JASSS
THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION

Social Simulation Models as Refuting Machines

Nicolas Mauhe[1], Luis R. Izquierdo[2], Segismundo S. Izquierdo[3]

[1] *GREThA, Université de Bordeaux, France*
[2] *Universidad de Burgos, Edificio A, Avda. Cantabria s/n, Burgos, 09006, Spain*
[3] *Escuela de Ingenierías Industriales, Doctor Mergelina s/n, 47011 Valladolid, 47011, Spain*
Correspondence should be addressed to *lrizquierdo@ubu.es*

Journal of Artificial Societies and Social Simulation 26(2) 8, 2023
Doi: 10.18564/jasss.5076 Url: http://jasss.soc.surrey.ac.uk/26/2/8.html

Received: 20-09-2022 Accepted: 08-03-2023 Published: 31-03-2023

Abstract: This paper discusses a prominent way in which social simulations can contribute (and have contributed) to the advance of science; namely, by refuting some of our incorrect beliefs about how the real world works. More precisely, social simulations can produce counter-examples that reveal something is wrong in a prevailing scientific assumption. Indeed, here we argue that this is a role that many well-known social simulation models have played, and it may be one of the main reasons why such well-known models have become so popular. To test this hypothesis, here we examine several popular models in the social simulation literature and we find that all these models are most naturally interpreted as providers of compelling and reproducible (computer-generated) evidence that refuted some assumption or belief in a prevailing theory. By refuting prevailing theories, these models have greatly advanced science and, in some cases, have even opened a new field of research.

Keywords: Social Simulation, Computer Simulation, Refutation, Modelling, Counter-Example, Markov Chain

Introduction

1.1 Social simulation models have been used increasingly in recent years, but their role in the scientific method remains the subject of intense debate (see e.g. Axelrod 1997a; Troitzsch 1997; Gilbert & Terna 2000; Edmonds 2001; Leombruni & Richiardi 2005; Edmonds & Hales 2005; Epstein 2006b, 2008; Gilbert & Ahrweiler 2009; Galán et al. 2009; Edmonds 2010; Arnold 2014; Squazzoni et al. 2014; Davidsson et al. 2017; Edmonds et al. 2019; Arnold 2019; Anzola 2019, 2021a). Nowadays there seems to be broad consensus that understands models as *tools* designed to achieve a certain purpose (Epstein 2008; Davidsson et al. 2017; Edmonds et al. 2019). This view highlights the *purpose of the model* as a key factor to determine how the model should be interpreted and evaluated. Edmonds et al. (2019) review seven distinct purposes for social simulation models (prediction, explanation, description, theoretical exposition, illustration, analogy, and social interaction), acknowledging the fact that their list is not exhaustive.

1.2 In this paper we elaborate on the role of (some) social simulation models as *refuting machines*. Under this role, models are interpreted as formal machines that allow us to explore the logical implications of some of the assumptions contained in a certain empirical theory (Here we use the term 'empirical theory' in a rather loose sense, as a shortcut to denote a set of beliefs about how certain aspects of the real world work, which are often not made explicit). Under this interpretation, the important point is not whether the formal model is an adequate representation of any real-world system, but rather, whether it fulfils the assumptions of a prevailing empirical theory, so the formal model can be used to explore the logical consequences of the empirical theory. In this case, if the model produces something that is at odds with the predictions of the theory, this can be understood as an indication that something may be wrong with the theory.

1.3 The role of models as refuting machines can be subsumed in Edmonds et al.'s (2019) list of common modelling purposes under the *illustration* category or, often more appropriately, under the *theoretical exposition* category,

depending mainly on the extent to which the model has been thoroughly analysed and proved to be robust and sufficiently general.

1.4 To explain the role of social simulation models as refuting machines, let us first clarify some of the terminology we will use. We say that a theory has been *refuted* if it has been proven wrong. If the refutation occurs within the empirical realm (i.e., by an empirical observation), we say that the theory has been *falsified* (Popper 2005). For example, the observation of a white raven falsifies the empirical statement 'All ravens are black'. If the refutation occurs within the realm of formal statements (devoid of empirical content), we say that the theory has been *disproved*. For example, proving that number 3 is odd disproves the statement 'All integers are even'.

1.5 In our view, social simulations cannot directly *falsify* scientific theories since they cannot produce genuinely new empirical evidence,[1] but they may *disprove* them by providing (new) formal statements that are inconsistent with some of the theory's assumptions or implications. In this way, social simulation models can be seen as *refuting machines* of scientific theories with empirical content.

1.6 The basic framework we wish to put forward in this paper is sketched in Figure 1. An empirical theory is created from a certain real-world process by identifying key entities, individual variables, interaction rules and aggregate variables. The empirical theory is defined by a certain set of assumptions ('shared assumptions + other assumptions' in Figure 1). Now let us consider a computer model that contains some of the assumptions in the empirical theory ('shared assumptions' in Figure 1) plus potentially some other assumptions that are not significant ('non-significant assumptions' in Figure 1). Non-significant assumptions are those that do not affect the logical consequences of the computer model significantly. If the computer model generates (formal) evidence that contradicts some of the assumptions in the empirical theory or some of its implications, then the empirical theory is shown to be inconsistent. This is because the computer model would be producing logical consequences of the shared assumptions, which are contained in the empirical theory.

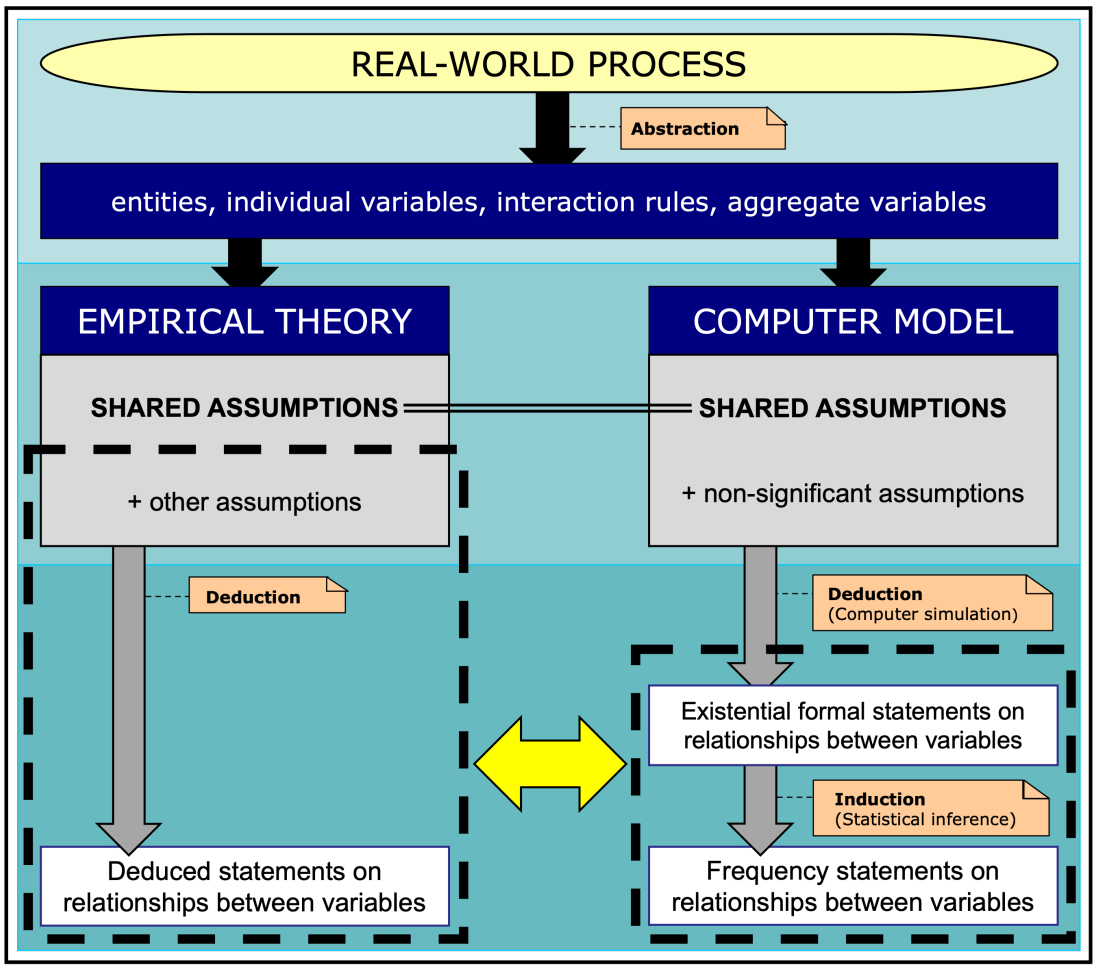

Figure 1: Sketch of how a computer model may disprove an empirical theory. The computer model contains a subset of the assumptions of the empirical theory plus, possibly, some non-significant assumptions (i.e., assumptions that do not affect the logical consequences of the computer model significantly). If we find a contradiction between (some of) the formal statements derived with the computer model and (some of) the assumptions or logical implications of the empirical theory, then we can say that the empirical theory is inconsistent.

1.7 We believe that many of the most celebrated models in the social simulation literature have played this refuting role, i.e., they were able to disprove a theory that was widely believed to be true at the time of publishing. To test this hypothesis, we examined several popular models in the social simulation literature and found that most of these models are most naturally interpreted, not as scientific theories that could be falsified by empirical observations, but rather as providers of compelling (computer-generated) logical statements that refuted a prevailing belief or assumption (see Appendix B for various examples). Rather than falsifiable theories, these models were refuting machines. And by refuting prevailing theories, these models greatly advanced science and, in many cases, they even opened a new field.

1.8 The remainder of the paper is organized as follows. In the next section, we present in broad terms the different types of statements that we may find in science (*empirical* vs. *formal*; and *existential* vs. *universal*), and we explain that simulation models can produce formal existential statements. In section 3, we discuss a key aspect of the scientific method: refutation by using a counter-example. Section 4 explains how formal models can indeed refute empirical theories by providing a counter-example that reveals an inconsistency within them. However, it should be noted that this type of refutation is in principle impossible if the empirical theory allows for exceptions. Thus, section 5 explains how formal models can refute, in practice, empirical theories that allow for exceptions. In section 6, we turn to empirical theories that are vaguely specified. Here, we argue that a necessary condition for a formal model to be able to refute a prevailing and vaguely specified empirical theory is that the significant assumptions in the formal model must be credible. In section 7, we note that most social simulation models that have played the role of refuting machines tend to be strikingly simple, and we offer an explanation within our framework of why that may be the case. The final section summarizes our view of social simulation models as refuting machines.

1.9 The paper has two appendices. Appendix A includes a NetLogo (Wilensky 1999) implementation of the Schelling-Sakoda model which can be run online, and Appendix B includes a review of some emblematic models to illustrate our view.

Simulation models as tools to produce formal existential statements

> *Computers are calculating machines and computer simulations are nothing but programed mathematical models that run on the computer. Therefore, computer simulations can just like models produce no other than purely inferential knowledge, that is, knowledge that follows deductively from the premises built into the simulation. In particular, computer simulations cannot produce genuine empirical knowledge like experiments or observations can.*
> – Arnold (2019, pp. 210-11)

Empirical vs formal statements. Existential vs universal statements

2.1 To fully understand the role of social simulation models in science, we have to understand that *statements* can be of different types. In particular, we have to distinguish between *empirical* and *formal* statements, and between *existential* and *universal* statements (Popper 2005, originally published in 1934).[2]

2.2 *Empirical* statements convey information about the real world, obtained from observation or experimentation. An example could be: 'There exists a black raven'. *Formal* statements, on the other hand, are expressed in a formal language devoid of meaning or context and say nothing about the real world. An example would be: '$1 + 1 = 2$'. Some formal statements –called axioms– are postulated as being true by assumption, while others are logically derived applying rules of inference to the axioms and to previously derived statements.

2.3 Independent of the distinction between empirical and formal statements, we can distinguish between *existential* and *universal* statements. *Existential* statements refer to one particular individual of a class (e.g., 'There exists a black raven'), while *universal* statements refer to all members of a class (e.g., 'All ravens are black'). More precisely, a universal statement refers to all members of a *universe*. In the raven example, the universe is composed of all current ravens. Using mathematical language, a universal statement can be written as follows:

$$\forall\, x \in \Omega : Cx$$

Here, the symbol $\forall$ means 'for all', C denotes a property, and Cx means that x possesses this property. If C means 'being black' and Ω is the universe of all ravens, then we have 'All ravens are black'. The statement is universal because it pertains to all members of a class.

2.4 In contrast, an existential statement can be written as follows:

$$\exists\, x \in \Omega : Cx$$

Here, the symbol $\exists$ means 'There exists'. Thus, the statement becomes 'There exists a raven which is black'. The statement is existential because it does not necessarily pertain to all members of a class, but at least to one.

2.5 Examples of the different combinations of formal/empirical and universal/existential statements are shown in Figure 2.

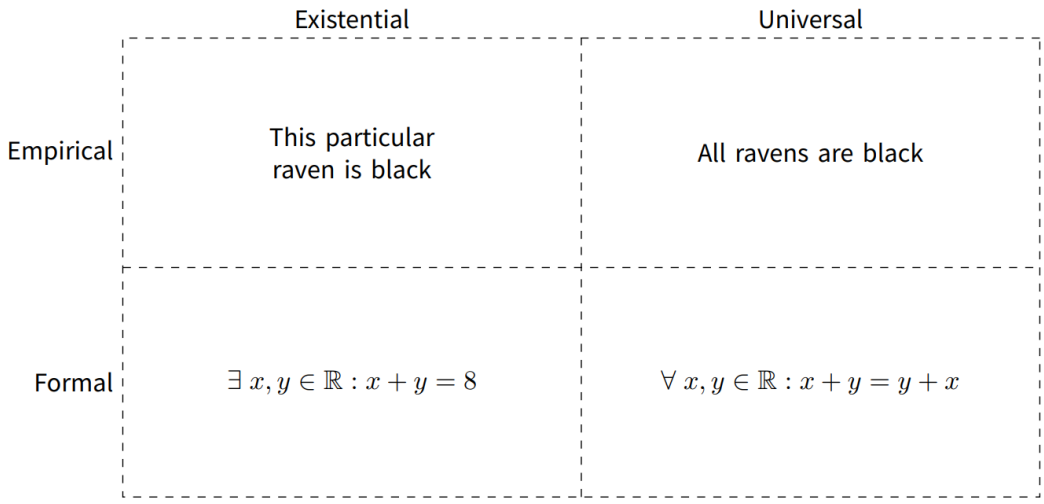

Figure 2: We can distinguish between *empirical* and *formal* statements, and between *universal* and *existential* statements.

2.6 What type of statements can social simulation models provide? It seems clear that they cannot produce genuinely new empirical statements because they are not part of the natural world (though they can generate empirical hypotheses). They are formal models –since they are expressed in a programming language and can be run on a computer (Suber 2007)– and, as such, can only provide formal statements. In this sense, they do not differ substantially from other formal models expressed in mathematical formalism. Indeed, social simulation models expressed in a programming language can be perfectly expressed using mathematics (Leombruni & Richiardi 2005; Epstein 2006a,b; Richiardi et al. 2006). This means that the exact same function that a social simulation model implements can be expressed in mathematical language, in the sense that both implementations would lead to exactly the same output if given the same input. Thus, the language in which a formal model is expressed is rather immaterial for our purposes.

2.7 In particular, the type of formal statement (existential or universal) that can be derived from a formal model does not depend on the language in which it is expressed. It depends, rather, on the approach followed to derive the statements.

2.8 In general, one can derive statements from formal models using two approaches: the mathematical analysis approach or the computer simulation approach (Izquierdo et al. 2013).[3] The *mathematical approach* to analyse formal models consists in examining the rules that define the model directly. Its aim is to deduce the logical implications of these rules for any particular instance to which they can be applied, i.e., for the whole domain of the model. Thus, using the mathematical approach we can derive universal statements about the formal model, i.e., statements that are true for all parameter values of the model (including initial conditions), or at least for all parameter values within a certain class.

2.9 However, most social simulation models are not analysed using mathematics, but rather following the so-called *computer simulation approach*. In contrast with mathematical analysis, the computer simulation approach does not consider the rules that define the formal model directly, but instead applies these rules to *particular* instances of the input space, by running the model on a computer (after having assigned a particular value to each parameter of the model). Each individual run of a model constitutes a formal *existential statement*, since it refers to one particular instance of the input space. In this sense, running a computer model once can be seen as the formal counterpart of conducting an empirical experiment or observation: observations provide direct information about empirical systems, while simulations provide direct information about computer models. Nonetheless, if a simulation model is based on an empirical theory and there is an equivalence between the assumptions of both systems (the theory and the computer model), then information about a computer model can provide information about the associated theory, as well as hypotheses about empirical systems.

2.10 Before presenting a more detailed example, let us introduce a formalism that is particularly useful to describe and analyse social simulation models, namely *time-homogeneous Markov chains*. Markov chains are stochastic processes that are useful to describe systems which evolve in discrete time steps and which, at each time step, find themselves in one of a set of possible states. The key property of a time-homogeneous Markov chain is that the probability that a system that is in state i at time step t will be in state j at the following time step $t+1$

depends exclusively on the states i and j (i.e., this probability does not depend on the particular time step t at which the transition takes place, or on the states that the system visited before reaching state i).

2.11 Given that most social simulation models are stochastic and run in discrete time steps, it is convenient to see them as time-homogeneous Markov chains (Izquierdo et al. 2009; Gintis 2013). When we run the model in a computer, we are effectively sampling one specific realisation of that stochastic process, i.e., we obtain just one sequence of states in time. The following section presents an example to clarify these arguments.

An example of a model and of formal statements derived with it

2.12 Consider the following model, which implements the main features from a family of models proposed by Sakoda (1949, 1971) and –independently– by Schelling (1969, 1971, 1978). Specifically, we focus on the model described in Schelling (1971, pp. 154-158). A NetLogo (Wilensky 1999) implementation of this model can be found (and run online) in Appendix A, together with some instructions on how to use it. Let us call this model **M**. The assumptions of **M** are (see Figure 3):

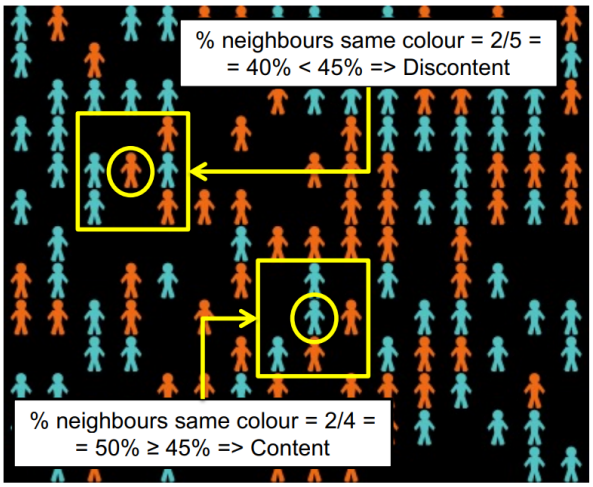

Figure 3: Illustration of the 13x16 grid of the Schelling-Sakoda model **M**, with $n = 138$ agents and intolerance threshold $\alpha = 45\%$.

- There is a 13x16 grid containing n agents. The number of agents n is assumed to be even. Half of the agents are blue and the other half orange. We assume $n \in [100, 200]$.
- Initially, agents are distributed at random in distinct grid cells.
- Agents may be content or discontent.
- Each individual agent is content if it has no Moore neighbours, or if at least $\alpha\%$ of its neighbours are of its same colour. Otherwise, the agent is discontent.
- In each iteration of the model, if there are any discontent agents, one of them is randomly selected to move to the closest empty cell where the moving agent would be content, if there is any available. If there is no such cell, the discontent agent will move to a random empty cell. If there is more than one closest cell where the agent would be content, one of them is chosen at random. We use taxicab (aka Manhattan) distance.

2.13 This is a stochastic model with two parameters: the number of agents, n, and the intolerance threshold α. Figure 3 shows a snapshot where $n = 138$ and $\alpha = 45\%$.[4] This value of α means that agents are content as long as 45% of their neighbours (or more) are of their same colour.

2.14 Model **M**(n, α) is completely specified and could be implemented in many different programming languages and also studied mathematically. To see it as a Markov chain, we can define the state of the system as a vector of dimension $13{*}16 = 208$, where each component corresponds to a certain cell in the grid. Every component of this vector can take 3 different values to denote whether the cell is empty, occupied by a blue agent or occupied

by an orange agent. With this definition, the number of possible states is $L = \binom{208}{\frac{n}{2}, \frac{n}{2}, 208-n}$; the two snapshots in Figure 4 show two different states for a model where $n = 138$ (which implies $L = 6.87382 \times 10^{96}$, a number greater than the estimated number of atoms in the observable universe).

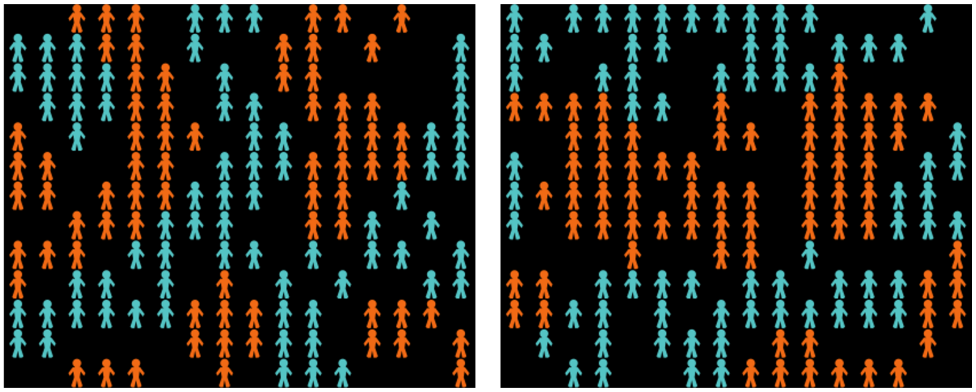

Figure 4: Illustration of two representative absorbing states of the Schelling-Sakoda model **M** with $n = 138$ and $\alpha = 45\%$.

2.15 What is interesting about the Schelling-Sakoda model $\mathbf{M}(n, \alpha)$ is that its dynamics most often lead to situations with clearly distinctive segregation patterns, even if agents are not particularly intolerant (i.e., $\alpha \geq 35\%$). Figure 4 shows two end states of $\mathbf{M}(n = 138, \alpha = 45\%)$, where every agent is content and therefore no more changes occur in the model.

2.16 To quantify the level of segregation, let us define –for each realisation of the model– the *long-run segregation index* w as the long-run average percentage of neighbours of the same colour across agents with at least one neighbour. Most realisations of the model reach an absorbing state, such as the two shown in Figure 4 for $\mathbf{M}(n = 138, \alpha = 45\%)$; in that case, the long-run segregation index is just the average percentage of neighbours of the same colour at that final state.

2.17 Thus, each simulation run i, parameterized with a fixed number of agents n_f and a fixed intolerance threshold α_f, will have a certain long-run segregation index w_i associated with it.[5] Note that two runs parameterized with the same values of n_f and α_f could perfectly have different values of the long-run segregation index, since each simulation run i involves a certain realisation of stochastic events ϕ_i –for a start, the initial distribution of agents in the grid, which is random, is most likely going to be different in each run. Thus, each run of the model i constitutes a formal *existential statement* of the following type:

$$\langle \mathbf{M}(n_f, \alpha_f), \phi_i \rangle \implies w = w_i,$$

where the arrow ($\implies$) denotes logical implication.[6] This is an existential statement in the sense that we are saying that *there exists* a realisation of model $\mathbf{M}(n_f, \alpha_f)$ for which the segregation index is w_i. To be clear, using Φ to denote the set of all possible realisations of a model, we could write the statement above using the existential quantifier $\exists$ as:

$$\exists\, \phi \in \Phi : \langle \mathbf{M}(n_f, \alpha_f), \phi \rangle \implies w = w_i.$$

2.18 By running many simulation runs of parameterized model $\mathbf{M}(n_f, \alpha_f)$, we can gather many formal statements of the type $\{\langle \mathbf{M}(n_f, \alpha_f), \phi_i \rangle \implies w = w_i\}$, where each value of i corresponds to a different simulation run. Each run would be showing us how model $\mathbf{M}(n_f, \alpha_f)$ *can* behave, while, by the law of large numbers, a sufficiently large set of simulation runs would show us how model $\mathbf{M}(n_f, \alpha_f)$ *usually* behaves.[7]

2.19 Note that the long-run segregation index of the stochastic Schelling-Sakoda model $\mathbf{M}(n_f, \alpha_f)$ –and any other well-defined statistic of the model– follows a certain probability distribution which, in principle, could be derived exactly. Let $\mathbb{W}(n_f, \alpha_f)$ denote this distribution. We do not know the exact expression of $\mathbb{W}(n_f, \alpha_f)$, but we know the distribution exists, and we can approximate it to any degree of confidence by running simulations. Figure 5 offers an approximation to $\mathbb{W}(n = 138, \alpha = 45\%)$, computed with 10^6 simulation runs, which we call $\hat{\mathbb{W}}(n = 138, \alpha = 45\%)$.

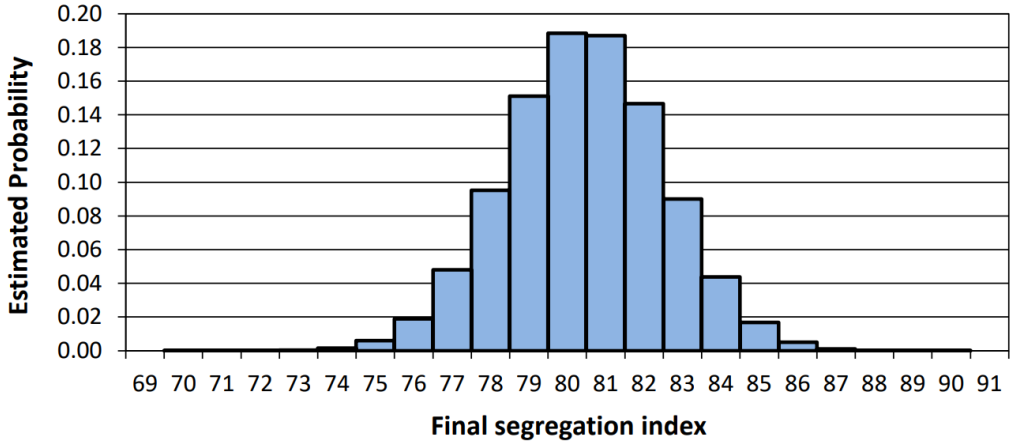

Figure 5: Estimated probability distribution of the long-run segregation index for a Schelling-Sakoda model **M** with $n = 138$ and $\alpha = 45\%$, computed running the model 10^6 times. All standard errors are below 10^{-3}.

2.20 Therefore, given some specific values n_f and α_f for parameters n and α respectively, we can establish two types of logical implications:

- Using the mathematical approach, in principle we could deduce the logical implication:

$$\mathbf{M}(n_f, \alpha_f) \implies w \sim \mathbb{W}(n_f, \alpha_f)$$

The problem here is that we may not be able to derive the exact distribution $\mathbb{W}(n_f, \alpha_f)$. In any case, note that this type of logical implication can lead to universal statements. For instance, if distribution $\mathbb{W}(n_f, \alpha_f)$ placed all its mass on values of w greater than 50%, then we could state that *all* realisations of $\mathbf{M}(n_f, \alpha_f)$ lead to segregation indices greater than 50%.

- Using the computer simulation approach, we can derive sets of formal existential statements such as the following:

$$\{\langle \mathbf{M}(n_f, \alpha_f), \phi_i \rangle \implies w = w_i\}_{i=1,\ldots,10^6}$$

We summarize the set of 10^6 implications above as:

$$\mathbf{M}(n_f, \alpha_f) \xrightarrow{10^6} w \sim \hat{\mathbb{W}}_{10^6}(n_f, \alpha_f),$$

where distribution $\hat{\mathbb{W}}_{10^6}(n_f, \alpha_f)$ is the finite-sample distribution corresponding to sample $\{w = w_i\}_{i=1,\ldots,10^6}$, and the single arrow $(\rightarrow)$ does not denote logical implication. Instead, $\mathbf{X} \xrightarrow{m} s$ denotes that statement s is true for a random sample of m simulation runs of model **X**. In particular, $\mathbf{X} \xrightarrow{m} y \sim \mathbb{Y}$ denotes 'Distribution $\mathbb{Y}$ for statistic y has been obtained by running m simulations of model **X**'; this is the type of 'frequency statement' that can be derived using the computer simulation approach (see Figure 1). Naturally, this type of relation $(\xrightarrow{m})$ is weaker than the logical implication $(\implies)$ derived with the mathematical approach.

Falsifications, disproofs and refutations

3.1 In the previous section we have seen that, following the computer simulation approach, we can use social simulation models to generate formal existential statements. Existential statements are useful because they can refute universal statements, both in the empirical realm (in which case we use the term 'falsification') and in the realm of formal statements (and then we use the term 'disproof').

3.2 As an example, consider the empirical statement 'A prokaryote has been found on Mars'. This existential statement is enough to falsify the universal empirical hypothesis 'There is no extra-terrestrial life'. An example within the realm of formal systems is the existential statement 'Number 3 is not even', which disproves the universal statement 'All integers are even'. A less obvious example –involving the use of computer simulation– is the

disproof of Polya's conjecture, which reads 'more than half of the natural numbers less than any given number have an odd number of prime factors'. This statement gained popularity over the years because it was unknown whether it was true or not, despite the fact that a single counter-example would have sufficed to settle the issue. Science had to wait... until computer simulations became available. Using this new approach, Haselgrove established in 1958 the existence of an integer that disproved Polya's conjecture, benefiting from the fact that '*now that electronic computers are available it is possible to calculate [...] over large ranges with considerable accuracy*' (Haselgrove 1958).

3.3 Another beautiful example was provided by Lander & Parkin (1966). This article is famous for being one of the shortest papers ever published, with only two sentences (Figure 6). Indeed, it can take just one (possibly very concise) formal existential statement to disprove a theoretical hypothesis. Mertens' conjecture is another case of a mathematical hypothesis disproved using computer simulation (Odlyzko et al. 1984).

COUNTEREXAMPLE TO EULER'S CONJECTURE ON SUMS OF LIKE POWERS

BY L. J. LANDER AND T. R. PARKIN

Communicated by J. D. Swift, June 27, 1966

A direct search on the CDC 6600 yielded

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5$$

as the smallest instance in which four fifth powers sum to a fifth power. This is a counterexample to a conjecture by Euler [1] that at least n nth powers are required to sum to an nth power, $n > 2$.

Reference

1. L. E. Dickson, *History of the theory of numbers*, Vol. 2, Chelsea, New York, 1952, p. 648.

Figure 6: Full paper by Lander & Parkin (1966), reproduced by permission of the American Mathematical Society.

3.4 To put all this in the context of social simulation, let us go back to our version of the Schelling-Sakoda model. Running this model once, we can obtain the following existential statement:

$$\exists\ \phi \in \Phi : \langle \mathbf{M}(n = 138, \alpha = 45\%), \phi \rangle \implies w = 82.16\%.$$

This statement can also be written as:

$$\mathbf{M}(n = 138, \alpha = 45\%) \xrightarrow{1} w = 82.16\%.$$

3.5 This existential statement can disprove statements such as 'The long-run segregation index of model $\mathbf{M}(n = 138, \alpha = 45\%)$ is necessarily less than 75%'. Such a universal statement refers to all possible realisations of the model and could be formalized as:

$$\forall\ \phi \in \Phi : \langle \mathbf{M}(n = 138, \alpha = 45\%), \phi \rangle \implies w < 75\%.$$

Another example of a (more general) universal statement that would be disproved by the same simulation run would be:

$$\forall n \in [100, 200] \text{ and } \forall \alpha \in [0, 50\%] \text{ and } \forall \phi \in \Phi,\ \ \mathbf{M}(n, \alpha) \implies w < 75\%.$$

The universal statement above includes all possible realisations of the model and also various combinations of parameter values.

How can formal statements refute empirical theories? A matter of consistency

> [Consistency] *can be regarded as the first of the requirements to be satisfied by* ***every*** *theoretical system, be it empirical or non-empirical.*
> – Popper (2005, originally published in 1934, section 24, p. 72)

4.1 It is clear that computer simulations can provide counter-examples to formal conjectures. But what about empirical conjectures? Very often, an empirical theory can be considered a story that captures, to some degree, the relationship between empirical entities by combining assumptions about these empirical entities and their interactions, and by showing some of the logical consequences of these assumptions. This last step, finding the logical consequences of some assumptions, is shared with computer simulations (see Figure 1).

4.2 Computer models can disprove empirical theories by showing that they are internally inconsistent, i.e., that they contain or imply statements that are logically incompatible. For instance, if a theory contains the statement 'all ravens are black', and it also contains the statement 'Austrian ravens are white', then the theory is inconsistent. Inconsistency is the worst flaw that a theory can have (Popper 2005, section 24): a theory with flawed reasoning may still produce conclusions that are true, but if a theory is inconsistent, then any conclusion and its negation can be logically deduced from the theory, making the theory totally useless.[8]

4.3 Crucially, to prove the inconsistency of an empirical theory, one does not necessarily require an empirical statement, a formal one may do just as well. Formal statements can refute empirical theories by uncovering inconsistencies in their logical foundations. This is the role that some simulation models have been able to play in the literature. They did not falsify any theory –since, being formal, that would be impossible–, but they refuted an empirical theory by revealing certain formal inconsistencies within it (see Figure 1). In layman terms, social simulations can provide counter-examples to formal preconceptions that lay the ground for an empirical explanation of the world.

4.4 As an example, consider an empirical theory which, implicit or explicitly, rests on the truthfulness of

Assumption 1:

'When individuals can choose their location freely, tolerant individual preferences for segregation lead to weakly segregated societies.'

Even though this assumption is (intentionally) expressed in vague terms, we believe that many scientists would concede that

Formal statement 1:

$$\mathbf{M}(n = 138, \alpha = 45\%) \xrightarrow{1} w = 82.16\%$$

4.5 constitutes evidence that seems to be at odds with *Assumption 1*, and it therefore challenges, at least to some extent, the consistency of any empirical theory that contains such an assumption. (Here, we are considering that being content in a neighbourhood where $(1 - \alpha) = 55\%$ of your neighbours are unlike you can be considered a 'tolerant preference', but some people may prefer the term 'mildly segregationist individual preferences'.)

4.6 However, there is a crucial condition that must be checked before we can say that *Formal statement 1* constitutes evidence that refutes *Assumption 1*: the significant assumptions used to produce *Formal statement 1* must be a subset of the assumptions implied in *Assumption 1* (i.e., tolerant individual preferences and free movement).

4.7 In general, if a computer model is to challenge the consistency of an empirical theory or belief, **the assumptions of the model responsible for the statements that challenge the empirical theory must be contained in the empirical theory**. Galán et al. (2009) define significant assumptions as those that are the cause of some significant result obtained when running the model. Using this terminology, our first principle states that the significant assumptions of the model must be contained in the empirical theory. In other words, any assumption of the computer model that is not shared with the empirical theory must be non-significant (see Figure 1).

4.8 Let us elaborate on this important issue. Note that the formal model may contain assumptions that are not present in the empirical theory it aims to refute. An example of such an assumption in our computer implementation of model **M** would be the use of floating-point arithmetic, instead of real arithmetic. Naturally, the

results obtained with the model can contradict the empirical theory only if they are a logical consequence of the assumptions that are also present in the empirical theory, and not of other auxiliary assumptions that are not in the empirical theory (such as the use of floating-point arithmetic in our example). Thus, in our example, we would have to make sure that the use of floating-point arithmetic is not driving the results.[9]

4.9 Interestingly, in this view, the model does not have to be a simplification or an abstraction of any real-world system. It may just be a description of some of the assumptions or beliefs used in the empirical theory, which, ideally, would be shown to contradict some other assumptions used by the theory, or some other statements that can be deduced from the theory. We believe that many well-known social simulation models are most naturally interpreted in this way, i.e., not necessarily as simplifications of any real-world system, but as counter-examples that *disprove* assumptions widely believed to be true. The following quote by Schelling about his own model offers some support for this interpretation:

4.10 "*What can we conclude from an exercise like this? We may at least be able to disprove a few notions that are themselves based on reasoning no more complicated than the checkerboard. Propositions beginning with 'It stands to reason that...' can sometimes be discredited by exceedingly simple demonstrations that, though perhaps true, they do not exactly 'stand to reason.' We can at least persuade ourselves that certain mechanisms could work, and that observable aggregate phenomena could be compatible with types of 'molecular movement' that do not closely resemble the aggregate outcomes that they determine.*" (Schelling 1978, p. 152)

Refuting empirical theories that allow for exceptions

5.1 We have seen that social simulations (or, more generally, formal statements) can reveal the inconsistency of a formal system *that forms the basis of* an empirical theory: *Assumption 1* above establishes that some aggregate event follows necessarily from certain individual behaviour –as a logical implication, not as an empirical fact–, but *Formal statement 1* is not consistent with *Assumption 1*.

5.2 Note, however, that social theories hardly ever rest on assumptions expressed in terms as definite as those in *Assumption 1*. It is far more common to encounter assumptions that allow for exceptions or are expressed in statistical terms, such as:

Assumption 2:

'**In most cases**, when individuals can choose their location freely, tolerant individual preferences for segregation lead to weakly segregated societies.'

5.3 Clearly, adding 'in most cases' changes the game completely. *Formal statement 1* above could be said to refute *Assumption 1* because both statements are incompatible, but it clearly does not refute *Assumption 2*. This second assumption leaves room for situations where tolerant preferences could indeed lead to distinctive patterns of segregation. Such situations may be exceptional, according to *Assumption 2*, but they can exist. Since *Assumption 2* does not rule out any specific event, it is clearly non-refutable (i.e., there is no existential statement that could ever contradict it).

5.4 A more formal example would be the following,

Probability statement 1:

$$\Pr\left(\mathbf{M}(n=138,\alpha=45\%) \xrightarrow{1} w<75\%\right)>0.9$$

5.5 In plain words, *Probability statement 1* says that, when running model $\mathbf{M}(n=138,\alpha=45\%)$, the probability of obtaining a segregation index w smaller than 75% is greater than 0.9. Since this is a probability statement that can accommodate any possible outcome, there is no single observation that could contradict it. The same issue arises when dealing with the (empirical) falsifiability of probability statements:

5.6 "*Probability hypotheses do not rule out anything observable; probability estimates cannot contradict, or be contradicted by, a basic statement; nor can they be contradicted by a conjunction of any finite number of basic statements; and accordingly not by any finite number of observations either.*" Popper (2005, originally published in 1934, section 65, p. 181)

5.7 Popper examined the problem of randomness at great length (Popper 2005, p. 133-208) and proposed a methodological rule according to which we could falsify probability statements in practice. The criterion is '*reproducibility at will*'. If we are able to show an effect that (i) is unlikely according to a probability statement and (ii) it is

'reproducible at will' (i.e., we can repeat it as many times as we like), we may consider the probability statement falsified in practice. The rationale is that the probability of observing a very large series of consecutive unlikely events is extremely unlikely (and the larger the series, the more unlikely it is). So, if we observe a series of events which, according to a certain probability statement, is extremely unlikely, then there is reason to doubt the truthfulness of the probability statement.

5.8 This type of reasoning is in line with Fisher's foundations for hypothesis testing and statistical significance,[10] and we can adopt the same 'statistical' approach for refuting (formal) assumptions via computer simulations.

5.9 "*[...] no isolated experiment, however significant in itself, can suffice for the experimental demonstration of any natural phenomenon; for the 'one chance in a million' will undoubtedly occur, with no less and no more than its appropriate frequency, however surprised we may be that it should occur to* ***us****. In order to assert that a natural phenomenon is experimentally demonstrable we need, not an isolated record, but a reliable method of procedure. In relation to the test of significance, we may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us a statistically significant result.*" (Fisher 1971, pp. 13-14, originally published in 1935)

5.10 Returning to our running example, it is clear that a single simulation run like *Formal statement 1*, i.e. $\mathbf{M}(n = 138, \alpha = 45\%) \xrightarrow{1} w = 82.16\%$, is not enough to refute *Probability statement 1*, but if, when running model $\mathbf{M}(n = 138, \alpha = 45\%)$, we systematically and consistently obtain evidence that is considered unlikely according to *Probability statement 1*, then there is much reason to doubt the truthfulness of *Probability statement 1*.

5.11 This is actually the case with model $\mathbf{M}(n = 138, \alpha = 45\%)$, since it rarely fails to give us the following result:

Reproducible-at-will statement 1:

$$\mathbf{M}(n = 138, \alpha = 45\%) \xrightarrow{1} w > 75\%$$

5.12 The reader is invited to check how 'reproducible at will' this statement is by running the online model provided in Appendix A.

5.13 The fact that we can reproduce the formal statement above 'at will', constitutes evidence strong enough to refute *Probability statement 1* in practice. But, is that enough to refute a more general (and vague) statement, such as *Assumption 2* (i.e., 'In most cases, when individuals can choose their location freely, tolerant individual preferences for segregation lead to weakly segregated societies')? We believe it is not, since *Reproducible-at-will statement 1*, which imposes $n = 138$ and $\alpha = 45\%$, seems too specific to contradict a statement that refers to 'most cases'. After all, the 'in most cases' in *Assumption 2* could well mean that it applies to all but a few special parameter combinations such as the one we have found.

5.14 A much more convincing case can be put forward noting that model $\mathbf{M}(n, \alpha)$ consistently leads to distinctive patterns of segregation not only in 13×16 artificial worlds populated by $n = 138$ agents with $\alpha = 45\%$ moving according to the '*closest-content-cell*' rule, but also for a wide range of different conditions. In particular, we consistently obtain strong patterns of segregation in any artificial world that is neither too sparse nor too crowded (i.e., for all $n \in [40\%, 90\%]$ of the number of cells in the world), populated by agents who have 'mildly segregationist preferences' (i.e., for all $\alpha \in [35, 50]$), and move according to different rules (such as '*random-content-cell*', '*random-cell*', or '*best-cell*' –see Appendix A; all these rules lead to even greater segregation). The reader can check that this is the case by running the online model provided in Appendix A. The model is also robust to various other changes in its assumptions, like e.g. the use of rectangular grid cells (Flache & Hegselmann 2001).[11]

5.15 The fact that model **M** and its multiple variations can consistently produce evidence against *Assumption 2* over such a wide range of situations suggests that *Assumption 2* (and any empirical theory that includes such an assumption) may not be adequate, and further investigation is needed.

5.16 In general, **we consider a formal model useful to refute an empirical theory that allows for exceptions if it covers a sufficiently important range of the situations to which the theory can be applied, and produces evidence that contradicts a logical assumption or deduced statement of the theory at will**. This second principle should be understood in relative terms, as it contains a number of graded concepts. Specifically, we do not see 'refutability', 'importance', or 'reproducibility as will' as all-or-nothing properties. Rather, we believe that these properties are a matter of degree. The model will be able to better refute the empirical theory (i) the greater the range, diversity and importance of the situations over which it offers contradicting evidence, and (ii) the more reproducible and compelling this contradicting evidence is.

5.17 In this regard, with hindsight, the refutation of an empirical theory is often best attributed not to one single model only, but also to the whole *set of models* that derive from the original model, complementing it by making slight variations on its assumptions, and proving that the refuting inference is robust to such variations.[12] This is Ylikoski & Aydinonat's (2014) *family of models thesis*:

5.18 "*The implication of these observations is that a more appropriate unit of analysis is the ongoing research initiated by Schelling's papers. It would be arbitrary to focus only on early models in this tradition or to limit one's attention only to models that have been proposed by Schelling himself. The research concerning Schelling's original insights is extensive both in a temporal and a social sense: checkerboard models have been studied for over forty years by a multitude of scholars in various disciplines. The research has produced many surprising findings that should not be read back to the original models. It is by analyzing the whole continuum of such models that one can make better sense of the real epistemic contribution of this family of models. This is our first cluster thesis: abstract models make better sense when understood as a family of related models, not as isolated representations.*" (Ylikoski & Aydinonat 2014, pp. 22-23)

How to refute unspecified empirical theories?

> *Schelling does not elaborate on what notions he has disproved. Possibly what he has in mind is the notion that either deliberate policy or the existence of strongly segregationist preferences is a necessary condition for the kind of racial segregation that is observed in American cities. His claim, then, is that he has discredited this notion by means of a counter-example.*
> – Sugden (2000, p. 10)

6.1 The point we want to make in this paper is that social simulation models can refute, and have been able to refute, empirical theories widely believed to be true by revealing certain formal inconsistency in them. However, the empirical theory to be refuted is hardly ever made explicit, so how can a model refute an empirical theory that is not even fully specified? In the following paragraph we argue that for a model to have a chance to refute a prevailing empirical theory, its significant assumptions (i.e., those that are driving the results) must be *credible*. For an assumption to be credible, it must be coherent with our background knowledge, with our intuition and with our experience (Sugden 2000, section 11).

6.2 The fact that an empirical theory, whatever it may be, is widely believed to be true implies that its assumptions must be credible, i.e., they must cohere with the background knowledge of the time.[13] On the other hand, if the model is to refute this theory on logical grounds, the assumptions in the model that are responsible for producing contradicting evidence (i.e., the significant assumptions) must be present in the empirical theory. Thus, **for a model to refute the empirical theory, the significant assumptions in the model must be credible**.

6.3 Let us discuss this last principle in the context of the Schelling-Sakoda model **M**. We saw that there are many assumptions in **M** that are not responsible for the emergence of segregation (e.g., the size of the world and the movement rule). Since these assumptions are not 'doing the work', there is no need to worry about them. In contrast, there are other assumptions that are crucial, like that α must be in the interval $[35, 50]$ –something that could be interpreted as agents having 'mildly segregationist preferences'. Since assuming this type of preferences is significant to obtain the results of the model, if the model is going to have some practical relevance, then we must make sure that such preferences are credible. In this regard, note that Schelling (1978, pp. 143-147) offers a number of stories whose "*[...] purpose, surely, is to persuade us of the credibility of the hypothesis that real people – it is hinted, people like us – have mildly segregationist preferences*" (Sugden 2000, p. 10). Even though Schelling does not make explicit the empirical theory he is refuting with his model, he makes sure that the assumptions in his model that are leading to the unexpected results seem credible. Otherwise, the model would not be telling us anything about the real world.

6.4 "*Furthermore, its implications have been buttressed by accumulated findings on preferences in multicultural settings which show that all major racial and ethnic groups hold preferences that are as strong as or stronger than the relatively mild preferences Schelling considered in his original two-group formulation.*" (Clark & Fossett 2008, p. 4109)

6.5 In Appendix B, we discuss several popular social simulation models which have been able to –at least partially– refute a prevailing belief:

- Schelling (1971) 'Dynamic models of segregation'

  Prevailing belief it refutes: Pronounced social segregation is the result of either deliberate public policy or strongly segregationist preferences.

- Epstein & Axtell (1996) 'Growing artificial societies: social science from the bottom up'

  Prevailing belief it refutes: To explain complex social phenomena such as trade or wealth distributions, one needs to consider complex patterns of individual behaviour and social interaction.

- Nowak & May (1992) 'Evolutionary games and spatial chaos'

  Prevailing belief it refutes: Sustaining cooperation in social dilemmas requires repetition and memory.

- Conway's 'Game of life' (Gardner 1970)

  Prevailing belief it refutes: Systems governed by a small number of very simple rules (for their individual agents and interactions) have a limit on the aggregate complexity that they can explain or generate.

- Gode & Sunder (1993) 'Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality'

  Prevailing belief it refutes: The efficiency of competitive markets rests on the rationality of economic agents.

- Reynolds (1987) 'Flocks, herds and schools: A distributed behavioral model'

  Prevailing belief it refutes: Complex phenomena such as the synchronized motion of a flock of birds require complex individual rules of coordination.

- Arthur (1989) 'Competing technologies, increasing returns, and lock-in by historical events'

  Prevailing belief it refutes: When there are several competing technologies, a market competition mechanism ensures that the most efficient technology will be the surviving one.

- Axelrod (1997b) model of 'dissemination of culture'

  Prevailing belief it refutes: Individuals' tendency to agree with social neighbours leads to global consensus in a society.

The value of toy models and emergence

> *When economic models are used in this way to explain casually observable features of the world, it is important that one be able to grasp the explanation. Simplicity, then, will be a highly desirable feature of such models. Complications to get as close as possible a fit to reality will be undesirable if they make the model less possible to grasp. Such complications may, moreover, be unnecessary, since the aspects of the world the model is used to explain are not precisely measured.*
> – Gibbard & Varian (1978, p. 672)

7.1 In our review of popular social simulation models, we have noticed that models (or families of models) that have successfully refuted a prevailing belief tend to be strikingly simple and idealized (see Appendix B). This kind of models are often called 'toy models' (Reutlinger et al. 2018). Is this a coincidence? Or is simplicity a desirable feature for models that are meant to refute an empirical theory? Here we explain why, *ceteris paribus*, toy models are in better conditions than complex and less idealized models to play the role of refuting machines.

7.2 First, note that the kind of refutation we are proposing here is by no means strict. It is a *subjective* matter of degree, influenced by the *perceived* credibility of the significant assumptions of the model, the *perceived* strength of the contradicting evidence they provide, and the *perceived* importance and variety of the situations to which they apply. Now, the simpler and more idealized the model, the easier it is to understand both its assumptions and how these assumptions lead to its logical implications.[14] This understanding enables the reader to gain intuition, or knowledge, about which assumptions are significant, and therefore should be perceived as credible. Simplicity also helps us to assess the extent to which the assumptions of the model collectively cover a sufficiently important range of situations over which the empirical theory can be applied. Without developing this intuition or knowledge, it would not be clear what sort of empirical theory the model is refuting, if any. Thus, to refute an empirical theory, simpler models seem to be favoured, even though they are not necessarily more general (see e.g. Evans et al. 2013; Edmonds 2018).[15]

7.3 Another feature shared by most refuting machines in the social simulation literature is that they show some kind of *emergent* phenomenon. This is the case, in particular, for all the models reviewed in Appendix B, and it makes plenty of sense in our framework. Let us consider why. Emergent phenomena are macroscopic patterns that

arise from the decentralized interactions of simpler individual components (Holland 1998). What characterizes emergent phenomena is that their appearance is not evident from a description of the system consisting of the specification of the behaviour of its individual components and the rules of interaction between them (Epstein 1999; Gilbert & Terna 2000; Gilbert 2002; Squazzoni 2008). In the most striking examples of emergent phenomena, rather complex macroscopic patterns arise from the interactions of very simple individual components. These emergent phenomena are, by definition for many scholars, surprising, i.e., counter-intuitive, i.e., they refute a prevailing belief. Thus, it is natural that models showing surprising emergent phenomena are likely to be able to play the role of a refuting machine.

7.4 A typical example of an emergent phenomenon is the formation of differentiated groups in the Schelling-Sakoda model; the emergence of clear segregation patterns is not explicitly imposed in the definition of the model, but emerges from the local interactions of individuals with surprisingly weak segregationist preferences.

Conclusions

8.1 In this paper we have argued that a prominent way in which social simulations can contribute (and have contributed) to the advance of science is by refuting some of our incorrect beliefs about how the real world works. More precisely, social simulations can produce counter-examples that reveal a logical inconsistency in a prevailing empirical theory. To do so, there are two conditions that the social simulation model should fulfil. Firstly, all the significant assumptions of the model (i.e., the assumptions that are leading to the contradicting results) must be included in the empirical theory, so they should be credible. Secondly, when trying to refute an empirical theory that allows for exceptions, the model should produce compelling evidence against the theory in a sufficiently important range of situations over which the empirical theory can be applied.

8.2 Naturally, many of the terms used in the two conditions above, such as 'refutability', 'credibility', or 'importance', should be understood as graded concepts that can be fulfilled to a partial extent. The model will be able to better refute the empirical theory (i) the more credible its significant assumptions are, (ii) the greater the range, diversity and importance of the situations over which it offers contradicting evidence, and (iii) the more reproducible and compelling this contradicting evidence is. This implies that refutation often builds up in time (as more assumptions are explored and more contradicting evidence is obtained), and it is best attributed to a whole family of models, rather than just to the first model that illustrated some counter-intuitive phenomenon.

8.3 We have also argued that simpler models are often in a better position to refute a prevailing belief, compared with more complex models. One reason for this is that in simple models it is easier to check the conditions that the model must fulfil so it can convincingly refute an empirical theory, i.e., the extent to which its significant assumptions are credible, and the extent to which they collectively cover a sufficiently important range of situations over which the empirical theory can be applied.

8.4 In the appendix, we show that many well-known models in the social simulation literature seem to have been able to –at least partially– refute a prevailing empirical theory, and they have done so with a strikingly simple model. In essence, these models are simple credible stories with an unexpected end.

Acknowledgements

Financial support from the Spanish State Research Agency (PID2020-118906GB-I00/AEI/10.13039/501100011033), from the Regional Government of Castilla y León and the EU-FEDER program (CLU-2019-04), from the Spanish Ministry of Science, Innovation and Universities, and from the Fulbright Program (PRX19/00113, PRX21/00295), is gratefully acknowledged. We are indebted to José M. Galán and José I. Santos, for countless wonderful and inspiring discussions, which have shaped and very much improved our understanding of the philosophy of modelling. We are also very grateful to two anonymous reviewers, whose critical comments significantly improved our exposition. Finally, Luis R. Izquierdo is grateful to the Center for Control, Dynamical Systems, and Computation at UC Santa Barbara, where part of this work was done, for their hospitality.

Appendix A: Online Schelling-Sakoda Model

Online model (Izquierdo et al. 2022) available at: `https://luis-r-izquierdo.github.io/schelling-sakoda-refuting-machine/`

Parameters

We use a green teletype font to denote `parameters` (i.e. variables that can be set by the user).

- Parameter `number-of-agents` is the number of agents n.
- Parameter `%-similar-wanted` is α, i.e. the minimum percentage of neighbours of the same colour needed to be content. An agent without neighbours is content.
- Parameter `movement-rule` determines how (discontent) agents move.
  - If `movement-rule` = *"closest-content-cell"*, the discontent agent will move to the closest empty cell in the grid where the moving agent will be content (with ties resolved at random), if there is any available. Otherwise it will move to a random empty cell. We use taxicab distance.
  - If `movement-rule` = *"random-content-cell"*, the discontent agent will move to a random empty cell in the grid where the moving agent will be content, if there is any available. Otherwise it will move to a random empty cell.
  - If `movement-rule` = *"random-cell"*, the discontent agent will move to a random empty cell in the grid.
  - If `movement-rule` = *"best-cell"*, the discontent agent will move to a random empty cell where the proportion of colour-like neighbours is maximal. If a cell has no neighbours, it is assumed that the proportion of colour-like neighbours is 1.

The value of parameters `%-similar-wanted` and `movement-rule` can be changed at runtime, with immediate effect on the dynamics of the model.

Buttons

- `setup`: Sets the model up, creating $\frac{\texttt{number-of-agents}}{2}$ blue agents and $\frac{\texttt{number-of-agents}}{2}$ orange agents at random locations.
- `go once`: Pressing this button will run the model one tick only.
- `go`: Pressing this button will run the model until this same button is pressed again.

Appendix B: Analysis of Models

Schelling's (1971) 'Dynamic models of segregation'

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'Either deliberate policy or the existence of strongly segregationist preferences is a necessary condition for the kind of racial segregation that is observed in American cities.' Sugden (2000, p. 10)

"*Schelling does not elaborate on what notions he has disproved. Possibly what he has in mind is the notion that either deliberate policy or the existence of strongly segregationist preferences is a necessary condition for the kind of racial segregation that is observed in American cities. His claim, then, is that he has discredited this notion by means of a counter-example.*" Sugden (2000, p. 10)

"*The model did not predict anything about the level of segregation, nor did it explain it. All it did was provide a counter-example to the current theories as to the cause of the segregation, showing that this was not necessarily the case.*" Edmonds et al. (2019, par. 6.6)

How is the prevailing belief refuted?

"*We might say that Schelling is presenting a critique of a commonly-held view that segregation must be the product either of deliberate public policy or of strongly segregationist preferences. The checkerboard model is a counter-example to these claims: it shows that segregation could arise without either of those factors being present.*" Sugden (2000, p. 9)

"*How can this counter intuitive result come about? Is it just an artifact of all the artificialities in the contrived model? Or does the model point to some fundamental flaw in our thinking about segregation? Modelers would rightfully claim the latter.*" Thompson & Derr (2009, par 1.7)

"*Schelling's (1971) segregation model is important not because it's right in all details (which it doesn't purport to be), and it's beautiful not because it's visually appealing (which it happens to be). It's important because—even though highly idealized—it offers a powerful and counter-intuitive insight.*" Epstein (2006a, pp. 65-66)

Did this paper open a new line of research?

"*We have also seen that Schelling's model has been explored in many ways in order to test the plausibility of its results. These explorations also show that Schelling's model opened up a new line of research that considers mildly discriminatory preferences as a possible cause of residential segregation.*" Aydinonat (2007, p. 446)

"*The Schelling segregation model has inspired a robust research stream where segregation mechanisms have been thoroughly explored, corroborating the evidence that formalized models encourage the cumulativeness of scientific progress.*" Squazzoni (2012, p. 210)

"*Schelling's model has become a classic reference in many (partially overlapping) scientific contexts: explanation of residential segregation, unintended consequences, micro-macro relations, clustering, attractors, social phase transitions, invisible-hand explanations, emergence of spontaneous order and structure. In philosophy of science Schelling's model is a (and often* the*) paradigmatic example for the study of mechanisms, or for reflections on the status of models more generally.*" Hegselmann (2017, p. 2)

Epstein & Axtell's (1996) 'Growing artificial societies: social science from the bottom up'

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'Complex phenomena such as migration, group formation, combat or trade are best explained as the result of coordinated interactions among highly informed and rational agents, and centralized organization'.

How is the prevailing belief refuted?

The authors design and simulate an artificial society of simple agents who live on a two-dimensional grid where there is a resource (sugar) that agents can harvest, consume, store and –in one of the chapters of the book– trade for a second commodity, namely spice. The rules that govern how agents move over this 'sugarscape' and interact (e.g. fighting, trading sugar for spice, and transmitting their 'culture' or their diseases) are local and extremely simple. Nonetheless, the interactions of the agents based on these local and simple rules give birth to very interesting patterns that resemble complex real-world phenomena, such as seasonal migrations, the formation of tribes, wars, and the presence of stable and highly-skewed wealth distributions.

"*Fundamental social structures and group behaviors emerge from the interaction of individual agents operating on artificial environments under rules that place only bounded demands on each agent's information and computational capacity. The shorthand for this is that we 'grow' the collective structures 'from the bottom up'.*" Epstein & Axtell (1996, p. 6).

"In short*, it is not the emergent macroscopic object per se that is surprising, but the generative sufficiency of the simple rules.*" Epstein & Axtell (1996, p. 52).

Did this book open a new line of research?

Epstein & Axtell's (1996) book can be seen as a pioneering, explicit and convincing proposal for a 'generative program for the social sciences' (Epstein & Axtell 1996, p. 177), an interdisciplinary research program –focused on providing *explanations*, rather than predictions– which has consistently grown since the publication of the book (see e.g. Epstein 2006a,b). This very journal is a testimony of this growth.

"*The aim is to provide initial microspecifications (initial agents, environments, and rules) that are sufficient to generate the macrostructures of interest. We consider a given macrostructure to be 'explained' by a given microspecification when the latter's generative sufficiency has been established. As suggested in Chapter I, we interpret the question, 'can you explain it?' as asking 'can you grow it?' In effect,* we are proposing a generative program for the social sciences and see the artificial society as its principal scientific instrument" Epstein & Axtell (1996, p. 177)

"*What constitutes an explanation of an observed social phenomenon? Perhaps one day people will interpret the question, 'Can you explain it?' as asking 'Can you* grow *it?' Artificial society modeling allows us to 'grow' social structures* in silico *demonstrating that certain sets of microspecifications are* sufficient to generate *the macrophenomena of interest [...]. We can, of course, use statistics to test the match between the true, observed, structures and the ones we grow. But the ability to grow them [...] is what is new. Indeed, it holds out the prospect of a new,* generative*, kind of social science*" Epstein & Axtell (1996, p. 20)

Nowak & May's (1992) 'Evolutionary games and spatial chaos'

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'The evolution of cooperation among self-interested individuals requires repeated interactions and some memory of these repeated interactions, i.e. individuals must be able to remember past encounters at least to some extent.'

"*The Prisoners' Dilemma is an interesting metaphor for the fundamental biological problem of how cooperative behaviour may evolve and be maintained. [...] Essentially all previous studies of the Prisoners' Dilemma are confined to individuals or organized groups who can remember past encounters, who have high probabilities of future encounters (with little discounting of future pay-offs), and who use these facts to underpin more-or-less elaborate strategies of cooperation or defection.*" Nowak & May (1992, p. 829)

How is the prevailing belief refuted?

Nowak & May (1992) study the evolution of a population of self-interested agents who play the Prisoner's Dilemma. Agents can only (unconditionally) cooperate or defect, i.e. they have no memory. If every agent in the population could interact with any other agent with the same probability (i.e. random matching, or well-mixed population), cooperation could not emerge in this setting. However, Nowak & May (1992) place their agents on a 2D grid, and let them interact only with their immediate neighbours. In this spatial context, they observe that (partial) cooperation could emerge and be sustained in their model.

"*In contrast, our models involve no memory and no strategies: the players are pure C or pure D. Deterministic interaction with immediate neighbours in a two-dimensional spatial lattice, with success (site, territory) going each generation to the local winner, is sufficient to generate astonishingly complex and spatially chaotic patterns in which cooperation and defection persist indefinitely.*" Nowak & May (1992, p. 829)

Did this paper open a new line of research?

Nowak & May's (1992) paper –which has been cited more than 4000 times, according to Google Scholar– opened a new line of research aimed at studying the influence of space –and, more generally, of population structure– on the evolution of cooperation. Nowadays we know that this influence turns out to be quite complex, as it generally depends on many factors that may seem insignificant at first sight, and whose effects interact in intricate ways (see Izquierdo et al. In progress, chapter 2). As Roca et al. (2009, pp. 14-15) eloquently put it: "*To conclude, we must recognize the strong dependence on details of evolutionary games on spatial networks. As a consequence, it does not seem plausible to expect general laws that could be applied in a wide range of practical settings. On the contrary, a close modeling including the kind of game, the evolutionary dynamics and the population structure of the concrete problem seems mandatory to reach sound and compelling conclusions.*"

Conway's 'Game of life' (Gardner 1970)

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'Life-like complexity cannot be generated by just a handful of extremely simple decentralized rules'

"*The game can also serve as a didactic analogy, used to convey the somewhat counter-intuitive notion that design and organization can spontaneously emerge in the absence of a designer.*" Wikipedia contributors (2022b)

How is the prevailing belief refuted?

John Conway presented a two-state, two-dimensional cellular automaton, whose rules are so simple that can be described in a few sentences. Each cell in the grid can be dead or live. Every cell simultaneously interacts with its eight Moore neighbours at discrete time steps. At each time step, a dead cell comes to life if it has exactly three live neighbours, and a live cell remains live only if two or three of its neighbours are live. All other cells go dead in each time step.

This stunningly simple setup gives rise to incredibly complex life-like patterns. Self-organization seems to emerge and generate an entire ecosystem of patterns. Many of these patterns have now names, such as *blinkers*, *oscillators*, *gliders*, or *pulsers*. Some patterns can produce other patterns and even self-replicate.

This model showed that very simple rules could generate a world whose complexity is yet to be fully understood and is somewhat reminiscent of the complexity of life itself.

"*Complexity arises from simplicity! That is such a revelation; we are used to the idea that anything complex must arise out of something more complex. Human brains design airplanes, not the other way around. Life shows us complex virtual 'organisms' arising out of the interaction of a few simple rules — so goodbye 'Intelligent Design'.*" Roberts (2020)

Did this model open a new line of research?

Conway was not the first researcher to study cellular automata, by any means. Cellular automata were discovered by John von Neumann and Stan Ulam in the 1940s. However, most cellular automata before Conway's were rather complicated as they had many states (e.g. von Neumann's universal constructor had 29 states). In contrast, Conway's automaton has only two states, which is a key factor to explain its popularity. Conway's automaton showed that *simplicity* can give birth to life-like complexity: to that effect, the presence of only two states (live and dead) was crucial. Thus, this model opened and popularized a line of research aimed at studying *simple* cellular automata.

One of the most renowned contributors to this line of research is Stephen Wolfram, who conducted a systematic study of all so-called *elementary cellular automata*, i.e. two-state, one-dimensional cellular automata where the rule to update the state of a cell depends only on its current state and on that of its two immediate neighbours (see Wolfram 1983, 2002). This work has shown that a wide range of complex behaviour –including chaos and computational universality (Cook 2004)– can be produced by extremely simple mechanisms.

"*[...] there were by the beginning of the 1980s various kinds of abstract systems whose rules were simple but which had nevertheless shown complex behavior, particularly in computer simulations. But usually this was considered largely a curiosity, and there was no particular sense that there might be a general phenomenon of complexity that could be of central interest, say in natural science. And indeed there remained an almost universal belief that to capture any complexity of real scientific relevance one must have a complex underlying model. My work on cellular automata in the early 1980s provided strong evidence, however, that complex behavior very much like what was seen in nature could in fact arise in a very general way from remarkably simple underlying rules.*" Wolfram (2002, p. 861)

Gode & Sunder's (1993) 'Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality'

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'The convergence of transaction prices to equilibrium levels and the high allocative efficiency observed in markets are both direct consequences of traders' cognitive abilities and their pursuit of profit.'

"*In the absence of these results, one might have attributed the high efficiency of the markets with human traders to their rationality, motivation, memory, or learning. Since our ZI traders, bereft of such faculties, exhibit comparable performance, the validity of such attribution is doubtful.*" Gode & Sunder (1993, p. 133)

"*Economic models assume utility-maximizing agents to derive market equilibria and their welfare implications. Since such maximization is not always consistent with direct observations of individual behavior, some social scientists doubt the validity of the market-level implications of models based on the maximization assumption. Our results suggest that such maximization at the individual level is unnecessary for the extraction of surplus in aggregate.*" Gode & Sunder (1993, p. 135-6)

How is the prevailing belief refuted?

Gode & Sunder (1993) implement a simulation model of a double-auction market populated by artificial Zero-Intelligence (ZI) traders who submit random bids and offers, with the only constraint that these bids and offers cannot imply a loss. Thus, ZI agents have no intelligence whatsoever, they do not maximize profits, they do not have memory and they do not learn. Gode & Sunder (1993) compare this artificial market of ZI agents with one populated by human traders over a series of 5 experiments. In both settings, they observe that transactions prices converge to equilibrium levels,[16] and that the aggregate allocative efficiency (i.e. the sum of producer and consumer surplus) is very high (almost 100%).

"*We show that a double auction, a non-Walrasian market mechanism, can sustain high levels of allocative efficiency even if agents do not maximize or seek profits.*" Gode & Sunder (1993, p. 120)

"*Our point is that imposing market discipline on random, unintelligent behavior is sufficient to raise the efficiency from the baseline level to almost 100 percent in a double auction. The effect of human motivations and cognitive abilities has a second-order magnitude at best.*" Gode & Sunder (1993, p. 133)

"*The primary cause of the high allocative efficiency of double auctions is the market discipline imposed on traders; learning, intelligence, or profit motivation is not necessary. The same market discipline also plays an important role in the convergence of transaction prices to equilibrium levels.*" Gode & Sunder (1993, p. 134)

"*[...] the convergence of price to equilibrium and the extraction of almost all the total surplus seem to be consequences of the double-auction rules.*" Gode & Sunder (1993, p. 135)

Did this paper open a new line of research?

Zero-Intelligence agents have been used in a variety of contexts to isolate the features of a situation that are due to agents' cognitive abilities or motivations. In this sense, they are very useful as a null model for agent behaviour, providing both a benchmark and a good starting point to identify the assumptions of a model that are responsible for certain observed properties.

"*This is part of a broader research program that might be somewhat humorously characterized as the 'low-intelligence' approach: we begin with minimally intelligent agents to get a good benchmark of the effect of market institutions and, once this benchmark is well understood, add more intelligence, moving toward market efficiency. We thus start from almost zero rationality and work our way up, in contrast to the canonical approach of starting from perfect rationality and working down*" Farmer et al. (2005, p. 2258-9).

Ladley (2012) reviews the Zero-Intelligence methodology for investigating markets and states that "*ZI models have proven one of the most successful applications of agent-based computational economics, with notable pieces of work appearing within high-quality economics (e.g.* The Journal of Political Economy*,* The Quarterly Journal of Economics*) and physical science journals (e.g.* Proceedings of the National Academy of Science*)*" Ladley (2012, p. 274).

Reynolds's (1987) 'Flocks, herds and schools: A distributed behavioral model'

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'The complex and synchronized motion of a flock of birds –and the way they avoid collisions between them and with other objects– must be the result of complex rules followed by the birds. Just a few simple local rules cannot generate the observed complex behaviour of the flock'.

How is the prevailing belief refuted?

Reynolds (1987) proposes a computer model to simulate the motion of flocks where each individual –called *boid*, for 'bird-oid'– follows just three simple local rules. In Reynolds's words,

- Collision Avoidance: avoid collisions with nearby flockmates
- Velocity Matching: attempt to match velocity with nearby flockmates
- Flock Centering: attempt to stay close to nearby flockmates

"*The model is based on simulating the behavior of each bird independently. Working independently, the birds try both to stick together and avoid collisions with one another and with other objects in their environment. The animations showing simulated flocks built from this model seem to correspond to the observer's intuitive notion of what constitutes 'flock-like motion.'*" Reynolds (1987, p. 33)

"*An interesting result of the experiments reported in this paper is that the aggregate motion that we intuitively recognize as 'flocking' (or schooling or herding) depends upon a limited, localized view of the world. The behaviors that make up the flocking model are stated in terms of 'nearby flockmates.'*" Reynolds (1987, p. 30)

The macro-behaviour of a flock of boids following these rules is very similar to the real motions of a flock of bird, making this model widely used in videogames and motion pictures. No centralized control or coordination, but just a few simple individual local rules are needed to reach this outcome.

"*The key point is that there is no choreographer and no leader. Order, organization, structure - these all* emerge *as by-products of rules which are obeyed* locally *and many times over, not globally.*" Dawkins (2009, p. 220)

"*The aggregate motion of the simulated flock is the result of the dense interaction of the relatively simple behaviors of the individual simulated birds.*" Reynolds (1987, p. 25)

"*[...] Boids is an example of emergent behavior; that is, the complexity of Boids arises from the interaction of individual agents (the boids, in this case) adhering to a set of simple rules. [...] Unexpected behaviours, such as splitting flocks and reuniting after avoiding obstacles, can be considered emergent. [...] At the time of proposal, Reynolds' approach represented a giant step forward compared to the traditional techniques used in computer animation for motion pictures.*" Wikipedia contributors (2022a)

Did this paper open a new line of research?

"*The classical paper* Flocks, herds and schools: a distributed behavioral model *of Reynolds was cited so many times and extended in so many different ways (see the next section for more details) that one could think that there is almost nothing that could be added.*" Hartman & Benes (2006, p. 199)

This paper also served as an inspiration to a new line of research called *particle swarm optimization*, which proposes algorithms based on flock or swarm behaviour to solve optimization problems. The foundational paper of this line of research is Kennedy & Eberhart (1995), who explicitly mention Reynolds (1987) in their introduction as a source of inspiration.

"*A number of scientists have created computer simulations of various interpretations of the movement of organisms in a bird flock or fish school. Notably, Reynolds (1987) and Heppner and Grenander (1990) presented simulations of bird flocking [...]. The particle swarm optimizer is probably best presented by explaining its conceptual development. As mentioned above, the algorithm began as a simulation of a simplified social milieu. Agents were thought of as collision-proof birds, and the original intent was to graphically simulate the graceful but unpredictable choreography of a bird flock.*" Kennedy & Eberhart (1995, pp. 1942-43)

Arthur's (1989) 'Competing technologies, increasing returns, and lock-in by historical events'

What is the prevailing belief (at the time of publication) that the model refutes?

Prevailing belief that the model refutes: 'In a situation where several competing technologies are available, a *laissez-faire* policy will ensure that the optimal technology will be eventually selected and prevail.'

"*The usual policy of letting the superior technology reveal itself in the outcome that dominates is appropriate in the constant and diminishing-returns cases. But in the increasing returns case laissez-faire gives no guarantee that the 'superior' technology (in the long-run sense) will be the one that survives.*" Arthur (1989, p. 127)

"*Where we observe the predominance of one technology or one economic outcome over its competitors we should thus be cautious of any exercise that seeks the means by which the winner's innate 'superiority' came to be translated into adoption.*" Arthur (1989, p. 127)

How is the prevailing belief refuted?

Arthur (1989) studies competing technologies under different regimes and shows that, in the presence of increasing returns, which technology is adopted in the end may well depend on seemingly insignificant events and on the order in which they occur. He illustrates these arguments with several dynamic models which –under increasing returns– show:

- *non-predictibility*, i.e. knowing the adopters' preferences and the technologies' possibilities is not enough to predict the final outcome.
- *potential inefficiency*, i.e. the selected technology may not be the one with the greatest long-run potential.
- *inflexibility*, i.e. no marginal adjustment to technologies' returns can alter the final outcome.
- *non-ergodicity*, i.e. seemingly accidental and insignificant events, and the order in which they occur, can determine the outcome.

"*Under increasing returns, [...] Insignificant circumstances become magnified by positive feedbacks to 'tip' the system into the actual outcome 'selected'. The small events of history become important.*" Arthur (1989, p. 127)

"*This paper has attempted to go beyond the usual static analysis of increasing-returns problems by examining the dynamical process that 'selects' an equilibrium from multiple candidates, by the interaction of economic forces and random 'historical events.' It shows how dynamically, increasing returns can cause the economy gradually to lock itself in to an outcome not necessarily superior to alternatives, not easily altered, and not entirely predictable in advance.*" Arthur (1989, p. 128)

Did this paper open a new line of research?

At the time of publishing of this paper, increasing returns were not very popular in Economics: "*As recently as the mid-1980s, many economists still regarded increasing returns with skepticism. In March 1987 I went to my old university, Berkeley, to have lunch with two of its most respected economists. What was I working on? Increasing returns. "Well, we know that increasing returns don't exist," said one. "Besides, if they do," said the other, "we couldn't allow them. Otherwise every two-bit industry in the country would be looking for a hand-out." I was surprised by these comments. Increasing returns* did *exist in the real economy, I believed. And while they might have unwelcome implications, that seemed no reason to ignore them.*" Arthur (1994, p. xi)

Arthur's work was certainly key in promoting an '*intense burst of activity in increasing returns economics*' (Arthur 1994, p. xi). The paper also contributed to the general use of dynamic models in Economics, by providing yet another example of a model where tipping effects, self-reinforcing mechanisms and out-of-equilibrium analyses provided neat insights. More generally, Arthur's work was part of a much greater research programme –fostered and nourished at the Santa Fe Institute– which saw '*the Economy as an Evolving Complex System*', and which would have a profound impact on the field of Economics in the years to come.

"*The increasing-returns world in economics is a world where dynamics, not statics, are natural; a world of evolution rather than equilibrium; a world of probability and chance events. Above all, it is a world of process and*

pattern-change. It is not an anomalous world, nor a miniscule one –a set of measure zero in the landscape of economics. It is a vast and exciting territory of its own. I hope the reader journies in this world with as much excitement and fascination as I have." Arthur (1994, p. xx)

Twenty years after the publication of Arthur's (1989) paper, Colander (2008) wrote in a review of the third volume of '*The Economy as an Evolving Complex System*', a book derived from the 2001 Santa Fe Institute Conference: "*Twenty years ago, contributions in this volume such as those by Samuel Bowles and Herbert Gintis on agents' genetic makeup being shaped by our environment, or Doyne Farmer et al.'s model of zero intelligence agents, would not have been seen as being even on the edge of mainstream economics. Today they are part of the mainstream conversation. The Santa Fe complexity work has played a part in making that happen.*" Colander (2008, p. 192)

Axelrod's (1997b) model of 'dissemination of culture'

What is the prevailing belief (at the time of publication) that the model refutes?

Axelrod (1997b) himself spells out four beliefs that his model refutes, at different levels:

- Prevailing belief that the model refutes: 'Intuition is a good guide for predicting the behaviour of simple dynamic models.'

  "*Perhaps the most important lesson of the social influence model is that intuition is not a very good guide for predicting what even a very simple dynamic model will produce.*" Axelrod (1997b, p. 219)

- Prevailing belief that the model refutes: 'If a practice followed by few people is lost, it surely means that the practice was inferior compared with other alternatives.'

  "*Thus the mere observation that a practice followed by few people was lost does not necessarily mean either that the practice had less intrinsic merit or that there was some advantage in following a more common practice.*" Axelrod (1997b, p. 220)

- Prevailing belief that the model refutes: 'The appearance of polarization requires some kind of divergent process.'

  "*Thus, when polarization is seen, it need not be due to any divergent process.*" Axelrod (1997b, p. 220)

- Prevailing belief that the model refutes: 'Observing correlation in cultural traits across geographic regions implies that these traits are intrinsically related in some way.'

  "*Likewise, when cultural traits are highly correlated in geographic regions, one should not assume that there is some natural way in which those particular traits go together.*" Axelrod (1997b, p. 220)

How is the prevailing belief refuted?

Axelrod proposes a simple model of social influence. In his model, the 'culture' of each agent (which is taken to be "*the set of individual attributes that are subject to social influence*") is formalized as a vector where each position of the vector denotes a specific cultural feature (e.g. language, religious belief, political views, etc.). Each agent will have a certain value, or 'trait', for each of the features (e.g. in the case of the feature 'language', possible traits could be 'English', 'Spanish', 'Chinese', etc.)

Axelrod studies a mechanism of social influence based on two simple principles: (i) agents are more likely to interact with those who are similar to them (i.e. those who share many of their cultural traits) and (ii) agents who interact become even more similar (the number of shared traits will increase after each interaction). Thus, greater similarity leads to more interactions, and more interactions lead to greater similarity.

Intuitively, one could think that this self-reinforcing dynamic would make all cultural differences eventually disappear, i.e. every agent will eventually have the same culture. However, it is often the case that different cultural regions emerge and coexist in the model. These different cultural regions are stable because members of adjacent regions have no traits in common, so they do not interact. Thus, Axelrod's model provides a striking example of cultural polarization emerging "*even though the only mechanism for change is one of convergence toward a neighbor*" (Axelrod 1997b, p. 220). Cultural polarization emerges *precisely* as a consequence of homophily (i.e. a preference to interact with similar people), because when we interact preferentially with similar people, we interact less with dissimilar people.

Axelrod considers this model an example of path dependence (Axelrod 1997b, footnote 8), since even if one had all the knowledge about the dynamics and the initial conditions of the model, the final number of cultural regions is very hard to predict.

Did this paper open a new line of research?

Axelrod's (1997b) model is considered one of the most influential models in the field of cultural dynamics:

"*A prominent role in the investigation of cultural dynamics has been introduced by a model by Axelrod (1997) that has attracted a lot of interest from both social scientists and physicists. The origin of its success among social scientists is in the inclusion of two mechanisms that are believed to be fundamental in the understanding of the dynamics of cultural assimilation and diversity: social influence and homophily. [...] From the point of view of statistical physicists, the Axelrod model is a simple and natural "vectorial" generalization of models of opinion dynamics that gives rise to a very rich and nontrivial phenomenology, with some genuinely novel behavior*" Castellano et al. (2009, p. 613)

"*Axelrod's (1997) model has had a significant impact in the scientific world, and several authors have analysed the model and some of its structural assumptions in depth.*" Izquierdo et al. (2009, Appendix B)

Notes

[1]We understand that social simulation models are formal entities created by researchers as they please and, as such, they belong to the universe of formal systems, not to the natural world (Rosen 2012, p. 45). There are, however, alternative ways of understanding the role of computer simulations in science (see e.g. Winsberg 2009; Barberousse & Vorms 2014; Parker 2017; Anzola 2021b; Alvarado 2022), which may be potentially more useful in other contexts.

[2]Popper (2005) actually uses 'logical' instead of 'formal'. 'Singular' can also be used instead of 'existential'.

[3]These two approaches are not mutually exclusive but complementary. There are plenty of synergies to be exploited by using the two approaches together (see e.g., Gotts et al. 2003; Seri 2016; García & van Veelen 2016, 2018; Hindersin et al. 2019; Izquierdo et al. 2019, In progress).

[4]In this model (which uses Moore neighbourhoods), setting $\alpha = 45\%$ is equivalent to imposing that every individual requires at least half of his neighbours to be like itself, which is the default case in Schelling (1971). This is so because the ratio $\frac{\#\text{alike neighbours}}{\#\text{total neighbours}}$ cannot take values in the interval $[0.44, 0.5)$.

[5]The segregation index w could be undefined if no agent has any neighbours in the final state of the simulation, but this is not a problem for our purposes.

[6]As Epstein (2006a, p. 56) clearly explains, 'every realization of an agent model is a *strict deduction*'.

[7]Seri & Secchi (2017) explain how power analysis can help to determine the appropriate number of runs one should conduct.

[8]According to Karl Popper, '*consistency is the most general requirement for a system, whether empirical or non-empirical, if it is to be of any use at all*' (Popper 2005, section 24, p. 72). Popper goes as far as stating that, for a theory, inconsistency is worse than being false (in the sense of having been falsified). Systems of statements known to be false can '*nevertheless yield results which are adequate for certain purposes*' (Popper 2005, section 24, p. 72). Popper cites Nernst's approximation for the equilibrium of gases as a case in point (the equation works well in some contexts, but fails to predict correctly the reality when the impact of ions must be taken into account.

Still, the equation remains useful, and when necessary, researchers switch to another equation such as the Goldman Equation.) Another clear example is Newton's law of universal gravitation, which is known to be inappropriate when dealing with very strong gravitational fields (i.e., there are situations where the theory can be falsified), but which is undoubtedly useful in many other contexts. False but coherent systems of statements can sometimes be useful, while inconsistent systems are entirely uninformative.

[9]Polhill et al. (2006) show a model, namely *CharityWorld*, where the use of floating-point arithmetic can significantly affect the results. See also Izquierdo & Polhill (2006).

[10]See e.g. Perezgonzalez (2015) and Szucs & Ioannidis (2017).

[11]Aydinonat (2007), Hatna & Benenson (2012), and Ylikoski & Aydinonat (2014) review more extensions to the model.

[12]In terms of Edmonds et al.'s (2019) modelling purposes, this may be interpreted as the evolution from *illustration* towards *theoretical exposition*.

[13]Background knowledge (of the time) is '*the complex of scientific theories generally accepted and well established at some stage in the history of science*' (Chalmers 1999, pp. 81-82).

[14]Reutlinger et al. (2018, pp. 1069-70) characterize toy models as models '*which the experts in a particular field of inquiry can cognitively grasp with ease*'.

[15]In a similar vein, Edmonds et al. (2019, appendix) identify the so-called KISS (Keep It Simple, Stupid) approach as one of the most important modelling strategies to design *illustrations* and *theoretical expositions*.

[16]Transactions prices converge to equilibrium levels more slowly in the artificial ZI setup.

References

Alvarado, R. (2022). Computer simulations as scientific instruments. *Foundations of Science*, *27*(3), 1183–1205

Anzola, D. (2019). Knowledge transfer in agent-based computational social science. *Studies in History and Philosophy of Science*, *77*, 29–38

Anzola, D. (2021a). Capturing the representational and the experimental in the modelling of artificial societies. *European Journal for Philosophy of Science*, *11*(3), 63

Anzola, D. (2021b). Social epistemology and validation in agent-based social simulation. *Philosophy & Technology*, *34*(4), 1333–1361

Arnold, E. (2014). What's wrong with social simulations? *The Monist*, *97*(3), 359–377

Arnold, E. (2019). Validation of computer simulations from a Kuhnian perspective. In C. Beisbart & N. J. Saam (Eds.), *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, (pp. 203–224). Cham: Springer International Publishing

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, *99*(394), 116–131

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. Ann Arbor, MI: University of Michigan Press

Axelrod, R. (1997a). Advancing the art of simulation in the social sciences. In R. Conte, R. Hegselmann & P. Terna (Eds.), *Simulating Social Phenomena*, (pp. 21–40). Berlin Heidelberg: Springer

Axelrod, R. (1997b). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, *41*(2), 203–226

Aydinonat, N. E. (2007). Models, conjectures and exploration: An analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology*, *14*(4), 429–454

Barberousse, A. & Vorms, M. (2014). About the warrants of computer-based empirical knowledge. *Synthese*, *191*(15), 3595–3620

Castellano, C., Fortunato, S. & Loreto, V. (2009). Statistical physics of social dynamics. *Review of Modern Physics*, *81*(2), 591–646

Chalmers, A. F. (1999). *What is This Thing Called Science?* New York, NY: Open University Press

Clark, W. A. V. & Fossett, M. (2008). Understanding the social context of the Schelling segregation model. *Proceedings of the National Academy of Sciences*, *105*(11), 4109–4114

Colander, D. (2008). Review of "The Economy as an Evolving Complex System III: Current Perspectives and Future Directions", Edited by Lawrence E. Blume and Steven N. Durlauf. *Economica*, *75*(297), 191–192

Cook, M. (2004). Universality in elementary cellular automata. *Complex Systems*, *15*(1), 1–40

Davidsson, P., Klügl, F. & Verhagen, H. (2017). Simulation of complex systems. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science*. Cham: Springer International Publishing

Dawkins, R. (2009). *The Greatest Show on Earth: The Evidence for Evolution*. New York, NY: Free Press

Edmonds, B. (2001). The use of models - Making MABS more informative. In S. Moss & P. Davidsson (Eds.), *Multi-Agent-Based Simulation*, vol. 1979, (pp. 15–32). Berlin Heidelberg: Springer

Edmonds, B. (2010). Bootstrapping knowledge about social phenomena using simulation models. *Journal of Artificial Societies and Social Simulation*, *13*(1), 8

Edmonds, B. (2018). A bad assumption: A simpler model is more general. Review of Artificial Societies and Social Simulation. Available at: `https://rofasss.org/2018/08/28/be-2/`

Edmonds, B. & Hales, D. (2005). Computational simulation as theoretical experiment. *The Journal of Mathematical Sociology*, *29*(3), 209–232

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

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, *4*(5), 41–60

Epstein, J. M. (2006a). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press

Epstein, J. M. (2006b). Remarks on the foundations of agent-based generative social science. In L. Tesfatsion & K. Judd (Eds.), *Handbook of Computational Economics*, vol. 2, (pp. 1585–1604). Amsterdam: North-Holland

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

Epstein, J. M. & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution Press

Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C. M., Merz, M., O'Malley, M. A., Orzack, S. H., Weisberg, M., Wilkinson, D. J., Wolkenhauer, O. & Benton, T. G. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution*, *28*(10), 578–583

Farmer, J. D., Patelli, P. & Zovko, I. I. (2005). The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences*, *102*(6), 2254–2259

Fisher, R. A. (1971). *The Design of Experiments*. London: Macmillan

Flache, A. & Hegselmann, R. (2001). Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. *Journal of Artificial Societies and Social Simulation*, *4*(4), 6

Galán, J. M., Izquierdo, L. R., Izquierdo, S. S., Santos, J. I., del Olmo, R., López-Paredes, A. & Edmonds, B. (2009). Errors and artefacts in Agent-Based modelling. *Journal of Artificial Societies and Social Simulation*, *12*(1), 1

García, J. & van Veelen, M. (2016). In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, *161*, 161–189

García, J. & van Veelen, M. (2018). No strategy can win in the repeated Prisoner's Dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI*, *5*, 102

Gardner, M. (1970). Mathematical games. The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, *223*(4), 120–123

Gibbard, A. & Varian, H. R. (1978). Economic models. *The Journal of Philosophy*, *75*(11), 664–677

Gilbert, N. (2002). Varieties of emergence. Proceedings of the Agent 2002 Conference on Social Agents: Ecology, Exchange, and Evolution

Gilbert, N. & Ahrweiler, P. (2009). The epistemologies of social simulation research. In F. Squazzoni (Ed.), *Epistemological Aspects of Computer Simulation in the Social Sciences*, vol. 5466, (pp. 12–28). Berlin Heidelberg: Springer-Verlag

Gilbert, N. & Terna, P. (2000). How to build and use agent-based models in social science. *Mind & Society*, *1*(1), 57–72

Gintis, H. (2013). Markov Models of social dynamics: Theory and applications. *ACM Transactions on Intelligent Systems and Technology*, *4*(3), 19

Gode, D. K. & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, *101*(1), 119–137

Gotts, N. M., Polhill, J. G. & Law, A. N. R. (2003). Agent-based simulation in the study of social dilemmas. *Artificial Intelligence Review*, *19*(1), 3–92

Hartman, C. & Benes, B. (2006). Autonomous boids. *Computer Animation and Virtual Worlds*, *17*(3–4), 199–206

Haselgrove, C. B. (1958). A disproof of a conjecture of Pólya. *Mathematika*, *5*(02), 141

Hatna, E. & Benenson, I. (2012). The Schelling Model of ethnic residential dynamics: Beyond the integrated - segregated dichotomy of patterns. *Journal of Artificial Societies and Social Simulation*, *15*(1), 6

Hegselmann, R. (2017). Thomas C. Schelling and James M. Sakoda: The Intellectual, Technical, and Social History of a Model. *Journal of Artificial Societies and Social Simulation*, *20*(3), 15. doi:10.18564/jasss.3511

Hindersin, L., Wu, B., Traulsen, A. & García, J. (2019). Computation and simulation of evolutionary game dynamics in finite populations. *Scientific Reports*, *9*(1), 6946

Holland, J. H. (1998). *Emergence. From Chaos to Order*. Reading, MA: Addison-Wesley

Izquierdo, L. R., Izquierdo, S. S., Galán, J. M. & Santos, J. I. (2009). Techniques to understand computer simulations: Markov chain analysis. *Journal of Artificial Societies and Social Simulation*, *12*(1), 6

Izquierdo, L. R., Izquierdo, S. S., Galán, J. M. & Santos, J. I. (2013). Combining mathematical and simulation approaches to understand the dynamics of computer models. In B. Edmonds & R. Meyer (Eds.), *Simulating Social Complexity*, (pp. 235–271). Berlin, Heidelberg: Springer

Izquierdo, L. R., Izquierdo, S. S., Galán, J. M., Santos, J. I. & Sandholm, W. H. (2022). Schelling-Sakoda model of spatial segregation. Software available at: `https://luis-r-izquierdo.github.io/schelling-sakoda-refuting-machine/`

Izquierdo, L. R., Izquierdo, S. S. & Sandholm, W. H. (2019). An introduction to ABED: Agent-based simulation of evolutionary game dynamics. *Games and Economic Behavior*, *118*, 434–462

Izquierdo, L. R., Izquierdo, S. S. & Sandholm, W. H. (In progress). *Agent-Based Evolutionary Game Dynamics*. Madison, WI: University of Wisconsin Pressbooks

Izquierdo, L. R. & Polhill, J. G. (2006). Is your model susceptible to floating-point errors? *Journal of Artificial Societies and Social Simulation*, *9*(4), 4

Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks

Ladley, D. (2012). Zero intelligence in economics and finance. *The Knowledge Engineering Review*, *27*(2), 273–286

Lander, L. J. & Parkin, T. R. (1966). Counterexample to Euler's conjecture on sums of like powers. *Bulletin of the American Mathematical Society*, *72*(6), 1079

Leombruni, R. & Richiardi, M. (2005). Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and its Applications*, *355*(1), 103–109

Nowak, M. A. & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*, 826–829

Odlyzko, A. M., te Riele, H. J. & Odlyzko, A. (1984). *Disproof of the Mertens Conjecture*. Centrum voor Wiskunde en Informatica

Parker, W. S. (2017). Computer simulation, measurement, and data assimilation. *The British Journal for the Philosophy of Science*, *68*(1), 273–304

Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology*, *6*, 2015

Polhill, J. G., Izquierdo, L. R. & Gotts, N. M. (2006). What every agent-based modeller should know about floating point arithmetic. *Environmental Modelling & Software*, *21*(3), 283–309

Popper, K. R. (2005). *The Logic of Scientific Discovery*. London: Routledge

Reutlinger, A., Hangleiter, D. & Hartmann, S. (2018). Understanding (with) toy models. *The British Journal for the Philosophy of Science*, *69*(4), 1069–1099

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. Computer Graphics (ACM SIGGRAPH '87 Conference Proceedings of the 14th annual conference on Computer graphics and interactive techniques)

Richiardi, M., Leombruni, R., Saam, N. & Sonnessa, M. (2006). A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation*, *9*(1), 15

Roberts, S. (2020). The lasting lessons of John Conway's Game of Life. The New York Times, December, 28. Available at: `https://www.nytimes.com/2020/12/28/science/math-conway-game-of-life.html`

Roca, C. P., Cuesta, J. A. & Sánchez, A. (2009). Effect of spatial structure on the evolution of cooperation. *Physical Review E*, *80*(4), 046106

Rosen, R. (2012). *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*. New York, NY: Springer New York

Sakoda, J. M. (1949). Minidoka: An analysis of changing patterns of social behavior. PhD Thesis, University of California

Sakoda, J. M. (1971). The checkerboard model of social interaction. *The Journal of Mathematical Sociology*, *1*(1), 119–132

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, *59*(2), 488–493

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, *1*(2), 143–186

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York, NY: Norton

Seri, R. (2016). Analytical approaches to agent-based models. In D. Secchi & M. Neumann (Eds.), *Agent-Based Simulation of Organizational Behavior*, (pp. 265–286). Cham: Springer International Publishing

Seri, R. & Secchi, D. (2017). How many times should one run a computational simulation? In B. Edmonds & R. Meyer (Eds.), *Simulating Social Complexity: A Handbook*, (pp. 229–251). Cham: Springer International Publishing

Squazzoni, F. (2008). The micro-macro link in social simulation. *Sociologica*, *1*, 2008

Squazzoni, F. (2012). *Agent-Based Computational Sociology*. Hoboken, NJ: John Wiley & Sons

Squazzoni, F., Jager, W. & Edmonds, B. (2014). Social simulation in the social sciences: A brief overview. *Social Science Computer Review*, *32*(3), 279–294

Suber, P. (2007). Formal systems and machines: An isomorphism. Available at: https://legacy.earlham.edu/ peters/courses/logsys/machines.htm

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*(1), 1–31

Szucs, D. & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience*, *11*, 2017

Thompson, N. S. & Derr, P. (2009). Contra Epstein, Good Explanations Predict. *Journal of Artificial Societies and Social Simulation*, *12*(1), 9. `https://www.jasss.org/12/1/9.html`

Troitzsch, K. (1997). Social science simulation - Origins, prospects, purposes. In R. Conte, R. Hegselmann & T. P. (Eds.), *Simulating Social Phenomena*, vol. 456, (pp. 41–54). Berlin Heidelberg: Springer

Wikipedia contributors (2022a). Boids. Available at: https://en.wikipedia.org/w/index.php?title=Boids&oldid=1094517009

Wikipedia contributors (2022b). Conway's Game of Life. Available at: https://en.wikipedia.org/w/index.php?title=Conway%27s_Game_of_Life&oldid=1102145663

Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. Available at: http://ccl.northwestern.edu/netlogo/

Winsberg, E. (2009). A tale of two methods. *Synthese*, *169*(3), 575–592

Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of Modern Physics*, *55*(3), 601

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media

Ylikoski, P. & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19–36