When can a Computer Simulation act as Substitute for an Experiment? A Case-Study from Chemisty

Johannes Kästner

Institute of Theoretical Chemistry, University of Stuttgart

Eckhart Arnold

Institute for Philosophy, University of Düsseldorf

July 2013

Abstract

In this article we investigate with a case study from chemistry under what conditions a simulation can serve as a surrogate for an experiment. We set out with a brief discussion of the similarities and differences between simulations and experiments. There are three fundamental differences: 1) Ability (of experiments) to gather new empirical data. 2) Ability to operate directly on the target system. 3) Ability to empirically test fundamental hypotheses. Given that there are such fundamental differences it becomes an important question if and under what conditions simulations can still act as surrogate for experiments.

We investigate this question by analysing a simulation of H_2-formation in outer space. We find that in this case the simulation can act as a surrogate for an experiment, because there exists comprehensive theoretical background knowledge about the range of phenomena to which the investigated process belongs and because any particular modelling assumptions as, for example, on the validity of approximations, can be justified. If these requirements are met then direct empirical validation of a "virtual experiment" may even be dispensable. We conjecture that in the absence of comprehensive theoretical background knowledge direct empirical validation of "virtual experiments" remains unavoidable.

Keywords Computer Simulations, Virtual Experiments, Epistemology of Simulations, Quantum Chemistry

1

Contents

1 Introduction

Since computer simulations have caught the attention of philosophers of science, there are ongoing debates concerning their ontological status [Barberousse *et al.* , 2009], their function in science and their epistemic reach [Humphreys, 2004], whether they are a novelty or merely introduce a more powerful but not essentially different form of modeling [Frigg & Reiss, 2009, Humphreys, 2009], how they affect scientific practice [Winsberg, 2010] and how they are related to other tools of science like "Gedankenexperimente" on the one hand or "material" experiments on the other hand.

In this paper we pick up the debate whether simulations are experiments. This debate appears to have reached a stalemate: The arguments pro and contra have been exchanged [Guala, 2002, Morgan, 2003, Morrison, 2009, Parker, 2009, Winsberg, 2009] and obvious misunderstandings have been clar-

ified, but no universally accepted solution has been reached. We believe that this stalemate is in part due to the nature of the question. For, given that simulations and experiments have some features in common but also differ in certain respects, the answer to the question depends on whether the common features or the differences are considered essential, which ultimately depends on the point of view of the philosopher investigating the question.

However, because experiments play a very distinctive role in science, the question is too important to be left just at that. In order to break the stalemate, we therefore suggest shifting the emphasis of this question. Quite frequently simulations are used in situations where experiments are not feasible for practical or ethical reasons. Instead of asking the question whether these simulations are truly experiments, we should ask:

1. Under what conditions can a simulation serve as a surrogate for an experiment?

2. What requirements must be met so that the results of a simulation are at least as reliable as the the results of an analogous experiment would be?

We will investigate this question with a case study from theoretical chemistry, the simulation of H_2-formation in outer space on the basis of quantum-mechanics [Goumans & Kästner, 2010]. Different aspects of the simulation process will be discussed with respect to the epistemic reliability of the simulation. We identify the existence of comprehensive and empirically well-confirmed background theories as a crucial requirement for the reliability of simulations of experimentally inaccessible phenomena.

2   Similarities and Differences between Simulations and Experiments

The debate whether simulations are experiments is motivated by the fact that simulations and experiments have many important features in common.

In the following, we will briefly highlight the most important common features and differences between simulations and experiments. Although we cannot enter into all details of the debate here,[1] it will become apparent that despite many similarities some crucial differences between simulations and experiments remain which set them apart with respect to their possible epistemic role in science.

2.1 Similarities of Simulations and Experiments

One obvious reason why computer simulations are often labeled as "computer experiments" is that the process of designing, setting up, running and evaluating a simulation is by its very appearance quite similar to that of designing, setting up, running and evaluating an experiment. Both simulations and experiment share the same structure: They operate on an object to learn something about a target system. With *object* we mean the entity on which a computer simulation or an experiment operates. With *target system* we mean that entity in nature that we want to learn something about with a simulation or an experiment. The object must in some way or the other be representative of the target system. In the case of an experiment, however, the object can also be an instance of the target system itself, while in a simulation the object is always a representation of the target system.[2]

Both simulations and experiments run in a controlled environment and both allow interventions on the object [Parker, 2009, p. 487]. In simulations the same tools are applied that were formerly thought of as typical for experimental data analysis like visualisation, statistics, data mining [Winsberg,

---

[1]A more detailed refutation of the arguments against the separation between simulations and experiments will be published separately Arnold [2013]. A criticism that is similar in spirit as the one given there, has been formulated by Peschard [2012 forthcomming]. Here we largely confine ourselves to setting out the positive reasons for distinguishing between simulations and experiments.

[2]Mary S. Morgan [2003] introduces a distinction between *being a representative of* and *being a representation of* to describe this difference. The term *representative* in contrast to *representation* is reserved by her for those cases where the object under study is either identical or an instance of the target system. A very similar distinction is suggested by Guala [2002] and picked up, though not endorsed by Winsberg [2009]. Emphasizing the relation to the target rather than the role of the object, Guala speaks of *material similarity* and *formal similarity*.

2010, p. 33]. This includes similar techniques of error management for simulations and experiments. Among these are: Validation of the set-up (or the apparatus) against cases with known results, testing for the responsiveness on interventions, replicating the results under slightly different conditions, testing for the conformance of the results with undisputed theoretical and phenomenological background knowledge. Also, simulations just as experiments allow us to learn something new and potentially surprising about their object and, if the object is truly representative of the target system, also about the target system itself.

Further shared characteristics between simulation and experimental practice are mentioned in the literature. It suffices to summarize them here: 1) the "constant concern for uncertainty and error" [Winsberg, 2010, p. 34], 2) that simulations – just like experiments according to Hacking [Hacking, 1983] – "have a life of their own" and are in part "self-vindicating" [Winsberg, 2010, p. 45], 3) that simulations and experiments share the same challenge of bridging the gap between their object and the target system, or, as one might also say, between the experimental set-up and the real world outside the experiment [Arnold, 2008, p. 174/175].

2.2 Differences between Simulations and Experiments

It would be rather surprising if despite these similarities there were no differences between simulations and experiments. For, if there were no substantial differences why would people still invest large sums into highly sophisticated experimental set-ups like that of a particle collider when they could just as well buy a super computer? Some philosophers believe that it is rather difficult to draw a sharp line between simulations and experiments [Morrison, 2009, Parker, 2009, Winsberg, 2010], but we believe there are at least three fundamental differences between simulations and experiments which are highly relevant for the epistemic status of either category.

First of all, experiments can provide us with new empirical data, computer simulations cannot. While it is true that computer simulations deliver results that may not have been known or expected by us beforehand, com-

puter simulations can by their very nature only produce results that are implied by the premises on which the simulation is built. It is important here to understand the difference between 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 also known to us. For category a, simulations cannot help us; only experiments can help. For category b, simulations and experiments can help us. And for category c, neither is needed, because we know it already. Another way of putting it would be to say that simulations can only deliver us results that fall within the deductive closure of our prior knowledge.[3]

Therefore, if the term "empirical data" is understood as data of empirical origin then computer simulations do not generate new empirical data. Sometimes the term "empirical" is used in a wider sense. Barberousse *et al.* [2009, p. 560], for example, speak of the data that is generated by simulations as data *about* empirical systems. But they, too, do not consider it to be new data of empirical origin.

Another important difference is that some experiments operate directly on the target system, while computer simulations never do. More precisely, the kind of relation that subsists between the experimental object and the target system is typically one of the following: identity, being an instance, being a part. For example, if a car tester wants to know whether a car can drive faster than 100 mph and, in order to find out, accelerates the car to that speed then the object is identical with the target system. If physicists want to know whether white light is composed of different colors, they can let a beam of light fall through a prisma to find out. In this case the object

---

[3]Winsberg apparently holds a different view when, referring to a particular example, he says: "To think it is true is to assume that anything you learn from a computer simulation based on a theory of fluids is somehow already 'contained' in that theory. But to hold this is to exaggerate the representational power of unarticulated theory. It is a mistake to think of simulations as tools for unlocking hidden empirical content." [Winsberg, 2010, p. 54] While it is true that the presuppositions of a simulation are not only formed by theories but by theories plus further modelling assumptions, there is no way around the restriction that a simulation cannot deliver results that go beyond what is implied in the presuppositions. In this sense simulations are indeed just tools for unlocking hidden content.

(a particular beam of light) is an instance of the target system (light). If an archaeologist intends to determine the age of an old building and takes a stone of that building to submit it to certain tests than the object is a part of the target system.

There are also experiments that do not operate directly on the target system. If one experiments with an electrical harmonic oscillator in order to learn something about a mechanical oscillator [Hughes, 1999, p. 138], then this experiment does not operate on the target system itself. But the fact that some experiments operate on the target system or on an instance or a part of the target system, suffices to set the two categories of experiment and simulation apart. Because experiments can operate on the target system, the experimental method has an epistemic reach beyond that of simulations.

As our case study below shows, there are also application cases where simulations have an epistmic reach beyond that of experiments. However, in this case the limitations of the experimental method are more a matter of practical constraints, while in the opposite case there are principle reasons why the epistemic reach of simulations cannot have the same extent as that of experiments.

Finally, experiments can be used for the testing of fundamental hypotheses (*experimentum crucis*), which, again, computer simulations cannot. It is obvious that a simulation cannot be used to test fundamental hypotheses. For, the outcome of the simulation would simply depend on the very hypothesis upon which the simulation is built. It would be impossible to replace an *experimentum crucis* like Young's double-slit experiment that demonstrated the wave nature of light by a simulation, because the results of the simulation would merely reflect which of the competing theories was programmed into the simulation.

Summing it up: Despite many striking similarities there are several features of experiments that clearly set the experimental method apart from the simulation method. This is true, even if in some cases simulations can act as an surrogate for experiments. We will discuss one such case further below.

In the following we look at one concrete example case and examine under what conditions simulations can serve as a surrogate or for experiments. The

question has both an epistemological interest and practical relevance. It has an epistemological interest, because it touches on the relation between theoretical reasoning, experimental testing and empirical observation in science. The question has practical relevance, because it is important to understand when one can trust the results of a simulation that is offered as a surrogate for an experiment or measurement.[4]

3 Case Study: Simulation of H-2-Formation in Outer Space

We examine the question under what preconditions a simulation can serve as a substitute for an experiment with the example of a recently published simulation of H_2-enrichment in outer space by Goumans & Kästner [2010]. This example was chosen, because it fits quite well with the idea of an "experiment-surrogate". Also, it is simple enough to highlight the epistemologically important aspects. At the same time it is a case study from real science and not merely a stylized example for didactic purposes.

3.1 Introductory Remarks on Simulations in Chemistry

One can safely say that by today computer simulations have a long standing tradition in chemistry. The interest in chemical simulations is, among other things like environmental considerations, motivated by the fact that simulations allow to study details of chemical reactions that cannot be obtained from experimental data or that are practically inaccessible by experiment at all. But also when experiments are possible, simulations can be used to double check the experimental results for their plausibility [Alexander *et al.* , 2002, Wang *et al.* , 2008].

---

[4]This question aroused public attention in the aftermath of the eruption of the Iclandic volcano Eyjafjallajökull in 2010, when airlines were ordered to stay on the ground on the basis of simulation of the spreading of the volcanic ashes. This was criticized by representatives of the airlines who claimed that computer simulations were an insufficient basis for this decision [tag, 2010]. Similar questions are raised by environmental protection policies justified by climate simulations.

In our example of the simulation of H_2-formation in outer space, the very slow reaction rate renders the details of the reaction mechanism practically inaccessible to experimental techniques [Goumans & Kästner, 2010, p. 7351]. This simulation can thus be considered as a stand in for an otherwise impossible experiment or as a *surrogate simulation.* A direct validation of the simulation results is not possible in this case. Here, we speak of "direct validation" of a simulation when the same or almost the same process that has been simulated has also been tracked empirically either by a) an experiment of the same process under the same conditions or by b) observations in case the same process occurs under the same conditions in nature and is directly observable in nature.

However, consequences of the results, e.g. H_2-enrichment in outer space, can be compared to observational data [Goumans & Kästner, 2010, p. 7351f.]. Also, indirect forms of experimental validation are possible for some aspects of the simulation and have indeed been applied (see section 3.4.3 on validation below). The simulations cannot rule out, though, that other mechanisms than those predicted by the simulation may explain the observations. The guiding question of our case study will be on what grounds we can consider the simulation a proper experiment-surrogate if no direct validation is possible.[5] In other words, where does this simulation get its credentials from or why is it trustworthy?

Simulations in chemistry are based on physical theories. Different approximations to those theories are used. The choice of the approximation depends on the particular reaction that is simulated and on the level of detail and accuracy that is desired as well as on the inevitable constraints in computing power. Two popular types of approximations can be distinguished. 1) Molecular dynamics simulations are cheap in terms of computing power but cannot describe changes in electronic structure, e.g. bond breaking and bond

---

[5]It should be noted that this question is different from that which is pursued by Barberousse *et al.* [2009] whose criticism of the the physicality-argument to bracket simulations and experiments is otherwise likeminded to our tenets. Barberousse *et al.* [2009] examine the semantic relation between simulations and their target system, which we take for granted here, but do not ask the question of epistemic justification, which is our main concern.

formation. 2) Quantum mechanical simulations treat electrons in much detail and, thus, allow to simulate breaking and formation of chemical bonds, charge transfer, and electronic excitation. They require significantly higher computational cost.

Our example belongs to the second category. Because the number of atoms involved in the simulated reaction (H_2-formation catalyzed by chemisorption of H on benzene) is small enough, a quantum mechanical simulation of the reaction is feasible.

3.2 The Role of Quantum Mechanics as Comprehensive Background Theory

In their daily work, theoretical chemists are not concerned with the physical theories upon which their simulations are based. They rather focus on the design, selection, and justification of models, approximations, and algorithms which allow to apply the background theory to specific chemical reactions at an affordable computational cost and with sufficient accuracy. Although the physical background theories are usually taken for granted without further question, a few words on their epistemic role are in order, because understanding the epistemic role that the background theories play for these simulations is important for their philosophical justification.

It is an important condition for the kind of simulations that are done in theoretical chemistry that they can rely on background theories that are well-approved and uncontested within the range of application. We can speak of theories that fulfill this condition as "comprehensive theories". With *comprehensive theories* we mean theories that correctly describe all causally relevant factors for a well-defined range of phenomena. Or, to put it in simpler words, everything that happens within this range of phenomena happens according to the theory. For such a theory to be well-approved and uncontested, three conditions must hold:

1. The theory has been empirically confirmed in many important instances.

2. The theory has not been disconfirmed in any instances. If any anomalies (i.e., contradictions of the theory with empirical facts and, thus, possible candidates for falsification of the theory) have occurred, then the sub-range of phenomena for which anomalies are to be expected can at least clearly be delineated.

3. Any alternative theory (i.e., a different theory that fully or partly covers the same range of phenomena) has identical consequences as the comprehensive theory within the overlap region of their respective ranges of phenomena and within reasonable bounds of precision.

The theory of *quantum mechanics* meets these requirements for the description of chemical reactions. Anything that can happen in a chemical reaction is – at least in principle – covered by quantum mechanics. Quantum mechanics can be formulated in different ways. For example, the Schrödinger picture is equivalent to the Heisenberg picture and to Feynman's path integral method [Jensen, 1998]. These formulations may be regarded as alternative theories as defined above.

In principle, quantum electrodynamics, i.e. quantum mechanics in a formulation which takes effects of the theory of special relativity into account, should be more accurately used as the "comprehensive" background theory. Then quantum mechanics can already be seen as the first approximation to quantum electrodynamics. An even cruder approximation is to use molecular dynamics as a comprehensive theory for the behaviour of molecules. Its accuracy may be sufficient in cases where no changes in the electronic structure occur. Thus, in these cases it may count as a comprehensive theory even though it is not the most fundamental theory. The range of phenomena it covers is then delineated by the non-occurrence of changes in the electronic structure.

Slightly simplifying, one could also say that a theory is comprehensive for a certain range of phenomena if we can safely expect it to produce results of sufficient accuracy within this range of phenomena. It should be noted that this is not circular, because we need to be sure beforehand (i.e. be able to "safely expect") that it produces results with sufficient accuracy. It should

11

also be noted that the property of being comprehensive as we understand it here is relational to particular classes or ranges of phenomena. Thus, when speaking of a theory as being comprehensive in this sense we do not mean to say that it is the most general theory about a certain subject matter in the sense in which, say, the theory of relativity is the most general theory about space, time and gravity.

Unfortunately, it is only some areas of some sciences where we really have comprehensive theories. But if we do, it has important epistemological consequences for the validation of models and simulations. Generally speaking, the existence of a comprehensive theory increases the trustworthiness of our models or simulations and it eases the burden of validation. For if a simulation is based on a comprehensive theory then the question whether the simulation's results are valid is reduced to the question whether the approximations and modelling techniques are passable. In case these can sufficiently be justified theoretically, an additional empirical validation of the simulation is not necessary any more, because we assume the theory to be correct and to describe the phenomenon in question comprehensively.

Contrast this with the situation when there is no comprehensive theory. In this case, even if we could justify all approximations and simulation techniques theoretically, we would still need direct empirical validation.[6] For, unless our simulation was confirmed by empirical validation we would not know whether the theoretical assumptions about the simulated phenomena hold true in a particular application case or not.

This situation frequently occurs, for example, in agent-based simulations in economics and other social sciences. Agent-based simulations cannot rely on any uncontested theory, because, typically, there exists either no theory at all for the phenomena simulated by agent-based simulations, in which case these simulations must rely on ad-hoc assumptions. Or there exist different and competing theories, in which case it is hard to justify the choice of a particular one of them without direct validation. Or the theories, like utility theory, are too sparse and have too little content to serve as a comprehensive

---

[6]By direct validation we mean validation of the simulation set-up with an experiment that closely mimics the simulation set-up.

theory. The contrast that exists between economic theory and physics in this respect is often overlooked, but it has been pointed out very clearly in Cartwright [2009, p. 48/49]. It is all the more unfortunate, therefore, that proper validation is not yet common practice in the field of agent-based-simulations [Heath *et al.* , 2009].

In science and engineering the favorable case occurs more often in which an uncontested and well-approved background theory does indeed exist. For example, simulations of chemical processes such as the H_2-formation simulation described below can rely on quantum mechanics as a comprehensive theory. As we shall see, this greatly reduces the need for direct empirical validation of their results and makes it possible to employ them as experiment surrogates.

3.3 The Motivation for Simulating the H-2-Formation in Outer Space

The simulation of H_2-formation in outer space described in the following is documented in Goumans & Kästner [2010]. The purpose of this simulation is to contribute to the explanation of H_2-enrichment in the interstellar medium. The simulation can best be described as a piece in the puzzle to explain this phenomenon. The point where the simulation study picks up the problem is defined by a number of previously established facts and existing astrochemical hypotheses:

1. It has been measured in astronomy that H_2 is abundant in the interstellar medium "despite inefficient gas-phase formation routes and H_2-destruction by cosmic rays and photons." [Goumans & Kästner, 2010, p. 7350]

2. To explain this fact, other H_2-formation routes must exist. One possible route is the chemisorption of hydrogen atoms (H) on dust grains made mostly of carbon [Cazaux *et al.* , 2008]. "Astrochemical models require facile chemisorption of H on carbonaceous dust grains at intermediate

temperatures" [Goumans & Kästner, 2010, p. 7350]. Intermediate temperatures are temperatures approximately between 100 K and 250 K. Such dust grains mainly consist of graphite and its smaller fragments, polycyclic aromatic hydrocarbons.

3. Is has been suggested that the H_2-formation rates must exceed 3×10^{-17} or 2×10^{-16} cm^{3}molecule^{-1}s^{-1}. [Habart *et al.* , 2004] (The rate is specified relative to the concentration of dust molecules which catalyse the process.)

4. The chemisorption of the first hydrogen atom to an aromatic hydrocarbon determines the rate. The addition of the second hydrogen atom is known to be much faster [Hornekaer *et al.* , 2006].

5. Hydrogen exists in the form of two stable isotopes, the lighter protium (^{1}H) and the heavier deuterium (^{2}H or D). Observations show that D is much more abundant in atomic hydrogen than in molecular hydrogen (H_2 vs. HD) [Cazaux *et al.* , 2008]. This suggests that atom tunneling is involved in the formation of H_2 because deuterium tunnels less efficiently than protium due to its higher mass. "D$_2$ has not been observed to date [in photon dominated regions]." [Cazaux *et al.* , 2008, p. 496]

The question that Goumans & Kästner [2010] seek to answer is whether chemisorption of H and D atoms to polycyclic aromatic hydrocarbons (as a model for dust grains consisting of carbon), and in particular the tunneling effect, can account for the H_2-enrichment in the interstellar medium. In order to answer this question the reaction rates of the chemisorption of H and D on benzene, the simple-most aromatic hydrocarbon, need to be determined. The reaction rates of the chemisorption of H and D on benzene can be determined experimentally only for temperatures that are much higher than those in in the interestellar medium in outer space. Therefore, the experimental determination of the reaction rate must be surrogated by numerical calculation. In the given low temperature setting the reaction rates depend crucially on the tunnel effect. If the tunneling rates can be brought into agreement with the observations and suggestions listed above, then this supports both

the assumption that H_2-formation in outer space is catalyzed by polycyclic aromatic hydrocarbons and that the tunneling effect plays a crucial role in this reaction.

In principle, the tunneling effect can also be observed experimentally, but practically this is well-nigh impossible in the given scenario, because the reaction rates are too low for experimental purposes due to the low temperatures [Goumans & Kästner, 2010, p. 7351]. The time scales relevant to the interstellar medium (10^5 years) can not even closely be reached in experiments. The more welcome therefore is the possibility to simulate this reaction in the computer. At the same time, because no direct experimental validation of the simulation is available, more strain is put on the justification of the theoretical and technical ingredients of this simulation which will be described in the following.

3.4   Modeling Techniques and their Credentials

Having seen what motivates the use for a computer simulation in outer space, we now describe in more detail, the abstractions that this computer simulation relies on and the modeling techniques it makes use of and how both of these are justified.

3.4.1   Abstractions

First of all, the simulation uses a simplified model to capture the essence of the chemical reaction that is considered to be most relevant for H_2-enrichment in outer space, namely the chemisorption of H and D on benzene. Benzene is the simplest aromatic hydrocarbon and it is expected that the results of the simulation regarding the reaction rate under consideration and the tunneling effect will not fundamentally differ for other polycyclic aromatic hydrocarbons, an assumption meanwhile confirmed by additional simulations [Goumans, 2011].

Furthermore, only the first part of the reaction, namely the chemisorption of H or D on benzene, is simulated but not the addition of the second H or D which would complete the formation of the H_2, HD, or D_2 . The

rationale for this abstraction lies in the fact that the "addition of a second H or D atom is barrier-less *para* to a chemisorbed H ... on the edge of a polyaromatic hydrocarbon" [Goumans & Kästner, 2010, p. 7351], wherefore it is the chemisorption of the first H atom that is rate-limiting.

3.4.2 Modeling Techniques

The H-tunneling simulation makes use of a number of approximations and modeling techniques within the realm of quantum mechanics. To describe the atomic motion, it uses instanton theory (also called harmonic quantum transition state theory, HQTST) which is an approximation to quantum mechanics. Its approximations are mathematically well defined, and it has been applied for decades to calculate tunneling rates [Takatsuka *et al.* , 1999, Vaĭnshteĭn *et al.* , 1982]. "This approach has been shown to be quite accurate in comparison to analytic solutions and results from quantum dynamics, especially low temperatures where other semi-classical tunneling approaches often underestimate transmission coefficients" [Goumans & Kästner, 2010, p. 7350]. Instanton theory is based on (an approximation to) Feynman's path integral method.

The electronic motion in turn was described by density functional theory, a different formulation of quantum mechanics. The theory itself is exact [Hohenberg & Kohn, 1964], but the functional involved is unknown and has to be approximated. Goumans & Kästner [2010] performed reference calculations with the CCSD(T)/CBS *ab-initio* method, another approximation to the exact quantum mechanical result with well established accuracy. Then they compared different forms of functionals, all of which are frequently used and were empirically found to be credible in many cases in chemistry. They used the MPWB1K functional for the simulation since it provided a satisfactorily good match to the CCSD(T)/CBS reference data, and because it was at the same time the relatively best match in comparison to several other tested functionals.

In all quantum chemical simulations, the quantum mechanical wave function has to be expanded in a finite basis set. The error introduced by this

expansion is already accounted for in the comparison of the functionals, because CCSD(T)/CBS does not show this error (CBS stands for complete basis set limit).

The computer programs used by Goumans & Kästner [2010] (NWChem, ChemShell, Gaussian, Molpro) are well established and used by many researchers throughout the world. Significant errors would likely have been found prior to their study and can, thus, be regarded as very unlikely. The program implementing the instanton theory was partially written specifically for the one study presented here. Before the production calculations, it was tested extensively against cases with known results. This is in line with the established best practices in the field and will usually not even be mentioned in the scientific papers.

Additional approximations, like the truncation of a number series or convergence criteria, were tested by extending them and monitoring the change of the result. The used values are reported to allow the exact reproduction of the calculations by other scientists.

Summing it up, in order to realize the simulation, the authors made use of a number of modeling techniques, including several levels of approximations. While one could say that this inevitably introduces some degree "motleyness" and "autonomy" [Winsberg, 2001], these characterizations are not very fitting in this particular example, because the simulation discussed here is diligently built to reflect the theory as closely as possible and does not draw on any independent phenomenological considerations. Also, many techniques are used to keep deviations from the theory in check, like the comparison of different functionals with reference calculations as well as various measures of testing and error checking. Ultimately, when interpreting the results it is taken into account that the simulation results may – due to the employed functional – deviate from the real values in a certain way: "Since the functional we used overestimates the classical barrier, the real rates [of H_2 formation] will be higher than our calculated ones."[Goumans & Kästner, 2010, p. 7352]

3.4.3 Validation

The tunneling rates calculated by Goumans & Kästner cannot, at least with the techniques currently available, directly be tested by experiment. Thus, they serve as real theoretical predictions. Limiting or similar cases, can be validated, though.

The approximations used can be divided into two classes: (1) instanton theory as an approximation to full quantum mechanics. (2) Density functional theory, the basis set, and the computer codes involved to describe the potential energy surface. The latter is needed as input to instanton theory.

Instanton theory, as mentioned above, was shown previously (by other scientists) to be accurate in a broad class of chemical reactions. Thus, the authors assume that it is also accurate enough for the chemisorption of hydrogen on benzene.

The rate of the chemisorption of hydrogen on benzene was measured experimentally at higher temperature than relevant for the interstellar medium (300–600 K). In this temperature interval, tunneling does not play a role, which facilitates the simulation of rates. Goumans & Kästner therefore were able to use the computationally expensive CCSD(T)/CBS method to calculate the rate in this temperature interval. They compared it to the experimental values [Goumans & Kästner, 2010, Figure S2 of the Supporting Information] and found satisfactory agreement. Additionally, they calculate the rate with a number of density functionals, and found again satisfactory agreement with the MPWB1K functional. The rate at these high temperatures (300–600 K) depends only on a small part of the potential energy surface. The part is crucial for the tunneling rate at low temperatures as well, though. Then they calculated the potential energy surface at the whole tunneling path both with the CCSD(T)/CBS reference method and the MPWB1K functional used for the tunneling rates. The agreement between these surfaces adds credibility to the tunneling simulations.

Thus, although direct empirical validation was impossible for this simulation, different means for indirect validation were available and have been made use of.

3.5 Experiment-likeness

The main conclusions drawn by Goumans & Kästner were the suggestion that "H atoms could chemisorb on PAHs in the moderately warm (100–200 K) regions of the interstellar medium, contributing to the catalytic formation of H_2" and that "D will chemisorb an order of magnitude slower [than protium]" [Goumans & Kästner, 2010, p. 7352]. A conclusion given more implicitly was that the reaction would be too slow were it not for the tunneling process which accelerates the reaction by many orders of magnitude (depending on the temperature). Thus, when drawing conclusions the simulation results are treated like experimental results by the authors. The simulation here acts as a stand-in for an otherwise impossible experiment.

Not surprisingly, the simulation of the H-chemisorption on benzene shows many characteristics that make it appear experiment-like: Just like an experiment, the calculations provide numbers which have to be post-processed in order to obtain rates, and interpreted in order to draw conclusions. The simulation also mimics an experiment with respect to its function of hypothesis confirmation. For, the "computer experiment" could either confirm or falsify the hypothesis that tunneling contributes to H_2-formation in space. In the end it confirmed the hypothesis.

The process of designing or setting up the simulation also exposes analogies to experiments. While in a material experiment, the techniques of measurements have to be chosen adequately, in the simulation various approximations have to be chosen adequately. A wrong choice will result in wrong results, generally, however, no results at all, in either case. Errors in these choices can in both cases be identified by reproducing the experiment or the simulation with different parameters.

And like an experiment the simulation can be replicated. In the case of computer simulations replication means: Reimplementing the same simulation under different conditions, like, for example, under a different system environment, with different but functionally equivalent software frameworks and libraries, with different but equally well motivated approximations or with different functionals that have a comparable reliability for the prob-

lem at hand. Just like the replication of an experiment the replication of a simulation serves the purpose of reassuring the researchers that the obtained results were not merely an artifact of the idiosyncrasies of a particular set-up.

Speaking of the relation between simulations and experiments in general, we have noted earlier that there are not only many striking similarities between simulations and experiments but also some important differences (see section 2.2 above). Whether a simulation can be considered as experiment-like on the phenomenological level and, beyond that, as a viable experiment-surrogate on the epistemological level, does also depend on how relevant these differences are in a particular case.

For our example of the H-tunneling simulation we can safely draw the conclusion that the fact that the simulation is not a material experiment and does not operate directly on the target system or generate any new empirical data, does not matter in this case, because the problem of determining the H-tunneling rates can be decided by theory alone, and does not require collecting new empirical data. Of course it must be taken for granted that the theory is true.

An experiment could – at least conceivably – also reveal that quantum theory is false or, say, not valid in outer space. Such an accidental finding (in an experiment that was not intended to test the theory) is impossible with a computer simulation. But given how very well tested quantum mechanics is, this seems extremely unlikely to occur in any such experiment and one would surely first take the possibility of all sorts of experimental error into account, before starting to doubt quantum mechanics.

Thus, in the example of our case-study, materiality is not really an issue and we can therefore consider the simulation not only as experiment-like in virtue of its many similarities to experimental procedures, but also as an experiment-surrogate in the stronger epistemological sense as well. Its epistemic reliability ultimately rests on the confidence in quantum mechanics as a comprehensive background theory. Because the simulated phenomena are completely covered by quantum mechanics we can be sure that no causal factors will be missed out by basing the simulation exclusively on quantum mechanics. Apart from quantum mechanics the epistemic reliability depends

also on the credibility of the justifications for the approximations and modeling techniques employed in the simulation. This, again, does not require materiality. Therefore it would be safe to label this simulation a "computer experiment" without running the danger of exaggerating its epistemic reliability.

4 Summary and Conclusions

As our case-study shows, computer simulations can in many respects be compared to experiments. Yet, as we argued earlier, being comparable to experiments does not mean that computer simulations are experiments. But it seems promising to pursue the question under what conditions a computer simulation can act as substitute for an experiment and what kind of research design a simulation of phenomena that are not directly accessible by experiment must follow.

From our case study we can learn that there is a good chance for employing simulations as an experiment-surrogate if the investigated phenomena fall in the realm of a comprehensive theory, i.e. a tried and tested theory that fully covers the phenomena that are simulated. If this is the case then the question of validating the simulation is reduced to justifying the employed modeling techniques and approximations which does not necessarily require empirical validation. The underlying research design of such simulations could roughly be described as "comprehensive theory plus well-approved modeling techniques". This research design appears to be a valid research design for experiment-surrogate simulations.

Here, we consider a research design as valid when, if executed properly, it has the chance of generating reliable scientific knowledge. In contrast, a research design is invalid or "broken" or "unsound" if, even when properly followed, it cannot generate reliable scientific knowledge about the investigated subject matter. It is important to understand that even if the research design is valid a particular research project can still fail, namely, if one or more of the required steps have not been executed properly. On the other hand, if the research design itself is unsound, any research project following

it will inevitably fail.

We conjecture that the same research design and therefore the same epistemic justification also underlies many other examples of simulation research in the natural sciences. A salient candidate for future investigation would be QM/MM simulations in chemistry. The situation in the case of QM/MM simulations is slightly more complicated than in our case, because QM/MM simulations make use of two background theories, molecular mechanics and quantum mechanics and introduce coupling terms to bridge the molecular mechanics and quantum mechanics part [Senn & Thiel, 2009]. This raises the question if our idea of a "comprehensive theory" can still be applied to describe the research logic of such simulations or if it needs to be adjusted.

Simulations in astronomy that, like simulations of the collision of galaxies, cannot directly be validated by observation might also be an example for a similar kind of research logic and epistemic justification. But they probably introduce a further problem that did not play a prominent role in our example. Our example was a simulation of a reaction of benzene and hydrogen. The structure and the properties of the involved molecules and chemical elements are very well known. But can the same be said of the initial data from which simulations in astronomy start? And, if not, what would be the consequences for their epistemological assessment?

A question that has not been touched in this paper is how simulations that do not or cannot rely on comprehensive theories at all get their credentials. One could say that a comprehensive theory and well-approved modeling techniques jointly form a sufficient condition for a proper experiment-surrogate. But does that also mean that being able to rely on a "comprehensive theory" is a necessary condition? While we have not investigated and therefore cannot exclude the possibility that there are other valid research designs for experiment-surrogate simulations, it appears to us that it would probably be much harder to establish the credibility of an experiment-surrogate simulation of phenomena that are not covered by a comprehensive theory. At least it is difficult to imagine how in this situation an experiment-surrogate simulation could claim credibility without direct empirical validation. For, how were we to know without direct empirical validation that our simulation

did not omit one or more relevant causal factors?

However, this is just a conjecture and we do not want to dogmatically exclude the possibility of valid research designs for experiment-surrogate simulations that do not rely on a comprehensive theory. It would indeed be unfortunate, if there weren't any, because, except for the exact natural sciences and their technological application fields, there are only few areas of science where we can rely on comprehensive theories.

But for the same reason a healthy amount of skepticism is also advisable when the use of numerical methods in the humanities is justified by reference to their success in the exact natural sciences – as it is sometimes done by those schools in the social sciences that try to repeat the success of the natural sciences by imitating their methods [Shapiro, 2005]. Successful research designs can only be transferred from one science to another if the conditions for their applicability have been properly understood. By presenting a case-study from theoretical chemistry we hope to have made a contribution to the better understanding of a particular kind of simulation research design for simulations that act as substitute for experiments.

Acknowledgement

References

2010 (4). *Tagesschau vom 18.4.2010: Nichts geht mehr - mindestens bis 20 Uhr. Sperrung des deutschen Luftraums verlängert.*

Alexander, Millard H., Capecchi, Gabriella, & Werner, Hans-Joachim. 2002. Theoretical Study of the Validity of the Born-Oppenheimer Approximation in the Cl + $H_2 \rightarrow$ HCl + H Reaction. *Science*, **296**, 7150–718.

Arnold, Eckhart. 2008. *Explaining Altruism. A Simulation-Based Approach and its Limits.* Heusenstamm: ontos Verlag.

Arnold, Eckhart. 2013. Simulations and Experiments. Do They Fuse? *Page preprint of:* Duran, Juan, & Arnold, Eckhart (eds), *Computer Simulations and the Changing Face of Scientific Experimentation.* Cambridge Scholars Publishing.

Barberousse, Anouk, Franceschelli, Sara, & Imbert, Cyrile. 2009. Computer simulations as experiments. *Synthese*, **169**, 557–574.

Cartwright, Nancy. 2009. If No Capacities Then No Credible Worlds. But Can Models Reveal Capacities? *Erkenntnis*, **70**, 45–58.

Cazaux, S., Caselli, P., Cobut, V., & Le Bourlot, J. 2008. The role of carbon grains in the deuteration of H_2. *Astron. Astrophys.*, **483**, 495–508.

Frigg, Roman, & Reiss, Julian. 2009. The philosophy of simulation: hot new issues or same old stew? *Synthese*, **169**, 593–613.

Goumans, T. P. M. 2011. *Mon. Not. Roy. Astron. Soc.*, submitted.

Goumans, Theodorus P. M., & Kästner, Johannes. 2010. Hydrogen-Atom Tunneling Could Contribute to H_2 Formation in Space. *Angewandte Chemie, International Edition*, **49**, 7350–7352.

Guala, Francesco. 2002. Models, simulations and experiments. *Pages 59–74 of:* Magnani, Lorenzo, & Nersessian, Nancy (eds), *Model-Based Reasoning: Science, Technology, Values.* Kluwer Acacdemic Publishers.

Habart, E., Boulanger, F., Verstraete, L., Walmsley, C. M., & des Forêts, G. P. 2004. Some empirical estimates of the H_2 formation rate in photon-dominated regions. *Astron. Astrophys.*, **414**, 531–544.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* Cambridge University Press.

Heath, Brian, Hill, Raymond, & Ciarello, Frank. 2009. A Survey of Agent-Based Modeling Practices (January 1998 to July 2008). *Journal of Artifical Societies and Social Simulation (JASSS)*, **12**(4), 9.

Hohenberg, P., & Kohn, W. 1964. Inhomogeneous electron gas. *Phys. Rev.*, **136**, B864.

Hornekaer, L., Rauls, E., Šljivancanin, Z., Xu, W., Otero, R., Stensgaard, I., Laegsgaard, E., Hammer, B., & Besenbacher, F. 2006. Clustering of chemisorbed H(D) atoms on the graphite (0001) surface due to preferential sticking. *Phys. Rev. Lett.*, **97**, 186102.

Hughes, R. I. G. 1999. The Ising model, computer simulation, and universal physics. *Chap. 5, pages 97–145 of:* Morgan, Mary S., & Morrison, Margaret (eds), *Models as Mediators. Perspectives on Natural and Social Science.* Cambridge University Press.

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

Humphreys, Paul. 2009. The Philosophical Novelty of Computer Simulations. *Synthese*, **169**, 615–626.

Jensen, F. 1998. *Introduction to Computational Chemistry.* Wiley, New York.

Morgan, Mary S. 2003. Experiments without Material Intervention. Model Experiments, Virtual Experiments, and virtually Experiments. *Pages 216–233 of:* Radder, Hans (ed), *The Philosophy of Scientific Experimentation.* University of Pittsburgh Press.

Morrison, Margaret. 2009. Models, measurement and computer simulation: the changing face of experimentation. *Philosophical Studies*, **143**, 33–57.

Parker, Wendy S. 2009. Does matter really matter? Computer simulations, experiments, and materiality. *Synthese*, **169**, 483–496.

Peschard, Isabelle. 2012 forthcomming. Is Simulation an Epistemic Substitute for Experimentation? *Pages 1–26 (online preprint) of:* Vaienti, S. (ed), *Simulations and Networks.* Paris: Hermann.

Senn, Hans Martin, & Thiel, Walter. 2009. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie. International Edition*, **48**, 1198–1229.

Shapiro, Ian. 2005. *The Flight from Reality in the Human Sciences.* Princeton and Oxford: Princeton University Press.

Takatsuka, K., Ushiyama, H., & Inoue-Ushiyama, A. 1999. Tunneling paths in multi-dimensional semiclassical dynamics. *Phys. Rep.*, **322**, 347.

Vaĭnshteĭn, A. I., Zakharov, V. I., Novikov, V. A., & Shifman, M. A. 1982. ABC of instantons. *Sov. Phys. Usp.*, **25**, 195.

Wang, Xingan, Dong, Wenrui, Xiao, Chunlei, Che, Li, Ren, Zefeng, Dai, Dongxu, Wang, Xiuyan, Casavecchia, Piergiorgio, Yang, Xueming, Jiang, Bin, Xie, Daiqian, Sun, Zhigang, Lee, Soo-Y., Zhang, Dong H., Werner, Hans-Joachim, & Alexander, Millard H. 2008. The Extent of Non–Born-Oppenheimer Coupling in the Reaction of Cl(^{2}P) with para-H$_2$. *Science*, **322**, 573–576.

Winsberg, Eric. 2001. Simulations, Models and Theories: Complex physical systems and their representations. *Philosophy of Science*, **68 (Proceedings)**(September), 442–454.

Winsberg, Eric. 2009. A tale of two methods. *Synthese*, **169**, 575–592.

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