Models we Can Trust: Toward a Systematic Discipline of (Agent-Based) Model Interpretation and Validation

Blue Sky Track

Gabriel Istrate

West University of Timişoara and the e-Austria Research Institute

Timişoara, Romania

gabrielistrate@acm.org

ABSTRACT

We advocate the development of a discipline of interacting with and extracting information from models, both mathematical (e.g. game-theoretic ones) and computational (e.g. agent-based models). We outline some directions for the development of a such a discipline:
- the development of logical frameworks for the systematic formal specification of stylized facts and social mechanisms in (mathematical and computational) social science. Such frameworks would bring to attention new issues, such as *phase transitions*, i.e. dramatical changes in the validity of the stylized facts beyond some critical values in parameter space. We argue that such statements are useful for those logical frameworks describing properties of ABM.
- the adaptation of tools from the theory of reactive systems (such as bisimulation) to obtain practically relevant notions of two systems "having the same behavior".
- the systematic development of an adversarial theory of model perturbations, that investigates the robustness of conclusions derived from models of social behavior to variations in several features of the social dynamics. These may include: activation order, the underlying social network, individual agent behavior.

KEYWORDS

agent-based simulation; game-theoretic models; robustness; logical frameworks; adversarial perturbations; bisimulation

1 INTRODUCTION

Despite the recent surge in experimental studies of human behavior induced by the availability of (mostly online) social data, a large percentage of work in the mathematical and computational social sciences is still devoted to *theorizing*, that is building *models* of social phenomena, rather than analyzing social data. Whether mathematical or computational agent-based ones, a dizzying variety of models is proposed and analyzed in the scientific literature.

And yet, controversy (if not outright dissatisfaction) about the status and true meaning of such models, and of the modeling process itself, is prevalent throughout the social sciences. An example is

the vivid debate on the role of (mathematical) models in economics [102]. Economic models can be interpreted as "credible worlds" [101], analogies [46], (thought) experiments [72], parables/fables [16], intermediate byproducts of robustness analysis [65], or ludic devices similar to children's toys [105]. In any case, the discussion about the *robustness of scientific models*, currently taking place in the Philosophy of Science literature [109], is highly relevant.

A similar debate takes place in the social simulation literature. The critical research problem is that of *verifying* and *validating* agent-based models (ABM) [4, 10] (in short, the *V & V problem*). Theoretical frameworks have been proposed that attempt to deal with this issue, such as the *generative approach to social simulations* [34, 35], model calibration, *docking/alignment* [5, 32], *replication* [112], model-to-model analysis [51], etc. But there is no consensus on what verification and validation mean (see also [13, 66, 78, 113]).

There are multiple reasons that make the V&V problem important and difficult. A first reason is *scale*: whereas Schelling [91] could conceive his celebrated segregation model using pen and paper only, recent simulations models and projects aim to reach global dimensions [6, 7, 12, 36, 106]. A second reason has to deal with the potential *social consequences*: social simulations increasingly serve as consultants to (and implicitly affect) public policy [73, 74, 80]. A dramatic illustration of this fact in the context of the global pandemic of 2020 has been the controversy around the recommendations of the Imperial College epidemiological model [40]. This has led to significant discussion in the social simulation community, illustrated e.g. by the programmatic article [98] and the subsequent comments (e.g. [18, 24, 30, 31, 48, 99]). A final reason that makes the V&V question difficult is the very nature of simulation models, incomplete abstractions of reality, subject to complex behavior [88] that often involves multiple types of emergence [45].

It has been noted [70] that the proposals put forward in the ABM literature often have an ad-hoc nature, and that a more systematic theory is needed. *The goal of this paper is to advocate the use of logic and formal methods as useful tools for the systematic development of such theories.* We discuss a number of ways in which this may happen, and outline several research challenges associated with our proposals. The distinctive feature of the kind of frameworks we advocate is that **they require a highly unusual combination of two areas with very different languages: logic and formal methods [20], on one hand, sociological theory [21], on the other.** Importantly, the logical frameworks we envision should actively seek to avoid becoming what Edmonds [29] called the "philosophical approach" to logic. Instead, they should attempt to formalize genuine aspects of social theory (e.g. organizational logic,

see e.g. [27, 49]), help with addressing issues related to V&V, and serve as "middleware" to the agent-based simulations, helping in advancing conclusions that are robust and believable.

2 FORMAL TOOLS FOR SOCIAL THEORY

Is there any role for logical formalizations in describing and analyzing social dynamics, in ABM in particular ? This is a question that seems to have been asked so many times, with so many different interpretations in mind that a complete survey of this literature would not be particularly enlightening. Early on, Elster [33] argued that "logical theory can be applied not only in the formalization of knowledge already obtained by other means, but that logic can enter in the creative and constructive phase of scientific work" (op.cit. pp. 1). He explored the role of *quantified modal logic* in describing social reality, with a particular focus towards developing his method as an alternative to Hegelian dialectics. Closer to present Hannan et al. [54] (see also [84, 85]) proposed a rational reconstruction of social theory (organization science in particular) using techniques based on first-order predicate logic. Logical methods are, of course, well-established in economics. To give just one example, the so-called *interactive epistemology* program [3] is by now a classical part of theoretical economics, and a key ingredient of a recent proposal for a common foundation of all social sciences [47].

The use of logic-based methods would certainly **not** be controversial to a large part of the AAMAS audience: In fact, one could justifiably ask what is novel in such a proposal. After all, formal methods based on temporal logic are a particularly significant success story - techniques such as *model checking* [20] and *runtime verification* [8] lie behind eliminating errors in designing computer circuits, in writing software for technological artifacts (from remote controls and mobile devices to airplanes) or the Mars Rover [15]. Logical methods are widely used in in the area of multiagent systems [95, 114, 115]. Model checking techniques are useful in the verification of software agents [68, 69, 116] and auctions [104].

Yet, **the above optimism seems not to be shared by the practicing social simulation community.** The mentioned advances in *software agents* do not necessarily translate into corresponding advances on simulating *social agents* [28]. The techniques developed in the former literature rely too little on existing sociological knowledge, and address to an insufficient degree the concerns of social scientists. Unsurprisingly, they have been criticized (Edmonds [29], see also [26, 37, 44]) as "not useful given the state of MAS" and "not [...] useful in either understanding or building MAS".

We believe that logical methods can indeed help in increasing the reliability of conclusions derived from social simulations. However, to be useful, **such logics have to be tailored to the needs of the social scientist, not defined as an object of intrinsic mathematical interest**, and have properties that make them useful:
- the logics to be developed should be expressive enough to help formalize not only game-theoretic aspects of social theories (see e.g. [81, 83, 86, 107]) but also a variety of aspects of *sociological theory* [21, 56]. We give in the sequel two examples of concepts that we would like to see formalized: *stylized facts* and *social mechanisms.*
- the study of logical frameworks we propose **should be driven by considerations related to their implementation in (and applications to) ABM**. Their primary goal should **not** be that of

enabling deductive reasoning about social phenomena. Instead, they should be used to **formally specify** the observed social facts, in a way that enables *the construction of automated "monitors" serving as "middleware" between the social simulation and the decision support level by recognizing (and signaling) the emergence of the given fact in a given simulation run.* Our proposal is naturally related to the recent call for the development of live simulations [103], i.e. continuously feeding a simulation model with real-world data. In contrast, however, our proposal is related to (automatically) *extracting data from the simulation*, and using it to understand in a more systematic fashion the unraveling of the social dynamics.
- **it is not that important whether deciding implication in the new logics is tractable** (we can just run the simulation to see if a certain fact becomes true). However the *model checking* problem (given a description of the state of the world, is a stylized fact true in it ?) should have efficient algorithms (see also [52]).
- one problem of significant importance is the *monitoring question* for a logical formula ϕ: given a sequence of "states of the world" W_i (corresponding to a simulation run) and a statement ϕ, how do we efficiently detect that ϕ becomes true at some point in (W_i) ? This is a question pertaining to *runtime verification* [9], so we should use the inspiration from this literature but, given the rooting of the logical frameworks in social theory, it is likely that a simple adaptation of existing logics will not be enough.
- several new research topics, motivated by our vision of studying "robust" stylized facts observed from simulation runs, may gain preeminence. We give an example: the study of "continuity" properties of parameterized families of logical statements, as we vary the parameters of a given model. The opposite scenario, that of emergence of *critical points (phase transitions)* in the properties of social systems (and in their logical description) is also interesting.

2.1 Parameterized logics of stylized facts: "continuous statements" and "phase transitions"

A first application domain for the logics we envision is the formal specification of *stylized facts*. There is no agreement what a stylized fact is (however, see [75], as well as [41] for some relevant philosophical work). To advance a working definition, according to the former paper, at least in microeconomics, "stylized facts are currently understood as broad, but **robust enough** statistical properties pertaining to a certain economic phenomenon".

The requirement that stylized facts are robust is crucial in deciding what is and what is not a *useful* stylized fact: consider e.g. the following trivial baseline scenario (only important as a pedagogical example): Each of *n* agents may be in one of two states, *A* and *B*. Each agent prefers state *A* to *B*. Agents are scheduled at random; when scheduled, each agent changes its state according to the *best-response dynamics*, moving to the state that gives it the highest utility. Hence, when scheduled they will turn to state *A* (and subsequently stay that way, even if scheduled again).

An obvious conclusion about the dynamics, and a candidate stylized fact, could be the following: *eventually every agent will play strategy A*. This is not, however, a *robust* stylized fact. This can be seen by parameterizing the baseline model and modifying agent behavior: we will assume a single parameter $\epsilon \geq 0$. When

scheduled, an agent will choose A with probability $1 - \epsilon$ and B with probability ϵ. The baseline model corresponds to the case $\epsilon = 0$.

It is easy to see that the proposed stylized fact ceases to be true for $\epsilon > 0$, i.e. as soon as we move away from the baseline model. In other words, *the conclusion that **every** agent eventually holds state A is not robust to even the slightest variation in agents' choice probability*, hence **it cannot be considered as a (robust) stylized fact**. A more robust formulation is one that claims that agents' state converges to a *stationary distribution* with each agent independently being in A w.p. $1 - \epsilon$ and B w.p. ϵ. Note that:
- to formalize the robust version of this stylized fact we don't deal anymore with individual statements, but with *parameterized families* of logical statements. They encode a (single) social fact, expressed slightly differently across variations of the model.
- in a very well-defined intuitive sense **the baseline fact (all agents eventually adopt state A) is "the limit", as $\epsilon \to 0$ of the corresponding parameterized statements for $\epsilon > 0$.** Existing logical frameworks cannot, however, deal with such examples: while probabilistic/continuous logical frameworks (and their model checking) exist and might be useful in ABM [67, 87], and parametricity is important in such settings [2], *at the metalevel* logic is still largely a discrete framework, with no concept of "distance between statements", or "continuous limits of statements"
- in other scenarios the continuous behavior of stylized facts is no longer true. Instead, *social systems display **phase transitions***: abrupt changes in the validity of certain stylized facts beyond some critical value of a given parameter. While the study of phase transitions is a well established topic in Complex Systems and A.I. [19, 92], with phase transitions apearing even in settings relevant to model checking [77], the *logical* study of such "phase transitions" is still a relatively underdeveloped area. An exception is the topic of "zero-one laws" in the theory of random graphs [96]. There are many social phenomena where such concepts seem relevant. An example is the discussion about *tipping points*. Whether one talks about natural or social phenomena [17] there is a considerable interest in anticipating such tipping points [89]. In the theory of random graphs *the characterization of monotone properties that have "phase transitions" is fairly well understood:* such properties have a "global" nature, depending crucially on the presence of most of the edges of the network [42]. In contrast "local properties", e.g. the existence of a fixed subgraph, lack a phase transition [1]. The nature of logical theories in which one formulates the stylized facts also impacts the detection of tipping points: for instance, the emergence of the giant component in a random graph cannot be "sensed" by first-order logic [93]. Finally, *similar results exist in scenarios with a dynamical flavor:* start with an empty graph, add random edges, measuring the time when a certain graph property appears. It may be possible to extend such results to settings relevant to ABM:

CHALLENGE 1. *Develop a theory of logical frameworks that admit "parameterized statements", and study "phase transitions" in such statements. Ideally this study would yield algorithmic methods to anticipate "tipping points" in agent-based social simulations. Having such methods would operationalize the discussion about the robustness of stylized facts: to argue whether a given stylized fact holds in reality one could ask whether the parameters of the real world lie in the region of the parameter space where the stylized fact varies continuously.*

2.2 Formalizing social mechanisms

It is not only *(stylized) facts* that are in need of a logical formalization. After all, in a social simulation we are not interested in facts only, but in illuminating the *causal reasons* that lead to their emergence. Often (e.g. in the area of Analytical Sociology [56]) such causal explanations involve *social mechanisms* [25, 55, 57].

There is little consensus what a social mechanism is: Hedström ([55] pp. 25) compiles a list of seven definitions (due to Bunge, Craver, Elster, Hedström and Swedberg, Little and Stinchcombe). Of these seven, the most useful is due to Machamer et al. [71] (also [22, 23]). As paraphrased in [55] *"mechanisms can be said to consist of* entities *(with their properties) and the* activities *that these entities engage in, either by themselves or in concert with other entities. These activities bring about change [...]. A social mechanism, as here defined, describes a constellation of entities and activities that are organized such that they regularly bring about a certain type of outcome. We explain a social phenomenon by referring to the social mechanism by which such phenomena are regularly brought about"*

Social mechanisms are complemented by other approaches: Hedström lists *covering-law explanations* [58] and statistical explanations. These alternatives are not mutually exclusive: social mechanisms can, e.g., be sometimes inferred from statistical considerations; they can have themselves stochastic/statistical ingredients.

In any case, whatever social mechanisms are, they seem to have a complex structure: they can appear in *families* [90], can *concatenate* [43] and be *hierarchically nested* [22]. It seems, therefore, that:
- Verifying and validating social models (including simulation models) needs to address issues pertaining to explanation and causality. Statistical testing guidelines pertaining to replication, such as those discussed in [4], or generative explanations such as those proposed in [34, 35] are necessary but not sufficient. On the other hand social mechanisms, being in one acception "interpretations in term of individual behavior of a model that abstractly reproduces the phenomenon that needs explaining" [90] naturally complete and complement these methods (see also [53]).
- The role of social mechanisms in validating social models could be informally described as follows: simulations should reproduce known social mechanisms that are part of the expert knowledge in the area of concern and, of course, perhaps suggest new ones.
- In accord with [97], "formalizing models is a prerequisite to illuminate social mechanisms" and may help in making this notion precise. As a consequence we propose the following

CHALLENGE 2. *Give logical formalizations of the various notions of social mechanism in Analytical Sociology, and use these formalizations for the automatic recognition and inference of concrete social mechanisms in ABM runs.*

2.3 Towards a systematic theory of adversarial model perturbations

It is clear by now that some form of *robustness analysis* [109] is crucial to the verification and validation of social models. The concept has been heavily discussed in the Philosophy of Science literature, and can be applied to both mathematical models (e.g. the *robust Volterra principle* [111]) and to ABM (see [110]).

In contrast there is relatively little work on approaches to robustness with a practical potential: it is known, for instance, that

scheduling order can severely impact the conclusions derived from game-theoretic and related models [60, 108]: indeed, a rich literature on this topic has developed in the cellular automata community (e.g. [14, 38, 39]). A more general direction, the *adversarial scheduling* approach put forward in [64] (see also [61, 62, 94]) advocates the study of mathematical and computational models under generalized models of agent activation, as a way to increase the robustness of conclusions derived from these models. Paraphrasing [64], adversarial scheduling is specified by the following principles:

- *Start with a "base case" stylized fact P, valid under a particular (scheduling) model, often random.* Then *attempt to "break P" by creating adversarial schedulers under which P no longer holds true.*
- *Analyzing perhaps these examples, identify structural properties of the scheduling order that causally impact the validity of P. Use these insights to generalize P "from below" by identifying classes of schedulers (including the random one) under which P is valid.*
- *In the process* **we may need to reformulate the original statement in a way that makes it hold under larger classes of schedulers**, *thus making it more robust.*

As described above, adversarial scheduling is obviously important in increasing the robustness of conclusions drawn from mathematical models: But could something like this be systematically implemented, and be useful for (logic-based specifications of) social simulations as well ? We believe that the answer is positive, and are going to give a pedagogical example, using the baseline scenario from Section 2.1. Indeed, one can logically describe the candidate stylized fact in temporal logic as $(\forall i)\Diamond[State(i) = A]$ ("every agent will eventually hold state A"). Is such a statement true under adversarial scheduling ? The answer is clearly no: informally, an adversarial scheduler which never schedules a particular agent x whose state is B will preclude the system from reaching the state "all A". In other words, to ensure that the baseline stylized fact remains true under adversarial scheduling we need to require the scheduler to be *fair*. The random scheduler is fair (at least with probability $1 - o(1)$, as the number of steps tends to ∞).

Could have we reached the above conclusion about the necessity of fairness in scheduling in a logical framework ? The answer is yes. To do so we need to consider a simple logical description of the *the effect axioms* corresponding to the baseline dynamics:

$$Scheduled(i) \rightarrow \Box[State(i) = A]$$

("if an agent i is scheduled then globally (from now on) the agent will have state A"; we formulated our axiom this way in order to avoid having to deal in this pedagogical example with the frame problem). Can we derive the statement $(\forall i)\Diamond[State(i) = A]$, expressing the baseline stylized fact from the action axiom described above ? The answer is negative: to do so we would also need $(\forall i)\Diamond Scheduled(i)$ ("eventually every agent is scheduled"). However, **backward chaining [100] applied to this example would identify the statement expressing scheduler fairness as a necessary condition** for the validity of the stylized fact. This is evidence that adversarial scheduling might be feasible even for ABM, thus we propose the following:

CHALLENGE 3. *Extend the theory of adversarial scheduling to more central models of social dynamics, including ABM.*

Scheduling is not the only aspect of a mathematical or computational model which could be studied from an adversarial perspective. Many other aspects are susceptible of a similar treatment. For instance, in many game-theoretic and agent-based models the underlying dynamics takes place on a social network. One may vary this social network and attempt to understand the robustness of the baseline result to changes in the social network. The same can be done with (adversarial perturbations of) initial conditions. Some results in this direction have recently appeared [118].

CHALLENGE 4. *Develop a theory of adversarial perturbation of social networks and initial conditions for models of social dynamics. Extend it and apply it to ABM.*

2.4 When are two models "the same" ?

The verification and validation problem is related to the the question in the title of this section: when can we really consider two such models, perhaps with different ontologies (e.g. system dynamics and ABM) as "equivalent" ? Again, there is little agreement what a right answer may be. [32] argue that it is not enough to "eyeball" the outputs of the two models. One of more interesting attempts at an answer is [110] (Chapter 8), where model equivalence is formalized as a "weighted feature-matching" problem.

The theory of reactive systems [76] provides an elegant mathematical notion of system equivalence: in this setting, the equivalence of two reactive systems is formalized by the notion of *bisimulation*. A seminal theorem due to Hennesy and Milner [59] states that two bisimilar systems satisfy the same statements in a certain modal logic M, and conversely. That is, bisimilar systems satisfy the same set of "stylized facts" formalizable in M. As impressive as this result is, there is a wide gulf between such theory and the realities of ABM. There are multiple reasons that bisimulation is inadequate for social simulation. The most important one is that **bisimulation is too "microscopic":** it requires the fact that *every single move* of one of the system is enabled in the corresponding state of the second system. In contrast, cross-validation of ABM is coarser and often qualitative [79]. *In an ABM we don't mean to reproduce the actions of every agents: it's only* **macro patterns** *that we care about.*

There is some hope, though, that such methods are relevant to the study of ABM after all: recent results [50, 82, 117], some even from AAMAS [11] have related bisimulation to game-theoretic scenarios. It is thus reasonable to propose

CHALLENGE 5. *Develop a theory of (bi)simulation of (social) systems aligned with (an relevant to) the practice of V& V in ABM.*

3 CONCLUSION

We believe that logical formalization plays an important rule in assessing (and increasing) the reliability of results in social simulations. We have highlighted a couple of research directions that (if successful) would orient and ground the current discussion on model validity in (computational) social sciences. We don't believe that the directions we outlined are going to completely solve this problem. But, besides the obvious intellectual interest of developing such concepts, they may contribute to turning simulation and modeling in social settings from largely being an art (which still is now) to an engineering discipline.

REFERENCES

[1] N. Alon, P. Erdős, and J. Spencer. 1992. *The probabilistic method* (second ed.). John Wiley and Sons.

[2] Davide Ancona, Angelo Ferrando, and Viviana Mascardi. 2017. Parametric Runtime Verification of Multiagent Systems.. In *AAMAS*, Vol. 17. 1457–1459.

[3] R.J. Aumann. 2000. Collected Papers-Vol. 1. *M.I.T. Press* (2000).

[4] R. Axelrod. 1997. Advancing the art of simulation in the social sciences. *Complexity* 3, 2 (1997), 16–22.

[5] R. Axtell, R. Axelrod, J. Epstein, and M. Cohen. 1996. Aligning simulation models: a case study and results. *Computational and Mathematical Organization Theory* 1 (1996), 123–141.

[6] C. Barrett, K. Bisset, S. Chandan, J. Chen, Y. Chungbaek, S. Eubank, Y. Evrenosoğlu, B. Lewis, K. Lum, A. Marathe, et al. 2013. Planning and response in the aftermath of a large crisis: An agent-based informatics framework. In *2013 Winter Simulations Conference (WSC)*. IEEE, 1515–1526.

[7] C. Barrett, K. Bisset, S. Eubank, X. Feng, and M. V. Marathe. 2008. EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*. IEEE, 1–12.

[8] H. Barringer, A. Goldberg, K. Havelund, and K. Sen. 2004. Rule-based runtime verification. In *Verification, Model Checking, and Abstract Interpretation*. Springer, 277–306.

[9] E. Bartocci, Y. Falcone, A. Francalanza, and G. Reger. 2018. Introduction to runtime verification. In *Lectures on Runtime Verification*. Springer, 1–33.

[10] C. Beisbart and N. Saam. 2019. *Computer simulation validation*. Springer.

[11] F. Belardinelli, R. Condurache, C. Dima, W. Jamroga, and A. Jones. 2017. Bisimulations for verifying strategic abilities with an application to threeballot. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 1286–1295.

[12] S. Bishop, D. Helbing, P. Lukowicz, and R. Conte. 2011. FuturICT: FET Flagship Pilot Project.

[13] R. Boero and F. Squazzoni. 2005. Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science. *Journal of Artificial Societies and Social Simulation* 8, 4 (2005), 6. http://jasss.soc.surrey.ac.uk/8/4/6.html

[14] O. Bouré, N. Fates, and V. Chevrier. 2012. Probing robustness of cellular automata through variations of asynchronous updating. *Natural Computing* 11, 4 (2012), 553–564.

[15] G. Brat, D. Drusinsky, D. Giannakopoulou, A. Goldberg, K. Havelund, M. Lowry, C. Pasareanu, A. Venet, W. Visser, and R. Washington. 2004. Experimental evaluation of verification and validation tools on martian rover software. *Formal Methods in System Design* 25, 2 (2004), 167–198.

[16] N. Cartwright. 2010. Models: Parables v fables. In *Beyond Mimesis and Convention*. Springer, 19–31.

[17] D. Centola, J. Becker, D. Brackbill, and A. Baronchelli. 2018. Experimental evidence for tipping points in social convention. *Science* 360, 6393 (2018), 1116–1119.

[18] E. Chattoe-Brown. 2020. The Policy Context of Covid19 Agent-Based Modelling. *Review of Artificial Societies and Social Simulation* (2020).

[19] P. Cheeseman, B. Kanefsky, and W. Taylor. 1991. Where the really hard problems are. In *Proceedings of the 11th IJCAI*. 331–337.

[20] G. Clarke, O. Grumberg, and U. Peled. 1999. Model Checking. *M.I.T. Press* (1999).

[21] James Coleman. 1994. *Foundations of social theory*. Harvard University Press.

[22] Carl F. Craver. 2001. Role functions, mechanisms, and hierarchy. *Philosophy of Science* (2001), 53–74.

[23] Carl F. Craver. 2006. When mechanistic models explain. *Synthese* 153, 3 (2006), 355–376.

[24] C. de Matos Fernandes and M. Keijzer. 2020. No one can predict the future: More than a semantic dispute. *Review of Artificial Societies and Social Simulation* (2020).

[25] P. Demeulenaere. 2011. *Analytical sociology and social mechanisms*. Cambridge University Press.

[26] F. Dignum, B. Edmonds, and L. Sonenberg. 2004. The use of logic in agent-based social simulation. *Journal of Artificial Societies and Social Simulation* 7, 4 (2004).

[27] V. Dignum and F. Dignum. 2012. A logic of agent organizations. *Logic Journal of the IGPL* 20, 1 (2012), 283–316.

[28] V. Dignum and F. Dignum. 2020. Agents are Dead. Long live Agents!. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 1701–1705.

[29] B. Edmonds. 2004. How formal logic can fail to be useful for modelling or designing MAS. *Regulated Agent-Based Social Systems* (2004), 1–15.

[30] B. Edmonds. 2020. Basic Modelling Hygiene–keep descriptions about models and what they model clearly distinct. *Review of Artificial Societies and Social Simulation* (2020). https://rofasss.org/2020/05/22/modelling-hygiene/

[31] B. Edmonds. 2020. Good Modelling Takes a Lot of Time and Many Eyes. *Review of Artificial Societies and Social Simulation* (2020).

[32] B. Edmonds and D. Hales. 2003. Replication, replication and replication: Some hard lessons from model alignment. *Journal of Artificial Societies and Social Simulation* 6, 4 (2003).

[33] J. Elster. 1978. *Logic and society: Contradictions and possible worlds*. Wiley New York.

[34] J. Epstein. 1999. Agent-based Computational Models and Generative Social Science. *Complexity* 4, 5 (1999), 41–60.

[35] J. Epstein. 2007. *Generative Social Science: Studies in Agent-based Computational Modeling*. Princeton University Press.

[36] S. Eubank, H. Guclu, V.S. Anil Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. Monitoring and Mitigating SmallPox Epidemics: Strategies Drawn from a Census Data Instantiated Virtual City. *Nature* 429, 6988 (2004), 180–184.

[37] M. Fasli. 2004. Formal Systems and Agent-Based Social Simulation Equals Null? *Journal of Artificial Societies and Social Simulation* 7, 4 (2004).

[38] N. Fates and M. Morvan. 2004. Perturbing the topology of the game of life increases its robustness to asynchrony. In *International Conference on Cellular Automata*. Springer, 111–120.

[39] N. Fates and M. Morvan. 1997. An Experimental Study of Robustness to Asynchronism for Elementary Cellular Automata. *Complex Systems* 11 (1997), 1–1.

[40] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al. 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. (2020).

[41] Luciano Floridi. 2008. The method of levels of abstraction. *Minds and machines* 18, 3 (2008), 303–329.

[42] E. Friedgut. 1999. Necessary and Sufficient Conditions for Sharp Thresholds of Graph Properties, and the k-SAT problem. With an appendix by J. Bourgain. *Journal of the A.M.S.* 12 (1999), 1017–1054.

[43] D. Gambetta. 1998. Concatenations of mechanisms. *in* Social mechanisms: An analytical approach to social theory (1998), 102.

[44] B. Gaudou, A. Herzig, E. Lorini, and C. Sibertin-Blanc. 2011. How to do social simulation in logic: modelling the segregation game in a dynamic logic of assignments. *Proceedings of MABS* (2011), 24–35.

[45] N. Gilbert. 2002. Varieties of Emergence. In *Proceedings of the Agent'02 Conference: Social agents: ecology, exchange, and evolution*. Chicago, IL.

[46] I. Gilboa, A. Postlewaite, L. Samuelson, and D. Schmeidler. 2014. Economic models as analogies. *The Economic Journal* 124, 578 (2014), F513–F533.

[47] H. Gintis. 2009. *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton University Press.

[48] U. Gostoli and E. Silverman. 2020. Sound behavioural theories, not data, is what makes computational models useful. *Review of Artificial Societies and Social Simulation* (2020), 22.

[49] D. Grossi, F. Dignum, M. Dastani, and L. Royakkers. 2005. Foundations of organizational structures in multiagent systems. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. 690–697.

[50] J. Gutierrez, P. Harrenstein, G. Perelli, and M. Wooldridge. 2017. Nash Equilibrium and Bisimulation Invariance. In *28th International Conference on Concurrency Theory*. 17.

[51] D. Hales, J. Rouchier, and B. Edmonds. 2003. Model-to-model analysis. *Journal of Artificial Societies and Social Simulation* 6, 4 (2003).

[52] Joseph Y Halpern and Moshe Y Vardi. 1991. Model checking vs. theorem proving: a manifesto. *Artificial intelligence and mathematical theory of computation* 212 (1991), 151–176.

[53] L. Hamill. 2010. Agent-Based Modelling: The Next 15 Years. *Journal of Artificial Societies and Social Simulation* 13, 4 (2010), 7. http://jasss.soc.surrey.ac.uk/13/4/7.html

[54] M.T. Hannan, L. Pólos, and G. Carroll. 2007. *Logics of organization theory: Audiences, codes, and ecologies*. Princeton University Press.

[55] P. Hedström. 2005. *Dissecting the social: on the principles of analytical sociology*. Cambridge University Press.

[56] P. Hedström and P. Bearman. 2009. *The Oxford handbook of analytical sociology*. Oxford University Press, USA.

[57] P. Hedström and R. Swedberg (Eds.). 2006. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge University Press.

[58] C. Hempel. 1965. Aspects of scientific explanation. *Free Press* (1965).

[59] M. Hennessy and R. Milner. 1985. Algebraic laws for nondeterminism and concurrency. *Journal of the ACM (JACM)* 32, 1 (1985), 137–161.

[60] B. Huberman and N. Glance. 1993. Evolutionary games and computer simulations. *Proceedings of the National Academy of Science of the USA* 90 (1993), 7716–7718.

[61] Gabriel Istrate. 2018. Stochastic Stability in Schelling's Segregation Model with Markovian Asynchronous Update. In *International Conference on Cellular Automata*. Springer, 416–427.

[62] Gabriel Istrate, Madhav V. Marathe, and S. S. Ravi. 2008. Adversarial scheduling analysis of game-theoretic models of norm diffusion. *Logic and Theory of Algorithms* (2008), 273–282.

[63] Gabriel Istrate, Madhav V. Marathe, and S.S. Ravi. 2001. Adversarial Models in Evolutionary Game Dynamics. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms (SODA'01)*. AMS-SIAM.

[64] Gabriel Istrate, Madhav V. Marathe, and S.S. Ravi. 2012. Adversarial Scheduling Analysis of Discrete Models of Social Dynamics. *Mathematical Structures in Computer Science* 22, 5 (2012), 788–815. (journal version of [63]).

[65] J. Kuorikoski, A. Lehtinen, and C. Marchionni. 2010. Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science* 61, 3 (2010), 541–567.

[66] G. Küppers and J. Lenhard. 2005. Validation of Simulation: Patterns in the Social and Natural Sciences. *Journal of Artificial Societies and Social Simulation* 8, 4 (2005), 3. http://jasss.soc.surrey.ac.uk/8/4/3.html

[67] Marta Kwiatkowska, Gethin Norman, and David Parker. 2011. PRISM 4.0: Verification of probabilistic real-time systems. In *International conference on computer aided verification*. Springer, 585–591.

[68] A. Lomuscio, H. Qu, and F. Raimondi. 2009. MCMAS: A model checker for the verification of multi-agent systems. In *Computer Aided Verification*. Springer, 682–688.

[69] A. Lomuscio and F. Raimondi. 2006. Model checking knowledge, strategies, and games in multi-agent systems. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*. ACM, 161–168.

[70] I. Lorscheid, U. Berger, V. Grimm, and M. Meyer. 2019. From cases to general principles: A call for theory development through agent-based modeling. *Ecological Modelling* 393 (2019), 153–156.

[71] P. Machamer, L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67, 1 (2000), 1–25.

[72] U. Mäki. 2005. Models are experiments, experiments are models. *Journal of Economic Methodology* 12, 2 (2005), 303–315.

[73] M. Marathe and A. Kumar Vullikanti. 2013. Computational epidemiology. *Communications of the ACM* 56, 7 (2013), 88–96.

[74] W.L. Martinez. 2006. Modeling and Simulation for Defense and National Security. *Statistical methods in counterterrorism: game theory, modeling, syndromic surveillance, and biometric authentication* (2006).

[75] M. Meyer. 2019. How to Use and Derive Stylized Facts for Validating Simulation Models. In *Computer Simulation Validation*. Springer, 383–403.

[76] R Milner. 1989. *Communication and Concurrency*. Prentice-Hall, Inc.

[77] C. Moore, G. Istrate, D. Demopoulos, and M. Vardi. 2007. A continuous-discontinuous second-order transition in the satisfiability of a class of Horn formulas. *Random Structures and Algorithms* 31, 2 (2007), 173–185.

[78] S. Moss. 2008. Alternative Approaches to the Empirical Validation of Agent-Based Models. *Journal of Artificial Societies and Social Simulation* 11, 1 (2008), 5. http://jasss.soc.surrey.ac.uk/11/1/5.html

[79] S. Moss and B. Edmonds. 2005. Sociology and simulation: Statistical and qualitative cross-validation. *American journal of sociology* 110, 4 (2005), 1095–1131.

[80] National Institute of Health. 2010. NIH Study Models H1N1 Flu Spread. (2010). Press release http://www.nigms.nih.gov/News/Results/h1n1_092110.htm.

[81] Rohit Parikh. 2002. Social software. *Synthese* 132, 3 (2002), 187–211.

[82] Marc Pauly et al. 1999. Game constructions that are safe for bisimulation. *JFAK-Essays Dedicated to Johan van Benthem on the Occasion of His 50th Birthday* (1999).

[83] Marc Pauly and Mike Wooldridge. 2003. Logic for Mechanism Design: A Manifesto In *Proceedings of the 2003 Workshop on Game Theory and Decision Theory in Agent Systems (GTDT-2003), Melbourne, Australia*.

[84] G. Péli, J. Bruggeman, M. Masuch, and B.O. Nualláin. 1994. A logical approach to formalizing organizational ecology. *American Sociological Review* (1994), 571–593.

[85] G.L. Peli, L. Pólos, and M.T. Hannan. 2000. Back to inertia: Theoretical implications of alternative styles of logical formalization. *Sociological Theory* 18, 2 (2000), 195–215.

[86] Andrés Perea. 2012. *Epistemic game theory: reasoning and choice*. Cambridge University Press.

[87] André Platzer. 2012. Logics of dynamical systems. In *2012 27th Annual IEEE Symposium on Logic in Computer Science*. IEEE, 13–24.

[88] K. Sawyer. 2005. *Social Emergence*. Cambridge University Press.

[89] M. Scheffer. 2010. Foreseeing tipping points. *Nature* 467, 7314 (2010), 411–412.

[90] T. Schelling. 1998. Social mechanisms and social dynamics. *in* Social mechanisms: An analytical approach to social theory (1998), 32–44.

[91] T. Schelling. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* 1 (1971), 143–186.

[92] Bart Selman. 1995. Stochastic search and phase transitions: AI meets physics. In *IJCAI (1)*. 998–1002.

[93] S. Shelah and J. Spencer. 1994. Can you feel the double jump? *Random Structures & Algorithms* 5, 1 (1994), 191–204.

[94] S. Shirazipourazad, B. Bogard, H. Vachhani, A. Sen, and P. Horn. 2012. Influence propagation in adversarial setting: how to defeat competition with least amount of investment. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 585–594.

[95] Y. Shoham and K. Leyton-Brown. 2009. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

[96] J. Spencer. 2001. *The strange logic of random graphs*. Vol. 22. Springer Science & Business Media.

[97] F. Squazzoni. 2008. The micro-macro link in social simulation. *Sociologica* 1, 2 (2008).

[98] F. Squazzoni, J Gareth Polhill, B. Edmonds, P. Ahrweiler, P. Antosz, G. Scholz, É. Chappin, M. Borit, H. Verhagen, F. Giardini, et al. 2020. Computational models that matter during a global pandemic outbreak: A call to action. *Journal of Artificial Societies and Social Simulation* 23, 2 (2020).

[99] P. Steinmann, J. Wang, G. van Voorn, and J. Kwakkel. 2020. Don't try to predict COVID-19. If you must, use deep uncertainty methods. *Review of Artificial Societies and Social Simulation* (2020).

[100] S. Russell and P. Norvig. 2020. Artificial intelligence-a modern approach 4th ed.

[101] R. Sugden. 2000. Credible worlds: the status of theoretical models in economics. *Journal of economic methodology* 7, 1 (2000), 1–31.

[102] R. Sugden. 2009. Credible worlds, capacities and mechanisms. *Erkenntnis* 70, 1 (2009), 3–27.

[103] S. Swarup and H. Mortveit. 2020. Live Simulations. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 1721–1725.

[104] E. Tadjouddine, F. Guerin, and W. Vasconcelos. 2008. Abstractions for model-checking game-theoretic properties of auctions. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*. 1613–1616.

[105] A. Toon. 2012. *Models as make-believe: Imagination, fiction and scientific representation*. Springer.

[107] Johan Van Benthem. 2014. *Logic in games*. M.I.T. Press.

[106] TRANSIMS 2011. TRANSIMS web page. http://code.google.com/p/transims/.

[108] C. Weimer, J. Miller, R. Hill, and D. Hodson. 2019. Agent scheduling in opinion dynamics: A Taxonomy and comparison using generalized models. *Journal of Artificial Societies and Social Simulation* 22, 4 (2019).

[109] M. Weisberg. 2006. Robustness analysis. *Philosophy of Science* 73, 5 (2006), 730–742.

[110] M. Weisberg. 2012. *Simulation and similarity: Using models to understand the world*. Oxford University Press.

[111] M. Weisberg and K. Reisman. 2008. The robust Volterra principle. *Philosophy of science* 75, 1 (2008), 106–131.

[112] U. Wilensky and W. Rand. 2007. Making Models Match: Replicating an Agent-Based Model. *Journal of Artificial Societies and Social Simulation* 10, 4 (2007), 2. http://jasss.soc.surrey.ac.uk/10/4/2.html

[113] P. Windrum, G. Fagiolo, and A. Moneta. 2007. Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation* 10, 2 (2007), 8. http://jasss.soc.surrey.ac.uk/10/2/8.html

[114] M. Wooldridge. 2000. *Reasoning about rational agents*. MIT Press.

[115] M. Wooldridge. 2002. *An introduction to multiagent systems*. Wiley.

[116] M. Wooldridge, M. Fisher, M.P. Huget, and S. Parsons. 2002. Model checking multi-agent systems with MABLE. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM, 952–959.

[117] Michael Wooldridge, Giuseppe Perelli, Paul Harrenstein, and Julian Gutierrez. 2019. Nash Equilibrium and Bisimulation Invariance. *Logical Methods in Computer Science* 15 (2019).

[118] A. Zehmakan. 2021. Majority Opinion Diffusion in Social Networks: An Adversarial Approach. In *Proceedings of the 35th AAAI conference (to appear)*. available as arxiv.org preprint arXiv:2012.03143.