Explicating the Epistemological Role of Simulation in the Development of Theories of Cognition

Markus F. Peschl (Franz-Markus.Peschl@univie.ac.at)
Department of Philosophy of Science and Social Studies of Science, University of Vienna
Sensengasse 8/10, A-1090 Wien, Austria

Matthias Scheutz (msx@cs.bham.ac.uk)
School of Computer Science, The University of Birmingham, Edgbaston
Birmingham, B15 2TT, UK

Abstract

In this paper, we argue that simulation introduces a completely new quality to the process of theory development. One of the main methodological characteristics of cognitive science (compared to other disciplines studying cognition) is the extensive use of simulation models. In the first part of this paper the foundations as well as implications from the perspective of epistemology as well as of philosophy of science will be developed. It will be shown how the method of simulation becomes an integral part for the process of theory construction in cognitive science. The second part of this paper is concerned with the question of identifying the adequate level of abstraction for computational models of cognition. The strength of cognitive models with high explanatory power lies in providing mechanisms on a *conceptual level*; i.e., on a level of abstraction which *respects* the structure of underlying (physical) mechanisms, but *reduces* the empirical details of these mechanisms in such a way that the resulting model sufficiently approaches the behavioral functionality.

1 Introduction

"Simulation" understood as "the process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behavior of the system or of evaluating various strategies (within the limits imposed by a criterion or set of criteria) for the operation of the system" (Shannon, 1975) underwrites an increasing body of research in cognitive science these days. Computer simulations are employed to study cognitive phenomena and used as supporting evidence for theoretical claims about cognition. Yet, to warrant this methodological strategy of using simulation models as explanatory devices, the role of simulation in cognitive science and its relation to cognitive theories in general need to be analyzed carefully. Such an analysis will help understand the epistemological role of computer simulations in the development of theories of cognition.

In this paper, we show that and how computer simulation is tied into an epistemic loop of scientific discovery. In particular, we explicate and examine closely the relation between cognitive phenomena, an empirical theory capturing some of their aspects, an abstract computational description thereof, and the concrete implementation of the computational description as simulation model. We point to the mutual restriction of the space of cognitive functions by mere causal structure from above (i.e., its abstract computational description) and physical constraints from below (i.e., its realization as physical system), and identify the level of abstraction at which a particular cognitive model has high *explanatory value*.

2 The Process of Theory Development

2.1 The Empirical Approach

What is it that makes science such a powerful tool for understanding and controlling almost every aspect of our world in such an efficient manner? By following a particular set of *methodological rules* a particular kind of knowledge is generated which becomes the basis for the endeavor of science; the goal of any scientific theory or model is to account for phenomena in the "observable world"; the term "observable" refers to those phenomena which can be detected by our sensory systems (possibly mediated by gauges of scientific apparati).

In the case of cognitive science we are confronted with behaviors, neural activities, etc. which are interpreted as being more or less cognitive/intelligent. Some of them can be observed directly, others only indirectly, but often both kinds are *effects* of processes on a deeper level, not directly accessible to our sensory system. It is exactly this "hidden character" of many cognitive processes which makes this domain so interesting as an object of scientific research. Hence, any scientific activity aims at constructing possible mechanisms which could serve as explanations for this hidden domain. Their explanatory value consists in establishing *causal relations* between (observable)

phenomena which are—under normal circumstances—only seen as a more or less coherent succession of states or behaviors over time.

The interesting question is how these causal links are constructed and which methods one should use in order to accomplish this task in an efficient manner. Normally one thinks of the traditional *empirical* approach as the standard means for developing knowledge or a scientific theory about a certain aspect of reality. In the natural sciences the classical epistemological feedback loop between the phenomenon (explanandum) and its theory is applied (see Figure 1 lower part): this cyclic process is based on the "epistemological tension" between a real phenomenon and its theoretical description; i.e., any kind of empirical knowledge (at any level of sophistication) will always be one step "behind" reality, because it is only an approximation of reality. The goal of any scientific endeavor consists in closing this epistemological gap by applying highly sophisticated methods for exploring the regularities of the phenomenon under investigation.

The classical method is to conduct experiments in which a theory or a hypothesis is tested. This theory/hypothesis either has its roots in our common sense understanding of this phenomenon or it is already the result of a change of an earlier theory. In short, the goal is to verify/falsify the given theory (e.g., Oreskes et al. 1994, Popper 1962); this is achieved by deducing a prediction for a particular case. By applying the methods determined by the theory an experiment is conducted to test the theory. This is the point where the (constructed) knowledge or theory is confronted with the "real phenomenon". Besides the fact that the design of the experiment and the process of measurement and observation influence the behavior of the investigated phenomenon, the environment reacts to the input triggered by the experiment according to its intrinsic dynamics. Using gauges environmental states and their changes over time are registered. In the process of observation and interpretation these results are transformed into quantitative values which can be compared with the predicted values of the original theory. A discrepancy between them indicates that changes to the original theory might become necessary. Combining results from other experiments and applying statistical methods leads to the inductive construction or adaptation of an alternative theory which then acts as a starting point for the next cycle in this feedback loop.

Hence, the environmental dynamics directly influences the construction/adaptation of the theory by virtue of the results of the measurements and their possible deviations from theoretical predictions. The goal is to establish a kind of "epistemological closure" between the chosen aspect of reality (be it a cognitive behavior or a neural process) and its description in a theory resulting in a satisfactory reduction of the tension/differences between these two poles is accomplished, i.e., the goal is accomplished, if the resulting theory satisfies the criterion of functional fitness by providing good predictions. Furthermore, such a theory has to provide sufficient explanatory power.

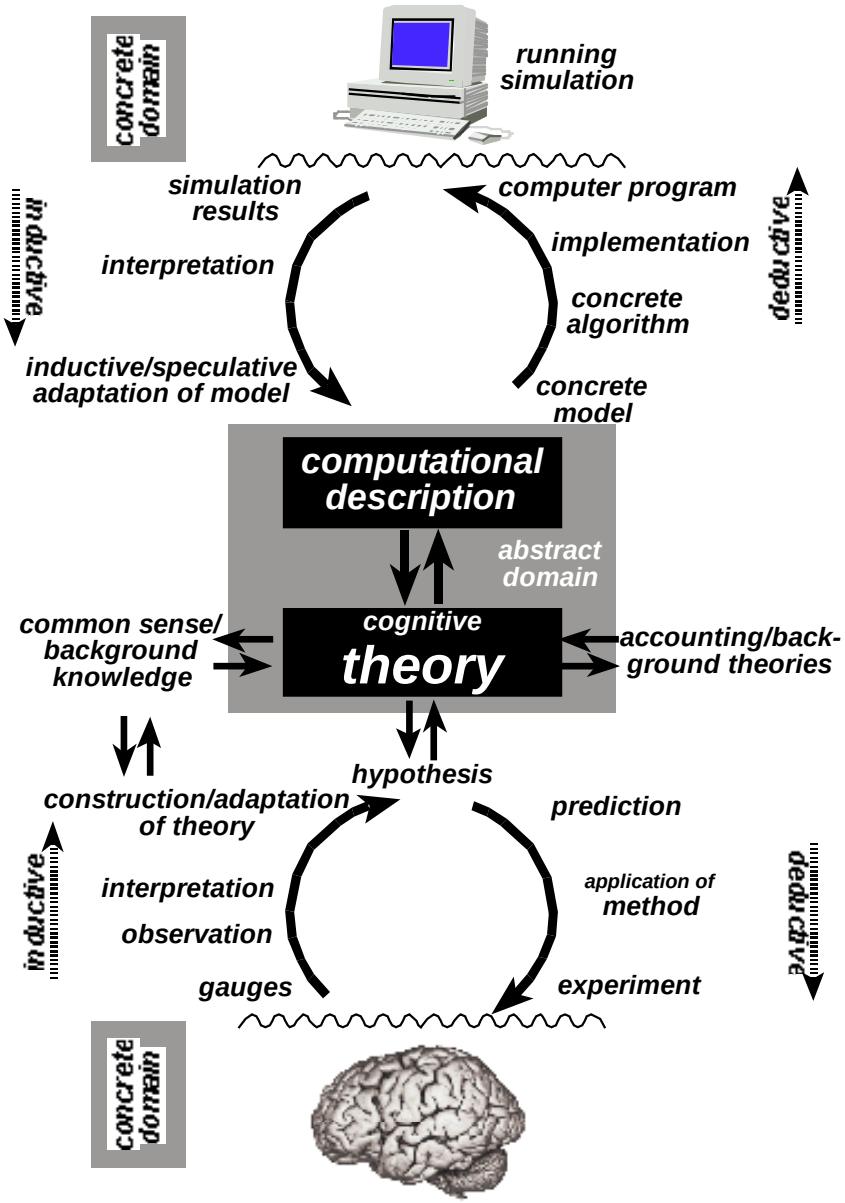

Figure 1: The process of theory construction in cognitive science: the classical feedback loop between a cognitive phenomenon in reality and its description in a theory, the "classical empirical loop" (lower part); the method of simulation is an extension establishing a second feedback loop for "virtual (simulation) experiments", the "virtual loop" (upper part).

2.2 Simulation as an Extension of the Classical Empirical Approach

Contrary to many other disciplines, which study cognitive phenomena by applying this classical empirical approach, one of the methodological characteristics of cognitive science consists in the fact that its theories/models in most cases result from *simulation experiments*. Simulation seems to be an integral part of cognitive science and its theories have a distinct character following a different strategy of construction. As can be seen in the upper part of Figure 1, the method of simulation introduces a *second feedback loop* which has a direct influence on the development of the particular theory. This extension of the classical method is a kind of mirror loop; i.e., the empirical loop gets extended/mirrored in the *domain of virtuality and computation*. The (empirically constructed) theory is transformed into a computational model and the empirical experiment replaced by a *virtual experiment* which is conducted by running a simulation of this model on a computer. The result of this virtual and cyclic

simulation process is twofold: (a) it creates predictions for "real world dynamics"; (b) if these predictions are not satisfactory, a possible change in the computational model may be required which, in turn, may necessitate changes in the original (empirically based) theory. In this case a rewritten version of the theory acts as the starting point for a new cycle of empirical and/or simulation experiments. Thus, simulation of a cognitive model does not only contribute to the development of the computational model, but can also influence the construction of the (empirical) cognitive theory.

3 Levels of Abstraction in Simulating and Modeling Cognitive Processes

After having examined the context of theory development and established the epistemological framework in which simulation models of cognition are embedded, we can now turn to the question of what the necessary steps for running a simulation are in the this cycle of theory-experiment-simulation-virtual experiment. Only a thorough understanding of the role of simulation and the epistemological status of what simulation models can establish, will allow us to justify the methodological role which simulation tacitly has assumed in cognitive science. In order to achieve such an understanding we need to analyze each (epistemologically relevant) step of abstraction involved in this process of simulation and theory construction in cognitive science. Seven steps can be isolated in most cases:

1. Construction of an empirical theory
2. Construction of abstract states
3. Construction of state transitions and automata
4. Construction of a concrete computational model
5. Implementation of the computational model of cognition on a computer
6. Executing the simulation on the computer
7. Verification with reality and feedback to empirical theory

In the following we briefly discuss the transitions between these steps and demonstrate them using an example from the field of neural networks.

3.1 From Physical Cognitive Systems to Empirical Theories

Constructing a theory by applying the "empirical loop" (cf. Figure 1, lower part) implies a first step of *abstraction*: the phenomenon under investigation is *reduced* to a set of "relevant magnitudes" (i.e., dimensions, parameters, variables, etc.). The task of the theory is to relate these parameters to each other in such a way that predictions are possible; furthermore, their interaction structure should provide some explanation in the form of possible mechanisms which are responsible for the generation of the observed cognitive behavior.

Example. Assume that the task is to investigate the neural basis of learning processes. The resulting empirical (neuroscientific) theory identifies changes in the neural architecture as the cause for these observed learning processes. The relevant parameters are a certain neural architecture, the activity and functioning of neurons, and their connections between each other. The theory relates these parameters in such a way that the externally observable behavior is caused by spreading activations and the learning processes are explained as changes in the architecture of the neural network.

3.2 From Empirical Theories to Abstract States

In most empirical theories these relevant magnitudes are properties of physical entities or processes in space and time. In this step towards a simulation model these magnitudes are expressed by variables (and the space of possible values they can assume) and "moved up" one level in the hierarchy of abstraction: they are transformed into a set of *abstract states*, in which the physical conditions and constraints (e.g., time, space, etc.) of the original system become almost irrelevant. I.e., almost no direct referential relationship between a particular abstract state and a particular environmental state can be found any more—in most cases a complex set of rules and operations is necessary for re-establishing this relationship.

Example. The neural network described by its architecture and the flow of activations between neurons is transformed into the mathematical/formal description of a vector space (activation- and weight space) or a weight matrix. Each point in this space refers to a certain pattern of activations or weight configuration, respectively. Note that at this point the physical properties are lost and the whole scenario has become a purely formal description

independent of the original substratum—the weight matrix or the vector space could be equally instantiated by a biological neural system or as any other physical system which fits into the description of these formalisms.

3.3 From Abstract States to State Transitions and Automata—the "Purely Causal and Computational Description"

By connecting these states with each other according to the rules determined by the (empirical) theory, a *dynamical* aspect is introduced into the model. In other words, the cognitive phenomenon under investigation is reduced to some sort of *automaton* whose state transitions represent the underlying mechanism for the generation of the dynamics of the observed cognitive behavior at a highly abstract level. I.e., this automaton provides an abstract mechanism which is capable of generating (a possibly coarse-grained version of) the observed behavioral dynamics. Reducing the cognitive phenomenon to such an automaton implies that the cognitive dynamics is represented as an abstract *computational process*.

At this level of abstraction we are confronted with the *problem* that the states and their transitions can potentially be instantiated by a large number of possibly very different physical systems, because references to the properties of the original physical system have been completely pruned. E are left with the *purely causal structures* of the original system being represented in a large automaton. Thus, in order to reintroduce some meaning to these states it will be necessary to step down in our hierarchy of abstractions.

At this level cognition is described as an *abstract computational process* irrespective of its implementation (in wet-, soft-, or hardware); the *physical restrictions* by which a physical cognitive system is normally constrained do *not* play any role any more on that level of abstraction. Rather, it is the abstract structure and the purely causal and functional relationships of the particular cognitive behavior which are the focus of interest here.

Example. Relating the activation- and weight space introduces a causal structure for generating behavioral dynamics in the activation space: each activation pattern represented as a point in the *n*-dimensional activation space receives implicitly directed edges—one directed edge for each possible input—which connect this point with its possible successor activation patterns. This operation introduces a dynamical aspect by transforming a static activation space into a "dynamic phase space": the total behavioral dynamics of the neural system is described abstractly by this highly structured activation space. One could even go one step further and arrange the points in activation space arbitrarily as states of an automaton. Note that in such an abstraction any referential relationship to the physical properties of the original (neural) system is lost: even its metrical relations within the vector space.

3.4 From Abstract Automata to Concrete Computational Models of Cognition

However, on this level of description/abstraction we have to ask ourselves the following questions: what is the use and explanatory value of such a description? Does it satisfy only some fundamentalist or rather theoretical claims about the relation between computation and cognition? Or does it help us to achieve a deeper understanding of a particular cognitive process? From a practical perspective such models/descriptions are rather rare in cognitive science. Why is that so?

Only, if this highly abstract automaton is transformed back into a *particular model* (e.g., a neural network architecture using a particular learning mechanism), these abstract computational processes are broken down into computational processes which are related/referring to concrete parameters of the original theory. Hence, by re-introducing a "meaning" to the entities of the computational/simulation model the explanatory value is increased on this level of decreased abstraction.

Example. The abstract description of the neural system as an automaton is broken down again in a particular neural network having a certain architecture—the spatial dimension is reintroduced. Work by Nolfi et al. (1991), Cangelosi et al. (1994), or Vaario et al. (1994, 1997) about growing neural networks shows, for example, that temporal as well as spatial issues play a crucial role in the understanding of the genesis of the architecture during the learning process (e.g., the speed and direction of an growing axon).

3.5 From Computational Models to Algorithms and Concrete Computer Programs

In the next step the model gets implemented as *particular algorithm* coded in a particular programming language. Interestingly, not all variables in this algorithm necessarily have to refer to particular physical states or entities of the (empirical) theory, as a large number of variables are necessary for controlling the flow of information for the execution of the program.

Example. The mathematical model describing the neural as well as its growth dynamics is transformed into algorithms, procedures, objects, etc. which are implemented in a particular programming language on a computer (e.g., in Mathematica).

3.6 From Programs to Simulation Models

Finally, this program is *executed* on a (concrete) computer and suddenly the original (empirical) theory of cognition seems to become "alive". The fascination of such a simulation model is based on the *illusion* that the observer ascribes cognitive abilities to a process which is nothing but a very cleverly orchestrated change of values in variables over time (combined with a suggestive graphical output or naming of variables). The *dynamic* aspect and properties of the original theory becomes explicit in this process of running the simulation.

Example. The learning dynamics being realized by the plasticity of neurons and the growth of axons (e.g., Cangelosi et al. 1994) is graphically displayed and each step of the development can be followed by the observer.

3.7 From Virtual Experiments Back to Reality

The execution of the program (i.e., the virtual experiment) yields results which have to be compared to predictions of the theory or already existing empirical data. If there are discrepancies, adaptations in the computational model might become necessary. As a consequence, the original empirical theory might turn out to be flawed which implies the necessity for changes in these theoretical concepts. Furthermore, the results from simulation models can have an "inspiring" implication for the development of completely new or alternative conceptual perspectives and/or experimental designs (e.g., Churchland et al. 1992, Gazzaniga 2000 give examples in which simulation models suggested alternative concepts of how a cognitive phenomenon can be understood and investigated). In any case a change in the empirical theory then initiates a new cycle of empirical and/or virtual experiments. These steps are repeated until the theory is such that it functionally fits into the constraints given by the dynamics of the observed cognitive.

4 Discussion

Whenever computational models for achieving a better understanding of a particular (cognitive) phenomenon are applied, one is confronted with the dilemma of identifying an adequate level of abstraction. Any description of a cognitive phenomenon as a computational process has to satisfy two criteria: (a) that of making good predictions (by using computational mechanisms of any degree of abstraction) and that of providing high explanatory power. For computational models one has to choose the accurate level of abstraction which (i) satisfies the requirement of the cognitive model to be computational and (ii) at the same time takes into consideration the constraints of the physical cognitive system (in time and space) in order to maximize the explanatory value. As has been shown the explanatory power of a highly abstract description as an automaton is rather limited (although highly interesting and significant from a purely computational and causal perspective) compared to the models we are normally using in cognitive science.

What makes less abstract cognitive models, such as neural networks, semantic networks, etc., as explanatory vehicles cognitively more accessible is the "grounding" of the simulation entities and processes in the physical system. This reference to particular entities, states, and processes in the physical cognitive system allows the theorist/observer to comprehend not only the abstract processes, computational properties, and functionalities involved, but also the particular *structure* which leads to this behavior.

So, what is it that characterizes successful cognitive models? The *explanatory strength* of convincing simulation models in cognitive science (in contrast to highly abstract computational descriptions) has its roots in the following properties:

- *Structural level*: they *preserve the relevant structural properties* of the real/physical cognitive system: only do they aim at generating the functionality of the externally observed behavior, but they *explicitly* take into account the structure of the physical system's internal mechanisms which are responsible for this behavior. Hence, there is not only a mapping between the real and the simulated functionality, but also—to a certain degree—between the mechanisms generating these functionalities. It is this stricter mapping between the physical and the virtual domain which increases the explanatory power of the model.
- *conceptual level*: successful cognitive models do not remain at the level of "micro details" of empirical theories, such as the description of molecular dynamics in neurons—this would distract from the overall goal of wanting to achieve a better understanding of cognitive processes. Even if all the structural properties of the responsible

mechanism have been preserved, such a level of description will be rather confusing, because the contribution of molecular processes to understanding, say, a reasoning process, is quite limited. Rather, the strength of cognitive models with explanatory power lies in providing mechanisms on a *conceptual level*; i.e., on a level of abstraction which *respects* the structure of underlying (physical) mechanisms, but *reduces* the empirical details of these mechanisms in such a way that the resulting model sufficiently approaches the behavioral functionality.

The concept of *computation* allows us to overcome this dilemma between getting stuck in physical micro details and preserving the structural properties of the underlying mechanisms, on the one hand, and reducing the observed cognitive dynamics to its pure functionality and causal relationships, on the other hand. The explanatory success of a cognitive model depends on identifying these concepts which represent an adequate balance between respecting the physical system's structure/constraints and constructing computational mechanisms which generate the observed cognitive dynamics in such a way that they are maximal cognitively accessible for the theorist.

References

Anastasio, T.J. and D.A. Robinson (1989). Distributed parallel processing in the vestibulo-ocular system. *Neural Computation 1*, 230—241.

Cangelosi, A., D. Parisi, and S. Nolfi (1994). Cell division and migration in a genotype for neural networks. *Network: computation in neural systems 5*(4), 497—516.

Churchland, P.S. and T.J. Sejnowski (1992). *The computational brain*. Cambridge, MA: MIT Press.

M.S. Gazzaniga (ed.) (2000). *The new cognitive neurosciences*. Cambridge, MA, MIT Press.

Nolfi, S. and D. Parisi (1991). Growing neural networks. Technical Report PCIA-91-18, Inst. of Psychology, Rome.

N. Oreskes, K. Shrader-Frechette, K. Belitz (1994). Verification, validation, and confirmation of numerical models in the earth sciences, Science, 263: 641–646.

K.R. Popper (1962). Conjectures and refutations; the growth of scientific knowledge, New York, Basic Books.

Shannon C. E. and Weaver W. (1975), The Mathematical Theory of Communication. Urbana: University of Illinois Press).

Vaario, J. (1994). From evolutionary computation to computational evolution. *Informatica 18*(4), 417—434.

Vaario, J., A. Onitsuka, and K. Shimohara (1997). Formation of Neural Structures. In P. Husbands and I. Harvey (Eds.), *The Proceedings of the Fourth European Conference on Artificial Life*, pp. 214—223. Cambridge, MA: MIT Press.