EXPERIMENTS AND SIMULATIONS: DO THEY FUSE?

ECKHART ARNOLD
UNIVERSITY OF STUTTGART – SRC SIMTECH

published in: Eckhart Arnold and Juan Duran (Eds.): Computer Simulations and the Changing Face of Scientific Experimentation, Cambridge Scholars Publishing, Newcastle 2013, pp. 46-75.

Introduction

In today's science, computers have become an indispensable tool. They are used for the evaluation of scientific data, for storing data, for the preparation of results, and for communication among scientists. However, computers are not only tools that help scientists to process and evaluate scientific data, but also they produce scientific data when they are used for running computer simulations. This raises the question of whether the data that computer simulations produce is the same as other kinds of scientific data, in particular experimental data. What speaks for this assumption is that the data produced by simulations are usually previously unknown to the scientists, often cannot be derived mathematically, and may yield the same or at least similar kinds of information about a simulated empirical system as an experiment yields. What speaks against this assumption is the fact that simulation data stems from a calculation performed with a computer and that it is not the result of an empirical measurement, or not directly the result. This is also the stance that I am going to take in this chapter.

I will set out the reasons for taking this stance in detail in the following section, when I review the debate on the relation of simulations and experiments. In particular, I will argue that computer simulations are not material in any sense that would liken them to experiments (as maintained by Parker, 2009) and that experiments are not intertwined with models to such a degree that the function of models in experiments becomes indistinguishable from the function of models in simulations (as maintained by Morrison, 2009).

But there is also a further possible line of reasoning against a strict separation of simulations and experiments that is not so easily dismissed. According to this line of reasoning, simulations and experiments cannot strictly be separated because, at least in some instances, the role that empirical data take can appear indistinguishable in simulations and experiments. The question arises for those simulations that do in one way or another make use of empirical input data, and for those experiments that in one way or another involve the computational post-processing of the measured data. In both cases, the computer produces some kind of output data by processing empirical input data. The question, then, is precisely: what kind of output data?

We can define those scientific procedures that involve both empirical input data and computational processing of these data collectively as *hybrid methods*. The problem of hybrid methods can then be formulated as follows:

> *What, if anything, distinguishes a computer simulation that makes use of empirical input data from a measurement that involves the computational refinement of empirical data?*

It is not entirely clear whether this question is the right way of formulating the problem. I will briefly discuss different alternatives in the third section of this chapter as well. The answer to the problem of hybrid methods that is advocated here treats it as a partly conventional matter whether the outcome of hybrids is considered as empirical data or as theoretical data (which includes simulation data). The convention proposed here is that *hybrids* should be considered as empirical methods, if

1. The output data represents quantities that are either causally responsible for the values of the input data or that are mathematically connected to them.

   It may appear paradoxical that the output should be causally responsible for the input, but a simple example suffices to explain what is meant: assume that you measure force with a simple spring. Then what you actually measure is the extension of the spring (input data) and the scale on the spring allows you to “compute” the force in Newton (output data). Now, it is of course the force (i.e. the output) that is causally responsible for the extension (i.e. the input). At the same time, it is true that the output value depends on the input value, but this dependence is computational and not causal. I hold that this pattern is typical for any measurement where the quantity that is measured is only indirectly accessible.

2. And the output data characterizes factors that operate in close spatiotemporal proximity to the input data or, more precisely, to the source data.

In order to defend this convention, I am going to argue that it is in harmony with the self-ascription by the scientists using these methods, with the traditional understanding of measurements, and with our intuition.

The Current State of the Debate

The philosophical debate on the epistemic status of computer simulations can be traced back at least until the early 1990s. One of the popular slogans that already appeared as early as that in the debate was that of simulations as a “third way of doing science” (Axelrod, 2006; Küppers and Lenhard, 2005; Rohrlich, 1990), indicating that computer simulations neither fully resemble material experiments nor conventional forms of theory or model building, but that they are something in between. While this is a fair characterization of the activity of conducting computer simulations, which in many ways resembles experimentation but also requires specific practical skills and virtues that differ from those of experimenters, it is doubtful whether computer simulations can be characterized as a “third way” in an epistemological sense. For scientists themselves it has been clear most of the time that computer simulations are not an empirical method of science, even though they resemble experiments, and that therefore computer simulations, just like theories and models, are in need of empirical validation themselves, rather than being able to confer empirical validation on theories (Gilbert and Troitzsch, 2005; Heath, Hill and Ciarello, 2009). This view is also reflected in much of the philosophical literature on computer simulations of the 2000s (Guala, 2002; Humphreys, 2004; Morgan, 2003).

However, in the latest installments of the philosophy of simulations, this view has come under attack. In the context of a sometimes confused debate about the alleged materiality of simulations, philosophers have denied that there is any fundamental or epistemologically relevant difference between simulations and experiments. Or, if there is, then at least “any epistemically relevant differences between experiment and simulation [are] very difficult to articulate” (Morrison, 2009, 48). I am convinced that this is a mistake. First, therefore, I am going to set out some of the core arguments against the epistemic difference between simulations and experiments and I will try to show why all of them are wrong, some of them quite obviously so. Then, I am going to put forward positive arguments for the differences between simulations and experiments. Finally, I explain why, in spite of the clear conceptual distinction, hybrids still provide a challenge for the epistemology of simulations.

Arguments Against the Difference between Simulations and Experiments

The philosophers who are the most critical of the attempts to draw a clear distinguishing line between simulations and experiments are Wendy Parker (2009), Eric Winsberg (2009, 2010) and Margaret Morrison (2009). Wendy Parker argues that simulations in a sense are also "material" and that at any rate what matters is not materiality but "relevant similarity" (Parker, 2009, 484), which can be quite independent from the material status of the experiment or simulation. Winsberg does not go quite as far as Parker, but he, too, argues that simulations and experiments cannot be sharply distinguished by their materiality or by any similar criteria. The only distinction he concedes is that the way in which scientists justify their belief that the object under study (in a simulation or an experiment) can stand in for the target differs between simulations and experiments. As we shall see, he cannot advocate this view without contradiction, because the justifications cannot differ without referring to some other difference on which the different justifications are based. But then, the different kind of justification is not the only difference any more.

Morrison, in contrast to Parker, does not diminish the difference between simulations and experiments by arguing that simulations are also somehow material and, thus, somehow like experiments. But, quite the contrary, she argues that experiments in advanced science are somehow like simulations, because “the way models function as the primary source of knowledge in each of the (…) contexts [simulation and experimental] is not significantly different” (Morrison, 2009, 43). As we shall see, she overlooks the simple fact that in simulations a model also functions as the source of data while in experiments, the data is at least coproduced by nature.

I will now explain the flaws of the central arguments by Parker, Winsberg, and Morrison in more detail.[1] Parker offers several arguments, which are partly independent from each other. As mentioned, one argument is that simulations like experiments are also “in a sense” material. The sense in which simulations are material is this:

> The experimental system in a computer experiment is the programmed digital computer – a physical system made of wire, plastic, etc. As described in the last section, a computer simulation study involves putting a computing system into an initial state, triggering its subsequent evolution (the simulation), and collecting information regarding various features of that evolution, as indicated by print-outs, screen displays, etc. It is those data regarding the behavior of the computing system that constitute the immediate results of the study. In a computer simulation study, then, scientists learn first and foremost about the behavior of the programmed computer. (Parker, 2009, 488ff)

But, obviously, the kind of materiality that computer simulations enjoy because they are run on a material system (i.e., the computer hardware) does not at all liken them to real material experiments. It is misleading to say that the data that is presented on the printouts and screen displays is “data regarding the behav-

---

[1] Still I have to confine myself to the most important points here. For an even more detailed criticism see the working paper by Kästner and Arnold (2012).

ior of the computing system.” For the data of a simulation usually does not convey any information about the computer on which it was produced, but only information about the simulated system. It would be equally awkward if someone makes a calculation with pen and paper to consider the resulting figure as data regarding the pen and the paper. In particular, the person could potentially perform the same calculation with the same result in her head, which would imply that the result written on the paper must also be data regarding the brain of the person. Clearly, this is absurd. But then it is also wrong to say that the data that results from calculations performed on a computer is data regarding the computer. If this is not true, then also Parker’s basic contention that “any computer simulation study classified as an experiment is first and foremost a material experiment” loses its ground.

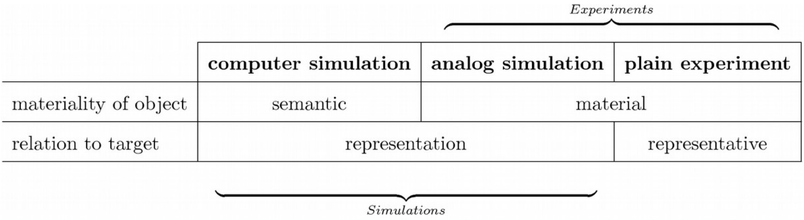
The same confusion of different levels of consideration (i.e., the symbolic or, if preferred, the “semantic level” (Barberousse, Franceschelli, and Imbert, 2009)) on which a computer simulation operates and the material level of the hardware on which it is implemented, is carried over by Parker to her reading of intervention. In Parker’s opinion, *intervention* in a computer simulation study occurs when the user sets up the simulation and puts it into an initial state, for which purpose the user has to interact materially with the computer. What Parker appears to misunderstand at this point is that it is not the interaction between the experimenter and the experimental machinery that is at stake when one speaks of *material experiments* in contradistinction to *computer simulations* or *computer experiments* but the interaction between the investigated experimental object and either the machinery or the experimenter or both. Now, in a computer simulation, the experimental object is either a fictional symbolic object or a symbolic (or “semantic” for that matter) representation of a material object. In any case, intervention on the “experimental” object of a computer simulation always occurs on the symbolic level (e.g., by assigning certain values to certain control variables). Thus, if one classifies computer simulation studies as experiments on the grounds that they involve intervention—which is, admittedly, one of several typical (though not exclusive) characteristics of experiments—then one still must concede that there exists an important difference between simulations and experiments regarding the type and kind of this intervention: in computer simulations, it remains purely symbolic and only in experiments it is material.

That is not to say that Parker is entirely unaware of the representational nature of computer simulations. At one point Parker even contrasts the representational quality of computer simulations with the property of involving interventions that experiments have:

> These characterizations imply at least the following fundamental difference between simulations and experiments: while a simulation is a type of representation — one consisting of a time-ordered sequence of states — an experiment is an investigative activity involving intervention. (Parker, 2009, 487)

However, apart from the fact that the there is at least a counterpart to the representational quality of the simulation model; namely, the representative quality of the experimental object, it is not at all clear why a simulation does not involve intervention. In both the simulation and the experiment, intervention consists in setting or changing certain conditions of the experimental system in a controlled way. Moreover, for both simulations and experiments there exist examples where this kind of intervention is achieved by: a) determining the boundary conditions through the setup before the experiment or simulation starts, or by b) user interaction during the simulation or experiment. While this line of reasoning might appear to strengthen Parker's point about the comparability of simulations and experiments as scientific methods, it still does not alleviate the counterargument that experiments operate on material objects while simulations operate on symbolic representations.

If we say that the experimental object is a representative, this means that it is a part or an instance of the target system of the experiment (i.e., the system in nature) the investigation of which was the purpose of the experiment. It is clear that the programmed model that represents the target system in nature in a computer simulation can never be a representative in this sense. On the other hand, there exist experiments where the object is also not a representative, but merely is some kind of representation. An example would be a ripple tank that is used to study such phenomena as reflection and interference of waves. Although the waves in the ripple tank are water waves, the ripple tank could also be used to learn something about waves of another kind, like sound waves or light waves. In this case, the waves in the ripple tank are not an instance of the target system and therefore the experimental object would not be called a representative of the target system. One can, in this special case, speak of the experiment as an *analog simulation* and consider the experimental object as a representation of the target system, just as in the case of a computer simulation. There still remains one obvious and one more subtle difference, nevertheless: the object of an analog simulation remains a material object, while the object of a computer simulation is always symbolic. This difference does not have any epistemic relevance in the case of analog simulations. The more subtle, but potentially epistemically relevant difference is that in the case of the analog simulations, there is still some kind of isomorphism involved between the object and the target, while in the case of computer simulations the relation remains purely representative.



	computer simulation	analog simulation	plain experiment
materiality of object	semantic	material	
relation to target	representation		representative

Figure 1: Conceptual relation of simulations and experiments

The different types of simulations and experiments that have just been described are summarized in Figure 3-1. Failure to distinguish properly between computer simulations and analog simulations is a constant source of error in both Parker's and Winsberg's treatment of simulations. For example, Parker complains "the proposed distinction implies that no study as a whole can be simultaneously both a simulation of some target system T and an experiment undertaken to learn about that same target system T, since the required relationships with T are mutually exclusive" (2009, 486). Then, she continues by presenting an example of a study that according to her interpretation is simultaneously an experiment and a simulation. Not surprisingly, her example of the San Francisco Bay Model concerns an analog simulation. However, this merely shows that the categories of simulations and experiments are not mutually exclusive in the first place. At the same time, it does not imply that there is no epistemically relevant difference between (computer) simulations and experiments that are not analog simulations, which is the conclusion that Parker suggests. In a similar vein, Winsberg (2009) complains that "if we can never be sure if something is an experiment or a simulation" it would not be worth knowing that, as Mary S. Morgan (2003) maintains, "experiments are more epistemically powerful than simulation" (Winsberg, 2009, 582). However, doubts whether something is an experiment or a simulation can arise only in the case of analog simulations. Even here is possible to distinguish analog simulations from plain experiments by their relation to the target system, as depicted in Figure 3-1.

Another point that Parker makes deserves more consideration; namely, that “what is ultimately of interest when it comes to justifying inferences about target systems is not materiality, but relevant similarity” (Parker, 2009). This is quite true, because material similarity does not automatically transform into epistemic reliability. In addition, numerical representations of nature in computer simulations can be quite accurate at times. Still, being of the same material stuff can be a good reason to assume relevant similarity (which Parker concedes); in some cases, it may be the sole reason. It must be expected that this is particularly true for those processes in nature about which we do not yet have comprehensive theoretical background knowledge in terms of either fundamental laws or at least well-tested phenomenological laws. Parker seems to be faintly aware of the connection between the existence of background knowledge and the possibility to simulate: “especially when scientists as yet know very little about a target system, their best strategy may well be to experiment on a system made of the ‘same stuff’” (Parker, 2009, 494). However, she does not seem to be aware that in this case it is not just an option (“best strategy”) but a necessity to conduct real material experiments. As the frontier of science is being pushed forward, one can assume that greater and greater regions of nature fall into the realm of what can reliably be simulated based on our scientific background knowledge. However, there will always remain scientific questions for which material experimentation is unavoidable.

Winsberg, in his paper entitled “A Tale of Two Methods” (2009), maintains that simulations and experiments can only be distinguished by how scientists argue for their validity. He does not notice that it would be impossible to argue in different ways for the validity of either simulations or experiments if there did not exist other differences on which the different arguments could be founded.[2] Indeed he implicitly admits this when he says of the experimenter that “She believes the inferences she will make are legitimate because she is prepared to argue that the two systems are, in relevant respects, the same kind of system, made out of the same material, and can be expected to exhibit relevantly similar behavior” (Winsberg, 2009, 590). However, this means that the experimenter relies on a relevant material similarity. So then, relevant material similarity must be another difference between simulations and experiments, besides the different justifications given for the respective methods. If it were not, it would not be understandable why the simulationist should not appeal to the same reason when justifying his or her procedure. Regarding the simulationist, Winsberg claims that he or she “will want to argue . . . that the computational model of his computer is relevantly similar to a good model of the behavior of the gas jets that interest him” (Winsberg, 2009, 590). However, this is an argument based on formal similarity, which means that formal similarity in contrast to material similarity must be an exclusive feature of simulations, if the justification based on formal similarity is to be exclusive to the simulationist. Otherwise, Winsberg’s thesis that simulations and experiments differ by the way they are justified would be empty. Thus, Winsberg is forced to admit the validity of Guala’s (2002) distinction between material and formal similarity that he tries to deny in his paper.

This is not the only contradiction in Winsberg’s paper. In order to explain his point, Winsberg sets out with the thought ex-

---

[2] Against this criticism of Winsberg, an anonymous referee objects, “two claims can be justified in different ways but have the same epistemic warrant.” However, since the epistemic justification of a scientific procedure usually consists in explaining or pointing out what its epistemic warrants are, it is hard to see how this is possible in this context. Moreover, as the passages quoted in the following pages from Winsberg demonstrate, he is unable to uphold his position that simulationists and experimenters rely on the same epistemic warrants when they justify their method.

periments of two physicists, one using a tank of fluid, the other using a digital computer to study fluid interaction. In other words, one scientist is conducting a material experiment; the other, a computer simulation. At one point he concretizes his story as follows: “what if we were to find that both of our original physicists’ primary area of interest is astrophysics? The systems that actually interest them are supersonic gas jets that are formed when gasses are drawn into the gravitational well of a black hole” (Winsberg, 2010). With respect to this setting, Winsberg remarks: “neither physicist, then, is actually manipulating his or her actual system of interest. Neither one is even manipulating a system of the same type, on any reasonably narrow sense of the term” (2010, 52). Thus, we are to assume that simulation and experiment cannot be distinguished by whether the actual system of interest is manipulated. However, only a few lines later Winsberg maintains exactly the opposite: “in some respects, the physicist’s tank is an instance of the system of interest, since it is in fact an instance of a supersonic interaction of a pair of fluids.” Now, how can a system that is not a “system of the same type, on any reasonably narrow sense of the term” be at the same time an “instance of the system of interest”? Winsberg denies that there exists a distinction between simulations and experiments that is more fundamental than the different kinds of justification for experiments and simulations respectively. It seems, however, that this denial rests in part on a self-contradictory analysis of the central thought experiment of his paper.

Another objection that Winsberg raises against the distinction is “on the Simon/Guala definitions of simulation and experiment, they are both success terms. An investigation will count as an experiment only if it is successful in the sense that the relevant material similarity between object and target actually obtain” (Winsberg, 2009). He concludes from this that in this definition if an experiment failed to establish a relevant material similarity then it would not be a failed experiment but it would simply fall into the other category (i.e. simulation), which seems wrong to Winsberg. With respect to this, he worries that "if experiment and simulation are success terms, then investigators may never be in a position to know if they are conducting a simulation or an experiment." However, Winsberg (2009), following a suggestion from Parker, already offers the obvious counterargument against his objection; namely that "simulation studies are characterized by the fact that the investigators aim for their objects to have relevant formal similarities to their targets and that ordinary experiments are characterized by the fact that the investigators aim for their objects to have relevant material similarities to their targets." Winsberg never answers this counterargument. Instead, he continues: "I do not think this works. I think the whole idea of formal versus material similarity is confused, no matter how much it is tempered by 'relevant,' 'aimed for,' or whatever." That is, Winsberg reasserts his opinion but does not offer an argument.

Margaret Morrison does not buy Parker's argument that computer simulations are also somehow material: "locating the materiality of computer experiments in the machine itself, however, carries with it no epistemological significance," she notes (2009). Nevertheless, she reaches the similar conclusion that "the modeling features of simulation are co-extensive with its experimental character making any epistemically relevant differences between experiment and simulation very difficult to articulate." More precisely, her claim is "that the way models function as the primary source of knowledge (...) is not significantly different" (Morrison, 2009). But this is obviously false, because in a simulation it is a model that produces the data, which is impermissible in a material experiment.[3] In a similar vein, Morrison maintains "experi-

---

[3] See also Peschard (forthcoming) who utters a very similar criticism of Morrison and nicely summarizes her complaints: "Admittedly, we 'know' of the features of the system that affect the instrument only in so far as we 'know' of the relation between these features and the state of the instrument; that is, only in so far as we

mental 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" (Morrison, 2009). However, one of her own examples, magnetic resonance imaging (MRI), suggests the opposite. For, in order validate that an MRI scanner works correctly, it is, among other things, tested with material objects. And when it is put to use in medicine, it is done so because it is able to reveal material features of the body or body part under examination and thus is able to validate or refute assumptions about health or illness by an appeal to materiality.[4] Because devices like an MRI scanner are diligently built to determine material properties of the objects under study, one could say that the "intricate network of models and inference" is tailored to the expression of the materiality of the object, rather than outstripping the appeal to materiality.

As we have mentioned earlier, with the scientific frontier moving onward, it is imaginable that increasing ranges of natural phenomena can be simulated, thereby potentially outstripping the need for experiments. This is, however, something completely different from maintaining that the appeal to materiality can be outstripped by models and inference in those cases where material experiments are still conducted. One might speculate that in future science there will be a growing dependence on observations

---

have and are justified in using a given model of the instrument. But to say that this mediating role of model makes causal interaction in experimentation epistemically irrelevant looks like saying that the role of language in expressing our sensory experience makes the sensory character of this experience epistemically irrelevant."

[4] According to an anonymous referee I have misunderstood the point that Morrison wanted to make with her example of MRI. I am aware that Morrison has several things to say about MRI. It is just this specific consequence about the relative epistemic weight of material factors and models that I intend to criticize. In the worst case my criticism only touches an unfortunate formulation by Morrison. Because Morrison formulates more or less the same idea in different ways at several points of her paper, I am inclined to believe that she means what she says at this point.

that are made with intricate and highly technicized measurement devices and continuously less reliance on ordinary sense perception. However, it is doubtful whether the point where sense perception becomes superfluous as a means of scientific investigation will ever be reached. One can say with Humphreys (2004) that this increases the epistemic opacity or that a greater and greater part of the epistemic processes that lead to knowledge will take place hidden from our eyes. But even then, humans will remain in the epistemic center, because it is humans that build and design the epistemic machinery that they make use of. However, the path —or, more likely, some of the paths—to the periphery where the epistemic machinery gets into contact with the world will continuously be extended.

Morrison may have been mislead into likening experiments to models by her own historical example, which she presents at the beginning of her paper. For the purpose of commenting on the contemporary discussion about models and experiments, this example unfortunately does not appear to be particularly well chosen. The example concerns Lord Kelvin's interpretation of electrodynamics. "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." (Morrison, 2009). This can be misleading if applied to the contemporary discussion, because it seems that Kelvin's notion of knowing an object "directly" rests entirely on an ontological commitment of Kelvin's in favor of mechanical models and explanations. Other than that, his jelly bowl (Morrison, 2009, 37) is just another example of what we call analog simulations and as such, it is just as remote from its target system as Maxwell's mathematical equations. Therefore, the example of

Kelvin is not a good example for showing, as Morrison seems to intend, that material experiments do not have a more direct relation to their target systems than simulations and that appeals to "knowing an object directly" through a certain kind of scientific method are badly founded. The appeal is merely badly founded in Kelvin's case. Incidentally, we see again how important the clear distinction between plain experiments and analog simulations is for the whole discussion.

Briefly summing it up: none of the arguments against the separation of simulations and experiments by Parker, Winsberg, and Morrison appear to be pervasive.[5] However, there is one point by Parker that ought be kept in mind; namely, that in any concrete case what ultimately matters is not the materiality of the procedure nor primarily whether the relation to the target system is a material or a formal similarity, but whether a relevant similarity can be established.

Arguments For the Difference between Simulations and Experiments

Having refuted the arguments against making a difference between simulations and experiments, the question remains: what positive arguments are there for drawing a strict distinction be-

---

[5] According to an anonymous referee, this misrepresents Winsberg's, Parker's and Morrison's position, because none of them believes that simulations and experiments are one and the same thing, but only that in some cases they may have the same epistemic warrants. My primary goal is not to criticize Winsberg, Parker and Morrison, but to refute those arguments that have been put forward against the difference between simulations and experiments. I have pointed out above some of the few concessions these authors make in the discussed papers in favor of the distinction between simulations and experiments. In no way do the discussed papers support the conclusion that Winsberg, Parker, and Morrison restrict themselves to *some* cases only. But even if restricted to some cases, most of their arguments remain false and seriously misleading.

tween simulations and experiments? There appear to be at least three fundamental and important differences between simulations and experiments, which I will discuss below.

Only experiments can operate on a representative of the target system

Operating on a representative of the target system means that the object that is manipulated and studied in the experiment is either a part of or an instance of the target system or is the target system itself. In contrast, both analog and computer simulations operate only on a representation of the target system. In the case of analog simulations, this is true in virtue of the definition of an analog simulation as an experiment that operates on a representation of, rather than on a representative of, the target system. In the case of computer simulations, this is true by necessity as long as the target system is a target system in nature.[6] Both the relation of being representative of and that of being a representation of a target system raise the analogous question of whether the respective relation truly holds. But this does not mean that both questions are one and the same. For establishing either of these relations provides a different challenge. Generally speaking, establishing the relation of representation requires comprehensive background knowledge about the target system, while the relation of being a representative can be established (though, as always, with a probability of error) on the basis of other indicators. For example, if one wants to know whether some kind of wood burns at 250°C it suffices to take a piece of that wood to establish the relation of *representative of* (in this case, in the sense of being part of

---

[6] One can also conceive of a model as a target system of a computer simulation. But this is a special case which in an epistemic connection is not at all comparable to the case where the target system is a system in the real world.

it). However, before one could be sure that a certain computer model of a piece of wood is truly a *representation of* that kind of wood, one would either need a comprehensive knowledge of the chemical structure of the kind of wood in question and of the chemical laws guiding oxidation, or one would at least need to know sufficiently detailed phenomenological laws about the burning of wood as to allow one to draw conclusions about the temperature at which the particular kind of wood in question starts to burn. Thus, the difference between *representation of* and *representative of* is a highly relevant epistemic difference.

This difference in relation to the target system can also be described as the difference between *material similarity* and *formal similarity* (Guala, 2002). Material similarity is the relation between the experimental system and the target system in the case of an experiment. Formal similarity holds between the simulation system and the target system in the case of computer simulations.

The case of analog simulations is ambiguous with respect to this terminology, and requires clarification as to whether *material similarity* also covers the similarity of different materials that obey the same laws. If this clarification is made or if the case of analog simulations is excluded, then Winsberg's (2009) criticism of this terminology can be circumvented. Another phrase that has been used to describe material similarity is the phrase "same stuff." This phrase is less ambiguous than the phrase "material similarity," because it clearly suggests that the material must be the same.

Only experiments can deliver knowledge to us that goes beyond what is implied in our background knowledge

Because computers are merely calculating machines, they cannot provide us with any knowledge about the world beyond what is implied in the premises of a computer simulation. As the premises must be rooted in our prior knowledge, the insights one can gain from computer simulations is limited to this prior knowledge and its implications.[7] The same does not necessarily need to be true of analog simulations. In order to be meaningful, an analog simulation only requires that the mapping relation (typically an isomorphism) between the object that serves as a stand-in for the target and the target system itself is known, but not that the laws of nature that govern the object are known as well. Therefore, the object could potentially reveal a behavior that is not merely a logical consequence of our prior knowledge. If we assume that the mapping relation is applicable nonetheless, then the novelty exposed by the object's behavior carries over to the target system as well. It may of course be disputed whether this assumption is true or whether it has much practical impact. But the case is at least imaginable.

Because of this limitation, computer simulations can be best thought of as tools for evaluating the consequences of an existing stock of knowledge. But only experiments (potentially including analog simulations in the hypothetical case just described) can break through the epistemic barrier that is determined by our prior knowledge and to which computer simulations are inevitably confined.

One can speculate whether one day our background knowledge will be so complete that we can deduce any possible further knowledge about the world from it. This, however, is pure science fiction and it seems as good as impossible within the limitations of the *conditio humana* that it should ever become real.

---

[7] It is important here to understand the difference among a) things that are not logically implied in our prior knowledge, b) things that are logically implied in our prior knowledge but unknown to us and c) things that are logically implied in our prior knowledge and known to us. For category a, simulations cannot help us, only experiments can help. For category b, simulations and experiments can help us. Finally, for category c neither is needed because we know it already.

Only experiments can be used to test fundamental theories

Can simulations be used to test hypotheses? They can, but only against the background of an existing theory. It may be the case that this theory can in turn be tested via simulations against another more fundamental theory. But at some point we reach a most fundamental theory, which cannot be tested by a simulation any more, because no theories or principles remain upon which such a simulation could be built. Thus, it is for basic reasons impossible to replace an *experimentum crucis* by a simulation. And this is true for both computer simulations and analog simulations, because an *experimentum crucis* requires that the investigated object be a representative of the target system, the particular nature of which is in question.

What counts as fundamental theory is, of course, historically relative. For example, Galileo's laws of motion and Kepler's laws of the movement of the planets were both fundamental theories at the time of their invention. Both, however, can be derived from Newtonian mechanics and, therefore, they lost the status of fundamental theories, which was then taken by Newtonian mechanics. Once Newtonian mechanics were accepted, Kepler's laws could also be tested by simulation (though this is strictly speaking unnecessary, because they could be derived mathematically already). But then this simulation does not replace an *experimentum crucis* of a *fundamental* theory anymore. Since at any past, present, or future point in the history of science there will exist at least one theory that is the most fundamental theory, material experiments will still be needed to test at least this fundamental theory. Even if we assume the hypothetical scenario above, where humanity has accumulated sufficient knowledge to derive everything else that is worth knowing from this knowledge, material experiments would still be needed to justify the fundamental theories that are part of this set of knowledge.

Further differences and conclusions

One can easily think of further differences between simulations and experiments: as mentioned earlier, experiments are material in the sense that the object under investigation is a material object. Simulations in contrast are virtual in the sense that the object that is investigated is a semantic representation. The criterion of materiality should not be confused with the relation of material similarity. Materiality as such concerns only the object under investigation and not the relation between object and target (see Figure 3-1). With respect to the relation of material similarity, materiality is a necessary but not a sufficient condition, because an analog simulation is also material but not of the "same stuff" as its target. Since it does not allow us to distinguish analog simulations from other experiments, materiality alone is a comparatively less important criterion for the distinction than, say, material similarity.

Yet, another difference is that experiments are an empirical method while computer simulations remain purely theoretical. Again, the case of analog simulations may be a cause of ambiguity, because by virtue of the materiality of their object, analog simulations could be considered empirical just like ordinary experiments, but they do not deliver empirical knowledge about the target system to us.

Overall, we find that there are sufficiently many and sufficiently important differences to warrant an epistemological distinction between simulation methods and experimental methods. This said, it cannot be denied that it is a fact that in modern science both methods, the experimental method and the simulation method, are frequently used in close connection with each other.

Does this mean that they merge into complexes where simulations and experiments become indistinguishable? We will now turn our attention to this question.

The Challenge of Hybrid Methods

In contemporary science, experimental methods are often closely intertwined with simulations or with simulation-like computational procedures. Simulations can be used to determine the optimal experimental design before experiments are carried out (Kramer and Radde, 2010). Computational methods can be used to select experimental data for further analysis while the experiment is run, as is done in particle accelerator experiments (CERN, 2011). They can furthermore be employed to post-process the raw data from measurements as, for example, in computed tomography (Lee and Carroll, 2010). In economics, experiments usually involve real human subjects that are placed in an artificial environment that differs substantially from the sort of real-world environments to which scientists try to apply results from the experiments and draw conclusions (Guala, 2002, 2012). Sometimes the artificial environment contains computer agents that interact with humans in the experiment. In the natural sciences, we also frequently encounter cases where empirical measurements and simulation methods jointly function as sources of data. Multiscale models of electrocardiac physiology, described by Annamaria Carusi, Kevin Burrage, and Blanca Rodriguez in another chapter of this book as model-simulation experiment systems, may serve as an example.

To give a name to these kinds of sophisticated procedures, we can speak of them as *hybrids of simulations and experiments*. Hybrid methods constitute a challenge for the philosophy of science in several respects. They challenge the distinction between simulations and experiments that has been defended above. Doing so, hybrid methods also challenge the logic of scientific research in general. For the logic of scientific research, as understood by most scientists and by many philosophers of science, rests on the testing of hypotheses against empirical data. This presupposes, one should assume, a clear distinction between the empirical and the theoretical. To put it in another way, if we cannot uphold the distinction between the theoretical and the empirical, then we would have to reconstruct the whole logic of scientific research.

The distinguishing features between simulations and experiments presented earlier do not really solve the problem of hybrids, because they only tell us what the difference between the categories of experiment and simulation are. However, they do not allow us in all cases to decide whether a particular procedure belongs to the class of simulations or to that of experiments. If we follow the reasoning of the first part of this chapter, then we know that only experiments can operate directly on the target system. But we may not be sure in a particular case whether some scientific procedure that makes scant use of some sort of empirical data and heavy use of computation falls into this category.

To solve the problem of hybrids, several quite different approaches are imaginable. One can even say that so far neither the framing nor the exact formulation of the question is clear. I am not going to attempt to give a comprehensive list of approaches to the problem of hybrids that have been proposed so far or that appear imaginable, but I will confine myself to the discussion of three approaches. Other authors have suggested two of these approaches; I briefly present them here since I consider these promising. After that I am going to present my own best guess at how the problem of hybrids could be solved.

Hybrids as Mixtures of Empirical and Virtual Data Sources (Zacharias/Lenel)

Guala (2002) considers as hybrid methods economic experiments where real human agents act in an artificial laboratory situation. Let us, for the sake of simplicity, imagine an experiment where human agents interact with computer agents. Generalizing from this case and adjusting it to the terminology developed in the first part of the paper, this leads to one possible definition of hybrids as procedures where the data source is partly empirical and partly virtual.

How does this relate to our earlier distinction between simulations and experiments in light of the material or formal similarity of object and target? Well, the example shows that both the object under study and the target can be complex entities that are made of different components. The material similarity that makes the method an experiment may hold only for some components of the object and target but not for others.

As a consequence of this, the differences between simulations and experiments that have been described earlier apply only insofar as such components of the object under investigation are concerned that do actually bear a material similarity to (parts of) the target system. One could classify hybrids (in the just-defined sense) as experiments, if one were willing to weaken the formulations of the differences a bit; for example, by allowing that it suffices that at least one component of the object is a part of or an instance of some part of the target system. However, this would be a somewhat strained attempt to keep up a strict dichotomy between simulations and experiments.

A much better solution has been proposed by Moritz Lenel and Sebastian Zacharias (unpublished). They give up the strict dichotomy in favor of a cross-classification of simulations and experiments (first dimension) and of laboratory and field methods (second dimension). In order to do, so they drop the idea of a monolithic target system. Instead, they differentiate between the target object and the target situation. Experiments and simulations are then distinguished by whether they operate directly on the target object or on a representation thereof. Laboratory research is distinguished from field research by whether it takes place in the target situation or in an artificially crafted laboratory environment. This classification scheme works quite well for economic experiments and simulations and for the social sciences in general. Economic experiments would most of the time fall under the category of laboratory experiments, but there is also room for laboratory simulations, field simulations, and field experiments.

It is an open question how well this or a similar scheme could work in the natural sciences. In addition, the case where human agents act together with computer agents in the same situation on an economic experiment might strain the classification. Still, it is so far one of the most convincing answers to the problem of hybrids.

Classification in Terms of the Degree of Materiality (Morgan)

A quite natural approach would be to examine to what extent the method employed depends on materiality (i.e., material data sources, material interaction, material output) throughout the course of the simulation or experiment in question. This is the approach that Mary S. Morgan (2003) has taken. Doing so, she reaches a fine-grained classification that ranges from lab experiments over “virtually experiments,” “virtual experiments” (which are not the same as “virtually experiments”!) to mathematical model experiments. Morgan takes into account the material status of input, intervention, and output, but also the relation between object and target where, again, she carefully distinguishes between "representative of," "representative for," and "representation of." Morgan's "Experiments without material intervention" (2003) is also one of the few attempts to explicitly deal with hybrid methods. I will not attempt to do justice to her careful and well-reasoned examination here. However, a few remarks are in order.

First, while it seems reasonable to consider the materiality or nonmateriality of the intervention for distinguishing degrees of virtuality, it is not equally clear why the material or nonmaterial status of the inputs or outputs should really matter. A simulation can start with empirical input data of some system and then calculate the future evolution of the system. However, this would not make the simulation any more experimental. The most that can be said is that materiality of input data is a necessary but not sufficient requirement for a procedure to be an experiment or empirical measurement. As will be argued below, it is, if anything at all, the relation between the input and the output what makes a hybrid an experiment or a simulation.

Mary Morgan's distinction between representative and representation is more convincing. Although it is very helpful for distinguishing experiments from simulations, it does not seem fit to solve the problem of hybrid methods, because—as has been argued above—the problem arises when both relations are present in the course of one and the same procedure. As sample cases, Morgan examines two different simulations of hipbones. They differ in the way the model of the hipbone is obtained on which the simulation is carried out. In one case, the model is obtained by cutting one particular hipbone into slices and determining the three-dimensional structure of the hipbone from these slices. In the other case, the scientists started with a stylized bone model that is then refined: "the individual side elements within the grid are given assorted widths based on averages of measurements of internal strut widths (taken from several real cow bones) and are gently angled in relation to each other by use of a random-assignment process" (Morgan, 2003, 222). Only in the first case is the input data clearly of empirical origin. The other case could— from the description given by Morgan—alternatively be interpreted as an example of a theoretical model that is adjusted or corrected with empirical data. For Morgan, the first simulation is therefore more like a material experiment than the second, and both lie somewhere between pure material experiments and pure mathematical modeling.

The stance I have adopted leads to a different evaluation, though. According to the view I advocate, both examples are clearly simulations. The reason is that the empirical origin of the input data alone is not sufficient to classify a procedure as experimental, or even partially experimental. In either of the two cases described by Morgan, it is only the input data what is empirical. The object that is manipulated during the study, however, is obviously a model. According to Morgan's description, "in both cases (...) the experiment consists of the 'application' of a conventionally accepted (...) mathematical version of the laws of mechanics (...) The computer experiment calculates the effects of the 'force' on individual elements in the grid and assembles the individual effects into an overall measure of the strength due to structure" (2003, 221).

The last description seems to fit one of our earlier characterizations of simulations in contrast to experiments quite well; namely that in a simulation it is a model and not a material object that produces the simulation data. This characterization is not as clear as it may seem at first glance though, because it requires that we can always distinguish the case where a model that is set up with empirical parameter values produces simulation data from cases of mere refinement of empirical input data, like, for example, by noise reduction algorithms. In the examples that Morgan presents, however, it seems clear enough that the data is produced by programmed models in a way that goes beyond the typical inferential patterns that can be found in measurements. That the models have been created from empirical data does not contradict this finding.

Classification in Terms of the Relation Between Input and Output

In the following, I present my own best guess at how to answer the problem of hybrids. As stated earlier, the best way of framing the question in my opinion is to ask how computer simulations that make use of empirical input data can be distinguished from empirical measurements that involve the computational refinement of raw data. The difference can, I believe, easily be made clear with the help of examples.

Think for example of a climate simulation: a climate simulation calculates the future development of the climate. In order to do so it is fed empirical data. Thus, both components of a hybrid —empirical input data and the computational processing of this data—are present. Yet, it is clear that a climate simulation is a simulation and not a measurement, because it is impossible to measure something that lies in the future.

Now take as another example an MRI scan: again both components of a hybrid are present: the object or the person in the scanner from which the empirical input data is recorded in form of electromagnetic waves that are emitted in response to the prior excitation of its H-atoms and the computational processing, which in this case produces a visual image of the internal structure of the object from the input data. While the classification may be not quite as indisputable as the example of the climate simulation, it still appears reasonably clear that this is an empirical measurement, because the object's structure is reconstructed from data that reflects this structure.

As clear as the example cases may be, it is more difficult to find general criteria by which to decide whether a particular method or procedure belongs to the class of simulations or to that of measurements (or experiments for that matter). In the following, I am going to attempt an answer in two steps. The uniting idea for both steps is the assumption that the difference between simulation-like hybrids and measurement-like hybrids can best be spelled out in terms of the relation that exists between quantities that the output data represents and the quantities that the input data measures.[8]

A first approach: The same-system formula

Following the idea that one feature that distinguishes experiments from simulations is that experiments can operate on the physical target system itself, one can formulate the following criterion:

> *Same-system formula*: a hybrid procedure is a measurement if its output data describes the same system in the same state as its input data.[9]

---

[8] The relation between input and output that is meant here is not to be confused with the transformation function that transforms the input data into output data. Rather it concerns the relation of the input and output values within the target system. Examining the nature of the transformation from input to output might provide yet another alternative way to deal with the problem of hybrids. Nevertheless, this alternative is not examined here.

[9] It might be worthy of notice that the input of the computational part of a hybrid always has a precisely and unambiguously defined magnitude; namely the digital data as it is entered into the computer (either by hand or by a digitizing device) before any calculations on this data have been carried out.

One can easily check that this criterion works well with the two examples given above: the output of the MRI scan is obviously data about the very system that the input data is taken from, and it is about the system in exactly that state in which the input data was recorded. Although in the case of the climate simulation one could say that the input and output system is the same; namely, the climate system, the output clearly concerns the system in a future state and therefore in another state than the input. The same-system formula therefore correctly places it in the class of simulations.

The same-system formula works well enough in many cases, but unfortunately not in all cases. Imagine a similar case as Mary Morgan (2003) discusses: we determine empirically the structure of a particular hipbone. Then, we run simulations where pressure is put on the hipbone in order to estimate the strength of this hipbone. The hipbone's strength is thus inferred by a calculation from its structure. Now, measurements often involve some kind of inference, but usually this is backward inference, where we measure the deeper causes of a phenomenon by some overt phenomenon (e.g., we measure the temperature by the extension of the liquid in a thermometer). However, in the case of the hipbone, the inference goes in the other direction. It therefore appears very doubtful whether one could call this a measurement of the hipbone's strength.

A second approach: The measuring-the-cause-by-its-effects pattern

Since the same system formula fails as a sufficient criterion for classifying hybrids, a subtler criterion is needed. Spelling out the same idea that only experiments operate on the physical target system itself, I propose the following two criteria for classifying hybrid procedures as measurements:

1. *Spatiotemporal concordance of source and output*: the output values have the same spatiotemporal location as the source values.

2. *Causal dependency of input on output*: the output values are either a necessary (!) cause for the input values, or the output values are linked by definitions or mathematical laws to the input values.

The first criterion makes sure that neither prognoses nor retrodictions (i.e., inferences about past events based on present observations) are accidentally classified as measurements. The second criterion reflects the well-known pattern of measuring a magnitude by its causal effects. For example, if one measures the force through the expansion of a spring. The further qualification that a link by definition or mathematical laws suffices is meant to capture such simple cases such as measuring the density by measuring and then dividing the weight and the volume of an object. If a hybrid procedure is found to be a measurement by these criteria, then we can also speak of the input data as *raw data* and the output data as *refined data*, thereby indicating that in the case of a (computationally enhanced) measurement, the input and the output data still concern one and the same thing. There exists an overlap between both criteria insofar they exclude prognoses, although this overlap is harmless. One can easily verify that neither criterion is superfluous in the sense of preempting the other criterion.

We speak here of “values” rather than “data,” because data is, strictly speaking, an entity located in a computer and causally linked to the software that processes it. What matters here, however, are the magnitudes in nature that the data informs us about. We understand “values” as always having the time, location, and causal connection to their occurrence in nature. In addition, it should be noted that in the first criterion we do not refer to *input* values but to *source* values.[10] This accounts for the fact that the measuring device can be located more or less remotely from its object. For example, a person observing an explosion may hear a noise and see a flash of light, both of which occur at a different time to the observer. Because of this, it would not be useful to require spatiotemporal concordance of the *input* values. Admittedly, introducing the concept of source values here raises questions regarding the relation between source values and input values. Since the source values cannot directly be observed, it requires at least a further inferential step to reconstruct the source values from the input values. It would take us too far afield to go into this problem here. Therefore, it must be noted as an open question.

In order justify the proposed criteria for classifying hybrid methods, we will briefly go through a number of typical examples of hybrid methods and try to show that the classification according to these criteria is sound in the sense of matching the intuitions one might have about the particular examples.

I have already mentioned climate simulations as probably the most well known example of simulations in science. Climate simulations are based on empirical input data, but clearly they do not constitute experiments or empirical measurements themselves. The output of climate simulations concerns the future development of the earth’s climate. It would seem awkward to consider climate simulations as a measurement of the possible future climate. As the output does not fall into the same spatiotemporal region as the source, climate simulations are also not measure-

---

[10] This distinction relates to Paul Humphreys’ distinction between source data and accessible data. See Figure 1-1 of Humphreys’ article in this volume.

ments according to our two criteria listed above. Thus, the classification of climate simulations according to our criteria is in harmony with our intuition and the self-ascription by scientists.

Another famous example of the most advanced kind of "technoscience" is the Large Hadron Collider (LHC). An interesting peculiarity of the Large Hadron Collider is that from the enormous number of events occurring during one second in the collider, only a number of events that is several magnitudes smaller is preselected[11] by automatic procedures for further examination (CERN, 2011). This nicely illustrates the idea of epistemic opacity, which, according to Humphreys (2004), is one of the characteristic features of modern computer-based science: It is the computer that decides which data will be selected and it is in principle impossible for any human agent to double-check each individual decision, even though the algorithms for that decision were of course developed by humans.

According to our criteria, which remain neutral with respect to the selection and preselection of data, the LHC data still counts as experimentally measured data. This is in accordance with the self-description of the LHC project, which also speaks of experiments. It is reasonable to do so, because the events selected by the computer for further analysis are still empirical events that occurred in the collider itself.

It is more difficult to decide how computational post-processing of data affects its status as empirical data. In magnetic resonance imaging, the raw data obtained from the electromagnetic signals emitted by the previously stimulated protons of the body are turned into an image by means of various highly sophisticated computations (Lee and Carroll, 2010). According to our criteria, magnetic resonance imaging falls still into the category of experimental measurement, because the output is an image of the struc-

---

[11] LHC terminology speaks of "reprocessing" of data. However, since the data is not changed but merely is a subset of data filtered from a larger set of data, we use the term "pre-selection" here to avoid misunderstanding.

ture of the body, but it is just that structure of the body that determines what the electromagnetic signals (i.e., the raw data) are like. In this sense, the output values are causally responsible for the input values. Simultaneously, both output values and source values lie in the same spatiotemporal region. But not only according to our criteria— intuitively it also makes sense to consider magnetic resonance imaging as a measurement. For it bears a strong similarity to photography. And it can be verified by dissection that the images it produces resemble the object under study and thus are not fabricated by a model.

Simulations are a very popular tool in astronomy. One reason for this is that it is impossible to carry out material experiments with stars and galaxies. However, the fact that it is impossible to study, say, the collision of galaxies experimentally does not turn a simulation of the collision of galaxies into an experimental procedure, other than in a purely metaphorical sense of the word "experimental." If we consider such examples, then these are not experimental measurements according to our criteria, because clearly the input data is not empirical, but is model data about hypothetical galaxies (Struck, 1997). In this case, the simulation would not even be classified as a hybrid in the first place.

There are of course other kinds of simulations in astronomy that make heavy use of empirical input data, like the Bolshoi simulation (HIPACC, 2011). The Bolshoi simulation is a simulation by our criteria because the output of the simulation (evolution of the universe or, rather, of regions of the universe) is not a cause of the initial state nor is it located at the same time and place. The classification of the Bolshoi simulation as a simulation and not as an experiment is in agreement with the self-ascription by its creators, and it is intuitively plausible that it is a simulation and not an experiment.

This brief survey of examples indicates that our criteria for distinguishing experimental measurements that involve the computational refinement of data from simulations based on empirical input data can account for many prominent examples of advanced science. This in turn suggests that the criteria articulate at least an implicit standing convention for distinguishing data-based simulations from empirical measurements. It still leaves open the philosophical question whether and how this practice can be justified epistemologically. However, this answer to the problem of hybrids builds on a structural feature that is already present in traditional measurement instruments and that has been described here as the measuring the cause by its effects pattern. Therefore, I conjecture that the problem of justifying it is either exactly the same or very similar to that of justifying traditional measurement or observation methods which rely on this pattern. For example, we say we measure the temperature, when in fact we are measuring the extension of the volume of a liquid in a thermometer and infer the temperature with the help of a scale. Still, we consider the temperature value as empirical data and I believe we do so because the kind of inference we make adheres to the two conditions stated above.

Summary and Open Questions

In this chapter I have argued that experiments and simulations and, by the same token, empirical measurements and theoretical calculations are clearly separate and well-distinguished categories. I have defended this distinction against what appears to me to be a strong tendency towards the contrary in the newer philosophy of simulation literature. However, the problem of hybrid methods (i.e., methods that combine empirical measurement of data with the computational processing of this data) raises conceptual problems that are not so easily solved. There are different possible approaches to solving these problems. In my opinion, the best way to frame these problems is by asking the question: what distinguishes a computer simulation based on empirical input data from an empirical measurement that involves the computational refinement of data? My answer consisted in transferring a typical of pattern of traditional measurement methods to the case of hybrids.

Several questions remain open, however. First, as the approach proposed by me is not the only possible or promising approach, it can still turn out that other approaches work better. Alternatively, it could turn out that no universal answer can be given, but only different answers for different subject areas. For the area of economic simulations, in particular, the approach proposed by Sebastian Zacharias and Moritz Lenel appears to be the best suited and promising.

However, there are also other open questions. The definition of hybrids that I have used more or less silently assumes that the output data really is computed from the input data and not ignored or dropped or the influence of the empirical component changing over time. However, plausible cases where this does not hold can at least be imagined: imagine, for example, a control device that regulates a machine based on data it receives from sensors. Let us assume that since the sensors tend to be unreliable from time to time, the regulatory device runs a simulation of the machine alongside the sensors. Whenever some kind of plausibility test shows that the sensors have delivered unreliable data, the machine switches to the simulation. Otherwise, it uses the sensor data as input and updates the simulation with the measured state of the machine. While it is not possible to tell whether the data produced by the device is empirical or not, this case turns out to be rather unproblematic upon closer inspection. For lack of another word, we could describe the data produced by this device as *potentially empirical data*. Now regarding the epistemic potential of this data, it is clear that this data can only be used in those contexts where in principle simulation data also would suffice (provided it is accurate enough), but not in those contexts, like empirical theory testing or model validation, where real empirical data is indispensable.

Similarly unproblematic is the case where a switch between empirical and simulation sources of input data does not occur, but where empirical and simulation sources are merged. This case is already covered by the theory of hybrids proposed here: as long the empirical data source has any significant influence on the computed output, the procedure can be classified as empirical data. In principle, it is suitable for all purposes for which real empirical data is needed. Of course, the details still matter. If a theory is to be tested, then the validity of any model that is required for producing (or better, revealing) the empirical data against which it is to be tested must be independent from the theory. This must of course already be considered in the case of conventional measurements. It does not constitute a novel or singular problem of computationally enhanced measurement techniques.

Acknowledgements

I would like to express my gratitude to the participants of the 2011 SimTech-workshop “Computer Simulations and the Changing Face of Scientific Experimentation.” The discussions have been very valuable for me. Some of the thoughts and questions raised in the section “The challenge of hybrid methods” and the following sections, in particular, have been inspired by discussions with Paul Humphreys.

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the SimTech Simualation Technology Cluster at the University of Stuttgart (http://www.simtech.uni-stuttgart.de/).

References

Axelrod, R. "Simulation in the Social Sciences." In *Handbook of Research on Nature Inspired Computing for Economics and Management* (Volume 1 of *Nature Inspired Computing and Social Sciences)*, edited by J-P. Rennard. Hershey, PA: IGI Global, 90–100, 2006.

Barberousse, A., Franceschelli, S., and Imbert, C. "Computer simulations as experiments". *Synthese* 169, (2009): 557–574.

CERN. *Worldwide LHC Computing Grid* 2012.
http://public.web.cern.ch/public/en/LHC/Computing-en.html

Gilbert, N., and Troitzsch, K. G. *Simulation for the Social Scientist.* McGraw-Hill, 2005.

Guala, F., "Models, Simulations and Experiments". In *Model-Based Reasoning: Science, Technology, Values*, edited by L. Magnani and N. Nersessian. Kluwer Academic Publishers, 59–74, 2002.

Guala, F. "Reciprocity: Weak or Strong? What Punishment Experiments Do (and Do Not) Demonstrate". *Behavioral and Brain Sciences* 35, (2012): 1–59.

Heath B., Hill, R., and Ciarello, F."A Survey of Agent-Based Modeling Practices (January 1998 to July 2008)". *Journal of Artifical Societies and Social Simulation (JASSS)* 12, no. 4, (2009): 9. http://jasss.soc.surrey.ac.uk/12/4/9.html

HIPACC (High-Performance Astronomic Computing Center). Bolshoi Cosmological Simulations, 2011. http://hipacc.ucsc.edu/Bolshoi/

Humphreys, P. *Extending Ourselves. Computational Science, Empiricism and Scientific Method.* Oxford University Press, 2004.

Kästner, J., and Arnold, E. “When can a Computer Simulation Act as Substitute for an Experiment? A Case Study from Chemistry”. Working Paper, 2012. http://www.eckhartarnold.de/papers/2012_Simulations_as_Virtual_Experiments/Simulations_as_Virtual_Experiments_long_version.html

Kramer, A., and Radde, N. “Towards Experimental Design Using a Bayesian Framework for Parameter Identification in Dynamic Intracellular Network Models”. In *Complex Adaptive Systems*, edited by C. H. Dagli, Elsevier, 1645–1653, 2010.

Küppers, G. and Lenhard, J. “Computersimulationen: Modellierungen 2. Ordnung”. *Journal for General Philosophy of Science* 36, (2005): 305–329. http://dx.doi.org/10.1007/s10838-006-2348-1

Lee, D. C., and Carroll, T. J. “Magnetic Resonance Imagin*g”.* In *Practical Signal and Image Processing in Clinical Cardiology*, edited by J. J. Goldberger and J. Ng. Dordrecht: Springer Verlag, 251–273, 2010.

Morgan, M. S. “Experiments Without Material Intervention. Model Experiments, Virtual Experiments, and Virtually Experiments”. In *The Philosophy of Scientific Experimentation*, edited by H. Radder. University of Pittsburgh Press, 216–233, 2003.

Morrison, M. “Models, Measurement and Computer Simulation: the Changing Face of Experimentation”. *Philosophical Studies* 143, (2009): 33–57.

Parker, W. S. “Does Matter Really Matter? Computer Simulations, Experiments, and Materiality”. *Synthese* 169, (2009): 483–496.

Peschard, I. “Is Simulation an Epistemic Substitute for Experimentation?” In *Simulations and Networks*, edited by S. Vaienti. Paris: Hermann 1–26 (online preprint), 2012.
http://ipeschard.free.fr/Is_simulation_an_epistemic
%20_substitute.pdf

Rohrlich, F. “Computer Simulation in the Physical Sciences”. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association. Volume 2: Symposia and Invited Papers*. University of Chicago Press, 507–518, 1990.

Struck, C. “Simulations of Collisions Between Two Gas-Rich Galaxy Disks with Heating and Cooling”. *The Astrophysics Journal Supplement Series* 113, (1997): 269–309.

Winsberg, E. “A tale of Two Methods”. *Synthese* 169, (2009): 575–592.

––– *Science in the Age of Computer Simulation*. Chicago and London: The University of Chicago Press, 2010.