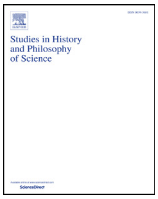
Studies in History and Philosophy of Science xxx (2015) 1–8

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 xxx

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.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

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

Please cite this article in press as: Lusk, G., Computer simulation and the features of novel empirical data, Studies in History and Philosophy of Science (2015), http://dx.doi.org/10.1016/j.shpsa.2015.10.005

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.

Please cite this article in press as: Lusk, G., Computer simulation and the features of novel empirical data, Studies in History and Philosophy of Science (2015), http://dx.doi.org/10.1016/j.shpsa.2015.10.005