Minds and Machines (2019) 29:127–148
https://doi.org/10.1007/s11023-019-09489-4

A Minimalist Epistemology for Agent-Based Simulations in the Artificial Sciences

Giuseppe Primiero[1]

Received: 27 April 2018 / Accepted: 4 January 2019 / Published online: 11 January 2019

Abstract
The epistemology of computer simulations has become a mainstream topic in the philosophy of technology. Within this large area, significant differences hold between the various types of models and simulation technologies. Agent-based and multi-agent systems simulations introduce a specific constraint on the types of agents and systems modelled. We argue that such difference is crucial and that simulation for the artificial sciences requires the formulation of its own specific epistemological principles. We present a minimally committed epistemology which relies on the methodological principles of the Philosophy of Information and requires weak assumptions on the usability of the simulation and the controllability of the model. We use these principles to provide a new definition of simulation for the context of interest.

Keywords Agent-based simulation · Artificial sciences · Multi-agent systems · Constructionism · Controllability · Usability

1 Introduction

Computer simulations play an essential role in many scientific enterprises, from engineering to geography, from biology to social sciences, supporting research and even determining breakthroughs. In the last two decades, this increasingly impactful role has been considered by philosophers of science interested in establishing epistemological principles of computer simulations and in drawing comparisons with the classic scientific method. Here and in the following the term *epistemological foundation* refers to a set of epistemological principles and methodological requirements formulated to qualify, clarify and guide the practice within a scientific field. One of the main issues in establishing an epistemological foundation for computer simulations is their variety in technology and applicability: in this respect, the analyses

✉ Giuseppe Primiero
giuseppe.primiero@unimi.it

[1] Department of Philosophy, University of Milan, Via Festa del Perdono 7, 20122 Milano, Italy

developed for equation-based simulation may differ from those required to analyse Monte-Carlo or agent-based simulations. This debate[1] only recently focused on the use of agent-based simulations and their specific technical and methodological problems. Moreover, even within the specific approach of agent-based simulations, the application considered may require different characterisations.

Agent-based modelling and the accompanying computer simulations are used in a large variety of fields, and for several distinct aims. What makes this technique so vastly applicable is its simplicity of use and adaptability. Economics, population analysis, natural and environmental sciences, and biology have all been supported by agent-based modelling and simulation via computers. Agent-based models are usually presented in the literature as mimicking the behaviour of *natural agents* in naturally occurring environments. On the other hand, multi-agent systems (MAS) refer to agents defined in simulation to create and act in novel artificial environments. Here one typically refers to *artificial agents*, with applications especially in robotics. Nonetheless, it is common to provide a general characterisation of agents to accommodate both applications:[2]

- *Autonomy*: agents are autonomous information processing and exchanging units, free to interact with other agents;
- *Heterogeneity*: agents may have different properties and be grouped according to similar characteristics;
- *Active*: agents are goal-directed, reactive, endowed with (bounded) rationality, interactive, mobile, adaptive, with a form of memory or learning;
- *Interdependence*: agents influence others in response to the influence that they receive, or indirectly through modification of the environment.

While the definition of agent is shared, the distinction between natural and artificial agents is crucial. Agent-based modelling intends to create a plausible model of an existing system, often with explanatory purposes; this environment has some given properties and the agents coordinate and relate with those properties. This is the case for the natural and social sciences, using simulations to discover and predict new information about systems of which we have only partial knowledge. Multi-agent systems, on the other hand, aim often at the creation of an entirely new model, definition and implementation of protocols: these models have mostly exploratory purposes. This is the case for the sciences of the artificial, like robotics and network theory. Despite this difference might appear obvious at first sight, the implications

[1] The literature on this topic is usually referred to as starting with Frigg and Reiss (2009).

[2] For definitions of agents including these properties see for example Crooks and Heppenstall (2012, p. 87) and Macy and Willer (2002, p. 146). Note that it is possible to argue that the following characterization of agents presents at least some overlapping with the definition of agents in other contexts, e.g. in game-theoretical analyses, although the latter would typically have stronger constraints like perfect rationality. The arguments presented in the following of this contribution rely strongly on the methodological principles underlying the simulation processes in which these agents are involved, their behaviour analysed and conclusions drawn, rather than on whether the assumed properties of agents are exclusive of Agent-based modelling and MAS.

for the corresponding epistemological foundations are extensive, and they have been neglected so far in the philosophical literature. The aim of the present contribution is to provide such a reflection on the epistemological foundation of MAS, or simulations for the sciences of the artificial like network theory and robotics. Our claim is that for this specific area of study, we can define a robust minimalist epistemology.[3]

The remaining of this article is structured as follows. In Sect. 2 we overview some of the literature on the Epistemology of Computer Simulation in general and of agent-based simulation in particular. In Sect. 3 we briefly overview some example uses of computer simulations, to extract some observations to guide the formulation of relevant epistemological principles. In Sect. 4 we approach our task from the point of view of the relation between artificial models and their implementation. In Sect. 5 we accomplish this task, by showing that such an epistemological foundation is already available in the larger setting of the Philosophy of Information. We use the resulting analysis to formulate a definition of simulation in the context at hand.

2 Some Positions in the Literature

The epistemology of simulation methods has received large attention in the last few decades,[4] and so has their relation with computer experiments and their epistemological nature in relation with laboratory practices.[5]

A first issue at stake is the definition of simulation. A broad sense of this notion is referred to by Frigg and Reiss:

> In the broad sense, 'simulation' refers to the entire process of constructing, using, and justifying a model that involves analytically intractable mathematics […] Following Humphreys, we call such a model a 'computational model'.[6]

Within this setting, the authors argue notoriously *against* the following claims:

- *Metaphysical claim*: Simulations create some kind of parallel world in which experiments can be conducted under more favorable conditions than in the real world.
- *Epistemological claim*: Simulations demand a new epistemology.
- *Semantic claim*: Simulations demand a new analysis of how models/theories relate to concrete phenomena.
- *Methodological claim*: Simulating is a *sui generis* activity that lies 'in between' theorizing and experimentation.

[3] The validity of this analysis for the simulation of natural agents should be put under strict scrutiny and shall not be considered here.

[4] See in particular Humphreys (1990, 2004); Hartmann (1996).

[5] See respectively Guala (2002), Morrison (2009), Winsberg (2010) and Barberousse et al. (2009), Tal (2011). For a brief overview of several debates concerning the epistemology of computer simulations at large, see Durán (2013).

[6] See Frigg and Reiss (2009). The reference to Humphreys is to his (2004, pp. 102–104).

According to this view, simulation does not offer more favourable conceptual results than experiments; it does not need to be explained and guided methodologically in any different way than standard scientific enterprises; it does not present a different relation between model and theory, and its methodological nature is not any more complex than what scientific practice knows from the standard theory-experiment relation. This position can be considered at one end of a conceptual spectrum: it maintains that computer simulations do not offer, from the epistemological viewpoint, any novelty when compared with standard experimental practices in the sciences. In doing so, we are looking at the relation between the computational model underlying the simulation and the corresponding mathematical or theoretical template abstracted from reality.[7]

An intermediate position on the relation between standard scientific knowledge and practice based on computer simulations can be characterised as follows:

> [Computer simulation] carries with it problems, techniques and methods which are clearly new, such as debugging methods. [...] The difficulties with sorting out the epistemology of experimental science are not yet adequately resolved; but there is no reason to believe that that epistemology won't have rich enough resources to accommodate what scientists are today doing with their computers.[8]

This position relies on a homomorphic relation between simulation and the simulated process: the latter is at the outset of the scientific research, the former is conceptually posterior to it. Under this assumption, computer simulation requires some specific epistemological characterization, representing a variant of standard experimental sciences. This status is justified by some characteristics:

- *Visualization*: according to this view, the process of setting up an experiment in a standard scientific setting and *observing* the behaviour of the system under given initial conditions has analogies with the practice of observing a simulated system through the use of visualization techniques dealing with massive amount of data (number of agents, environment conditions and so forth); obviously, under this reading, analytical tools miss this aspect because no observational process is involved in the resolution of equations providing predictions for a given model.
- *Approximation*: distortions are true of any scale model and, more generally, of any physical system not strictly identical to the target system; in this sense, a simulation approximates the reality of the simulated system in a manner comparable to the approximation of experiments in the standard scientific practice.
- *Discretization*: in a real-world experiment, both the experimental and target processes may well both be continuous processes, but the experimenter will use them (in either manipulation or observation) only with some finite degree of error.

[7] See Humphreys (1990, pp. 499–500).

[8] See Korb and Mascaro (2009, Section 6).

- *Calibration*: calibration in simulation serves the same purpose as in a physical experiment, of finding the settings to support previously observed measurements of a target system under given initial conditions.

These aspects shared by simulations and experiments are related to the process of *verification*, i.e. the act of determining whether the simulation correctly implements the theory being investigated, requiring processes like design verification, debugging, and consistency checks.[9] In conclusion, experimentation with computer simulations is 'full-blooded experimentation', but it also shows new problems related to the techniques in use and it is moreover limited by computability theory, rather than by physical limits.

On the opposite side of the conceptual spectrum we find the position maintaining that computer simulations are a true novelty with respect to standard experimental science, with which it has only a metaphorical or analogical relation:

> computer simulations often use elements of theories in constructing the underlying computational models and they can be used in ways that are analogous to experiments.[10]

According to this position, simulation offers major epistemological novelties compared to other experimental approaches:

- *Epistemic Opacity*: a process is epistemically opaque relative to a cognitive agent X at time t just in case X does not know at t all of the epistemically relevant elements of the process; this is to say that within simulation a specific set of variables is chosen and the methodological validity of the process is confined to such limited set of elements, other aspects and their influence remaining inaccessible to the investigation;
- *Semantics*: the way in which simulations are applied to systems is different from the way in which traditional models are applied: while the latter ones are required to denote the model of reality under analysis, in simulation the relation is less rigid and proceeds more by approximation;
- *Temporal Dynamics*: while in a traditional scientific setting one requires a temporal representation of the dynamical development of the system under observation, in the case of simulations there is additionally a temporal process involved in actually computing the consequences of the underlying model, thus inducing a different, over-imposed temporal dynamics;
- *Practice*: finally, the computational setting in which simulations occur, illustrates a separation between what can be computed in theory and what can be computed with the available resources; this aspect must be considered also in the opposite direction, with computational means allowing more than what the system of reference can.

[9] See Korb and Mascaro (2009, Section 4.2).

[10] See Humphreys (2009, p. 625).

With respect to the problem of epistemic opacity, the level of knowledge that the cognitive agent *X* can exhibit with respect to the epistemically relevant elements of the process can vary, depending on the level of access and competence that *X* has with the implementation, and in particular depending on whether *X* is the designer, the programmer or only the user. This analysis can be better formulated by qualifying which level of access is granted to which agent. From the semantic point of view, the gap between system of reference, model and implementation suggests that a layer of complexity is added by simulations being different technical artefacts than their models: in this case, our analysis should consider whether the implementation is posterior to the model and whether the relation is one of isomorphism, analogy or just similarity. For the temporal characterization, a simulation compared with an underlying (theoretical) model, for which it acts as inferential engine, will import a different notion of time and the relation between the temporal representation of the process of interest at the two levels needs to be addressed. The last aspect, concerning the distinction between applicability in practice and in principle must be considered in view of the design and the implementation.

2.1 Positions on Agent-Based Simulation

A similar tripartite positioning of views can be identified in the literature on the epistemology of agent-based computer simulations, with particular attention to their explanatory power.

A first position[11] maintains that artificial models based on agents cannot provide full explanations of the phenomena they investigate, because their models cannot be validated. In particular, it is not possible to exert explanatory potential from the agents' behavioural rules applied in a precisely specified environment. These rules could not exclude other sets of rules generating the same explanandum, as well as from agents defined by different properties. Hence, simulations have none of the essential qualifications of the potential sources for evidential support: direct observation, well-confirmed theory, or results from externally valid behavioural experiments. What simulations cannot provide, therefore, is ground to believe that the differences between the experiment and the target system do not create an error in the transfer of results from one to the other. At most, computer simulations can provide candidates or contributions to explanations, and in general are not useful because too permissive. Here the problem of epistemic opacity returns in the form of a critique of permissible generalizations through simulation.

An intermediate position maintains that agent-based simulations are explanatory but do not provide predictions:

> In the social sciences, [in] generative explanation […] macroscopic explananda […] emerge in populations of heterogeneous software individuals (agents)

[11] Exemplified in Grüne-Yanoff (2009).

> interacting locally under plausible behavioral rules […]. I consider this model to be explanatory, but I would not insist that it is predictive.[12]

What this specific type of simulations can offer is to guide data collection, to create abstractions capturing qualitative behaviours, to suggest analogies for the identification of correct models and finally to help rising new questions directing research. It is notable here the stress on the role that simulations have in *shaping* the model of analysis, not only in explaining it.

A final approach maintains a positive stand towards the novelty represented by agent-based modelling and simulation in their relation with experimental sciences and the ability to provide explanations for them.[13] This approach strongly criticizes the first one for making any abductive inference from simulation, identifying limits that are general of the social sciences (like data inputting, partiality and unreliability) as proper of agent-based modelling. While this approach rejects the capacity of agent-based modelling to provide causal explanation, it identifies in mechanistic explanation the result of simulation:

> Seeing the social sciences as concerned with mechanisms means to not allow ‘black-box explanations’ such as statistical correlations. Although statistical correlations can be used as evidence for causal associations, they are not an explanation in themselves as they do not lay open the ‘cogs and wheels’ operating to produce the phenomenon in question. […] A mechanism approach neither reduces social entities to physical entities nor sees social mechanisms as the same as physical mechanisms.[14]

By the use of mechanisms, i.e. through the identification of a set of patterns in particular contexts with associated entities and activities at work, agent-based modelling offers the ability to generate predictions which can be tested in experimental settings. Also, when predictions turn out to be false, mechanisms are revisable for local faults. In summary, according to this last view:

- Data problems are shared by all approaches to social sciences;
- Retro-fitting is required by the dynamics of running the model, but it does not mean falsification;
- Level-distinctions are required by agents’ behaviours, but they are only a part of the explanation;
- Mechanistic explanation accounts for functional explanation as well and helps us generate predictions and identify faults.

Here again, we wish to stress the acknowledgement of a dynamic relation between simulation and model (retro-fitting) which seems to question the priority order between the two, or at the very least to illustrate the different influence that

[12] See Epstein (2008, sec.1.10).

[13] See e.g. Elsenbroich (2012).

[14] See Elsenbroich (2012, sec.2.7).

agent-based simulations (and MAS in particular, as we will argue) seem to exert on the design of models.

This brief review shows that the variously held positions on the epistemological novelty of computer simulations in general rely on one main assumption: there exists a system of reference which requires an explanation, from this system a model is extracted by abstraction and as such its definition is prior to the implementation in a simulation for explanatory purposes. This assumption seems to be much weaker in the debate on the epistemological and methodological basis of agent-based simulations. One aim of the following sections is to argue that such an assumption is especially misleading in the case of the artificial sciences, like network theory and robotics. We argue that in these contexts, only in a partial sense one can talk of a model defined prior to the implementation. While all modelling exercises are influenced by the results of experiments, we argue that the explorative nature of simulations strongly contributes to shaping the model itself when such a model is an artificial one which is aimed at for optimal results design. On this basis, our quest will be to determine a minimal epistemological committment for agent-based simulations in the artificial sciences. This will also allow us to return to the arguments and critiques exposed in this section.

3 Some Examples

In this section we want to provide some examples to compare the working relation between model and simulation as it happens in both natural and artificial sciences. Our claim is that in the case of the natural sciences, this relation is more static, with a given computational model from which the analysis starts and which is then explored through the design and use of (computer) simulations. On the other hand, in artificial sciences like robotics and network theory and analysis, the dynamics between the studied phenomenon, the constructed theoretical model of such phenomenon and the simulation of the model is very different. In particular, we argue that in such cases:

1. an artificially designed and constructed model establishes the reality of reference; in the strongest case, this can be a formal model (e.g. a logic);
2. the implemented simulation feeds back into the model design: in this sense, the model of reference is not a given structure to simulate, but it is dynamically redefined by the results provided by the simulation;
3. all resulting properties (of both model and implementation) are limited in applicability to a limited class of systems they help shaping: while in the natural sciences certain behaviours can be sometimes applied to a large class of systems that share certain structural properties (e.g. defined by the same physical or biological properties), in the artificial sciences there are often stronger initial constraints established by the intended application and defining the model of reference and guiding its implementation; the set of systems that share the same initial constraints is often very limited.

We argue that the conceptual priority of the theoretical (mathematical) model over the simulation is only partial. The definition of a logical or mathematical model as the first step in this process is an abstraction on reality: what the modeller does in this case is to select a set of axioms and rules which provide an interpretation of the model assumptions under which the analysis is performed. The application of rules to the axioms (i.e. syntactically the definition of a given set of derivable sentences, or semantically a consequence set) defines the expected validities of that model, i.e. the model prediction. While this process in the definition of a logic can be shared between modellers of natural and artificial phenomena, the former ones are constrained by the model assumptions, as these are what is to be modelled in the first place. Instead, provided some actual constraints that the designer of an artificial model needs to preserve in view of the intended application (e.g. the environment of interest, the available technology used, etc.), her initial model assumptions are often stronger than those required: the simulation then can be used not just to predict behaviours of the given model, but rather to tune the model assumptions in order to obtain the intended behaviours. The ability to manipulate the simulation offers freedom in the design of the actual agents, and this provides guidance for optimal results.

To support these arguments, we briefly describe below a known case in physics and two cases in multi-agent systems. Thereby, we hope to provide supporting evidence to identify essential principles for an epistemological foundation of simulation in the artificial sciences.

3.1 An Example of Computer Simulations in Physics

Consider, as an example, a computational simulation of the orbital motion of a planet around the sun and the corresponding implementation on a computational system.[15] Such simulation will require:

- the choice of appropriate coordinates (e.g. the angle and the distance between the centers of the Sun and the Earth, with appropriate abstractions like the rotating of the Sun around its mass);
- the selection of the relevant definitional equations (e.g. for kinetic energy, potential energy, gravitational constant and Lagrangian equation);
- the determination of significant and useful known equations for derivability purposes (e.g. the Euler–Lagrange equation and appropriate derivatives);
- the inference of relevant equations (e.g. the equations of motion);
- the computation of the values of interest (e.g. of angle and distance at some point).

The initial conditions of the system are set to some significant value (e.g. the average distance between the two bodies), with the first time derivative and speed

[15] For a concrete example, see e.g. https://evgenii.com/blog/earth-orbit-simulation/.

set to zero, an arbitrary initial value for the angle and a fixed angular speed. The simulation of the orbital motion then consists in successively computing position and velocity at discrete time intervals based on the given initial conditions and the chosen relevant equations. The outputs are then represented as the planet's motion, converting the atemporal properties of the mathematical model in the temporal computation of the simulated movement: this, according to Humphreys, represents the relation between the drive and constraints of experiments in the physical sciences by the development of tractable mathematics. Always according to Humphreys, the temporal nature underlying the target system is reflected by computer simulations, despite their limitations, while the representation offered by theories and models is more limited in this respect.[16] In doing so, the simulation resembles instruments in allowing humans amplifying their epistemic reach beyond what is for them naturally feasible.

3.2 An Example of Computer Simulations in Network Science

Consider an agent-based model of ideal information distribution in networks of agents. The model can be formulated first at the abstract level as a logical system of axioms and rules establishing respectively the properties and behaviour of agents, then translated to an algorithmic protocol and implemented in a simulator. The analysis of the network can be concerned with problems like: conditions for consensus-reaching transmissions; epistemic costs induced by confirmation and rejection operations; the influence of ranking of the initially labelled nodes on consensus; complexity results.[17]

To start with, it is essential to note here that an investigation of this kind does not have an absolutely stable phenomenon to account for: there is no observed system whose behaviour is modelled and reproduced through simulation. Instead, several network topologies can be simulated to analyse different models of information transmission to determine which one is the most effective to maximise a certain property (e.g. consensus in the network) and under which conditions a behaviour with certain characteristics occurs (e.g. trust in information coming from nodes with higher ranking).

The second observation concerns the formal model. The design of a logic aims at determining necessities in the model, i.e. formulas expressing properties that should always be displayed by the model (conclusions) whenever the corresponding initial conditions (premises) hold. As such, a logic is also a limiting tool, in that such validities are always bound to the axioms or rules defined. This model is not a reference

[16] See Humphreys (2004, p. 109).

[17] In Primiero et al. (2017) such a model is offerd where relations are characterised by positive and negative trust. Several topologies of networks are explored, where agents are ranked and have different epistemic attitudes, roughly corresponding to lazy agents (accepting information without control) and sceptic agents (accepting information under a computational cost corresponding to a verification process). Positive trust is a property of the communication between agents required when message passing is executed bottom-up in the hierarchy, or as a result of a sceptic agent checking information. Negative trust results from refusing verification, either of contradictory information or because of a lazy attitude.

to be reproduced by the simulation, it is rather an idealised benchmark against which to compare the results of the experimental runs of the simulation. Values matching against the provable formulas of the calculus can be interpreted as a confirmation of the simulation with respect to the formal model, rather than one of the model against some outside reality. When experiments provide values that conflict with the formal results, these can be seen as properties of the current implementation which are not in the scope of the formal model. Note that this latter case does not necessarily mean that those behaviours are always undesirable. In the context of properties optimization (e.g. when one aims at knowing under which conditions it is easier to obtain consensus), the simulation provides exploratory information with respect to the model. It is possible that a series of simulation runs offers an indication that a certain intended behaviour is not covered by the formal properties designed by the logic, and so certain choices are made to modify the latter in order to accommodate the desired results. If this happens, the designer typically moves back to the formal model, to modify rules and axioms in order to provide the intended behaviour: this can also happen in the simulation *first*, i.e. by implementing this in the code, checking what the resulting model offers and *then* translating it at the higher level of abstraction provided by the formal model.

Differently from other explanations of the model-simulation relation, our understanding refers to a strong initial model formulated as a logic with valid or refutable properties. The logic can either be used as a practical benchmark (i.e. this formula in the logic is derivable and any implementation of the logic should preserve it) or as a variable system of hypotheses (i.e. the behaviour in simulation is desirable, hence we should change the logic so as to accommodate it). This seems to be a fairly different understanding than the notion of observable system which needs to be modelled by simulation in order to be explained.

3.3 An Example of Computer Simulations in Robotics

As a second example, consider a classical problem in swarm robotics: an unsupervised community of robotic agents relying only on local rules to explore the world and communicate to other agents, until a member of the swarm can assess that a certain property ϕ holds or not for the world, and on that basis trigger a collective action. Also in this case, the development can start from a logic, followed by its implementation in a simulation to provide an experimental setting in which to test the validities of the logic and verify which further properties can be added to the model.[18]

A first observation on such an example concerns a possible deviation of the implementation from the formal model in the representation of the agents' memory. While in the formal model one looks at formal structures (e.g. logical derivations) to establish at which point in the derivation a certain proposition holds true; in the implementation, the agents' memory can be represented as an hashmap table

[18] For example, in Battistelli and Primiero (2017) this problem is treated in terms of a multi-agent temporal logic of weighted beliefs, with rules for distributed knowledge formation and conditional action.

to record their computed beliefs. In other words, while the formal model requires memory to be extracted from properties of its structure, the simulation makes this a property of agents.

Another interesting observation concerns the temporal dynamics of the model, a point already stressed by Humphreys. A formal logic can express time evolution as a set of indices on propositions or as predicates. In a simulation like the one referred here, timestamps can be used to indicate the starting of the swarm action, the reaching of a given threshold, the reaction to such threshold being obtained and the termination of the action. This implementation is an abstraction on a continuous perception of time in the real world. In other words, in order to structure and render the results intelligible, a fictional and less continuous notion of time for the agents can be created in the simulation, such that it can help identifying characteristics of interest of the constructed model. Notoriously, this is a problem affecting several modelling techniques and many analytical methods that focus on the equilibria of a model are unable to account for the dynamics leading to them. These models also lack a clear correspondence to real time, but this weakness remains hidden. In general, models that might lack a meaning of time often allow at least to derive hypotheses about the relative duration of processes and to compare process durations under different parameter conditions.[19] The present observation, nonetheless, does not represent a critique to the treatment of time by formal models in general, but rather a consideration on the artificiality of formal models in expressing appropriate continuous and layered notions of time, more easily modelled by the simulation. In this sense, the latter offers a better and more reliable analysis of this property.

Finally, and similarly to what happens with temporal properties, also epistemic properties (like beliefs) can be expressed in simulation by numerical values on properties holding for family of agents, which can be modified depending on applications, but can also be investigated separately for different groups of agents. Belief degrees can be arbitrarily set to determine when agents start performing actions, thus indirectly determining the state of their environment. These design choices concern mainly the intended investigation, but are not dictated by some fixed model of reference that the simulation has to faithfully represent: on the contrary, it is the simulation which can help assessing which of the several possible parameters is the most helpful in reaching the intended optimal results.

4 On Artificial Models and Their Implementations

The observations extracted from the two examples above in agent-based simulation, and for MAS in particular, are useful to formulate some remarks on the nature of the relation between models and implementation.

A standard way to define a simulation in its narrowest sense is by referring to the program that is run on a computer and that uses step-by-step methods to explore the approximate behaviour of a mathematical model: usually, this corresponds to

[19] I wish to thank an anonymous reviewer for this specific comment.

a model of a real-world system. On the contrary, in the type of systems we have considered, the target is only a hypothetical system to be engineered. This means the model has properties that are designed rather than discovered: for example, the temporal evolution of the system is determined by possibly ad hoc thresholds and parameters. The corresponding simulation is developed to obtain the optimal (and not in some sense real) configuration of the intended model. Moreover, as we have seen, it can be the simulation to offer insights on the features that the model has to take into account.

This seems to suggest that agent-based simulations for the artificial sciences cannot be included in a famous definition by Humphreys:

> any computer-implemented method for exploring the properties of mathematical models where analytic methods are not available.[20]

Humphrey's narrow definition of simulation appears problematic in the context of the sciences of artificial: it assumes the existence of a static mathematical model, whose properties can be explored by the computer implemented method, and because an analytic one cannot be provided. If one accepts the methodological process illustrated by the two examples above, the simulation does not just explore the model but rather it *contributes* to its design; hence, a full definition of the model assumptions is not conceptually prior to the simulation, but rather results from the analysis of the model predictions (at any given stage of its design) and their feeding back into the model assumptions, until the optimal design is reached. The narrow sense of simulation given by this definition seems in this sense unsatisfactory.

A broader definition of simulation notoriously refers to a comprehensive method for studying systems, which includes:

1. choosing a model;
2. finding a way of implementing that model in a form that can be run on a computer;
3. calculating the output of the algorithm;
4. and studying the resultant data (possibly aided by some visualization technique).

The crucial difference with the cases mentioned above is that there is not necessarily a target system for which inferences are to be drawn through the execution of the simulation. In our case, the presence of a logic underlying the implementation allows to establish which inferences are valid, and hence which instances of the model would be ideal. The task is then to find out which of these validities can be satisfied by an implementation, assuming it is possible (if not likely) that such implementation might not reflect all the properties of the artefact model. As illustrated above, sometimes the execution of the implementation provides insights that feed back in the design model, thus allowing to adjust it.

The first immediate consequence is on the notion of reliability. Consider the following remark by Winsberg:

[20] See Humphreys (1990, p. 500).

> Successful simulation studies do more than compute numbers. They make use of a variety of techniques to draw inferences from these numbers. Simulations make creative use of calculational techniques that can only be motivated extra-mathematically and extra-theoretically. As such, unlike simple computations that can be carried out on a computer, the results of simulations are not automatically reliable. Much effort and expertise goes into deciding which simulation results are reliable and which are not.[21]

In the case of simulations in the artificial sciences, the relation with the model is dynamic, and the reliability of the implementation cannot be asserted only by comparison with the model.

An intermediate position in the literature understands computer simulations to be about the *use* of computers to (approximately) model a system (either real or hypothetical). Then a simulation is any system that is believed, or hoped, to have a dynamical behavior that is similar enough to some other system such that the former can be studied to learn about the latter. According to this view, a simulation

> imitates one process by another process. In this definition the term 'process' refers solely to some object or system whose state changes in time.[22]

A more comprehensive definition, taking into account the above constraints, is due to Humphreys:[23]

Definition 1 (*Simulation*) A system *S* provides a core simulation of an object or process *B* just in case *S* is a concrete computational device that produces, via a temporal process, solutions to a computational model [...] that correctly represents *B*, either dynamically or statically. If in addition the computational model used by *S* correctly represents the structure of the real system *R*, then *S* provides a core simulation of system *R* with respect to *B*.

Our question is whether this definition matches the intuition of the type of simulation theory and practice illustrated by the examples above. One limitation seems to be the unidirectionality of the process, whereby the simulation is understood to provide computable solutions to the model offered by *B* of the system *R*. The aim of the next section is to formulate epistemological principles that can suggest a rephrasing of the above definition for the specific case of artificial systems.

[21] See Winsberg (2003, p. 111).

[22] See Hartmann (1996, p. 83).

[23] See Humphreys (2004, p. 110). With *core simulation* Humphreys refers to the temporal part of the computational process, which differentiates it from the underlying model consisting of atemporal logical or mathematical representations.

5 Epistemological Principles for Simulation in the Artificial Sciences

In the present section we formulate a number of epistemological principles to clarify and support the observations made above. These principles reflect the methodological approach of the Philosophy of Information (Floridi 2011) which subsumes that the analysis of any system is expressed in terms of semantic data. The methodology of the Philosophy of Information can be summarised by the following principles:

Principle 1 (Minimalism) *Models should be controllable, implementable, predictable. Problems are relative to a given problem space.*

Principle 2 (Levels of Abstractions) *Models are relative to a set of interpreted variables; several Levels of Abstractions (LoAs) over the same set of observables are possible (part of a so-called Gradient of Abstraction—GoA); a LoA allows to analyse the system and elaborate a related model.*

Principle 3 (Constructionism) *Because the model is constructed, it can be controlled.*

Let us briefly consider these principles. The duality between model and reality is given by the epistemic status of an external observer with respect to data: the observer is the privileged knower of the model in virtue of being its creator; this act of creation invests the knower with epistemic abilities: the model can be controlled, implemented and is predictable. The design of the model needs to be restricted to a given set of variables of interest in order to be functional. Accordingly, the selected LoA establishes the limits of the observer's ability to modify the reality and to control phenomena in it: the model is the only element that can be directly controlled. Discovery proceeds from the constructed model to the reality, not the other way around. As the exploratory activities of the observer are limited to the model, any epistemic statement concerns only the model and refers to reality only in an indirect way. Other associated principles concern the conceptual economy that a good modelling activity should always guarantee: resources defining the model should be no more than those used to analyse its results; inferences from the model analysis should not be generalised beyond the limits of the model itself.

In the remaining of this section, we shall illustrate how these principles are satisfied in the context of our analysis, organising our arguments in three main areas:

1. the relation between reality, model and simulation;
2. the verification of the simulation and the validation of its model;
3. the explanatory ability of simulations.

5.1 Designing the Model to Understand Reality

One central aspect in the epistemology of computer simulations understood in their broad sense is the relation between the implemented simulation, its model and the modelled reality. It is a generally accepted view that the modeling process often reveals relationships with—and helps our understanding of—reality: in other words, constructed models allow to qualify relationships between some elements of the reality which would otherwise remain hidden. In this sense, the simulation of a complex situation often provides a solution to a problem formulated in the space of that situation, even if it is not an analytical but a numerical solution, created by a computer. To this aim, the reliability of the results provided by a simulation standardly has to go through the design of a good model.

The standard theoretical approach to building a good model consists in starting from observations of the real world, transform them in formal expressions, implement the formalization in a system that allows to analyse the dynamic aspect of the model (eventually in the code of some simulation software) and finally evaluate the results of the simulation and compare to expected outputs:

> the simulations play a key heuristic role in the refinement and development of models. In this process, however, a crucial constraint is that adjustments in the model have to result in numerical solvability of the model.[24]

In this context, simulations for the artificial sciences are characterised often by the absence of initial observations. The principle stating that only the model is known, while the modelled is only hypothesised, assumes here its strongest meaning: there is no reality to be known, the model produces a *possible interpretation* of a world to be built.[25]

In this process, an essential step is the characterization of the level of detail to implement, directly following from the choice of interesting properties and behaviours for the agents as a function of the intended application. Formally, this corresponds to setting the Levels of Abstraction (LoA) one wants to see realised in the implementation. While often the artificial sciences can rely on a wealth of data (think of the amount of data that can be extracted from networks to design optimal protocols to address several issue), it is also not entirely strange to start from insufficient, incomplete data, up to no data at all, in the case of an entirely artificial model. Sometimes, the data available is limited to the environment in which the system is to function, but there is no data available *about* the system, which is still to be formulated and implemented. This means that one almost always faces a sub-optimal understanding of the working conditions of the system one is trying to simulate, i.e. of its intended behaviour in that environment. In these cases, a purely formal model

[24] See Humphreys (1995, p. 507).

[25] For example, in the multirobots system from Battistelli and Primiero (2017), the simulation allows to explore several possible configurations of reality that can be obtained, in order to choose the one that best resolves the intended task. Strictly speaking, the model is used to investigate the hypothesis [also known as Principle of Constructability in Floridi (2011)].

(as in our case) has the aim of providing *optimal* benchmarks, which can be approximated through a simulation implementing corresponding rules.

An optimality criterion reflects a balance between too little and too much detail, between too many and too few LoAs. This dynamics of optimal conditions, expressed by the formal model and approximated by simulation from the sub-optimal initial understanding conditions, reformulates the standard description from theories with fewer details and greater generality (potentially useless) to more detailed potential simulations (of possibly uncommon theories, or impractical or uninteresting).[26] Under such description, a model that accounts for too few LoAs, and accordingly implements too litte details, is understood as a tool to explore the behaviour of the system, imposing as little constraints as possible; the more LoAs and details are added, the more the model is determined and so the role of the simulation becomes explanatory; a model with a full set of details (for a given Gradient of Abstraction) becomes descriptive (and hence predictive) of the behaviour of the system with respect to the set of variables of interest.

A further clarification is required for the theory under which the model is constructed. In the process under consideration, there is no assumption of truth about the theory, only a correctness requirement. Also this dynamics is expressed in terms of levels of abstraction: the designer chooses which variables the formal model has to include and the simulation must be able to implement them. In its exploratory work, the simulation may provide additional variables, which are in turn to be added to the model. In this sense, there is no stable homomorphism between model and simulation, and it is possible that properties expressed by the simulation are not available in the model but still desired by the designer, and hence added to it a posteriori.

5.2 Verification and Validation as Controllability

In the previous section, we have argued that the standard view on the modelled reality of simulations should be discarded in the case of the artificial sciences: our thesis is that a reality of reference should not be assumed in general to exist before and independently of the construction of the model; rather, it should be intended as emerging from the processes of modelling and implementation.

If this view is accepted, also another aspect of the standard epistemology of computer simulations fails, namely the two-steps process composed by validation and verification:

> Verification is the process of making sure that an implemented model matches its design. Validation is the process of making sure that an implemented model matches the real-world.[27]

If validation is obtained by checking homomorphisms between model and reality,[28] we need to have a full, stable understanding of the simulated system and as

[26] Cf. Korb and Mascaro (2009, p. 10).

[27] See North and Macal (2007, pp. 30–31).

[28] This is the view held in Korb and Mascaro (2009).

such the simulating model cannot function either as explanatory nor as predictive: the model needs to be descriptive. In view of such obvious critique, graded validation can be admitted:

> The existence of an approximate homomorphism is crucial: it underwrites the relevance of the simulation for the system being simulated and, in particular, its use both for explaining events in the real world and in predicting them.[29]

This position[30] claims that no 'perfect mimesis' is required between simulations and physical systems. As an alternative, an account of how validation proceeds can be offered as follows:[31]

- first testing of low-level submodels are performed, which describe non-emergent phenomena in the simulation;
- then higher-level systems are considered, including properties of the simulation that emerge from interactions between submodels; at this stage simulated versions of controlled experiments are considered.

For the artificial sciences, validation cannot be defined by comparison with an existing real-world system, as we are working under the assumption that such a system is shaped dynamically by the explorative indications offered by the simulation itself. Instead, validation must be understood as the process of checking that the model abstracted from the current implementation approximates (up to some admissible degree of variation, see more below on this) the *intended* system. The latter can be described as (possibly a subset of) the validities of a formal model. This means that the appropriate level of abstraction is chosen for the model, which has to be endowed with the right semantic description and the appropriate inferential power to extract information for the relevant variables. This gives us the minimal indication for good model building in this context:

Definition 2 (*Validation*) A model is valid *sensu lato* if it is:

- valid (*sensu stricto*), i.e. defined at the appropriate level of abstraction (semantics);
- correct, i.e. providing the right inferential values to the relevant variables (syntax).

Verification is usually defined as the test checking that the simulation properly implements the model.[32] In the presence of an idealised formal theory which sets optimal benchmarks (but it does not reflect a reality to be mimicked), verification

[29] See Korb and Mascaro (2009, p. 9).

[30] Also maintained in Winsberg (2003).

[31] Cf. Railsback and Grimm (2011, p. 316).

[32] See e.g. Korb and Mascaro (2009, p. 13).

requires unit and integration testing to establish that the program does what it is supposed to do. This is assessed against the selected parameters of the formal models that are considered essential for the simulation to exemplify properties of the system of interest. Here the appropriate criterion of evaluation is *fitness-for-purpose* of the simulation, i.e that the simulation encodes the same (relevant) levels of abstraction of the model, while design choices can be made to provide the model with characteristics which arise only in simulation. But the simulation has to reflect also *usability*, i.e. that a sufficient level of well-functioning is guaranteed. This process illustrates how the designer in this case does not aim at discovering a given pre-existing model, but rather at calibrating or fine-tuning the simulation with respect to a particular context, to increase its level of fit for the resolution of a given task. To sum up:

Definition 3 (*Verification*) A simulation is verified if it is:

- *usable*, i.e. it guarantees a minimal level of well-functioning;[33]
- *fit-for-purpose*, i.e. defined at the LoAs appropriate to the corresponding model.

The combination of (model) validation and (simulation) verification does not aim at checking whether the theory represents some reality of reference, a relation usually named *confirmation*. The combined roles of verification and validation is to aim at the *controllability* of the constructed model, i.e. the ability of changing the parameters of reference and lead the simulation in a desired direction, or in other words to provide *predictability*.

5.3 Problem Solving Instead of Explanation

The principles of Economy and Context-Dependency in the Philosophy of Information state that conceptual resources in formulating the model need to be less than those used to obtain the result of the model and that the isomorphism between the model and the modelled reality is local, not global. They help clarifying our epistemological analysis for the artificial sciences in the context of the debate on the explanatory power of simulations.

One aspect of this debate is related to the role of simulations in exploring the deductive consequences of theories; another one is their role in empirical sciences.[34] In our analysis, we have assumed a formal theory to provide optimal benchmarking for the simulation's results. In this sense, simulations respect the principle of economy set out by the designer in terms of *information containment* with respect to the benchmarking offered by the theory. If the simulation is good at approximating the optimal benchmarking of the model, we can say accordingly that the design of the model is optimal:

[33] Usability is to be considered weaker than reliability, which substitues truth in Winsberg (2006).

[34] See e.g. Korb and Mascaro (2009, p. 11).

Definition 4 (*Optimality*) The design of a model is optimal if it maximizes the amount of correct data inferred from the simulation while preserving the appropriate level of abstraction.

On the other hand, we have left it open to the simulation to provide the theory dynamically with new properties: in doing so, the simulation acts as an empirical experiment, in the special characterization of exploratory experiments.[35] In Definition 1, this aspect was captured by the ability of simulation S to control the intended model B of an artificial system R. In this sense, simulations reflect a principle of *information expansion* justified by a local isomorphism between experiment and model, i.e. that the experiment cannot be considered valid in all contexts in which the theory can be. If the simulation is good at offering useful information without exceeding the limits of its model, we say that the design of the model is efficient:

Definition 5 (*Calibration or Efficiency*) The design of a model is efficient or calibrated if it minimizes the amount of correct data required by the simulation to be useful while still preserving the appropriate level of abstraction.

Bringing together these two trends of economy and context-dependency, reflected by information containment and information expansion, we can reconsider the role of agent-based simulation in providing explanation. As we have given up their role in a theory-driven understanding of experimentation, simulations in the artificial sciences perform in the first place an explorative role in shaping the model. Only when a stable model is reached (optimal and efficient), the corresponding simulation analysis can be said to provide an explanation of such model.[36] The role of simulations in the exploratory phase is better expressed in terms of their ability to solve well-formulated problems that fall within the benchmarking given by the formal model, and eventually modified by the recursive design of the simulation. This is, essentially, a characteristic of the sciences of the artificial, where a computational model can be formulated but often no system is already available against which the simulation can be assessed.

6 A Novel Definition and Final Remarks

To conclude, in this section we recollect briefly the main aspects of our analysis to evaluate their impact on the definition of simulation in the artificial sciences.

First, we have highlighted how the relation of representation betwen computational model and the represented object or process has a different, if not inverted, conceptual order, in that the object or process of reference may not exist before

[35] The characterization of experiments as exploratory in the artificial sciences and in robotics in particular is due to Schiaffonati (2016).

[36] This position complements the semantic interpretation of simulations given in Barberousse et al. (2009) through the addition of the essential exploratory phase.

the development of the model. Second, the confirmation of the model is substituted by its (dynamic) controllability by the simulation. Third, model controllability may be construed locally from partial relations within the simulation and it has to reach the appropriate equilibrium between semantic expressiveness and inferential power. Finally, the aim of simulation is to provide solutions to problems formulated in terms of the variables at the levels of abstraction of interest.

We note that these observations significantly depart from the notion of simulation considered in Definition 1. On this basis, we offer a tentative definition of simulation for an artificial system:

Definition 6 (*Simulation of an artificial system*) A system *S* provides a core simulation of an artificial system *R* just in case

1. *S* is a concrete computational device implementing a valid and correct computational model *B* of *R*, and
2. *S* is a verified simulation of a model *B*, offering a usable and fit-for-purpose interpretation of the system *R*, and
3. *S* controls the intended model *B* of *R*, and
4. *S* provides solutions to problems formulated within the model *B* of *R*.

In this paper we have considered the epistemological foundation of computer simulations for the artificial sciences. We have argued how the latter have a specific characterization, which is not necessarily shared by natural sciences and which in turn determines our understanding of computer simulations in their context. In particular, such characterization is due to the peculiar relation that modelled, model and implementation present in sciences like robotics and network theory. Our aim has been to illustrate an appropriate epistemological foundation in terms of the principles formulated within the Philosophy of Information. These principles are identified as knowability and constructability of the model in terms of appropriate levels of abstraction; controllability and confirmation of the simulation in terms of fit-for-purposeness and predictability; and finally economy and context-dependency of the simulation-model relation, in terms of information containment and information expansion, where problem-solving is a more appropriate context than explanation to investigate.

Future work in this area will focus on a precise, formal characterization of the isomorphism relations and their scope between simulation, model and system of reference. Note that this allows in turn to qualify the proper meaning of a simulationist theory in the context of the Artificial Sciences.

Acknowledgements The author wishes to thank the participants to the Summer School On Computer Simulation Methods for engaging discussions and observations around the topics treated in this contribution and two anonymous reviewers for important critiques which have contributed improving this paper. The author was partially supported by the Project PROGRAMme ANR-17-CE38-0003-01.

References

Barberousse, A., Franceschelli, S., & Imbert, C. (2009). Computer simulations as experiments. *Synthese*, *169*(3), 557–574.

Battistelli, L., & Primiero, G. (2017). Logic-based collective decision making of binary properties in an autonomous multi-agent system. Technical report, Middlesex University London. https://doi.org/10.13140/RG.2.2.31902.18246

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-Based Models of Geographical Systems* (pp. 85–105). Dordrecht: Springer.

Durán, J. M. (2013). A brief overview of the philosophical study of computer simulations. *American Philosophical Association Newsletter on Philosophy and Computers*, *13*(1), 38–46.

Elsenbroich, C. (2012). Explanation in agent-based modelling: Functions, causality or mechanisms? *Journal of Artificial Societies and Social Simulation*, *15*(3), 1.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12.

Floridi, L. (2011). A defence of constructionism: Philosophy as conceptual engineering. *Metaphilosophy*, *42*(3), 282–304.

Frigg, R., & Reiss, J. (2009). The philosophy of simulation: hot new issues or same old stew? *Synthese*, *169*(3), 593–613.

Grüne-Yanoff, T. (2009). The explanatory potential of artificial societies. *Synthese*, *169*(3), 539–555.

Guala, F. (2002). Models, simulations, and experiments. In L. Magnani & N. J. Nersessian (Eds.), *Model-Based Reasoning*. Boston, MA: Springer.

Hartmann, S. (1996). The world as a process. In R. Hegselmann, U. Mueller, & K. G. Troitzsch (Eds.), *Modelling and simulation in the social sciences from the philosophy of science point of view* (pp. 77–100). Dordrecht: Springer.

Humphreys, P. (1990). Computer simulations. In *PSA: proceedings of the biennial meeting of the Philosophy of Science Association, 1990* (pp. 497–506).

Humphreys, P. (1995). Computational science and scientific method. *Minds and Machines*, *5*(4), 499–512.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford: Oxford University Press.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, *169*(3), 615–626.

Korb, K. B., & Mascaro, S. (2009). The philosophy of computer simulation. In *Logic, methodology and philosophy of science: proceedings of the thirteenth international congress* (pp 306–325). Springer.

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 14366.

Morrison, M. (2009). Models, measurement and computer simulation: The changing face of experimentation. *Philosophical Studies*, *143*(1), 33–57.

North, M. J., & Macal, C. M. (2007). *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. Oxford: Oxford University Press.

Primiero, G., Raimondi, F., Bottone, M., & Tagliabue, J. (2017). Trust and distrust in contradictory information transmission. *Applied Network Science*, *2*(1), 12.

Railsback, S. F., & Grimm, V. (2011). *Agent-based and individual-based modeling: A practical introduction*. Princeton: Princeton university press.

Schiaffonati, V. (2016). Stretching the traditional notion of experiment in computing: Explorative experiments. *Science and Engineering Ethics*, *22*(3), 647–665.

Tal, E. (2011). From data to phenomena and back again: Computer-simulated signatures. *Synthese*, *182*(1), 117–129.

Winsberg, E. (2003). Simulated experiments: Methodology for a virtual world. *Philosophy of Science*, *70*(1), 105–125.

Winsberg, E. (2006). Models of success versus the success of models: Reliability without truth. *Synthese*, *152*(1), 1–19.

Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago, IL: University of Chicago Press.