VALIDATING SIMULATION MODELS

Klaus G. Troitzsch
Universität Koblenz–Landau
Postfach 201 602, D-56016 Koblenz-Germany
E-mail: kgt@uni-koblenz.de

KEYWORDS

Multi-agent model, mathematical model, stochastic model, simulation model, validation.

ABSTRACT

This paper discusses aspects of validating simulation models designed to describe, explain and predict real-world phenomena. It starts with a short review of arguments used in the simsoc mailing list discussion on theory, simulation and explanation a few months ago, deals with the use of quantitative and qualitative computational models to make quantitative and qualitative predictions or rather to draw conclusions from complex antecedents, and then discusses different types of explanation and prediction (and the relation between these two), It closes with an overview of topics in validity and validation from the point of view of the structuralist programme in the philosophy of science.

INTRODUCTION: THEORY, SIMULATION, EXPLANATION AND OBSERVATION

A few months ago, the simsoc mailing list experienced a longish discussion[1] which originated from Thomas Kron's question "about the relation of computer simulation and explanation, especially sociological explanation". More than fifty contributions to this discussion followed within less than three weeks, and contributors discussed the role of simulation in theory building (mostly, but not only) in the social, economic and management sciences — as well as the relation between observation on one hand and computer-assisted theory building (Hanneman 1988) on the other. Scott Moss came back to his presidential address at the 1st conference of the European Social Simulation Association, Groningen, September 2003, in which he said "that if social simulation with agents is to be anything other than another in the long line of failed approaches to social science, it will be a positive departure only because it facilitates the dominance of observation over theory" and continued that the great successful scientists (outside the social sciences) built their generalisations around observation, developing new theoretical structures based on and validated by new evidence (quoted from his contribution to the simsoc mailing list as of November 14, 2003). Well in the line of this trait of thinking is the role of simulation or computational modeling which can be found in Gilbert and Troitzsch 1999 which was recently extended by Alexis Drogoul (Drogoul *et al.* 2003: 5) and can be seen in Figure 1.

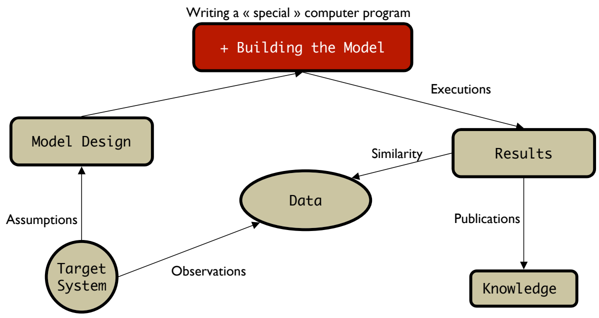

Figure 1: Drogoul's and his colleagues' interpretation of Gilbert's and Troitzsch's methodological proposition on the role of simulation

This diagram does not even contain the word 'simulation', but in its centre we find 'data' which are taken from observation and compared with results from simulation runs, for their similarity. Gilbert's and Troitzsch's original diagram describing "the logic of simulation as a method" (Gilbert and Troitzsch 1999: 16, 54, see also Troitzsch 1990: 2) is much the same: A model is built by abstraction from a target system, it is translated into a computer programme which can then be run and delivers results in the form of simulated data which can, and have to, be compared to data gathered from the same kind of target systems in the real world from which the model was abstracted.

Being aware that observation (as contrasted to just looking around in the world) presupposes at least some primitive form of theory (which tells us which entities and which of its properties to observe and which relations between them to register to find out whether there are some regularities), we should admit that our assumptions and our observation are not independent from each other (although Figure 1 insinuates this). And we should admit that in most cases computational (and other) models do not directly start from observation data but from a theory which in turn should build on, but often does not refer explicitly to observation data. Instead, we often start from a verbal theory which expresses our (or other authors') belief in how reality works, comparing simulation results with stylised facts instead of observation data.

A good example of this strategy is Sugarscape where the question "can you explain it?" is interpreted as "can you grow it?", and where "a given macrostructure [is] 'explained' by a given microspecification when the latter's generative sufficiency has been established." (Ep-

[1] The discussion can be found in the November 2003 section of `http://www.jiscmail.ac.uk/archives/simsoc.html`, topics "simulation and explanation" and "theory and simulation".

Proceedings 18th European Simulation Multiconference
Graham Horton (c) SCS Europe, 2004
ISBN 3-936150-35-4 (book) / ISBN 3-936150-36-2 (CD)

stein and Axtell 1996: 177)

At the other extreme, we might have microanalytical simulation which starts from a large collection of observational data on persons and households and the population as a whole. The model is initialised with empirical estimates of transition probabilities, age-specific birth and death rates and so on. Tens of thousands of software agents are created with data from real world people. And all this aims at predicting something like the age structure or kinship networks of this empirical poulation in the far future (see for instance Harding 1996).

In what follows we want to discuss the use of quantitative and qualitative computational models to make quantitative and qualitative predictions or rather to draw conclusions from complex antecedents and discuss different types of explanation and prediction (and the relation between these two) and close with an overview of topics in validity and validation.

QUALITATIVE AND QUANTITATIVE SIMULATION

Although most simulation uses quantitative procedures — doing calculations with numbers, often real valued, which make believe that the properties of the target system are quantitative, metric properties —, most of our mental models and verbal theories which are the predecessors of most of our simulation programmes do not talk about numbers and numerical values, but rather of properties which are categorical or, at best, ordinal. "However we claim that the use of numbers in this way is often simply a result of laziness — we often use numbers as a stand-in for qualitative aspects that we do not know how to program or have not the time to program." (Edmonds and Hales 2003: 3)

Example: Gender desegregation among staffs of schools

The following example — which is taken from (Gilbert and Troitzsch 1999: 108–114) and earlier papers — tries to "explain" how the process of overcoming gender segregation in German schools went on in the 1950s and 1960s. The modeling process started from a large collection of empirical data showing the proportion of male and female teachers in all grammar schools in the federal state of Rhineland-Palatinate (approximately 150 in number) from 1950 to 1990 (see Figure 2, left graph). The model reproducing the empirical distribution of this proportion over time quite well was designed as parsimonious as possible, just assuming three hypotheses:

1. that all teachers leaving their jobs are replaced by men and women with equal overall probability (Article 2 linea 2 of the German Basic Law),
2. that men stay in their jobs approximately twice as long as women (an empirical observation), and
3. that new women are assigned to an individual school with probability $P(W|\xi) = \nu(t)\exp(\kappa\xi)$ according to the percentage ξ of women among its teachers (a theoretical assumption); κ is 0.5 in this simulation run, and $\nu(t)$ is such that at all times men and women have the same overall probability of replacing retired teachers, to comply with hypothesis 1.

The simulation is initialized with a gender distribution close to the empirical distribution of 1950. With $\kappa > 1$, gender segregation would continue and even become stronger.

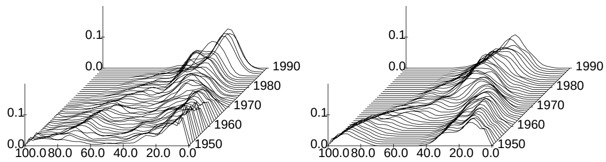

Figure 2: Distribution of percentages of women among teachers at 150 secondary schools in Rhineland-Palatinate from 1950 to 1990; left: empirical data, right: simulation

The simulation model reproduced the qualitative result that in the early 1970s the staff of all these 150 schools became mixed after twenty years of segregation where there were schools with high proportions of either male or female teachers but nearly no schools with between 40 and 60 per cent female teachers. And this reproduction / retrodiction was effected with the help of quantitative simulation, calculating probabilities of assigning teachers to schools. But did the model explain how and why this happened? Obviously not — since it is clear that the school authority, in fact officers in the ministry of education, did not cast dice or draw random numbers to select candidates for particular schools. Perhaps these officers saw to it that the overall proportion of men and women in school staffs was sufficiently equal to give women an equal chance, but even this has not been observed — instead we know that the process of desegregation of school staffs had entirely different origins: it was only the consequence of desegregation among girls and boys which in turn was due to the fact that most small towns could not afford separate schools for boys and girls (the percentage of girls in grammar schools rose steeply in the 1950s and 1960s). To summarise: a nice prediction (or at least retrodiction), but a poor explanation.

Example: Artificial eutrophication of a lake

Another example which is at the borderline between quantitative and qualitative simulation is the following. It was derived from a purely quantitative System Dynamics simulation in the tradition of Meadows and Forrester (Anderson 1973) which was used to quantitatively predict the consequences of bringing fertiliser into the soil in the neighbourhood of a lake and of actions taken to avoid these consequences by, for instance, harvesting algae or dredging the ground of the lake. This was, as it were, a simulation machine to predict the outcomes of real-world experiments or perhaps to replace such experiments. Anderson's model was not designed to predict how farmers, fishers, tourist offices, local authorities around the lake would act when they realised that dead fish was swimming on the surface of the lake or when its water reeked of decay: this was only introduced in a revised model where local authorities — modelled as software agents — could decide which action to take when they were informed about the state of the lake, and where local farmers — also modelled as software agents — could decide whether it was more profitable for them to pay taxes for using artificial fertiliser on their fields and to grow more crop or to waive fertilising, not to pay fertiliser taxes and to be satisfied with lower yield (Möhring and Troitzsch 2001; see Figure 3).

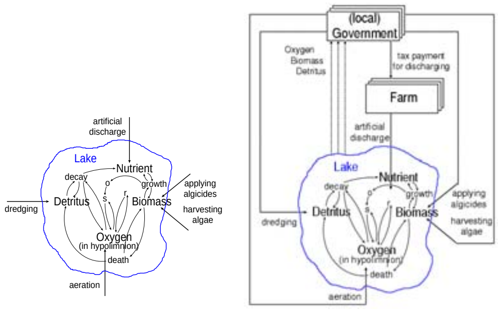

Figure 3: Two versions of Anderson' model of the eutrophication of a lake (Anderson 1973); left: a model of a lake subject to simulation experiments, right: a model of the lake and its (agri-) cultural environment

The difference between the two approaches is twofold:

- First, in the original approach Anderson took a formal model of the physical, chemical and biological processes running in a lake to simulate what could happen if these processes were disturbed by external influences imposed on the model by the simulating experimenter — whereas the extended model embeds the model of a lake into its social, political and economic environment and models influences external to the lake as internal to the model, thus taking into account that the lake and its socioeconomic environment interact and are interdependent.

- Second, the original approach starts from physical, chemical and biological theory providing the equations between the main variables describing the lake and generates quantitative simulation results (predictions) which might be compared to further observation data and help improve (fine tune) the theory of the biochemical processes occurring in a lake — whereas the extended model takes the behaviour of the lake for granted and adds a number of assumptions about the behaviour and actions of a number of human actors which are (or at least could have been) based on everyday observation and discussions with stakeholders, and these assumptions are not aimed at generating quantitative predictions about the effect of the tax rate imposed on artificial fertilisers on the oxygen concentration in the lake, but only in qualitative predictions answering the question under which tax regime and under which additional measures taken by government and other neighbours of the lake its eutrophication could or could not be avoided.

Both approaches would serve as explanatory models: the restricted model would explain how and why under certain external or internal circumstances a lake would be eutrophicated and suffocate and what was the physical and biochemical mechanism behind the processes leading to total consumption of oxygen at the ground of the lake — and it would at the same time recommend countermeasures and allow for an estimation for their potential success; and the extended model would explain under which conditions such countermeasures would be taken by the population living around the lake and what incentives one part of the population would have to promise another part of the population to take the necessary countermeasures.

EXPLANATION AND PREDICTION

There is a long discussion about the question whether explanation and prediction are equivalent, or, to put it in other words, whether a theory which predicts empirical observations correctly at the same time explains what it predicts. Grünbaum (1962) pleaded for the equivalence while Scriven (1969) pleaded that both were "non-symmetrical". If we consider prediction and explanation equivalent then our first example above would have explained the gender desegregation in German schools observed in the second half of the 20th century (although this was only retrodiction, but in principle, the three assumptions could have been stated in 1950), but this explanation is of the same quality as the explanation Mesopotamian priests could give 2,500 years ago for their (mostly correct) predictions of solar eclipses. In both cases, some scepticism in in order: from our research into the history of school staffs we know that desegregation had different causes than those stated in the assumptions, and the Mesopotamian theories of planetary movements were superseded 400 years ago by new theories which are substantially more valid.

The controversy between Grünbaum and Scriven, however, was different: Scriven had argued the other way round: "Satisfactory explanation of the past is possible even when prediction of the future is impossible." (Scriven 1969: 117; Grünbaum 1962: 126) while we argued above that even when prediction of the future is possible with the help of a theory, this does not mean that this theory satisfactorily explains what happened (another theory could yield the same prediction and deliver a better explanation).

Without going into the details of this old controversy we should instead discuss what explanation and prediction could mean in the context of (social) simulation. Epstein and Axtell argued that explanation of a phenomenon is achieved once the phenomenon could be reconstructed or generated ("grown"). From this point of view, the development of the distribution of percentages of female teachers in German grammar schools is explained by the three assumptions mentioned above, since this time-dependent frequency distribution as a macrostrucure could be reconstructed quite well from the microstructure defined in the three assumptions. Of course, this reconstruction is by no means quantitatively precise: the two graphs are similar, but not identical (perhaps due to some simplifications in the assumptions, perhaps due to the fact that the random number generator in the simulation run which generated the time-dependent frequency distribution was not perfect, or for any other reasons) — and, of course, Sugarscape explanations are of the same, non-quantitative type.

What simulation models like these are designed to predict is only how a target system might behave in the future qualitatively; what we want to know is whether any macrostructures might be observed and what these macrostructures might look like, given that on a micro level some specific rules are applied or some specific laws hold. This is what we should call a qualitative prediction which at best would tell us that a small number of categorical outcomes can be expected with their respective probabilities. But this is not the type of prediction as the objective of simulation "which most people think of when they consider simulation as a scientific technique" (Axelrod 1997: 24) — "most people think of" attempts at simulating planetary formation (Casti 1996: 14) instead of "simulating the movement of workers or armies". But if we use prediction in a non-quantitative sense, predictions delivered by simulations might still be useful "for the discovery of new relationships and principles" which Axelrod finds "at least as important as proof or prediction". They might answer questions like "Which kinds of macro behaviour can be expected from a given micro structure under arbitrarily given parameter combinations and initial conditions?" The definition of this micro structure will typically be derived from observations on the micro level, and the simulated macro structures will typically be compared to macro structures in the target systems (which perhaps have not even be discovered). And a simulation model which generates a macro structure which resembles real-world macro structures from simulated micro structures which resemble micro structures observable in the real world might be accepted as a provisional explanation of real-world macro structures.

In a second step we might apply simulation to proceed to a second stage of qualitative prediction, where we are not interested in the general behaviour of a certain *class* of target systems, but in the future behaviour of a particular *instance* of this class of target systems — say, the future market shares of a number of competing products in a market, trying to answer the question whether most trademarks will survive with reasonable market shares or whether most of them will survive only in small niches whereas one product will gain an overwhelming share of the whole market; this would still be a qualitative answer: we might not be interested in which trademark will be the winner, and we might not be interested in how many per cent of the market it will win (this would be only the third use of simulation, namely to predict quantitatively and numerically, as in microanalytical simulation and, perhaps, also in the simulation of climatic changes where we would not be content with the outcome that mean temperatures will rise but wanted to know when, where and how fast this process would have effects on nature and society). Or, to return to the example of the lake, its eutrophication and the countermeasures taken by its neighbours, we would

- first apply simulation to the very general question whether an artificial society "living" around an artificial lake which functions much like an empirical lake could ever learn to avoid eutrophication (something like a tragedy-of-the-commons simulation),
- then apply simulation to an empirical setting (describing and modelling an existing lake and its surroundings) to find out whether in this specific setting the existing lake can be rescued, and
- eventually to apply simulation to the question which political measures have to be taken to make the lake neighbours organise their economy in a way that the best possible use is made of the lake — and obviously this would be a discursive model in which stakeholders should be involved to negotiate and find out what "best possible use" actually means for them.

And to involve stakeholders in the development of a simulation model like this it will be necessary to validate the model (which could be done in the first two steps described just above) — otherwise stakeholders would not believe it was worthwhile to work with the simulation model.

TYPES OF VALIDITY

With Zeigler we should distinguish between three types of validity:

- replicative validity: the model matches data *already acquired* from the real system (retrodiction),
- predictive validity: the model matches data *before* data are acquired from the real system,
- structural validity: the model "not only reproduces the observed real system behaviour, but truly reflects the way in which the real system operates to produce this behaviour." (Zeigler 1985: 5)

Zeigler here addresses three different stages of model validation (and development). Social science simulation does not seem to have followed this path in all cases: Since often data are very poor in the social sciences, early models, too, tried to be structurally valid and did not bother much about replicative or predictive validity. "Data already acquired from the real system" were not available in a form that could be matched to the bulk of data simulation models had generated. There are several reasons for this difference between natural and social sciences: Data collection is a very expensive task in the latter, and in most cases it is even impossible to generate long time series for individual or group behaviour — individual attitudes, e.g., may be changed by the very measurement process, and groups may have changed in their composition before they were able to generate a time series which would have been long enough to allow for parameter estimation. On the other hand, the different kinds of influences non-living things exact upon each other are very much limited in their number, such that a structurally valid model can much more easily be found for the target systems natural sciences deal with than for social systems.

When talking about structural validity, a digression on structuralism might be in order: Structuralism as defined by Sneed (1979) and Balzer *et al.* (1987) sees both simulation models and observations as models of a theory which in turn — for them — is a mathematical structure consisting of (among others) three sets of such models. And these models — full models, potential models, and partial potential models — are defined as lists of terms and functions and (in the case of full models) invariants. Observations in this structuralist programme in the philosophy of science are intended applications of a theory, they are a subset of the set of its partial potential models in a sense that we can talk about them in terms which are non-theoretical with respect to a theory **T** in question ("T-non-theoretical terms", for short). Elsewhere it was shown that a simulation model *"of a theory"* is "analogous to a structuralist reconstruction of this theory", and that such reconstructions can easily be translated into simulation models and vice versa (Troitzsch 1994), provided the simulation language is object-oriented and functional (in other simulation languages the translation might be less straightforward). Simulation models would then be translated into full models in so far as they contain both T-non-theoretical terms (those we can use for talking about the target system irrespective of whether the theory is validated or not) and its T-theoretical terms — those which are only introduced by the theory, "in the sense that their meaning depends on **T**", (Balzer *et al.* 1987: 40) — and, thirdly, the axioms or invariants the theory postulates — whereas observations (or rather: intended applications, to keep to the terminology of structuralism) are only partial potential models listing just the terms which are non-theoretical with respect to this theory. Thus, simulation is "richer" than observation.

Validation of simulation models is thus the same (or at least analogous) to validation of theories. In the sense of structuralism, we can interpret validation as the attempt at finding whether there exist intended applications of a theory (observations to which the theory refers) which belong to the content of the theory — which means that it should be possible to make an observation (in T-non-theoretical terms) which complies with the axioms of the theory (which in turn might be expressed in T-theoretical terms, but then these must be linked to T-non-theoretical terms).

What does this mean for agent-based simulations in the range defined in the introduction? Sugarscape agents and plants correspond to T-theoretical terms, and the rules which the agents obey correspond to the axioms of this theory. But is there any empirical claim of the theory behind Sugarscape? If this theory predicts that — with a given parameterisation and initialisation — macrostructures emerge from the microstructures programmed into "its" models, and the emerging macrostructures sufficiently resemble observable macrostructures, we could admit that this observable macrostructure together with its microstructure (provided it resembles the model's microstructure) is an intended application of the theory behind Sugarscape and that it complies with its axioms.

In the case of the empirical examples sketched above, the case is even simpler. Our model of a lake and its socioeconomic environment was based on observation, but it would still contain a number of terms which can only be used within a theory of, say, ecological consciousness: There would be some link between the state of the lake (its smell or colour) and the state of ecological consciousness of a particular person living near the lake (something like "the worse the water smells, the more am I willing to protect the lake from sewage") and the action this person takes, and we could only observe the direct link between the observable smell of the lake and the observable actions taken, so the two "internal" links (as functions with their numerical coefficients, or as fuzzy rules with their membership functions) would remain theroetical with respect to such a theory — but the computer programme used for this simulation would still be a full model of this theory, because it would contain a function or rule representing this link, and that part of the simulation output which could be compared to empirical observational data would be the partial potential model of the theory. Stakeholders, however, might find that the T-theoretical links between the observable state of the lake and the observable actions on one hand and the T-theoretical state of ecological consciousness comply with what they think how ecological consciousness (if ever such a thing exists) works. And this could be the special value simulation could have in participatory modelling approaches (cf. the last few paragraphs of El hadouaj *et al.* 2001).

CONCLUSION

We dealt with the question about the relation of computer simulation and explanation, especially sociological explanation. and came to the conclusion that computer simulation programmes can be seen as models of theories from the point of view of the structuralist programme in the philosophy of science. This means that computer simulation should always have an empirical claim the same way as any theory should have an empirical content. Empirical claims of computer simulations come in different forms — from quantitative predictions of future measurements down (or up?) to qualitative descriptions of possible scenarios. Both can be used to validate (the theory behind) the simulation model.

"Good validation of social simulation requires prediction" (Moss 2001: 9), but a good prediction is not always a sufficient indicator for validity. And "descriptiveness" is also a good indicator "for the validity of ... models" (Moss 2001: 10): When we model social processes in a participatory context, then agreement of the participating stakeholders on the validity of the model can be a reasonable indicator for the validity of the model.

References

Anderson, J. M. (1973) The eutrophication of lakes. In D. Meadows and D. Meadows (eds), *Toward Global Equilibrium*, pp. 171–140. Wright Allen, Cambridge MA.

Axelrod, R. (1997) Advancing the art of simulation in the social sciences. In R. Conte, R. Hegselmann and P. Terna (eds), *Simulating Social Phenomena*, pp. 21–40. Springer-Verlag, Berlin.

Balzer, W., Moulines, C. U. and Sneed, J. D. (1987) *An Architectonic for Science. The Structuralist Program*, Synthese Library, Vol. 186. Reidel, Dordrecht.

Casti, J. L. (1996) *Would-Be Worlds. How Simulation Is Changing the Frontiers of Science*. Wiley, New York.

Drogoul, A., Vanbergue, D. and Meurisse, T. (2003) Simulation orientée agent: où sont les agents? In *Actes des Journées de Rochebrune, Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels*. Megève, France. janvier 2003, `http://www-poleia.lip6.fr/~drogoul/papers/Drogoul.rochebrune03.pdf`, last access 28 February 2004.

Edmonds, B. and Hales, D. (2003) When and why does haggling occur? Some suggestions from a qualitative but computational simulation of negotiation. In *First conference of the European Social Simulation Association*. Groningen. September 2003, `http://www.uni-koblenz.de/~kgt/ESSA/ESSA1/essa-edmonds.pdf`, last access 28 February 2004.

El hadouaj, S., Drogoul, A. and Espié, S. (2001) How to combine reactivity and anticipation: The case of conflicts resolution in a simulated road traffic. In S. Moss and P. Davidson (eds), *Multi-Agent-Based Simulation*, pp. 82–96. Springer, Berlin.

Epstein, J. M. and Axtell, R. (1996) *Growing Artificial Societies – Social Science from the Bottom Up*. MIT Press, Cambridge, MA.

Gilbert, N. and Troitzsch, K. G. (1999) *Simulation for the Social Scientist*. Open University Press, London.

Grünbaum, A. (1962) Temporally-asymmetric principles, parity between explanation and prediction, and mechanism and teleology. *Philosophy of Science*, 29: 162–170.

Hanneman, R. A. (1988) *Computer-Assisted Theory Building. Modeling Dynamic Social Systems*. Sage, Newbury Park, CA.

Harding, A. (ed.) (1996) *Microsimulation and Public Policy*, Contributions to Economic Analysis, Vol. 232. Elsevier North Holland, Amsterdam.

Möhring, M. and Troitzsch, K. G. (2001) Lake anderson revisited. *Journal of Artificial Societies and Social Simulation*, 4/3/1. `http://www.soc.surrey.ac.uk/JASSS/4/3/1.html`.

Moss, S. (2001) Editorial introduction: Messy systems — the target for multi agent based simulation. In S. Moss and P. Davidson (eds), *Multi-Agent-Based Simulation*, pp. 1–14. Springer, Berlin.

Scriven, M. (1969) Explanation and prediction as non-symmetrical. Explanation and prediction in evolutionary theory. In L. I. Krimerman (ed.), *The Nature and Scope of Social Science. A Critical Anthology*, pp. 117–125. Appleton-Century-Crofts, New York. (First published in 1959 in *Science* 130: 477-482).

Sneed, J. D. (1979) *The Logical Structure of Mathematical Physics*. Synthese Library 35. 2nd edn. Reidel, Dordrecht.

Troitzsch, K. G. (1990) *Modellbildung und Simulation in den Sozialwissenschaften*. Westdeutscher Verlag, Opladen.

Troitzsch, K. G. (1994) Modelling, simulation, and structuralism. In M. Kuokkanen (ed.), *Structuralism and Idealization*, Poznań Studies in the Philosophy of the Sciences and the Humanities, pp. 157–175. Editions Rodopi, Amsterdam.

Zeigler, B. P. (1985) *Theory of Modelling and Simulation*. Krieger, Malabar. (Reprint, first published in 1976, Wiley, New York).