Validation of Computer Simulations from a Kuhnian Perspective

Eckhart Arnold, Bavarian Academy of Sciences and Humanities

August 2018

erscheint in: Beisbart, C. & Saam, N. J. (eds.), Computer Simulation Validation - Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives, Cham: Springer 2019.

Abstract

While Thomas Kuhn's theory of scientific revolutions does not specifically deal with validation, the validation of simulations can be related in various ways to Kuhn's theory: 1) Computer simulations are sometimes depicted as located between experiments and theoretical reasoning, thus potentially blurring the line between theory and empirical research. Does this require a new kind of research logic that is different from the classical paradigm which clearly distinguishes between theory and empirical observation? I argue that this is not the case. 2) Another typical feature of computer simulations is their being "motley" (Winsberg, 2003) with respect to the various premises that enter into simulations. A possible consequence is that in case of failure it can become difficult to tell which of the premises is to blame. Could this issue be understood as fostering Kuhn's mild relativism with respect to theory choice? I argue that there is no need to worry about relativism with respect to computer simulations, in particular. 3) The field of social simulations, in particular, still lacks a common understanding concerning the requirements of empirical validation of simulations. Does this mean that social simulations are still in a pre-scientific state in the sense of Kuhn? My conclusion is that despite ongoing efforts to promote quality standards in this field, lack of proper validation is still a problem of many published simulation studies and that, at least large parts of social simulations must be considered as pre-scientific.

Keywords: Computer Simulations, Validation of Simulations, Scientific Paradigms

Contents

1 Introduction

Thomas Kuhn (1976) famously introduced the term *paradigm* to characterize the set of background beliefs and attitudes shared by all scientists of a particular discipline. According to Kuhn these beliefs and attitudes are mostly centered around *exemplars* of good scientific practice as presented in the textbook literature, but classical texts, specific methodological convictions or even ontological commitments can also become important for defining a paradigm. Furthermore, paradigms comprise shared convictions as well as unspoken assumptions of the group of researchers (Kuhn, 1976, postscript). An important function of paradigms is that they both define and limit what counts as relevant question and legitimate problem within a scientific discipline.

Kuhn's concept of a paradigm is closely connected with his view of how science develops. According to Kuhn phases of *normal science* where science progresses within the confinements of a ruling paradigm are followed by *scientific revolutions* which, in a process of creative destruction, lead to a paradigm-shift. Scientific revolutions are triggered by the accumulation of problems that are unsolvable within the ruling paradigm (so called *anomalies*). With an increasing number of anomalies scientists grow unsatisfied with the current paradigm and start to look for alternatives – a state of affairs that Kuhn (1976, ch. 7/8) describes as the *crisis* of the ruling paradigm. Then, a paradigm-shift can occur that consists in a thoroughgoing conceptual reorganization of a scientific discipline or, as the case may be, the genesis of a new sub-discipline. Unless there is a crisis, the search for alternative paradigms is usually suppressed by the scientific community.

This theory could be relevant for computer simulations and their validation. Because computer simulations are sometimes characterized as a revolutionary new tool that blurs the distinction between model and experiment, the question can be asked if this tool brings about or requires new paradigms of validation. Under *validation* I understand a process which allows to test whether the results of a scientific procedure adequately capture that part of reality which they are meant to explain or to enable us to understand. It is widely accepted that for theories or theoretical models, the process of validation consists in the empirical testing of their consequences by experiment or observation, which in this context is also often described as *verification* or *falsification* or, more generally, as *confirmation*.[1] The question then is, if the same still holds for computer simulations, that is, if computer simulations also require some form of empirical validation before they can be assumed to inform us about reality.

For the purpose of this paper, I understand empirical validation in a somewhat wider sense that does not require strict falsification, but merely any form of matching theoretical assumptions with empirical findings. In this sense, a historian checking an interpretation against the historical sources can also be said to validate that interpretation. However, I assume that proper validation always includes an empirical component and I therefore use the terms "validation" and "empirical validation" interchangeably in the

[1] In the realm of computer simulations the term *verification* is, somewhat confusingly, reserved for checking wether the simulation software is free from programming errors (so called "bugs") and whether it is faithful to the mathematical model or theory on which it is based. The term *validation* is used for the empirical testing of the simulation's results. See also Chapter 4 (Murray-Smith, 2019) in this volume.

following.

In the following, I first summarize Kuhn's philosophy of science (Sec. 2). Then I list some of the dramatic changes that computer simulations have brought about in science and – in order to forestall possible misunderstandings – explain why these changes are not scientific revolutions in the sense of Kuhn (Sec. 3). In the main part of this chapter (Sec. 4), I then examine the validation of simulations from a Kuhnian perspective. Relating to the discussion about the relation between computer simulations and experiments I argue that computer simulations can clearly be distinguished from real experiments and, therefore, do not require a new paradigm of validation. In principle, validating simulations is just like validating theory. I continue by examining whether computer simulations aggravate the problem of theory choice that is associated with the so called "Duhem-Quine-thesis" (S. G. Harding, 1976), which I deny. Finally, I examine some of the issues that the validation of social simulations and in particular agent-based-models raises from the point of view of Kuhn's philosophy of science. For the lack of commonly accepted standards of validation, it seems unclear whether this field has already reached a state of "normal science" with established paradigms of validation. Because the practices of validation vary greatly in this field, a general conclusion is not possible, however. I therefore confine myself to discussing the issue with respect to selected examples.

2 Kuhn's philosophy of science

A crucial aspect of Kuhn's concept of scientific revolutions is the alleged *incommensurability* of paradigms (Kuhn, 1976, ch. 12, postscript 5.) (Sismondo, 2010, ch. 2) (Bird, 2013, sec. 4.3f.). Incommensurability means that theories rooted in different paradigms cannot easily be compared with respect to their scientific merits, because of

1. *methodological incommensurability*, which means that the criteria of evaluation depend on and change with the paradigm,

2. the *theory-ladenness of observation*, due to which an assessment based on empirical evidence may not be able to resolve the dispute,

3. *semantic incommensurability*, which means that the differences of the respective conceptual reference frameworks and taxonomies may render the translation between the nomenclatures of different paradigms difficult and error-prone.

Kuhn did not go as far as the proponents of the strong programme of sociology of science who maintain that the resolution of interparadigm-disputes is primarily, if not exclusively, determined by social factors such as group allegiance and power-structures (Bird, 2013, sec. 6.3). However, he did deny that the choice between different theories is guided by a scientific meta-method such as systematic falsification or by any other particular set of rules. In this respect one can describe Kuhn's stance as a *mild relativism*. Kuhn's relativism is restricted by his belief that a common ground for theory choice can still be found in such general characteristics as empirical accuracy, consistency, breadth of scope, simplicity or parsimony, fruitfulness for future research (Kuhn, 1977, ch. 13). And he furthermore holds that the comparison and mutual evaluation of paradigms is possible on the pragmatic basis of their problem-solving capacity.

Although Kuhn regarded scientific revolutions and the paradigm shifts they bring about as scientifically perfectly legitimate processes, that is processes that are primarily driven by a scientific motivation and not just by social power, he nonetheless found that in almost any paradigm change some things get lost – if only that certain questions will not be considered worthwhile any more. An example is the question how physical bodies influence each other over a distance, which cannot be answered by Newton's theory of Gravity and therefore simply was not asked any more, although, before Newton it was considered important (Kuhn, 1976, ch. 12). The phenomenon that accepted questions, problems and even solutions can become orphaned after a paradigm shift has subsequently been called *Kuhn loss* (Bird, 2013, sec. 2).

Also, even though Kuhn allowed for paradigm-shifts to make sense scientifically, this does not always need to be the case, but one should expect that sometimes paradigm-shifts are primarily due to social factors. Not in the least because of the popularity of Kuhn's theory of scientific revolutions, it has become seductive for scientists to stage a paradigm shift to promote their scientific agenda. In order to distinguish illegitimate paradigm-shifs terminologically, the derogatory term *scientific imperialism* can be used, which has been coined to describe the take-over of a branch of science by a single paradigm (Dupré, 1994) by unfair means. Following Kuhn's line of thought the problem solving capacity could be a criterion by which to qualify a paradigm shift as either legitimate or imperialistic. Because of the incommensurability issues described before, an objective judgment about this can, of course, be difficult.

A contemporary of Kuhn that is often mentioned in the same breath, is Paul Feyerabend, who is (in-)famous for the slogan "anything goes". In popular folklore this is sometimes understood as meaning that Feyerabend advocated that in science any method is as good as any other. However, what Feyerabend actually demonstrated in his book "Against Method: Outline of an Anarchist Theory of Knowledge" (Feyerabend, 1975/1983) and other works was that even from the most humble historical beginnings, a serious scientific theory or school of thought can still emerge. Feyerabend's work gains its thrust from the fact that he can show that some of the game changers in the history of science such as, for example, Galileo's theory of motion, violated accepted scientific standards of their time (Feyerabend, 1975/1983, ch. 9). Just as Kuhn he denies that the historical development of science is or can be guided by methodological or epistemological rules. Similar to Kuhn, Feyerabend's philosophy has a certain relativistic flair, which Feyerabend other than Kuhn was ready to accept (Preston, 2016, sec. 5).

Nonetheless, despite of what the subtitle of his major work suggests, Feyerabend's analyses do not warrant a strong relativism. Almost all of Feyerabend's examples concern theories that – later in their historical development – would be considered as scientific even by conventional standards. Thus, what we can learn from Feyerabend is a certain tolerance against the methodological chaos of new scientific approaches in their infant stages. This can be important, for example, when evaluating social simulations, which according to some authors suffer from a lack of proper empirical validation (Heath, Hill, and Ciarallo, 2009). The question is then not so much whether these simulations adhere to a particular scientific standard but rather whether the respective scientific community learns from its failure to do so and will be able to develop appropriate methodological standards in the future.

Another point that deserves clarification, because it is – at least in the philosophical discussion – almost habitually mentioned in context with Kuhn, is the *Duhem-Quine-thesis* (S. G. Harding, 1976). The Duhem-Quine-thesis draws on the fact that if the logical consequence of a whole system of premises turns out to be false then it is still unclear which one or more of the premises are false.[2] This means that if a theory is empirically disconfirmed, we do not (yet) know which part of the theory is wrong. The Duhem-Quine-thesis can be seen as supporting a certain degree of arbitrariness, if not relativism in theory choice. And it corresponds well to Kuhn's view that the way scientists cope with anomalies is not strictly guided by methodological rules. It may be a matter of creative choice. As we shall see later, this choice is in practice much less arbitrary than it may appear in the formal logical representation of a theory as a system of propositions.

Despite all reservations, Kuhn's picture of the history of science is still one of linear development, where normal science and revolutionary phases follow each other in time. For Kuhn the prolonged co-existence of several competing paradigms was the mark of a pre-scientific stage where much intellectual energy is wasted in disputes between rivaling schools of thought. Recent research, however, has emphasized that the co-existence of different paradigms within one and the same science is much too common to be dismissed as pre-scientific (Kornmesser, 2014; Schurz, 2014). This is particularly true of the social sciences, where hardly ever one paradigm can claim to solve all puzzles so successfully that it is able to gather the entire scientific community under its flag. That Kuhn may have underestimated the amount of co-existence of paradigms in science does not invalidate his analyses, though. The concepts of *normal science* and *scientific revolutions* can still be employed as ideal-types to characterize the scientific proceedings within an established paradigm on

[2] See also chapter 39 (Lenhard, 2019) in this volume.

the one hand and the discourse between different co-existing paradigms on the other hand.

3 A revolution, but not a Kuhnian revolution: Computer simulations in science

Kuhn's theory of scientific revolutions is so popular that his concept of a paradigm has by now become part of the common vocabulary. Inevitably, it is often used in a sense that is different from what Kuhn had in mind. It may therefore help to make clear what is not a revolution or paradigm change in Kuhn's sense. A most salient example in this context is that of the introduction of computer simulations to science, because it can with some justification be said that computer simulations have revolutionized many areas of science. these really Kuhnian revolutions?

Computer simulations can roughly be defined as the imitation of a natural process (or, in the case of social simulations, a social process) by a computer program (Hartmann, 1996). Undoubtedly, computer simulations have brought about considerable changes in scientific practice and theoretical outlook. Here are but some examples:

- In engineering, simulations have been used before long to simulate the properties of machinery and processes. A large class of simulations is based on the method of finite elements which has as far reaching applications as structural engineering, car crash tests and even cardiovascular simulations (Carusi, Rodriguez, and Burrage, 2013).
- In chemistry simulations are employed in order to simulate chemical processes on a quantum-mechanical bases, some of which are even outside the reach of direct experimentation (Kästner and Arnold, 2013).
- In climate science the simulations are used to simulate the possible future development of the world climate. Naturally, experimentation with the world climate is not possible. By the same token, unfortunately, these simulations cannot be validated directly.

- The theory of non-linear dynamical systems ("chaos theory") can even be said to owe much of its origin to computational methods (Gleick, 2011). At any rate its development has certainly been propelled by the use of computers, though it might not necessarily have been computer simulations in the narrower sense of imitations of a natural process in the computer.

- In social science there exists a now already long standing tradition of simulating social processes. However, the social simulations community still struggles for the acceptance within the broader social sciences community (Squazzoni and Casnici, 2013).

Some of these examples certainly warrant the characterization as "revolutionary". Are they revolutionary in a Kuhnian sense, though? And would it be reasonable to call simulation-based science in general a new paradigm of science?

For one thing, the way Kuhn used the term paradigm, paradigms are always tied to specific scientific disciplines. Even though we are not tied to Kuhn's definition and the term *paradigm* has indeed been used more liberally by other authors since its original introduction, it would appear a bit vague to speak of a paradigm of computer simulations, because it is not at all clear what would be the content of this paradigm.

Even more importantly, Kuhn reserves the concept of scientific revolutions for changes that are caused by a crisis of the conceptual framework of a scientific discipline and that lead to a reconstruction of the conceptual system that is incommensurable with the previous reference framework. Not any dramatic change in science is a revolution in the Kuhnian sense. A prominent example for a dramatic change that is not a Kuhnian revolution is the discovery of the structure of the DNA-molecule by Watson and Crick. While this discovery was a door-opener for molecular genetics, it neither required nor effected a conceptual reconstruction and there was no question of it being incommensurable with the previously held views on hereditary biology. Quite the contrary, it fit in nicely with the existing body of knowledge. The discovery of the DNA was normal science at its best, not a Kuhnian revolution.

Similarly, the introduction of computer simulations into a particular branch of science alone is not a Kuhnian revolution, no matter how dramatic the changes in scientific practice and the extension of our knowledge through computer simulations might be. Only, if the use of computer simulations leads to a revision of established fundamental concepts, it is a Kuhnian revolution. A possible candidate from the list above might be chaos theory, in so far as it has modified the received picture of causality.

4 Validation of Simulations from a Kuhnian perspective

Can Kuhn's concept of paradigm illuminate the validation of computer simulations? And, if so, how? In the following, I am going to state several questions that can be raised in this context and then try to give answers to these questions based on the current discussion on computer simulations in the philosophy of science. The questions that in my opinion deserve consideration are:

1. Notwithstanding the question (discussed earlier) to what extent computer simulations have prompted paradigm shifts *in* science, another question is, whether computer simulations have lead to, or require new paradigms in the logic of scientific discovery. Classical research logic assumes a clear distinction between theoretical research based on deductive inference and empirical research based on experiment and (potentially theory-laden) observation.[3] Most importantly, there is a hierarchy between the theoretical and empirical realm. Theoretical

[3]Because theory-ladenness of observation is an often misunderstood topic, two remarks are in order: 1) Theory-ladeness of observation as such does not blur the distinction between theory and observation. At worst we have a distinction between pure theory (without any observational component) and theory-laden observation. 2) Theory-ladeness of observation does not lead to a vicious circle when confirming theories by empirical observation. This is true, as long as the observations are not laden with the particular theories for the confirmation of which they are used. – There are areas in science where no sharp distinction between theoretical reasoning and reporting of observations is made. However, as far as computer simulations are concerned, it is clear that because Turing Machines do not make observations, a computer program is always a theoretical entity

assumptions are confirmed or disconfirmed by empirical tests – not the other way round. Computer simulations are sometimes depicted as being located somewhere between empirical and theoretical research, and – as the common metaphor of "computer experiments" suggests[4] – blurring the lines between the two (Morrison, 2009).

2. In a similar vein, computer simulations often rely on a rich mixture of assumptions and technicalities that are drawn from diverse sources. In the philosophical literature on simulations this has been described as their being "motley"(Winsberg, 2015) and not simply falling from theory. This can raise worries concerning the prospects of empirical validation of computer simulations. In particular, the question can be asked if the sort of problems associated with the Duhem-Quine-thesis increase with computer simulations: You may know that your simulation contains many abstractions, simplifications and presumptions, but you cannot be sure which of these are potentially dangerous.

3. Finally, some thoughts shall be given to the validation of simulations in the social sciences. Because the social sciences are multi-paradigm-sciences the validation of simulations raises specific problems in this area. Given that it is still not common practice to validate simulations, one can even ask whether the field of social simulations has already emerged from a pre-scientific state.

4.1 Do computer simulations require a new paradigm of validation?

While Kuhn's theory of scientific revolutions is mainly concerned with the supersession of scientific theories, his concept of paradigms can also be applied to other aspects of scientific practice. For example, it might be

- notwithstanding the fact that a computer program may represent an empirical setting or make use of empirical data. In the latter respect it can be compared with a physical theory that may in fact represent empirical reality as well as contain natural constants (i.e. empirical data).

[4]See also chapter 37 (Beisbart, 2019) in this volume.

applied to changes in the logic of scientific research. The question whether computer simulations bring about (or require) a new kind of research logic is particularly salient, because it has been argued recently that computer simulations somehow blur the line between models and experiments (Winsberg, 2009). But if this means that computer simulations are – just like experiments – somehow empirical, the question naturally arises whether the validation of computer simulations can still be understood along the lines of what has earlier been described as classical research logic. Or, if a new paradigm of validation is necessary to assess whether a simulation adequately captures its target system or not?

Before the recent discussion about the relation of simulations and experiments, this question seemed to be rather trivial and its answer obvious: Computers are calculating machines and computer simulations are nothing but programmed mathematical models that run on the computer. Therefore, computer simulations can just like models produce no other than purely inferential knowledge, that is, knowledge that follows deductively from the premises built into the simulation. In particular, computer simulations cannot produce genuine empirical knowledge like experiments or observations can. It is true that computer simulations can produce new knowledge, because they yield logical consequences of the built-in premises that were not formerly known to us (Imbert, 2017, sec. 1.3.4). It is also true that computer simulations can – like any model – produce knowledge about empirical reality, because the premises built into them have empirical content and so have their logical consequences. But this is far cry from the empirical knowledge that experiments or observations yield and which – because it is of empirical origin – is genuine. But then computer simulations have just the same epistemic status as theories and models and therefore follow the same research logic and require just the same kind of validation. Now, in order to validate a model or a theory it must be tested empirically, and so must computer simulations.

What I have just described is more or less the picture of computer simulations that was pertaining in the general literature on simulations up to the beginning of the millennium. It had by that time been fleshed out with two distinctions that make the difference between computer simulations and

Experiments

<table>
<tr><th></th><th>computer simulation</th><th>analog simulation</th><th>real experiment</th></tr>
<tr><td>materiality of object</td><td>semantic</td><td colspan="2">material</td></tr>
<tr><td>relation to target</td><td colspan="2">representation (formal similarity)</td><td>representative</td></tr>
</table>

Simulations

Figure 1: Conceptual relation of simulations and experiments (Arnold, 2013b)

empirical research procedures extraordinarily clear: Firstly, by the distinction of the modus operandi. Is it a *formal* procedure (computer simulation) or a *material* process (experiment)? Secondly, by the distinction of their relation to the target system. Accordingly, this relation could be characterized as one of *formal similarity* (Guala, 2002) with the object of the simulation being a *representation* (Morgan, 2003) of the target system or, in the case of experiments, one of *material similarity* with the object of experimentation being a *representative* of the target system.

In recent years, however, there has been a persistent discussion among philosophers of science during the course of which the distinction between simulations and experiments has been seriously called into question. Most notably, some authors have claimed that it is impossible to make a sharp distinction between simulations and experiments – at least as far their epistemic reach or inferential power is concerned. (Winsberg, 2009; Parker, 2009; Morrison, 2009; Winsberg, 2015). Others have advocated the weaker claim that while there is a distinction between the two categories, the transition between them is smooth and that there are borderline cases for which it is difficult to determine into which category they fall (Morgan, 2003).

Now, if this were true, then the generally accepted research logic of empirical science, which relies on the ability to distinguish clearly between empirical observation and theoretical reasoning would find itself in a serious crisis and we would have to expect and, in fact, need to hope for new paradigms of research logic and, in particular, for the validation of computer simulations to emerge.

However, the case for the non-discriminability of simulations and experiments rests almost entirely on conceptual confusions and an ambiguous use of the term "experiment". The examples with which supporters of the non-disicriminabilty thesis demonstrate their claim concern almost exclusively atypical kinds of experiments, where the object of experimentation is not really a representative of the target system. For example, Winsberg (2009, p. 590), discusses "tanks of fluid to learn about astrophysical gas-jets" as an instance of an experiment. But this is an atypical experiment, because the tanks of fluid are not representatives of the target system (astrophysical gas-jets). This kind of experiment is indeed in no better position to produce genuine empirical knowledge about the target system than any computer model. But the fact that there are such atypical experiments does not contradict the fact that there exist real experiments that can produce genuine empirical knowledge about their target system and that this is a feature that distinguishes real experiments from models.

The conceptual confusion that exists in the philosophical discussion about the relation of simulations and experiments can easily be clarified by the schema on figure 1, which depicts the overlap in the use of the words "simulation" and "experiment". The kind of experiments that Winsberg and other authors advocating the non-discriminability between simulations and experiments discuss over and over again, has been termed "analog simulation" in the schema. As all experiments do, "analog simulations" operate on a material object, but this object does not have a material similarity to its target system and therefore is only a representation, but not a representative of its target system. The latter is required for an experiment to produce genuine empirical knowledge about its target system.

That simulations are not experiments – save for the ambiguity and overlap in the use of words – becomes furthermore clear if we consider the kind of experiments that give rise to anomalies and which in retrospect are declared crucial experiments that decide the choice between conflicting theories. Because the laws of the scientific theories are programmed into computer simulations, they cannot be used to test these very theories. If it really was as difficult to distinguish between simulations and experiments as some philosophers of science believe, then it should – at least in principle – be possible to substitute experiments with simulations in any context.

However, if we draw the demarcation-line between analog simulations and real experiments and not, as the authors advocating the non-discriminability-thesis implicitly do, between computer simulations and analog simulations, then we are able to distinguish clearly those scientific procedures that can generate genuine empirical knowledge about their target system from those that cannot. Simulations and, in particular, computer simulations belong to the latter category and therefore have – with respect to validation – the same epistemic status as theories and models. They need to be validated empirically, but they cannot provide empirical validation.[5]

Summing it up, computer simulations do not break the received paradigm of research logic of empirical science. Therefore, a new paradigm of validation specifically for simulations is not needed.

4.2 Validation of simulations and the Duhem-Quine-thesis

Another point frequently emphasized in the philosophy of simulation literature is that computer simulations can become highly complex. This is also one of the major differences between computer simulations and thought experiments, to which they are otherwise quite similar. At least in the natural sciences computer simulations can often be based on comprehensive and well tested theories, such as quantum mechanics, general relativity, Newton's of gravitation or – in engineering – the method of finite elements. But even in the natural sciences simulations cannot always be based on a single theory, but they sometimes rely on different theories from different origins. Climate simulations are a well-known example for this. And even where simulations are based on a single theory, they usually also draw on various sorts of approx-

[5]In simulation-science the term *empirical* is sometimes used to distinguish simulation and numerical methods from mathematical analysis. (Phelps (2016) is an example of this.) But this is just a different use of words and should not be confused with "empirical" in the sense of being observation-based as the word is understood in the context of empirical science.

imations, local models and computational techniques. None of these can be derived from theory, so that they need independent credentials. This situation has been described in the philosophy of simulation literature as their being motley and partly autonomous (Winsberg, 2003). This description echos a recent trend in the philosophy of science which emphasizes the importance and relative independence of models from theory (Morgan and Morrison, 1999; Cartwright and Press, 1983).

So, if simulations are knit together from many independent set pieces of theories, models, approximations, algorithmic optimizations etc., then the Duhem-Quine-thesis could point out a potential problem. A possible reading of the thesis assumes that if validation fails (for example, because an empirical prediction was made that turned out to be wrong), then one cannot know which part of the chain of theoretical reasoning failed that leads to the empirical prediction. In the case of computer simulations this means that one does not know whether the theory on which the simulation is based, the simplifications that may have been made in the course of modeling or, finally, the program code has failed.

By the same token, if this reading of Duhem-Quine is accurate, simulation scientists would – for better or worse – enjoy a great freedom of choice concerning where to make adjustments if a simulation fails, i.e. if it leads to unexpected, obviously false or no results at all. Some philosophers have even argued that scientists sometimes deliberately employ assumptions that are known to be false to make their simulations work. Among these are artificial viscosity (Winsberg, 2015, sec. 8), or – another often cited example – "Arakawa's trick" (Lenhard, 2007). Arakawa based a general circulation model of the world climate on physically false assumptions to make it work, which by the scientific community was accepted as a technical trick of trade.

However, this reading of Duhem-Quine paints a somewhat unrealistic picture of scientific practice, because in case of failure there usually exist further contextual cues where the error causing the failure has most likely occurred. While in the abstract formal representation of theories that is sometimes used to explain Duhem-Quine, the premises are represented as propositions with no further information, scientists usually have good reasons to consider the failure of some premises as more likely than others. In science and engineering, the premises are usually ordered in a hierarchy that starts with the fundamental physical, chemical or biological theories, ranges over various steps of system description and approximation down to the computer algorithms and, ultimately, the programm code. If a simulation fails one would start to examine the premises in backward order. And this is only reasonable, because prima facie, it is more likely that your own program code contains a bug than, say, that the theory of quantum mechanics is false or that some of the tried and tested approximation-techniques are wrong. Though, of course, this is not completely out of the question, too.[6] It should be understood that the credibility of the various premises occurring in this hierarchy does not follow their generality, but depends on their respective track record of successful applications in the past. It can safely be assumed that this situation is typical for normal science.[7]

It must be conceded, though, that during a scientific revolution or within cross-paradigm-discourse, there might be no hierarchy of premises to rely on, because some of the premises higher up in the hierarchy, like the fundamental theories, are not generally accepted any more. In this situation, there might, as Kuhn suggested, only be vague meta-principles left to rely on and we must face the possibility of not being able to resolve all conflicts of scientific opinion.

What about the conscious falsifications like artificial viscosity and "Arakawa's trick" that – according to some philosophers of science – are introduced by simulations scientists in order to make their simulations work? This reading has not gone unchallenged, and it has been called in to question whether the artificial viscosity that Winsberg mentions is more than just an-

[6]See Kästner and Arnold (2013, sec. 3.4) for a case-study containing a detailed description of this hierarchy of premises.

[7]But see Lenhard (2019) in chapter 39 in this book, who paints a very different picture. I cannot resolve the differences here. In part they are due to Lenhard using examples where " 'due to interactivity, modularity does not break down a complex system into separately manageable pieces.' " To me it seems that as far as software design goes, it is always possible – and in fact good practice – to design the system in such a way that each unit can be tested separately. As far as validation goes, I admit that this may not work as easily because of restrictions concerning the availability of empirical data.

other harmless approximation (Peschard, 2011) or whether "Arakawa's trick" not merely compensates for errors made at another place, which would make it an example of a simulation the success of which is badly understood rather than one that is very representative of simulation-based science (Beisbart, 2011, 333f.). It seems that these philosophically certainly interesting examples concern exceptions rather than what is the rule in the scientific practice with simulations. For the time being that is to say, because it is well imaginable that in the future development of science these tricks become more common.

Summing it up, with respect to the Duhem-Quine-thesis there are neither additional challenges nor additional chances for the validation of simulations. Under *normal science*-conditions it does not play a role at all. Other than that it merely reflects the greater methodological imponderabilities during a revolutionary phase or in an inter-paradigm context.

4.3 Validation of social simulations

Most of the discussion so far and all of the examples were centered around science and engineering. Therefore, in the following I am going to briefly discuss questions concerning the validation of simulations that are more specific for the social sciences.

4.3.1 Where social simulations differ

In the context of validation of social simulations two features of the social sciences become relevant that distinguish them from most natural sciences: Firstly, the social sciences are multi-paradigm-sciences. It is the normal state of these sciences that there exist multiple more or less mutually incommensurable paradigms at the same time. This multi-paradigm-character is well described in the textbook by Moses and Knutsen (2012). For Kuhn such a state of affairs was a sign of a pre-scientific phase. But given that the social sciences are – within inevitable confinements – nonetheless able to produce convincing explanations at least for some social phenomena, the qualification as pre-scientific seems inadequate. Also, if considered in isolation, most of these paradigms expose typical features of normal science, like a textbook-literature, role models and exemplars etc.

Deviating from Kuhn, I therefore suggest, that the qualification as pre-scientific should be reserved to those sciences or branches of a science that – given their state of development – have not yet been able at all to produce results that can be validated or confirmed by some reasonable procedure. The qualification as pre-scientific is in so far justified as without a common understanding and practice of validation one can never be sure whether the results are indeed reliable.

Secondly, the social sciences include qualitative paradigms, including paradigms that rely on hermeneutical methods. It is safe to assume that these can neither be completely ignored nor always be resolved to quantitative or otherwise formal methods and paradigms.[8] As computer simulations are quantitative, the decision to use computer simulations is also a decision for a quantitative paradigm.

Here, I understand the term "quantitative" in a wide sense, including anything that is described in a formal language. This can be formal logic, mathematics, or a programming language. This wide sense of using the term "quantitative" is motivated by the fact all formal descriptions share the same epistemic risks of either losing important information, because the expressive power of formal languages is limited in comparison to natural language, or adding arbitrary assumptions in form of modeling decisions. A simulation model forces its author to provide detailed mechanics of all processes that are included in the model, because otherwise the model would not run. However, if the mechanics are not known, this amounts to theoretical speculation. A purely verbal description, in contrast, allows its author to remain silent or at least adequately vague about underlying mechanics the details of which are not known. On the other hand, because of their strict specification, formal models cannot as easily be misunderstood as verbal descriptions. And they

[8] There are scientists who deny even this and who also believe that without formal models no explanation of any sort is possible in history or social science. I am a bit at loss for giving proper references for this point of view, because I have mostly been confronted with it either in discussions with scientists or by anonymous referees of journals of analytic philosophy. The published source I know of that comes closest to this stance is the keynote "Why model?" by Joshua Epstein (2008), which I have discussed in Arnold (2014).

enforce logical consistency.

Both of these features affect the validation of social simulations. Because, when trying to validate a simulation study, say, on the evolution of cooperation, it might become necessary to compare its findings with those of biological field research or, depending on the envisaged application cases, those of cultural history. Thus, different scientific disciplines with different paradigms might be affected. And, it might become necessary to translate between a qualitative descriptive language used in empirical research and the formal languages used in simulation research.

One possible objection when discussing social simulations in the connection with Kuhn, is that it is not a scientific discipline, but a field that runs across several disciplines. However, since this field is shaped by shared attitudes, well-known exemplars (Axelrod, 1984; Axtell et al., 2002; Epstein and Axtell, 1996; Schelling, 1971) and an emerging textbook-literature (Railsback and Grimm, 2012; Gilbert and Troitzsch, 2005), looking at it from a Kuhnian perspective does not seem too far-fetched.

4.3.2 Are social simulations still in a pre-scientific stage?

One of the most surprising features to the outside observer of the field of social simulations in general is the widespread absence of empirical validation, sometimes combined with a certain unwillingness to see this as a problem.

In a meta-study on agent-based-modeling (ABM), which is one very important sub-discipline of social simulations, Heath, Hill, and Ciarallo (2009) find that the models in 65% of surveyed articles have not properly been validated, which they consider "a practice that is not acceptable in other sciences and should no longer be acceptable in ABM practice and in publications associated with ABM" (4.11). While some of these not-validated simulations can serve a purpose as thought experiments that capture some relevant connection in an idealized and simplified form (Reutlinger, Hangleiter, and Hartmann, 2017), many of them are merely follow-ups to existing simulations and bear little relevance of their own. The practice of publishing simulations without empirical validation and seemingly little (additional) theoretical relevance is so widespread that it has been termed the YAAWN-Syndrome where YAAWN stands for "Yet Another Agent-Based Model ... Whatever ... Nevermind" (O'Sullivan et al., 2016). The fact that such a term has been coined is an indication that the ABM-community is growing weary of unvalidated or otherwise uninteresting simulations. Thus, the situation may change in the future. For the time being, lack of validation is still a problem.

To be sure, agent-based-modeling is a broad field. On the one hand side there are very theoretical simulations that set out from abstract concepts but without any particular application case in mind. And on the other hand side there exist simulations that are right from the start related to a particular empirical setting. The latter kind of simulations is typically found in corporate or political consulting. I am going to look at the theoretical simulations first and then consider the more applied kinds of simulations later.

Naturally, unvalidated simulations are much more prevalent among the theoretical simulations, where the lack of empirical validation is sometimes not even perceived as a problem. This may be illustrated by a quotation from an interview with a philosopher who has produced models of opinion dynamics (Hegselmann and Krause, 2002) that have frequently been cited in other modelling-studies but that have not been empirically validated:

> None of the models has so far been confirmed in psychological experiments. Should one really be completely indifferent about that? Rainer Hegselmann becomes almost a bit embarrassed by the question. "You know: In the back of my head is the idea that a certain sort of laboratory experiments does not help us along at all." (Grötker, 2005, p. 2)

But if laboratory experiments do not help us along, how can models that have never been confirmed empirically either by laboratory experiments or by field research help us along? This lack of interest in empirical research is all the more surprising as opinion dynamics concern a field with an abundance of empirical research. Naively, one should assume that scientists have a natural interest in finding out whether the hypotheses, models and theories they produce reflect empirical reality. That this is obviously not always the case, confirms Kuhn's view that the criteria by which scientific research is judged are also set by the paradigm that guides the thinking of the researchers and that there is no such thing as a "natural" scientific method independent of paradigms. However, even Kuhn's mild relativism would rule out science without any form of empirical validation as unrewarding.

The lack of empirical concern within the field of social simulations can furthermore be attributed to another working mechanism of paradigms that Kuhn identified, namely, the role of *exemplars*. As mentioned earlier, according to Kuhn scientific practice is not guided by the abstract rules of a logic of scientific discovery. Instead, scientists follow role models or *exemplars* of good scientific practice.

Some very influential role models in the field of social simulations concern simulations that have never successfully been validated. The just mentioned opinion-dynamics simulation by Hegselmann and Krause is one example for this kind of role model. But the arguably most famous unvalidated model that serves as an exemplar in Kuhn's sense is Robert Axelrod's "Evolution of Cooperation" (Axelrod, 1984). Despite the fact that the reiterated Prisoner's Dilemma simulations that Axelrod used as a model for the evolution of cooperation had turned out to be a complete empirical failure by the mid 1990s (Dugatkin, 1997) and despite the devastating criticism Axelrod's approach had received from theoretical game theory (Binmore, 1994; Binmore, 1998), it continues to be passed down as a role model of social simulations until this day. In a journal article from 2010 in the prestigious *Science*-journal, where a similar research design as Axelrod's was employed, it is mentioned as a role model that has been "widely credited with invigorating the field" (Rendell et al., 2010, 2008f.). And one can easily find recent studies (Phelps, 2016) that naively pick up Axelrod's study as if no discussions concerning its robustness, its empirical validity or its theoretical scope had ever taken place in the meantime. If simulation-research-designs without proper validation such as Axelrod's continue to be treated as exemplars, it is no surprise that many social simulations lack proper validation.

Now, there are two caveats: Firstly, in some cases unvalidated simulations can serve a useful scientific function, among other things as thought-experiments. Of a thought experiment one usually does not require empirical validation. Thus, if Axelrod's evolution of cooperation or Hegselmann's and Krause's opinion dynamics could be considered thought experiments their status as role models in connection with their lack of empirical validation could not be taken as an indication that social simulations still remain in a pre-scientific stage. However, the way that both these simulations functioned as role models was not by their (potential) use as thought experiments, but as a research programme. Indeed, it would be hard to justify the literally dozens if not hundreds of follow-up simulations to Hegselmann-Krause or Axelrod as thought experiments without invalidating the category of a thought experiment as a useful scientific procedure. But it has to be kept in mind that not any kind of unvalidated simulation is an indication of pre-scientific fiddling about.

Secondly, and more importantly, not all simulation traditions have, of course, remained as disconnected from empirical research as Axelrod's Evolution of Cooperation and Hegselmann's and Krause's opinion dynamics simulations. One example is the Garbage-Can-Model (GCM) by Cohen, March, and Olsen (1972) which describes decision making inside organizations with a four component model, taking "problems", "solutions", "participants" and "opportunities" into account. This model is highly stylized and, because of this, would be difficult to validate directly. Nevertheless, it is frequently referred to in studies on organizational decision making, including empirical studies.

But why, one may ask, could the connection to empirical research, or more generally, other kinds of research on organizational decision making be established in this case while it failed in the aforementioned cases? There are several possible reasons:

- Modeling organizational decision making is a much more restricted topic than, say, modeling evolution of cooperation in general. This makes it easier to find the right abstraction level for modeling. While biologists complained about simulations of the evolution of cooperation that "Most repeated animal interactions do not even correspond to repeated games" (Hammerstein, 2003, p. 83), researchers from organizational science have no such difficulties in relating to the Garbage Can Model in their case studies (Fardal and Sørnes, 2008; Delgoshaei and Fatahi, 2013).

- Within organization theory working with stylized descriptions is generally accepted. Thus, the target that the simulation-model had to match was an already highly stylized verbal description. (Nonetheless, the simulation model did not represent the verbal description faithfully (Fioretti and Lomi, 2008, p. 1.4).) It is much easier to cast a stylized verbal description convincingly into a simulation-model than, say, a thick historical narrative as in one of Axelrod's suggested application cases (Northcott and Alexandrova, 2015; Arnold, 2008).[9]

- For the study of organizational decision making the Garbage Can Model seems to serve as a kind of vantage point. It helps to analyze and communicate organizational decision making problems by relating a particular decision making situation to the model – even if the model is only used as a conceptual reference framework and the actual simulation results are ignored.[10] Because of its popularity the Garbage-Can-Model could even be considered an exemplar in Kuhn's sense. To serve as a vantage point, a model does not need to be empirically validated or even testable. It stands to reason, though, that it still needs to be "realistic enough" in some weaker sense to serve this purpose.

- While for the latter purpose (vantage point) a stylized verbal description could suffice, simulation models have the advantage that they can be run.

[9] I am indebted to Julian Newman for pointing out to me the excellent paper by Northcott and Alexandrova (2015) on the Prisoner's Dilemma. It contains the so far best analysis why Alexrod's reinterpretation in terms of the Prisoner's Dilemma of truces in WWI ultimately fails. And because the author's have obviously not been aware of my own research on the topic, I consider it as an independent confirmation of my own critical conclusions regarding Axelrod's chapter on WWI (Arnold, 2008, ch. 5.2.2).

[10] This seems to be the standard case for applying the GCM in organizational science. See Fardal and Sørnes (2008) and Delgoshaei and Fatahi (2013) for example. It will be interesting to see whether the more refined simulation models of the GCM that have been published more recently (Fioretti and Lomi, 2008) will bring about an increased use of simulation models in applied studies referring to the GCM or not.

> This allows to generate hypotheses about the simulated process which can help to establish the basic plausibility of the model, if the simulation itself and its results are plausible in view of the prior knowledge about the simulated process.[11] In the case of the GCM the model establishes the connection between a certain structure of the decision making process and certain characteristics of the outcome, like how efficiently problems will be solved. In a verbal description this connection can be maintained, but not be demonstrated. A simulation can show that such a connection exists, even if only within the model.

In view of the possible functions of communication and hypotheses-generation, one can argue that models like the Garbage Can Model can be useful in the context of empirical research even without being empirically validated themselves. Still, the question remains what characteristics a model of this kind must have to be considered useful or suitable, or how one can tell a good model from a bad model. There seems to exist an intuitive understanding within the scientific communities habitually using these models, but it is hard to find any explicit criteria. This strengthens the impression that a paradigm of validation is not yet in place, at least not for the more theoretical simulations.

What about applied simulations, though? Agend-based-models are, among other things, used to give advice about particular policy measures, like introducing a new pension plan (A. Harding, Keegan, and Kelly, 2010) or determining the best procedures for research funding (Ahrweiler and Gilbert, 2015). Obviously, validation is of considerable importance if simulations are used for political consulting. So, how do scientists who apply social simulations get around the restriction that the simulation results often cannot directly be compared with measurable empirical data? In particular, how can simulations be validated that are meant to evaluate the possible consequences of policy measures that might never be implemented?

[11] This is precisely where Axelrod's simulations was lacking, because a) his tournament of reiterated Prisoner's Dilemmas is too far removed from the phenomenology of either animal or human interaction to be prima facie plausible, and b) his results were - unbeknownst to him - highly volatile with respect to the simulation setup and thus also lack plausibility.

In their discussion of the validation of the SKIN-model, which simulates knowledge dynamics in innovation networks, Ahrweiler and Gilbert (2015, section 1.1.2) do not even assume that there exist objective observations independent of a concrete research goal or question.[12] At least for the sake of the argument they even accept the view that the observation of a social process is a construct of this process or "what you observe as the real world" (Ahrweiler and Gilbert, 2015, section 1.2), just like the simulation of the same process is another construct of this process. However, since the authority over what is observed as the real world lies with the "user community" (Ahrweiler and Gilbert, 2015, section 1.3), the output of a simulation can meaningfully be compared with the observations.

Since the construction of the simulation as described by Ahrweiler and Gilbert (2015, section 2.4) is a process in which the user community is deeply involved, it is tempting to raise the question how unbiased this kind of validation really is. After all, an administration assigning the task of examining the potential for enhancement of their administrative procedures to a team of simulation scientists might be more interested in the vindication of certain administrative procedures than in their unbiased assessment. However, the "user community view" as described by Ahrweiler and Gilbert (2015) depicts only the outline of the construction and validation process of applied agent-based-models. A more detailed analysis of the validation of applied agent-based-models as provided by A. Harding, Keegan, and Kelly (2010) reveals that there exists a whole array of validation procedures which, if executed properly, limits the risk of producing biased or arbitrary results. For the Australian Population and Policy Simulation Model A. Harding, Keegan, and Kelly (2010) report, among other measures: i) the calibration and benchmarking of the simulation with available cross-sectional and longitudinal data, ii) the comparison of the simulation model's projection with that of other models, iii) the modular structure and separate evaluation of each module, iv) the examination, if both the individual agent's simulated life

[12]They discuss this unter the heading of "theory-ladenness of observations", though their examples suggest that the issue at stake is rather different interpretations of observations or a focus on different observations depending on the research questions than different observations due to a different theoretical background.

histories and the summary statistics yield reasonable results. The impact of proposed policy measures as revealed by the simulation can by its very nature not beforehand be compared with empirical data. However, one can contend that in the context of policy advise a simulation is sufficiently validated, if it leads to policy decisions that are better grounded than they would be without running a simulation model.

Where does this leave us? Are social simulations still in a pre-scientific stage with respect to their validation? On the one hand there is a widespread lack of proper validation and the impression that the increasing number of published agent-based models does not necessarily pay off in terms of further deepening our understanding of the simulated processes. While other quality issues of agent-based models, such as their reproducibility and mutual comparability, have been addressed in recent years,[13] there is still no common understanding concerning how agent-based models should be validated. So far, the textbooks on agent-based simulations have little to say about validation. With the central issue of validation still being unresolved, the field of social simulations does yet seem to have matured into a normal science in the sense of Kuhn. The situation can positively be a described as a phase of humble beginnings in the sense of the interpretation of Feyerabend's anarchic epistemology that was given earlier.

On the other hand, scientists that apply agent-based-models to particular empirical processes typically invest considerable time and effort into the validation of their simulations and employ a diverse set of validation procedures to ensure the credibility of their simulations. So, we might indeed be witnessing a paradigm of validation of applied agent-based-models in the making. It is, so far, only in the making, because the various validation procedures and criteria used by the practitioners do not yet seem to have been consolidated to a degree where they become textbook knowledge.

[13] A most notable initiative in this respect has been the introduction of the ODD Protocol for the standardized description of agent-based-models (Railsback and Grimm, 2012).

5 Summary and Conclusions

Putting it all together, we arrive at fairly conservative conclusions: Kuhn's theory of scientific revolutions and his concept of a paradigm does not have any particular consequences for the validation of simulations. At least it does not have consequences that are any different from those it has for the validation of theories or non-simulation models. And neither do computer simulations require us to reconsider Kuhn's theory or related topics like the Duhem-Quine-thesis. This result is somewhat unspectacular, but it may be clarifying. With regard to the discussion about the novelty of computer simulations it means that, whatever the novelty may be, neither the introduction of computer simulations nor their validation is or requires a Kuhnian revolution.

The co-existence of multiple paradigms in the social sciences is a challenge for Kuhn's theory in its original form. But, again, the validation of simulations does not raise any specific problems in this context. Presently, many social simulations suffer from the fact that for the lack of proper validation they are quite uninformative about their target system. Although, there are also examples where social simulations do contribute to the understanding of the target system, the field as a whole does not yet seem to have become normal science in the sense of Kuhn. This is most notably due to the fact that – as of now – there exists no commonly shared understanding of the validation requirements of social simulations.

References

[1] Petra Ahrweiler and Nigel Gilbert. "The Quality of Social Simulation: An Example from Research Policy Modelling." In: *Policy Practice and Digital Science: Integrating Complex Systems, Social Simulation and Public Administration in Policy Research.* Ed. by Marijn Janssen, Maria A. Wimmer, and Ameneh Deljoo. Cham: Springer International Publishing, 2015, pp. 35–55. ISBN: 978-3-319-12784-2. DOI: 10.1007/978-3-319-12784-2_3.

[2] Eckhart Arnold. "Experiments and Simulations: Do They Fuse?" In: *Computer Simulations and the Changing Face of Scientific Experimentation*. Ed. by Eckhart Arnold and Juan Duran. Newcastle: Cambridge Scholars Publishing, 2013b, pp. 46–75. ISBN: 978-1443847926.

[3] Eckhart Arnold. *Explaining Altruism. A Simulation-Based Approach and its Limits*. Heusenstamm: ontos Verlag, 2008. ISBN: 978-3110327304.

[4] Eckhart Arnold. "What's wrong with Social Simulations?" In: *The Monist* 97.3 (2014), pp. 361–379. ISSN: 0026-9662. DOI: 10.5840/monist201497323.

[5] Robert Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[6] Robert L. Axtell et al. "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley." In: *Proceedings of the National Academy of Sciences* 99.suppl 3 (2002), pp. 7275–7279. ISSN: 0027-8424. DOI: 10.1073/pnas.092080799.

[7] Claus Beisbart. *A Transformation of Normal Science. Computer Simulations from a Philosophical Perspective*. unpublished, 2011.

[8] Claus Beisbart. "What Is a Computer Simulation and What Does This Mean for Simulation Validation?" In: *Verification and Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*. Ed. by Claus Beisbart and Nicole J. Saam. 2019. Chap. 37.

[9] Ken Binmore. *Game Theory and the Social Contract I. Playing Fair*. Fourth printing (2000). Cambridge, Massachusetts / London, England: MIT Press, 1994.

[10] Ken Binmore. *Game Theory and the Social Contract II. Just Playing*. Cambridge, Massachusetts / London, England: MIT Press, 1998.

[11] Alexander Bird. "Thomas Kuhn." In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2013. Metaphysics Research Lab, Stanford University, 2013.

[12] N. Cartwright and Oxford University Press. *How the Laws of Physics Lie*. Clarendon paperbacks. Clarendon Press, 1983. ISBN: 9780198247043.

[13] Annamaria Carusi, Blanca Rodriguez, and Kevin Burrage. "Model Systems in Computational Systems Biology." In: *Computer Simulations and the Changing Face of Scientific Experimentation*. Ed. by Eckhart Arnold and Juan Duran. 2013. Chap. 6.

[14] M.D. Cohen, J.G. March, and J.P. Olsen. "A Grabage Can Model of Organizational Choide." In: *Administrative Science Quarterly* 17 (1972), pp. 1–25.

[15] Bahareh Delgoshaei and Msoud Fatahi. "Garbage Can Decision-Making in a Matrix Structure. A Case Study of Linköping University." Linköping University / Department of Management and Engineering, 2013. DOI: `urn:nbn:se:liu:diva-95612`.

[16] Lee Alan Dugatkin. *Cooperation among Animals*. Oxford University Press, 1997.

[17] John Dupré. "Against Scientific Imperialism." In: *Philosophy of Science Association Proceedings* 2 (1994), pp. 374–381.

[18] Joshua M. Epstein. "Why Model?" In: *Journal of Artificial Societies and Social Simulation* 11.4 (2008), p. 12. ISSN: 1460-7425.

[19] Joshua M. Epstein and Robert L. Axtell. *Growing Artificial Societies. Social Science from the Bottom Up*. MIT Press, 1996.

[20] Harald Fardal and Jan Oddvar Sørnes. "IS Strategic Decision-Making. A Garbage Can View." In: *Issues in Informing Science and Information Technology* 5 (2008).

[21] Paul Feyerabend. *Wider den Methodenzwang*. Suhrkamp Verlag, 1975/1983.

[22] Guido Fioretti and Alessandro Lomi. "An Agent-Based Representation of the Garbage Can Model of Organizational Choice." In: *Journal of Artificial Societies and Social Simulation* 11.1 (2008), p. 1. ISSN: 1460-7425.

[23] Nigel Gilbert and Klaus Troitzsch. *Simulation for the Social Scientist*. New York: Open University Press, 2005.

[24] James Gleick. *Chaos: Making a New Science*. Open Road Media, 2011.

[25] Ralf Grötker. *Reine Meinungsmache*. German. Technology Review (heise Verlag). May 2005. URL: `http://%5C-www.%5C-heise.%5C-de/%5C-tr/%5C-artikel/%5C-Reine-%5C-Meinungsmache-%5C-277359.%5C-html`.

[26] Francesco Guala. "Models, simulations and experiments." In: *Model-Based Reasoning: Science, Technology, Values*. Ed. by Lorenzo Magnani and Nancy Nersessian. Kluwer Acacdemic Publishers, 2002, pp. 59–74.

[27] Peter Hammerstein. "Why Is Reciprocity So Rare in Social Animals? A Protestant Appeal." In: *Genetic and Cultural Evolution*. Ed. by Peter Hammerstein. Cambridge, Massachusetts / London, England: MIT Press in cooperation with Dahlem University Press, 2003. Chap. 5, pp. 83–94.

[28] Ann Harding, Marcia Keegan, and Simon Kelly. "Validating a dynamic population microsimulation model: Recent experience in Australia." In: *International Journal of Microsimulation* 3.2 (2010), pp. 46–64.

[29] Sandra G. Harding, ed. *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis*. kluwer, 1976.

[30] Stephan Hartmann. "The World as a Process: Simulations in the Natural and Social Sciences." In: *Simulation and Modelling in the Social Sciences from the Philosophy of Science Point of View*. Ed. by Rainer Hegselmann et al. Aug. 1996, pp. 77–110.

[31] Brian Heath, Raymond Hill, and Frank Ciarallo. "A Survey of Agent-Based Modeling Practices (January 1998 to July 2008)." In: *Journal of Artificial Societies and Social Simulation (JASSS)* 12.4 (2009), p. 9.

[32] Rainer Hegselmann and Ulrich Krause. "Opinion dynamics and bounded confidence: models, analysis and simulation." In: *Journal of Artificial Societies and Social Simulation* 5.3 (2002), p. 1.

[33] Cyrille Imbert. "Computer simulations and computational models in science." In: *Springer Handbook of Model-Based Science*. Ed. by Lorenzo Magnani and Thomas Bertolotti. 2017, pp. 735–781. DOI: 10.1007/978-3-319-30526-4.

[34] Johannes Kästner and Eckhart Arnold. "When can a Computer Simulation act as Substitute for an Experiment? A Case-Study from Chemisty." In: *Homepage Eckhart Arnold*. Ed. by Eckhart Arnold. preprint, 2013.

[35] Stephan Kornmesser. "Scientific Revolutions without Paradigm-Replacement and the Coexistence of Competing Paradigms: The Case of Generative Grammar and Construction Grammar." In: *Journal for General Philosophy of Science* 45.1 (Apr. 2014), pp. 91–118. ISSN: 1572-8587. DOI: 10.1007/s10838-013-9227-3.

[36] Thomas S. Kuhn. *Die Struktur wissenschaftlicher Revolutionen*. Suhrkamp, 1976.

[37] Thomas S. Kuhn. *The Essential Tension. Selected Studies in Scientific Tradition and Change*. The University of Chicago Press, 1977.

[38] Johannes Lenhard. "Computer simulation: The cooperation between experimenting and Modeling." In: *Philosophy of Science* 74.2 (2007), pp. 176–194. DOI: 10.1086/519029.

[39] Johannes Lenhard. "How Does Holism Challenge the Validation of Computer Simulation?" In: *Verification and Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*. Ed. by Claus Beisbart and Nicole J. Saam. 2019. Chap. 39.

[40] Mary S. Morgan. "Experiments without Material Intervention. Model Experiments, Virtual Experiments, and virtually Experiments." In: *The Philosophy of Scientific Experimentation*. Ed. by Hans Radder. University of Pittsburgh Press, 2003, pp. 216–233.

[41] Mary S. Morgan and Margaret Morrison, eds. *Models as Mediators. Perspectives on Natural and Social Science*. 1999.

[42] Margaret Morrison. “Models, measurement and computer simulation: the changing face of experimentation.” In: *Philosophical Studies* 143 (2009), pp. 33–57. DOI: DOI10.1007/s11098-008-9317-y.

[43] Jonthon W. Moses and Torbjørn L. Knutsen. *Ways of Knowing. Competing Methodologies in Social and Political Research*. 2nd (first edition 2007). London: palgrave macmillen, 2012.

[44] David J. Murray-Smith. “Verification and Validation.” In: *Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*. Ed. by Claus Beisbart and Nicole J. Saam. 2019. Chap. 4.

[45] Robert Northcott and Anna Alexandrova. “Prisoner’s Dilemma doesn’t explain much.” In: *The Prisoner’s Dilemma*. Ed. by MartinEditor Peterson. Classic Philosophical Arguments. Cambridge University Press, 2015, pp. 64–84. DOI: 10.1017/CBO9781107360174.005.

[46] David O’Sullivan et al. “Short Communication. Strategic directions for agent-based modeling: avoiding the YAAWN syndrome.” In: *Journal of Land Use Science* 11.2 (2016), pp. 177–187.

[47] Wendy S. Parker. “Does matter really matter? Computer simulations, experiments, and materiality.” In: *Synthese* 169 (2009), pp. 483–496. DOI: 10.1007/s11229-008-9434-3.

[48] Isabelle Peschard. “Review of Eric Winsberg’s “Science in the Age of Computer Simulation”. University of Chicago Press, 2010.” In: *Notre Dame Philosophical Reviews* (Mar. 31, 2011).

[49] Steve Phelps. “An Empirical Game-Theoretic Analysis of the Dynamics of Cooperation in Small Groups.” In: *Journal of Artificial Societies and Social Simulation* 19.2 (2016), p. 4. ISSN: 1460-7425. DOI: 10.18564/jasss.3060.

[50] John Preston. “Paul Feyerabend.” In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University, 2016.

[51] Steven F. Railsback and Volker Grimm. *Agent-Based and Individual-Based Modeling. A Practical Introduction*. Princeton University Press, 2012.

[52] Luke Rendell et al. "Why Copy Others? Insights from the Social Learning Strategies Tournament." In: *Science* 328 (Apr. 2010), pp. 208–213. DOI: DOI:10.1126/science.1184719.

[53] Alexander Reutlinger, Dominik Hangleiter, and Stephan Hartmann. "Understanding (with) Toy Models." In: *The British Journal for the Philosophy of Science* (2017), axx005. DOI: 10.1093/bjps/axx005.

[54] Thomas C. Schelling. "Dynamic models of segregation†." In: *The Journal of Mathematical Sociology* 1.2 (1971), pp. 143–186. ISSN: 0022-250X. DOI: 10.1080/0022250X.1971.9989794.

[55] Gerhard Schurz. "Koexistenz und Komplementarität rivalisierender Paradigmen: Analyse, Diagnose und kulturwissenschaftliches Fallbeispiel." In: *Die multiparadigmatische Struktur der Wissenschaften*. Ed. by Stephan Kornmesser and Gerhard Schurz. Wiesbaden: Springer Fachmedien Wiesbaden, 2014, pp. 47–62. ISBN: 978-3-658-00672-3. DOI: 10.1007/978-3-658-00672-3_2.

[56] Sergio Sismondo. *An Introduction to Science and Technology Studies*. 2nd ed. Wiley and Sons, 2010.

[57] Flaminio Squazzoni and Niccolò Casnici. "Is Social Simulation a Social Science Outstation? A Bibliometric Analysis of the Impact of JASSS." In: *Journal of Artificial Societies and Social Simulation* 16.1 (2013), p. 10. ISSN: 1460-7425.

[58] Eric Winsberg. "A tale of two methods." In: *Synthese* 169 (2009), pp. 575–592. DOI: 10.1007/s11229-008-9437-0.

[59] Eric Winsberg. "Computer Simulations in Science." In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2015. Metaphysics Research Lab, Stanford University, 2015.

[60] Eric Winsberg. "Simulated Experiments: Methodology for a Virtual World." In: *Philosophy of Science* 70 (Jan. 2003), pp. 105–125.