function that relates the height of the mercury column (quantity p) to the temperature of the object measured (quantity q). Measuring in this way is not always recognized as associative: users do not typically need to measure the height of the mercury column fundamentally and then compute the related temperature using the function; this work is done by instrument manufacturers during the design process, and the information is embedded in the scale printed on the glass tube of the thermometer. Nevertheless, the two steps can be epistemically distinguished.

It can be quite difficult to determine the function to be used in associative measurement. The simplest (and least accurate) way of determining what function to employ is to treat the measurement device as a black-box that simply maps inputs to outputs. One then uses the device to measure a series of objects whose quantity values are already known. This procedure maps the known-quantity values onto the indications of the device. This mapping can be used to define a function that is then inverted. The function that results from this inversion would take the indications of the device as input and yield quantity values as output. This method of computing the function used in associative measurement assumes that external influences on the measurement device are constant or negligible, and thus, the black-box strategy is best employed when the device and environment are well understood, and high accuracy is not required.

The more comprehensive approach to determining the function is to treat the measurement procedure as a "white-box." Essentially, the idea of white boxing is to break the measurement process up into its components and track how each component influences the result. Treating the measurement procedure in this way entails modeling the components of the measurement process as individual modules. These modules may represent each component of the measuring device, the sample to be measured, background effects, and the human operators (Tal, 2012, 152). Each module individually employs parameters, laws of temporal evolution, and characterizations of inter-module interactions representing each of the above factors (Tal, 2012, 151). These modules, taken together, constitute a model of the measurement process. This model is then used to articulate a series of equations that represent how the measurement outcome depends on the various components. If an analytic solution exists, these equations are solved to arrive at a specification of how the output of the device is deponent upon the input. The point to recognize is that associative measurement employs prior data, and performs a transformation on that data to obtain an estimate of the quantity of interest. The form of that transformation is determined by modeling the measurement process and its environment.

The last form of measurement is derived measurement, which, like associative measurement, is a form of measurement that relies on a previously measured quantity. What differentiates derived

[3] The typology presented here borrows heavily from Ellis (1968, chaps. 5–7) and is not meant to be exhaustive. Ellis focuses primarily on the nature of the measurement scale created by measurement operations, whereas I emphasize the assumptions and inferences in each form of measurement. I have chosen to employ this typology because it was an existing classification that nicely illuminates the role of assumptions and inferences in particular kinds of measurement. Such a classification could also be built from the more modern model-based account (Tal, 2011, 2012, 2013).

[4] The causal connection in this instance can be established either through physical interaction or through coextension. A physical interaction can be used to line align the ruler with the end of the object to be measured, or the ruler can be held away from the object and aligned visually.

[5] Chang (2004) calls these "ontological assumptions." The necessity of these assumptions is also discussed in Tal (2012) and Batitsky (1998).

measurement from associative measurement is the form the inference involved takes. In associative measurement, the inference is specifically tailored to the measurement context insofar as the measurement device, environment, etc., are all modeled explicitly and used to compute an inference from one quantity to another for that particular situation. In derived measurement, the inference from one quantity to another takes the form of a law of nature applicable to the measurement context. Such laws typically specify how several quantities relate to each other, and may involve a constant. Using the law of nature, one may compute the quantity of interest provided that all the other relevant quantities are known or have been measured. Like the previous form of measurement, derived measurement takes as input the values of quantities from prior measurement and performs a transformation to infer the quantity of interest. Given their similarity, we can refer to associative and derived measurement collectively as non-fundamental measurement.

Interestingly, while data produced by all of the above measurement types are the result of causal connections between a measurement device and the target system, they are not all the direct products of such interactions. The non-fundamental forms of measurement need not employ causal interactions directly; they can produce data that displays empiricality simply by performing transformations on existing data via inference. For example, fundamental measurement data may be gathered, and in an associative or derived measurement procedure, an inference is performed that results in different data that represents a different quantity. Both the data produced by the fundamental measurement and the subsequent non-fundamental measurements display empiricality, but only the former measurement needs to include a causal interaction within the measurement activity. The transformation that occurs in non-fundamental measurement maintains empiricality.[6] The question is, under what conditions, do inferences or transformations maintain empiricality.

While the three kinds of measurement described above differ in regards to the kind of assumptions and inferences they employ, they are similar in that each one requires that those assumptions and inferences accurately represent the measurement context. It is this correspondence between the measuring context and the assumptions or inferences that bestows upon the data empiricality. But how do scientists justify their assumptions or inferences? Scientists are aware that they can never know the true value of a quantity exactly, and that any assumptions or inferences are in some sense idealized. Hence, measurement outcomes do not denote individual values, but instead take the form of an estimated range of values that the measured quantity could take.[7] This estimated range is captured by "uncertainty", and in metrology, the term specifically refers to the likely dispersion of the values attributed to the quantity measured if the measurement was performed correctly (International Vocabulary of Metrology, 2012, 2.26). This range is determined when scientists test their measurement procedures against measurement standards whose accuracy is well established (Tal, 2011).

Demonstrating the reliability of a measurement procedure and the accuracy of the assumptions and inferences it contains involves performing an accuracy assessment to analyze that the investigation proceeded as expected, and then to analyze possible sources of error to determine a range of uncertainty to be associated with the outcome. The accuracy assessment occurs through "calibration", which typically proceeds by demonstrating the results of the measurement are consistent with the quantity values of well-known objects.[8] For most fundamental measurements and associative measurements that employ black boxing, an accuracy assessment involves measuring objects with known quantity values to see if the outcome of the measurement matches what would have been predicted under the assumptions or inferences employed. The variations are then modeled statistically (Tal, 2012, 150). The information about statistical variation is used to construct a range of uncertainty to be associated with the measurement outcomes. Uncertainty classified in this way is labeled type-a (sometimes said to arise from "uncontrolled" or "random" error), and uncertainties of this type are assessed the same way when a white box strategy is employed. Type-b uncertainties arise from error in the estimation of the parameters or other quantities employed in a white box strategy.[9] One way of estimating the uncertainty associated with these errors is to assess each quantity individually, and then use the model of the procedure to propagate the effect of the errors through to the measurement outcome. If the errors are independent of each other, they can be tallied up in an uncertainty budget (see Tal, 2011). If they are dependent on each other, a simulation may be employed to estimate the uncertainty. The point is: in order for measurement to produce data with the feature of empiricality, the assumptions or inferences employed should be accurate, and the justification that they are accurate rests on an accuracy assessment and uncertainty analysis.

We now are in a position to articulate the conditions—beyond the mere presence of a causal connection—under which measurement produces empirical data. Measurement requires a causal connection to the system being measured. In fundamental measurement, that interaction occurs under a set of assumptions, and it is the correspondence between those assumptions and the measurement procedure that bestows empiricality upon the data. In associative or derived measurement, the causal connection need not be direct; it can be established by employing data from a previous procedure in the associative or derived measurement. An associative or derived measurement performs transformations on the previously measured data. The data that is subsequently produced has empiricality because these transformations accurately represent the measurement context. What matters for empiricality is the presence of a causal connection within the chain of data production, and that the accuracy of any assumptions and inferences employed in that chain regarding the measurement context can be justified. This justification involves an accuracy assessment and the determination of a range of uncertainty to be associated with the data produced. It is under these conditions that measurement produces empirical data. We will now turn to computer simulation to see if the same conditions that produce empirical data in measurement can be found there as well.

2.2. Computer simulation data and empiricality

Computer simulation involves running a suitably programmed computer to produce results that are intended to represent a physical target system. This activity typically starts with a continuous mathematical model that is discretized to create a simulation

[6] The idea that successful inferences or transformations result in empirical data is not new. For example, Kosso (1988, 1989) and Chang (2004) argue that successful inference is involved in scientific observations.

[7] van Fraassen (2008, pp. 164) makes this point when he says that measurement locates objects in regions of logical space.

[8] See Tal (2013) for an overview of the philosophical literature on calibration.

[9] This classification is found in the Guide to Uncertainty in Measurement (2008). Philosophical treatments of the difference between, and the significance of estimating, these uncertainties can be found in Boumans (2013) and Tal (2011).

model. An algorithm instructs a computer to solve the discretized model. Iterative solutions to the model for successive time steps represent the time evolution of the states of the target system being simulated. The solutions provide values for variables that appear in the simulation model, which often represent physical quantities in the target system. These values constitute the data produced by simulation. While the program that instructs the computer to execute the algorithm is an abstract object that can be multiply realized, the computer simulation itself is a concrete process constituted by the operation of the digital computer.

Arguments that computer simulations are inherently different from measurements often focus on the (non-)empiricality of simulation, claiming that computer simulations do not interact with the target of the investigation. A causal connection is necessary for empirical data, the results of simulation could never be considered empirical. This position can be captured with a simple argument, which I will dub the "causal separation argument":

(P1) Empiricality requires that data be the result of a causal connection with the target of interest.

(P2) Computer simulations are concrete processes that are physically and causally separated from their target systems; the data they produce is not the result of a causal connection with the target.

(C) Computer simulation data cannot possess the feature of empiricality, and such data is therefore not empirical.

One response to the above argument is to simply deny that causal connections are necessary to produce data that displays empiricality. As some have argued, models are measuring instruments (Boumans, 2005), and it is typically the model that is responsible for the accuracy and precision of experiments and measurements (Morrison, 2009). If models measure independently of instruments that physically interact with the target, then simulations, as iterative numerical approximations of mathematical models, constitute measurements. However, this position is frequently dismissed. As Giere (2009) notes, though models play a significant role in experiments and measurements, there is no reason to extend the notion of measurement instrument to objects that do not interact with their target. In a similar vein, Beisbart and Norton (2012, p. 408) argue that although inferences are employed in empirical investigations, they are not essential: empirical investigations are powered by a causal connection to the target. I believe there is a different response to the causal separation argument, one that demonstrates that computer simulation data can have the feature of empiricality in the same way that measurement data does, without claiming that simulations are measurement instruments.

According to our analysis in the last section, measurement produces data with the feature of empiricality by either causally connecting to the target directly, or employing a prior measurement that does and then gauging the accuracy of any assumptions or inferences through an uncertainty analysis. The first step to demonstrating that computer simulation data has the feature of empiricality is to show that it relies on a causal connection with the investigative target. Accomplishing this step entails demonstrating that computer simulation data is produced in a way analogous to one of the kinds of measurement articulated above. If this can be shown, then there is no way to use the casual separation argument to claim that computer simulation data fails to rely on a causal connection, while simultaneously claiming that measurement outcomes do. In fact, the first step towards empiricality can be established by showing that some simulations utilize inferences and the products of causal connection in the same way as many measurements. As we will see, proponents of the causal separation argument face a dilemma: they either need to accept that computer simulation data is produced through a causal connection to its target, or deny that many measurement results are the products of causal connection. This dilemma is forced upon them because some simulations can be said to employ causal connections in the same way that some measurements do. The rest of this section will show how empiricality can be bestowed on simulation results in spite of the causal separation argument.

How could computer simulation data be said to result from a causal connection even when it is admitted that the digital computer never makes physical contact with the target system? Simulations employ several elements that are products of causal connections to their targets. By performing inferences on those products, simulations, *to the same extent as many measurements*, are generating data through the employment of causal interactions, and thus produce empirical data in the same way. The most obvious way that computer simulations could rely on the products of causal connections is by employing initial conditions that are empirically determined. Initial conditions are values of the quantities of interest from which the simulation will begin (i.e., the values of dependent variables at time $t = 0$). One way of establishing initial conditions is to use observed values obtained through measurement. Therefore, measurements might serve as the starting values for a computer simulation, and a simulation might perform a number of transformations of those values in order to produce its results. Extending an example from Giere (2009) that will be examined again later, a simulation might use observations of planetary locations and the exceedingly well-known regularities that govern planetary orbits to produce data regarding the positions of those planets prior to the observation. Such uses of simulation do not establish new causal connections, but they operate on existing causal connections in order to produce data. Hence, a consequence of this view is that standalone computer simulations—those that do not employ prior measurements—do not produce data that displays empiricality because no causal connection is involved in the investigation. On this view, computer simulations are analogous to the inference steps carried out in measurement: they are the transformations that occur between prior measurement and data.

However, proponents of the causal separation argument would want to deny that performing a transformation on a prior measurement would be sufficient to claim that the resulting data is the product of a causal connection. This move is not open to them if they want to maintain that most measurement data is the result of causal connections and meets the empiricality requirement. Recall that non-fundamental measurements were multi-step processes that relied on prior measurements. The important thing to notice for our purposes is that non-fundamental forms of measurement involve (at least) a two-step process that includes an inference on a previously measured quantity. The previously measured quantity is the result of a causal connection with the target and the second step is a transformation of this initial product. This two-step process employed in measurement mirrors the setup of a computer simulation that uses empirically gathered initial conditions: in both cases there is an initial product which is the result of a causal connection, and then a transformation is performed on that product to produce the final result. The transformations are themselves similar in that they both are often formulated from models of the physical context being investigated. The conclusion that should be drawn from this is that, since the results of associative and derived measurement are commonly understood as the products of a causal connection, the results of computer simulations that use empirically gathered initial conditions should be as well.

One could challenge this conclusion and claim that P1 is not appropriately strong. They might reformulate P1 as:

(P1*) Empiricality requires that data be the result of a direct causal connection with the target of interest.

Proponents of P1* would deny that performing transformations on extant data is sufficient to produce data that displays empiricality. By appealing to the directness of the causal interaction one excludes problem cases like simulation and limits the activities that generate empirical data to only experiment and measurement. Though, this move is motivated by a desire to carefully restrict what qualifies as empirical data, it also entails that many measurement results do not to produce data that displays empiricality. Recall that fundamental measurements were the only form that did not rely on prior measurements and inference to produce data. Thus P1* also entails that the products of non-fundamental measurement do not display empiricality. This is the problem generated by demanding direct causal connections: most, if not all, measurements performed in the physical sciences are non-fundamental; if associative and derived measurement results do not have the feature of empiricality, and as a consequence are not empirical, then there is very little empirical data to be had in natural science. Either one accepts that simulation data is the result of a causal connection to the investigative target via inference, or one is forced to admit that most, if not all, of the results produced through the contemporary practice of measurement in fact fail to display empiricality. This later option is unappealing, and the former should be preferred.

Simply demonstrating that simulation data is the result of a causal connection to its investigative target is not sufficient to show that it displays the feature of empiricality. As we saw when analyzing different forms of measurement, the assumptions and inferences employed should be accurate, and a degree of uncertainty should accompany the data. In order for simulation data to have the feature of empiricality, there must be a way to assess its accuracy and determine how uncertainty affects the results.

In principle, an accuracy assessment can be performed for computer simulation in much the same way that as it is performed for measurement. In fact, as Parker (2008) and Winsberg (2010) have argued, many of the same strategies that are used to sanction the reliability of procedures in experiment and measurement—the reproduction of known values, the comparability of results between devices, and responding as expected to interventions—can also be employed to sanction computer simulations. Perhaps a more systematic way, and indeed a way that has a lot in common with uncertainty assessment in measurement, is the strategy called verification and validation.

Verification and validation are two conceptually separable stages of accuracy assessment for computer simulations.[10] Verification assesses the correctness of the software and the numerical accuracy of the solution to a mathematical model. To do so, one checks the code for mistakes, and performs convergence tests to estimate the uncertainty attributable to the discretization process. Verification is primarily concerned with the simulation model's fidelity to the mathematical model upon which it is based, and not the accuracy of the simulation's results to the real world. However, ensuring the simulation works correctly can be seen as a prerequisite to further assessment.

Validation is explicitly concerned with the accuracy of simulation results in comparison to real world outcomes, and with determining how uncertainties affect the accuracy of simulation data. Validation experiments are designed explicitly for this purpose. Such experiments are performed on systems that a simulation represents. The aim of the experiment is to precisely and accurately record the states of the physical system under investigation in a situation analogous to the one to be simulated. This characterization involves careful measurements of the quantities (and their variation) that will be used in the simulation, including initial conditions, any boundary conditions, and the simulation's output quantities. Scientists then use metrics to quantify the accuracy of the simulation data compared to the data produced by the validation experiment. Validation experiments also provide crucial information for uncertainty quantification. For example, they produce distributions for initial and boundary conditions. These distributions characterize the uncertainty arising from such conditions, and the uncertainty can be propagated through the model so that its affects on the solution can be determined. A similar propagation approach can be taken with parameter uncertainty. Complex models can be broken down into subsystem components and, as in the white box strategy, the effect of each component and their interactions can be characterized individually. The intricacies of verification and validation are well rehearsed in the scientific (Oberkampf & Roy, 2010) and philosophical literature (Winsberg, 2010, chap. 2 and Morrison, 2014), and expanding more on the details would lead us astray. The point is: strategies exist for accuracy assessment and uncertainty quantification for computer simulations.

All is now in place to claim that the data produced by some computer simulations can display empiricality in the same way that some measurement data does. Computer simulation data can be the result of a causal connection with the investigative target in virtue of the computer simulation playing the role that inference plays in measurement. To do this, the computer simulation must use prior measurements as inputs and perform transformations on those inputs. Like in measurement, the accuracy of these transformations must be assessable, and a degree of uncertainty to accompany the data must be calculated. As I have argued above, strategies exist for carrying out accuracy assessments and uncertainty analyses for computer simulations. Therefore, when a computer simulation relies on a prior measurement, has undergone an accuracy assessment, and the relevant uncertainties can be quantified, the data it produces will display the feature of empiricality.

What should be noted is that the requirements of accuracy assessment and uncertainty quantification are actually quite high, but not impossible for a simulation to meet. For example, at least one opponent of epistemic parity between simulation and measurement acknowledges that some simulation data can meet this standard. Giere (2009, p. 60) admits that the results of some simulations—he cites simulations of planetary orbits—can be reliably substituted for actual measurements and that in certain instances simulations are more accurate than actual measurements. Though, Giere would reject that simulation results constitute novel empirical data, his position supports two important points that in turn support the claim that simulations results display empiricality. First, there are instances where scientists are confident that their simulation models capture the actual behavior of the system, and that they do so is exceedingly well established. Second, in such cases, the computer simulation data is just as accurate as measurement data and can be used for similar purposes. It can be inferred that for examples like Giere's, the uncertainty associated with the output of such simulations can be estimated, because otherwise the outputs of the simulation would not be deemed reliable replacements for measurements which themselves have estimated degrees of uncertainty. Giere provides an example where in practice, if the simulation used prior measurements as inputs to produce data regarding the previous location of the planets, the resulting data would display the feature of empiricality.

[10] The view of verification and validation described here follows Oberkampf and Roy (2010).

Though, I have just argued that simulations can produce data with the feature of empiricality, I do not mean to imply that it is easy for simulations to do so. In fact, I believe that we frequently lack the ability to justifiably assess a simulation's accuracy and the uncertainty to be associated with the data. Performing a validation experiment, for example, can only be done when the system under study is manipulable and assessable, and many systems that we represent with simulations are not of that sort. For many systems we simulate, we will not be able to gather the right kind of prior measurements to use as initial conditions. The sheer complexity of many simulations will make quantifying uncertainty problematic. Furthermore, in certain (usually large and complex) simulations, not every piece of data will be the result of transformations made on prior measurements. Only the data that is influenced by the prior measurement will have empiricality. The difficulties of producing data that displays empiricality will depend on the nature of the simulation and the system being simulated. This may suggest that in many situations, it is our ability to assess the accuracy of simulation-produced data, not the data's inability to "go directly to the world," that prohibits it from being empirical. Nonetheless, there are situations in which computer simulation data can display empiricality.

3. The second feature: novelty

Those that argue that simulations cannot produce novel empirical data might claim that simulation results that display empiricality could never also display novelty: given that empiricality is established in computer simulation by using extant pieces of empirical data as initial conditions, the result can never be considered novel because the simulation always (and only) manipulates what is already known. Simulation's inability to provide new information would disqualify its results as novel. Thus, at best, computer simulation data is only derivatively empirical. In the remainder of this section, I argue that a similar dilemma posed in regards to empiricality appears again when considering novelty: one must either affirm that certain simulation results can be novel, or one must deny that the results of many measurements are novel.

Let us make the notion of novelty evoked above more precise: novel data is that which enlarges the set of scientific information, and could not be produced from existing information using a number of formal or logical steps. As a consequence of this view, transformations of existing information never produce anything that could be considered novel. Call this the "non-entailment view." Non-entailment is clearly what commenters like Beisbart and Norton have in mind when they say Monte Carlo simulations "can only return knowledge of the world external to them in so far as that knowledge is introduced in the presumptions used to set up the simulation" (2012, p. 404). The problem for simulation on this view is that even if it is acknowledged that simulation data displays empiricality, it does so because the simulation employs previously measured initial conditions. Since the initial conditions are already known prior to the running of the simulation, and the simulation itself is just a set of finite transformations on those conditions, the simulation data is therefore wholly determined by the initial conditions and the simulation model; the resulting data is not novel.

The problem for this argument from non-entailment is that when applied consistently, the data resulting from some measurements do not fare any better than the data resulting from some computer simulations. Computer simulation data that displays empiricality is produced through inference and, therefore, on the non-entailment view, is not novel. However, non-fundamental measurement is not methodologically different when it comes to this point: associative and derived measurement rely on inference to infer the quantity of interest from input values (that have been previously fundamentally or non-fundamentally measured). In these forms of measurement, the input values can be seen as analogous to the initial conditions in simulations, and the function analogous to the simulation model. Viewing associative and derived measurement as containing two elements, one the more direct result of a causal connection and another an inference on that component, shows that under the non-entailment view, only the former component is actually providing novel information, and the rest of the measurement is merely an inference on that information. Therefore, the results of associative or derived measurement on this view are also determined purely by existing information, and cannot be considered novel.

Since non-fundamental measurement is often seen as an exemplar of novel data production, is there a way to save this account of novelty and understand non-fundamental measurement such that its data displays novelty? One suggestion is to include the determination of the values that serve as inputs as part of the non-fundamental measurement activity. Therefore, the inputs employed would not be considered information that existed prior to the measurement activity, and thus, the results of the activity could be considered novel. However, if we extend the bounds of non-fundamental measurement to include the determination of input values, we should do the same for simulation. To remain consistent, we should include measurements that establish initial conditions as part of a simulation activity.[11] Thus, simulation would be capable of producing novel data as well. What this discussion reveals is that whether simulation data displays novelty is a function of how we separate one activity from another, and there are multiple legitimate ways to do this. However, one would think that the epistemic character of the data should not change depending on whether prior measurements are included as part of an activity, or excluded from an activity but still employed within it. When two activities are combined, it does not matter where you draw the line between them; the character of the end product is the same. What should be stressed is that even on a strict view of novelty like non-entailment, some simulations and some forms of measurement can be grouped together in a consistent fashion. This suggests that the non-entailment notion of novelty does not capture an epistemically significant difference between measurement and simulation data in all circumstances.

It should be noted that this result does leave room to deny the novelty of simulation data by simply biting the bullet and agreeing that non-fundamental measurement is not a producer of novel data. It is important to note that this position admits parity between some measurement data (produced non-fundamentally) and computer simulation data. However, there are good reasons to resist this move. Associative and derived measurements are widely considered to produce novel data, and in fact, non-fundamental measurement results are often cited by philosophers (Bogen & Woodward, 1988; McAllister, 2011; Woodward, 2011) and scientists (Oberkampf & Roy, 2010) as exemplars of novel empirical data. Additionally, very few fundamental measurements are possible, and thus, associative and derived measurements make up a vast majority of the measurements performed. Often they are the only way to estimate many quantities of interest. To deny that scientists ever produce novel

[11] Views that encourage thinking about simulation as an activity broader than executing a program on computer already exist in the literature (see Parker, 2009; Winsberg, 2010). For example, Parker discusses *simulation studies* that include the steps necessary for placing the computer in the initial state.

empirical data regarding these quantities is to assert that we never gain novel empirical information regarding associative quantities like temperature or acceleration. While some philosophical positions may find this tenable, it is by no means forced upon us, and there are good reasons to resist doing so.

4. Conclusion: simulation as a producer of novel empirical data?

This paper examined whether computer simulation data could display two epistemically important features: empiricality and novelty. It had been previously argued that computer simulation data could never be both novel and empirical, primarily because computer simulations did not make causal contact with their investigative targets and could not go beyond existing information to return new knowledge. What I have shown is that insofar as certain common forms of measurement interact with their target and produce new knowledge of their target system, simulations, under certain conditions, can as well. I have argued for this conclusion by demonstrating that many common forms of measurement rely on previous measurements combined with accurate inferences to produce empirical data. I then showed that the non-entailment view of novelty, which would prevent simulation data from being empirical and novel, would also prevent most measurement results from being empirical and novel. Simulations, at least those that use prior measurements as initial conditions and those for which an uncertainty analysis can be performed, bestow on their products the same features as these common forms of measurement.

What is the upshot of this analysis? It shows that the epistemic differences between some simulation data and some measurement data are not as clear-cut as was previously supposed. In fact, it gives us reason to believe that the results of some measurements and some simulations fare the same. That instances of these two practices fare the same in the face of these features indicates that, on the basis of these two features, one cannot automatically discount some simulation data as less than empirical or novel, without also doing the same for data produced by associative or derived measurement.

Acknowledgments

I would like to thank Margaret Morrison, Wendy Parker, Joseph Berkovitz, and the two anonymous reviewers, for their detailed comments on this paper. This paper also benefitted greatly from feedback from participants at the MS6 conference, and particularly from Paul Humphreys and Johannes Lenhard. Melissa Charenko, as always, was an amazing editor. Any and all mistakes or inaccuracies are my own.

References

Barberousse, A., & Vorms, M. (2013). Computer simulations and empirical data. In J. M. Duran, & E. Arnold (Eds.), *Computer simulations and the changing face of scientific experimentation*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Batitsky, V. (1998). Empiricism and the myth of fundamental measurement. *Synthese, 116*(1), 51-73. http://dx.doi.org/10.2307/20118070.

Beisbart, C., & Norton, J. D. (2012). Why Monte Carlo simulations are inferences and not experiments. *International Studies in the Philosophy of Science, 26*(4), 403-422. http://dx.doi.org/10.1080/02698595.2012.748497.

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

Boumans, M. (2005). Measurement outside the laboratory. *Philosophy of Science, 72*(5), 850-863. http://dx.doi.org/10.1086/508103.

Boumans, M. (2013). The role of models in measurement outside the laboratory. *Measurement, 46*(8), 2908-2912.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. USA: Oxford University Press.

Ellis, B. (1968). *Basic concepts of measurement*. Cambridge University Press.

Giere, R. N. (2009). Is computer simulation changing the face of experimentation? *Philosophical Studies, 143*(1), 59-62. http://dx.doi.org/10.1007/s11098-008-9314-1.

van Fraassen, B. (2008). *Scientific representation: Paradoxes of perspective*. Oxford University Press US.

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

Guide to the expression of uncertainty in measurement.(2008). Joint Committee for Guides in metrology.

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

Humphreys, P. (2013). What are data about? In J. M. Duran, & E. Arnold (Eds.), *Computer simulations and the changing face of scientific experimentation* Newcastle upon Tyne: Cambridge Scholars Publishing.

International vocabulary of metrology.(2012). Joint Committee for Guides in metrology.

Kosso, P. (1988). Dimensions of observability. In *Observability and observation in physical science* (pp. 29-50). Springer. http://link.springer.com/chapter/10.1007/978-94-009-2434-5_3

Kosso, P. (1989). Science and objectivity. *The Journal of Philosophy*, 245-257.

McAllister, J. W. (2011). What do patterns in empirical data tell us about the structure of the world? *Synthese, 182*(1), 73-87. http://dx.doi.org/10.1007/s11229-009-9613-x.

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: University of Pittsburgh Press. http://eprints.lse.ac.uk/9506/

Morrison, M. (2009). Models, measurement and computer simulation: The changing face of experimentation. *Philosophical Studies, 143*(1), 33-57. http://dx.doi.org/10.1007/s11098-008-9317-y.

Morrison, M. (2014). Values and uncertainty in simulation models. *Erkenntnis, 79*(5), 939-959. http://dx.doi.org/10.1007/s10670-013-9537-1.

Oberkampf, W. L., & Roy, C. J. (2010). *Verification and validation in scientific computing*. Cambridge University Press.

Parker, W. S. (2008). Franklin, Holmes, and the epistemology of computer simulation. *International Studies in the Philosophy of Science, 22*(2), 165-2013.

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

Reed, D. A., Bajcsy, R., Fernandez, M. A., Griffiths, J.-M., Mott, R. D., Dongarra, J., … (2005). *Computational science: Ensuring America's competitiveness*. President's Information Technology Advisory Committee. DTIC Document http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA462840

Tal, E. (2011). How accurate is the standard second? *Philosophy of Science, 78*(5), 1082-1096.

Tal, E. (2012). *The epistemology of measurement: A model-based account*. University of Toronto. http://www.academia.edu/download/30891281/Tal_Eran_201211_PhD_thesis.pdf

Tal, E. (2013). Old and new problems in philosophy of measurement. *Philosophy Compass, 8*(12), 1159-1173. http://dx.doi.org/10.1111/phc3.12089.

Winsberg, E. (2009). Computer simulation and the philosophy of science. *Philosophy Compass, 4*(5), 835-845. http://dx.doi.org/10.1111/j.1747-9991.2009.00236.x.

Winsberg, E. (2010). *Science in the age of computer simulation*. University of Chicago Press.

Woodward, J. (2011). Data and phenomena: A restatement and defense. *Synthese, 182*(1), 165-179. http://dx.doi.org/10.1007/s11229-009-9618-5.