Strongly Empirical Modelling

Bruce Edmonds[1][0000-0002-3903-2507], Dino Carpentras[2][0000-0001-8471-2352] and Edmund Chattoe-Brown[3][0000-0001-8232-6896]

[1]Centre for Policy Modelling, Manchester Metropolitan University, UK
bruce@edmonds.name
[2]Department of Humanities, Social and Political Science, ETH Zurich, CH
dino.carpentras@gess.ethz.ch
[3]School of Media Communication and Sociology, University of Leicester, UK
ecb18@leicester.ac.uk

Abstract. The paper argues for, and discusses, the practice of "Strong Empirical Modelling". This involves a shift in effort from the internals of a model towards how the model relates to observational data. This kind of modelling is contrasted to theory-based modelling which, it is argued, will not be sufficient for making substantive progress understanding social phenomena in a scientific manner, due to the inherent weakness of social theory. However, there are different ways of relating data to modelling, so these are reviewed and criteria for their strength and the difficulties discussed. A sketch of how to define the empirical strength of a model is discussed, and the piece ends with a suggested system of "Modelling Ready Levels" to eliminate confusion about the state of modelling projects.

Keywords: Agent-Based Modelling, Social Simulation, Empirical Grounding, Evidence, Data, Theory

1 The mapping between model and what it represents

The word "model" is heavily overloaded in terms of its meaning. It can refer to many kinds of entity, including: a diagram, a species, a set of equations, an object or some computer code. However, a model has to be more than a thing because the purpose of a model is to represent something. A random rock, chunk of code or set of equations is not a model because it does not model anything. Thus, properly speaking, a model is composed of a thing plus a relationship to what it models. This target for the modelling can also be many different things, including: a set of ideas, an observed system, a design, the processes or structures in a theory, some data or even another model. Added to this, a model can be built for many different purposes including: prediction, to support an exploration or the theoretical exploration of a set of mechanisms [11]. Many of the confusions around modelling methodology seem to come from an assumption that all models are basically the same – that there is "one methodology to rule them all" – but it might be closer to the mark to say the situation is more complicated. Wartovsky [19] defined modelling as using A to get understanding about something else, B, by a mapping, C, between them. We adopt this definition here.

Whilst the context of this paper is agent-based social simulation, our focus is on the relationship between a model and what it represents. This relationship often gets much less attention than the specification and internal working of models as an object and can, itself, be of various kinds. The relationship might be precisely defined, so one knows which part of the model represents what exactly but might be much more loosely defined. In some models the relationship is not really defined at all, but merely implied by the labels given to its parts (e.g. variable names). Some modelling relationships are formally described leaving the modeller with no choice as to how it is applied in any circumstance, but in others the modeller can adjust the relationship to make the model apply in different ways for different cases. Some modelling relationships have been independently and repeatedly confirmed (e.g. how the gas laws relate to volume) and in others the intended relationship turns out to be simply wrong (e.g. relating phlogiston to observed cases) so that the relationship cannot be established. In some cases, the meaning of parts of the modelling relationship are straight forward (e.g. how many people have died of a particular disease) and, in other cases, the relationship itself is built upon other models (e.g. what temperature means and how one measures it).

To make its message clear, this paper does not attempt to unravel all the possible complications and variations in the modelling relationship but contrasts two very different cases: (a) where a model is intended to map to abstract ideas, processes, structures etc. and (b) where the model maps to empirical data in a well-defined manner. We argue that, if social simulation is to succeed in its goal to significantly increase reliable understanding of observed social phenomena, we need far more modelling that of type (b) than (a). Furthermore, we suggest that the first type of model (a) might often frustrate progress because it deceives as to the nature of social phenomena and diverts researcher effort from modelling that will be, ultimately, more productive.

The paper starts by illustrating the difference between theoretical and empirically grounded models and then goes on to explain the difficulties inherent in most existing social theory, difficulties that make it a hard tool to use to obtain progress towards empirical adequacy. Having motivated the empirical route, we then briefly survey the different ways that data and models can be related and some criteria for judging these. We end with a specific suggestion in terms of "Modelling Readiness Levels" to facilitate greater clarity with respect to different modelling projects.

2 Theoretical vs Empirical Grounding

To make the exposition clearer we define what we mean by "Empirical Grounding". A model is completely empirically grounded if all its assumptions, structures, theories, outcomes etc. either:

- Map to a set of data in a sufficiently convincing manner (which relates to the purpose of the model),
- Are uncontroversial – that is, it is not contested or seriously doubted by other researchers and could easily be empirically shown (e.g. that cars drive on roads),
- Are themselves empirically grounded (using the same definition).

There are several things to note about this. *Firstly*, this is a recursive definition. That is, if you chase down what the model relates to, then what those are based upon and then what those are based upon etc. you come to data in a precise manner or something that is so obvious that one would be wasting one's time getting data to prove it. *Secondly,* that it relies upon the judgement of researchers, namely: what counts as a sufficient mapping to data for its purpose and what is uncontroversial. Thus, the extent of empirical grounding is up for debate, but the grounds for such debate are relatively clear. It may well be that some components of a modelling enterprise are empirically grounded and others not. It will inevitably be the case that even where elements are empirically grounded, they are so to a degree and only given some assumptions or context. As with all judgement, mistakes are possible but a process of chasing the grounding and making the basis of a model clear, can make those more apparent and flag those for future evaluation if doubts emerge – *there is a clear process by which these judgements could themselves be empirically investigated.*

We contrast two cases: a model that is not empirically grounded and one which is.

1. A model whose specification and outcomes relate ultimately only to ideas. These ideas could be in the form of: assumptions, theories, processes, structures, stories or analogies, but these ideas are not themselves empirically grounded. The ideas might well relate to what is observed, but in a loose, analogical manner. Such a model might be used for exploring the theoretical consequences of the ideas it is based upon, as an analogy for thinking about a class of systems or simply an illustration [11]. For convenience we will call this a "*Theoretical Model*".
2. A model whose specification and outcomes ultimately relate to data – they are empirically grounded. Such a model might be used to support complex explanations, test hypotheses (in which case all but the hypothesis being tested is empirically grounded), predict possible outcomes or simply act as a kind of description of its data [11]. For convenience we will call this an "*Empirical Model*".

We realise these are extreme cases and that many models might have a mixture of empirical grounded and other elements. In this paper, we are arguing that, if social simulation is to make real progress in understanding observed social phenomena we need *more* empirically grounded models, *stronger* grounding in those models that have some grounding and *more* empirical grounding in all models. This is what we mean by arguing for "Strongly Empirical Modelling".

3 The Weakness of Social Theory

But what is wrong with basing one's model on theoretical ideas and not seeking to ground it in empirics? The short answer is that while such theories can be superficially attractive and convincing, they often have multiple, critical weaknesses. Here we briefly list some of the critical weaknesses of social theory.

- *Vagueness*. Social theory is very often vague. This is evidenced by the fact that when simulation modellers try to make a model that implements such theories, they have to make lots of modelling choices – choices where it is not clear which was intended, and which have more empirical support [1]. Each such implementation of the theory is then precise and is a better basis for use in modelling, but each implementation will come with its own additional assumptions, so the result is not easier to test.

- *Lack of Clarity in Terms of its Empirical Support*. Social theory is often supported using a variety of sources, including: argument, coherence with other theory, plausibility, qualitative agreement with evidence, case studies and checkable agreement with empirical data. It is often unclear which of these are critical to a theory and which offering merely additional support.
- *Undefined Scope*. One of the aspects of the vagueness of social theory is that, not only are the workings of the theory vaguely described but also the conditions under which it is posited as holding. In other words, the scope of much social theory – when it should and should not hold – is not made clear. Indeed, in many cases the scope is not described at all, leaving it entirely implicit.
- *Context Conflation*. Much numerical data derived from observations of social phenomena (e.g. Likert values from surveys) are shorn of information about the kind of context they were collected in (e.g. the situation of the persons surveyed). To use such data in the support of a social theory (e.g. using a statistical model) derived from different contexts – where people might be acting in fundamentally different ways – masks the sub-cases where the theory does not hold (or even where the relationship is in the opposite direction).
- *Effect Weakness*. Due to the complexity of most social phenomena, when the effect predicted by a theory is compared against empirical data, the effect size is often small – that is the theory only explains a small proportion of the observed variation, with the rest being random or external to the theory (in a correlation analysis this would be measured in terms of the R^2 statistic).
- *Suggestibility*. Finally, due to the ease with which theory expressed in natural language can be interpreted, there is a danger that such theory is used as an analogy – a way of thinking about some phenomena rather than explaining it in any more precise manner. Thus, social theory can be far more suggestable than the evidence warrants. Formal theories and models are also suggestable, but less easy to interpret.
- *Indirect nature of the mapping to observable phenomena*. Mapping a model to theoretical elements that might then relate to observed phenomena (or, even worse, map to other theoretical elements that then map to something observed), adds an intermediate stage in the journey from model to observed phenomena. This 'extra stage' inevitably adds more flexibility and imprecision to the total mapping and thus weakens the empirical grounding of the model.
- *Non-scientific attractiveness of non-empirical models*. One of the most insidious weaknesses of mapping to theory is that it is very attractive to researchers compared to trying to maps to evidence. It is a lot easier to relate to (since it provides an accessible narrative), it is easier to do (real data being notoriously difficult to align to), and more likely to get you published (somehow reviewers feel accept a theoretical basis when they question any modelling not based on a theory). Also, it gives an impression of progress, since one can imagine the future applicability of one's model because the model supports an analogy with which to think about phenomena [10].

> Ultimately, we need to develop lineages of models that map in a well-founded manner to observed phenomena, rather than their surface attractiveness, so that is what we should use to select from the available models [8].

To summarise this section, for the above reasons, theoretical elements tend to add huge amounts of unreliability into a modelling enterprise. This might be tolerable if either (a) the theory is reliable, that is empirically well established and has a well-defined way that it is related to observable phenomena or (b) the theory is what is being tested using a simulation, where all other aspects of the simulation are strongly empirically grounded (otherwise any 'test' of the theory's reliability is only relative to all the other non-empirically grounded aspects of the model). In either case, if one wants to make some progress in modelling observed phenomena then one needs to reduce the theoretical elements in a model as far as one can.

4 Different Approaches to Relating Models to Evidence

There are many ways to map a model to evidence at a variety of different modelling stages and to demonstrate different objectives. We will not go through all the possibilities here but, rather, discuss some of the 'dimensions' in which these differ and then some criteria by which one could address their adequacy.

4.1 Mapping to specification, parameters or outcomes

There are different 'parts' of a model that one can map to evidence.

One can use evidence to inform the processes and structures that are built into the model, thus constraining the model to the evidence as part of its design and implementation – this is essentially the "KIDS" approach [9]. The evidence used in this way might be quantitative (e.g. reading in map data) but may also be qualitative. In particular, qualitative evidence has been used to determine what kinds of decision-making processes an agent might use in a simulation (e.g. [3]).

One can adjust otherwise unknown values of a simulation to fit evidence – called 'calibration'. Often this is with respect to 'free parameters', which are parameters that cannot be derived from measuring or observing the target phenomena. Unless all other aspects of a simulation are empirically well-founded this does not infer the actual values, but it does show the model could be mapped to the evidence and thus potentially applicable to the situation from which the evidence came. Whilst calibration is usually done on parameter values, other unknown simulation data can be determined in this manner. An example of this is [2].

Validation is when the output of a simulation is compared to evidence. This is supposed to be a check that the model is not misguided. This is a harder test than making a simulation based on evidence or adjusting unknown parameters etc. since those are not strong constraints from the evidence – regardless of how wrong a simulation is, those are usually possible to fit. However, validation comes in all sorts of strengths and can be against known – so called 'out-of-sample' – data.

4.2 Different kinds of precision

A model is rarely supposed to be 100% precise concerning what is being modelled. There are many reasons for this, including measurement noise, intrinsically random processes in the target phenomena and processes that are simply not going to be included in a model [7]. The required kind of fit of model aspect to evidence depends upon what one is using the model for. Classically, model output might be compared numerically to a corresponding set of real-world measurements, but this is not the only kind that might be appropriate. For example, one might compare some distributional properties, as in [18] which describes a method to formally compare the shapes of data sets. Even less precisely, model and evidence might be compared in a qualitative manner, such as "are both monotonically increasing as a certain parameter increases" or "tend to decrease over time". Human narrative accounts can be used to inform what mechanisms are included in the model specification, e.g. as in [14].

4.3 Single vs. Multiple Dimensions

If a model is designed to focus on a particular output value (or sets of that value over time), e.g. if it was a model to predict that value, then the appropriate validation would only involve that value (e.g. a graph comparing simulation and empirical values). However, many agent-based models have many inputs and parameters which may be lacking data to fill. This means that any 'fit' with observed data might be due to deliberate or implicit 'tuning' of these and not due to a more fundamental ability of a simulation to adequately represent what it models (in the relevant aspects). In these cases, to establish the empirical reliability of a model, it will be necessary to compare the model output to available data in several dimensions simultaneously. This simultaneous comparison makes it far less likely that any fit is accidental if each dimension compared constrains different aspects of the model. If some aspects of the model are not constrained, then those aspects cannot be said to be validated. Generally, the more complicated the model is and more imprecisely the match between model outputs and empirical data, the more dimensions need to be compared. Pattern-oriented modelling advocates for a less precise comparison but over many independent dimensions [12].

4.4 For different modelling purposes

Establishing that a particular model achieves its purpose [11], will depend upon the kind of purpose. For example, if a model is designed to predict a particular target value given a set of parameter settings, then comparing to data that includes measured values that correspond to the target and parameters. For prediction, checking the other internal aspects of the model are not immediately important to the prediction (although they might be important to work out how a prediction happens or goes wrong), but it is important that this data is unknown to the modeller at the time of prediction (so there is no accidental fitting) and that it is checked repeatedly over the operating range of the model. For supporting a possible explanation of complex phenomena in terms of the workings of the model, it is acceptable to know the data beforehand and to actively try to 'fit' the target data, but enough aspects of the simulation need empirical checking to ensure that the model correctly supports the explanation given.

4.5 Criteria for the strength of the mapping

What is important about all this is how effective the mappings are for establishing that the model is adequate for its purpose [11]. This might involve evaluating the extent that the mappings:

1. *Show that the model is applicable to the observed case being studied.* Just because we can imagine that a model might be applicable to a particular case (e.g. due to its plausibility or that the model was developed with that in mind), does not mean it *actually* is – the apparent applicability might be 'delusional' and due to using the model as an analogy – a way of thinking about the target – rather than corresponding to anything empirical [10]. Calibrating the model to plausible input/parameter values and checking the output corresponds to some empirical data will establish this, as well as making clearer the assumptions behind the mapping from model to data in these respects. The applicability is shown by the extent of the adaption of free parameters/inputs to make the model fit the target data compared to the how constrained the model is by the empirical data available.
2. *Limit the uncertainty in the model.* Traditionally models were seen as prediction machines, and if prediction is the purpose of your model, then it is important to eliminate the uncertainty of the predictions the model supports. There are, at least, two ways of doing this: (a) simplify the internal construction of the model to use fewer inputs or so uncertainty in input is not propagated/amplified for the outputs, and (b) fix some of the inputs and parameters using data. Here, the extent to which the uncertainty is eliminated can be judged.
3. *Eliminate the possibility that the fit with the available evidence is accidental.* This criterion can be seen as an extension of (or stronger version than) criterion (1) above. The usual way to achieve this is via some kind of validation – a comparison of model outcomes with some sets. This is stronger if the validation is independent, that is the modeller did not have knowledge of the data/evidence involved when they made, or even calibrated, the model. Here the key comparison is between the 'flexibility' of the model compared to the constraints provided by the available data.
4. *Reduce the chance that an alternative model would be better.* More fundamentally there is the problem that although one's model seems to explain an observed situation well, this does not mean it is the best model for doing this – there may be multiple viable models. Comparing alternative models (or model versions) against available data to see which is the better explanation can be very informative (or more subtly *which* model fits/explains *which* observed situations best). Frequently there is no comparison of alternative models to data – not even a 'null' model (e.g. where interactions are random rather than via the network).

5 Comparing Models on Empirical Strength

Given that it is a desirable goal that models progress towards greater empirical grounding, the question arises as to how to compare models in this regard. In this there are several different ways of doing this, including the following.

1. How much of the variation/patterns observed in the outcome data is explained
2. How certain we can be that the match of model and data is not a matter of chance
3. How broad is the scope of the model, i.e. the set of circumstances where the model works well against the data

The *first* of these roughly corresponds to R^2 with traditional regression models, but could be much broader with ABMs, for example one model might get some of the levels roughly right, whilst another might get the levels as well as the kinds of distribution and a third get the levels, the general distribution and some of the short-term temporal dynamics. The *second* might be related to the ‘p values’ used in Null Hypothesis testing in statics, but might be more about the extent that a model can be fitted to any data set. The *third* is the conditions of application under which the model is expected to be reliable, roughly what might be called its level of generality.

At any stage, which of the above routes towards greater empirical grounding should be sought when comparing models is a difficult question to answer in general. However, there is considerable scope for improving all of them relative to the current state of the art and we suspect that it might be more productive to start with models with a very narrow scope but that are strong in terms of criteria 1 and 2 and then compete in terms of scope rather than pretend that models are vaguely applicable everywhere and look for weak signals in terms of these.

6 Some of the Difficulties of Empirical Modelling

Any particular kind of scientific project will face difficulties and potential limitations, and Strongly Empirical Modelling is no different. In this section we review some of these difficulties.

- *The Sheer Difficulty of Empirical Modellin*g. There is no doubt that making a model fit to sets of empirical data is far more difficult than merely being compatible to a theory. Thus, if one simply wants to play around with ideas, then theory-based modelling is appropriate.
- *Unearthing Hidden Contextual Assumptions*. All modelling rests upon a raft of assumptions, only a few of which we have good evidence to support. Some of these assumptions will be implicit, that is we are not aware of these (e.g. the various possible assumptions behind representations of space). Whilst many of these may be unproblematic (they are either common-sense or do not impact upon the results much), others of these may be important.
- *Finding Enough Data of the Right Kind*. It is a frequent experience of empirical modellers that there is either some desired data is not available, or it is not of the

right kind/quality. However, it is acceptable to leave parts where there is no good data, using synthetic data or good guesses for these as long as one makes that clear. Future researchers might then have that data. Progress towards SEM is a collective affair and can not just be down to what is possible in single projects.

- *Modelling ill-defined or ill-measured phenomena*. Whilst many aspects of a simulation may lack good data, a more fundamental problem is when the mapping from model elements to what they model are ill-defined. This is particularly true of constructs that theory proposes, e.g. social capital. However, part of the progress in empirical modelling involves making these explicit and testing which measurement methods are reliable.
- *Opening oneself up to criticism*. Adding data and other evidence into the modelling mix means that there are more aspects to criticise. Leaving these difficult aspects to the future restricts the range of possible criticism to internal modelling aspects. This will necessitate a change in habit from reviewers – instead of forming judgements based on how many worries/weaknesses one has and rather adopting an attitude that is more like "well some data is better than none". Otherwise, we risk giving more status to the safely unfinished over modelling work which engages with the messy world of data.
- *Over-convincing oneself*. It is a common phenomenon that the more one is engaged with one's model, the more one sees the world *through* the model. Finding good relationships with data can further convince one in this manner. Thus, there is a danger of being over-convinced of the reliability of a model (for its purpose). This is especially a danger if the specification of a model is largely evidence-based. Thus, some kinds of strong and independent validation processes are essential before we trust a model – calibration processes are not enough for this.
- *Getting published*. It is the experience of the authors that it can be more difficult to get SEM research published compared to theory-based work. Somehow having a "theoretical framework" reassures reviewers, resulting in data- or evidence-driven modelling being criticised as being "under theorised". Whilst the debate between theory-driven and evidence-driven modelling approaches is far from settled, a greater tolerance of these different approaches is necessary.

However, one of the potential issues – that of a lack of methodology for relating data to agent-based models – is, we feel, *not* a problem. For some of this methodology see [5, 6, 9, 11, 1418].

7 Being Clear about the Level of Modelling Achievement

After the "Challenger" disaster in 1986, and President Reagan's 1988 Directive on National Space Policy (which encouraged cooperation between NASA and other partners) NASA sought to formalise how ready technology was for deployment in space missions. In this context, clarity concerning the state of a technology is clearly crucial since the mission may have to rely upon it. If the description of some technology gives an impression that it is more mature than it is, then either this could endanger the crew, or require substantially more time and resources to further develop it. Thus, the idea of “readiness levels” was introduced to add clarity and prevent such false impressions [16] (these are showing in **Table 1**). These levels are relative to the requirements for reliability for their use, with safety critical technologies having to reach higher levels than others before being used. These levels were then further elaborated and similar systems of levels adopted by others, notably the EU. [13] describes this history of these levels and critiques its later interpretation.

Table 1. The Original NASA TRL Definitions (1989)

TRL	Criteria for Achievement
1	Basic Principles Observed and Reported
2	Potential Application Validated
3	Proof-of-Concept Demonstrated, Analytically or Experimentally
4	Component/Breadboard Laboratory Validated
5	Component/Breadboard Validated in Simulated/Realspace Environment
6	System Adequacy Validated in Simulated Environment
7	System Adequacy Validated in Space

Although it is rare that social simulations are as obviously critical as those in space missions, they are being increasingly looked to help inform the consideration of policies that can affect people’s lives (e.g. during COVID [17]). Maybe due to the pressure on academics to be “relevant”, some papers on social simulation work can give the impression that they are more ready to be reliably applied than is warranted by their substantiated progress. Thus, it may be useful to have an equivalent of these levels for the modelling domain, maybe as summarised in **Table 2**.

Table 2. Possible Modelling Ready Levels (MRL)

TRL	Criteria for Achievement
1	Concepts for an ABM described
2	Detailed specification for an ABM described
3	ABM is implemented, at least one run is shown
4	ABM assumptions etc. are all fully documented and the code is available
5	ABM is verified against specification and sensitivity analysis done
6	ABM is shown to be applicable to a situation, e.g. compared to some data/evidence or calibrated
7	ABM is sufficiently validated against evidence/data to show it is reliable for its declared purpose
8	ABM is shown to work for its intended use/situation, in practice
9	ABM is proven to work for the situation/problem described - repeatedly by users/stakeholders

Such a system of levels would aid those who would want to use/rely on social simulations in a policy context, and in grant calls/applications (e.g. “We are looking for research currently at level 3-4, for development to level 7”). However, it would also be useful for other modellers trying to understand the level of reliability of models described in scientific papers (e.g. "The model is currently at level 5. We think it could be shown to work at level 6 with access to the right data, but this has not been done yet"). This would help other researchers assess other work, in particular if they are looking to build upon that work by reusing aspects of the model or method.

8 Conclusion

Strongly empirical modelling is hard and faces a number of obstacles. However, if we really want to make scientific progress the extra effort cannot be avoided. One cannot rely upon theory-based modelling until it has been empirically tested in rigorous ways. Part of the rigour that is needed is a transparency as to what has been achieved empirically and what are merely future hopes. We hope that this paper helps lay out the groundwork for such an approach.

Acknowledgements

DC is grateful for support by the project "CoCi: Co-Evolving City Life", which received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 83316.

References

1. Antosz, P., Birks, D., Edmonds, B., Heppenstall, A., Meyer, R. Polhill, J.G., O'Sullivan, D. & Wijermans, N. (2023) What do you want theory *for*? - A pragmatic analysis of the roles of "theory" in agent-based modelling. *Environmental Modelling & Software*, 168, 105802. DOI: 10.1016/j.envsoft.2023.105802
2. Banisch, S. and Shamon, H. (2024) Validating Argument-Based Opinion Dynamics with Survey Experiments. *Journal of Artificial Societies and Social Simulation* 27(1), 17. http://jasss.soc.surrey.ac.uk/27/1/17.html.
3. Bharwani, S. (2004) Adaptive Knowledge Dynamics and Emergent Artificial Societies: Ethnographically Based Multi-Agent Simulations of Behavioural Adaptation in Agro-Climatic Systems. Doctoral Thesis, University of Kent, Canterbury, UK. http://cfpm.org/qual2rule/Sukaina%20Bharwani%20Thesis.pdf
4. Boero, R. and Squazzoni, F. (2005). Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science. *Journal of Artificial Societies and Social Simulation*, 8(4), 6. https://www.jasss.org/8/4/6.html
5. Chattoe-Brown, E. (2014) Using Agent Based Modelling to Integrate Data on Attitude Change. *Sociological Research Online*, 19(1)16. http://www.socresonline.org.uk/19/1/16
6. Chattoe-Brown, E. (2023). Comparing Opinion Dynamics Models with Data: Opportunities and Challenges. SocArXiv Papers. https://doi.org/10.31235/osf.io/tnrbs
7. Edmonds, B. (2009) The Nature of Noise. In Squazzoni, F. (Ed.) *Epistemological Aspects of Computer Simulation in the Social Sciences*. Lecture Notes in Artificial Intelligence, 5466:169-182.

8. Edmonds, B. (2010) Bootstrapping Knowledge About Social Phenomena Using Simulation Models. *Journal of Artificial Societies and Social Simulation*, 13(1), 8. http://jasss.soc.surrey.ac.uk/13/1/8.html
9. Edmonds, B. and Moss, S. (2005) From KISS to KIDS – an 'anti-simplistic' modelling approach. In P. Davidsson et al. (Eds.): Multi Agent Based Simulation 2004. Springer, *Lecture Notes in Artificial Intelligence*, 3415:130–144. http://cfpm.org/cpmrep132.html
10. Edmonds, B., Carpentras, D., Roxburgh, N., Chattoe-Brown, E. and Polhill, G. (2024) Delusional Generality – how models can give a false impression of their applicability even when they lack any empirical foundation. *Review of Artificial Societies and Social Simulation*, 7 May 2024. https://rofasss.org/2024/05/06/delusional-generality
11. Edmonds, B., le Page, C., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Montañola-Sales, C., Ormerod, P., Root H. & Squazzoni. F. (2019) Different Modelling Purposes. *Journal of Artificial Societies and Social Simulation*, 22(3):6. http://jasss.soc.surrey.ac.uk/22/3/6.html. DOI: 10.18564/jasss.3993
12. Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., ... & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310(5750), 987-991.
13. Héder, M. (2017). From NASA to EU: the evolution of the TRL scale in Public Sector Innovation. *The Innovation Journal*, 22(2), 1-23.
14. Moss, S. and Edmonds, B. (2005) Sociology and Simulation: - Statistical and Qualitative Cross-Validation, *American Journal of Sociology*, 110(4) 1095-1131.
15. Moss, S. and Edmonds, B. (2005a). Towards Good Social Science. *Journal of Artificial Societies and Social Simulation* 8(4), 13. http://jasss.soc.surrey.ac.uk/8/4/13.html
16. Sadin, S. R., Povinelli, F. P., & Rosen, R. (1989). The NASA technology push towards future space mission systems. In *Space and Humanity* (pp. 73-77). Pergamon.
17. Squazzoni, F., Polhill, J. G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., Chappin, E., Borit, W., Verhagen, H., Giardini, F. and Gilbert, N. (2020) Computational Models That Matter During a Global Pandemic Outbreak: A Call to Action. *Journal of Artificial Societies and Social Simulation*, 23(2), 10 <http://jasss.soc.surrey.ac.uk/23/2/10.html>. doi: 10.18564/jasss.4298
18. Thorngate, W. & Edmonds, B. (2013) Measuring simulation-observation fit: An introduction to ordinal pattern analysis. *Journal of Artificial Societies and Social Simulation*, 16(2), 14. http://jasss.soc.surrey.ac.uk/16/2/4.html
19. Wartofsky, M. W. (1979). The Model Muddle: Proposals for an Immodest Realism. In *Models: Representation and the scientific understanding*. Dordrecht: Springer Netherlands, pp. 1-11. DOI: 10.1007/978-94-009-9357-0_1