



---

Claudio Cioffi-Revilla (2010)

## A Methodology for Complex Social Simulations

*Journal of Artificial Societies and Social Simulation* **13** (1) 7

<<http://jasss.soc.surrey.ac.uk/13/1/7.html>>

Received: 13-Mar-2009 Accepted: 31-Aug-2009 Published: 31-Jan-2010

---



### Abstract

Social simulation—an emerging field of computational social science—has progressed from simple toy models to increasingly realistic models of complex social systems, such as agent-based models where heterogeneous agents interact with changing natural or artificial environments. These larger, multidisciplinary projects require a scientific research methodology distinct from, say, simpler social simulations with more limited scope, intentionally minimal complexity, and typically under a single investigator. This paper proposes a methodology for complex social simulations—particularly inter- and multi-disciplinary socio-natural systems with multi-level architecture—based on a succession of models akin to but distinct from the late Imre Lakatos' notion of a 'research programme'. The proposed methodology is illustrated through examples from the Mason-Smithsonian project on agent-based models of the rise and fall of polities in Inner Asia. While the proposed methodology requires further development, so far it has proven valuable for advancing the scientific objectives of the project and avoiding some pitfalls.

**Keywords:** Agent-Based Modeling Methodology, M2M, Social Simulation, Computational Social Science, Social Complexity, Inner Asia

---



### Introduction

**1.1** Simple social simulations, such as Conway's life model (Gardner 1970), Schelling's (1971) segregation model, Heatbugs (in Swarm, Netlogo, Repast, or MASON), or Sugarscape (Epstein and Axtell 1996; Bigbee et al. 2007), require relatively less methodological planning and developmental stages than more complex social simulations, such as large agent-based models of socio-natural or socio-technical systems. The target system of the latter class of social simulations typically comprises significantly greater social, natural, or technological complexity, multiple spatio-temporal scales, extensive validation (both internal and external), and the coordinated effort of teams of investigators from various relevant and diverse disciplines (e.g., Kohler and van der Leeuw 2007). To date, however, computational social scientists have mostly lacked a systematic methodology for developing complex social simulations, comparable to the

methodology for developing statistical or mathematical models in social science (e.g., Lave and March 1993). Such a systematic methodology would also contribute to the M2M (model-to-model) comparative literature (Hales et al. 2003; Rouchier et al. 2008).

- 1.2 With some rare exceptions, computational social science in general, and social simulations in particular, have not developed much interest in the methodology of research programs from a philosophy of science perspective. This is in contrast to mathematical social science in an earlier generation, which developed a significant interest on this same topic (e.g., Gillespie and Zinnes 1975; Cioffi-Revilla 1998; Cioffi-Revilla and Dacey 1988). However, as noted by Szmatka et al. (2002), some areas of computational social research have valued methodological concerns (e.g., Berger et al. 1972, 1993). Meeker's (2002) discussion of the methodology of "complex" social simulations (albeit less complex than the models considered in this paper) is constructive in the sense of highlighting significant challenges, but does not focus on agent-based modeling, an emerging framework for complex social simulations—together with social networks.
- 1.3 This article proposes a viable methodology for complex social simulations, specifically for large projects requiring multiple disciplines with the coordinated objective of modeling socio-natural systems with multi-level architecture. This same approach is also viable for other complex social simulations, such as computational models of socio-technical systems of coupled social-artificial-natural systems. The methodology is based on conceptualizing and developing a succession of models with increasing complexity as they approximate the target system, a procedure with foundations in but moving beyond the late Imre Lakatos' notion of a "research programme". The proposed methodology is grounded on specific features of the relevant class of complex social simulations, such as significant complexity (social and natural; structural and interactive), as well as multiple spatial and temporal scales. The methodology is illustrated with the Mason-Smithsonian project on developing agent-based models of the rise and fall of polities in Inner Asia (Cioffi-Revilla et al. 2007). While this methodology requires further development, at present it has proven valuable for advancing the scientific objectives of the project and for avoiding some pitfalls.



## **A Proposed Methodology for Complex Social Simulation**

- 2.1 A viable methodology for complex social simulation should be informed by the specific requirements of this type of modeling project, as well as by earlier valuable efforts. A viable methodology should also be explicit about its main operational features and guiding principles, communicating in a language that is as domain-neutral as possible in order to support and advance fruitful patterns of multi- and interdisciplinary collaborations that are typical in complex social simulations. Social scientists, computer scientists, and natural scientists need a common framework to collaborate successfully in these complex projects.
- 2.2 In this section I discuss general methodological requirements, Lakatos' legacy, and the proposed methodology.

### **General methodological requirements**

- 2.3 A general methodology for complex social simulations should provide systematic guidance for developing models—ideally from start to finish—and must, therefore, be informed by the main defining features exhibited by such a class of modeling projects. It is also valuable for a general methodology to indicate potential pitfalls and degenerative or unproductive directions to avoid.

As mentioned in the Introduction, complex agent-based simulation projects have a number of defining features that distinguish them from other social simulation projects. (Subsequently I identify additional features of social simulation—general social simulation characteristics—that are also useful for arriving at the general methodology being proposed. The three features in this section are more specific to complex social simulation projects.) More specifically:

1. *Socio-techno-natural complexity*. The target system is fundamentally complex, consisting of a social system interacting with some natural environment with physical and biological dynamics. Alternatively, the target system may also include technical systems, in addition to social and natural components. In other words, the ontology includes a variety of live and inert components: numerous components (*structural complexity*) consisting of people, beliefs, groups, history, as well as land, vegetation, water, and weather, with numerous interactions among relevant parts (*interaction complexity*). A social-artificial-natural system is more (normally far more) than the sum of its parts. By contrast, in simple social simulations (e.g., Schelling's segregation model), the structure and interactions lack this degree of social *and* natural complexity.
2. *Multi-scale complexity*. The target system of a complex social simulation also consists of multiple levels of social, technical, and natural organizational scales (*micro-macro emergence*), as well as multiple time scales (from the shortest to the longest time units in the simulation). In the real world (target system), agent decisions can occur within the range of milliseconds to seconds, whereas climate change and cultural variation can range from centuries to millennia. This yields, for each 1 KY (base time-scale), or  $10^3$  years,  $10^5$  days,  $10^6$  hours,  $10^8$  minutes,  $10^{10}$  seconds,  $10^{13}$  milliseconds, equivalent to a minimal range of 13 orders of magnitude in terms of time-scale. Simple social simulations typically lack this feature.
3. *Multi- and inter-disciplinary complexity*. As a corollary of the above requirements, the team of investigators in a typical complex simulation project is composed of scientific personnel from various disciplines, namely social science, computer science, and environmental science. In addition, the participation of applied mathematicians, statisticians, cartographers and GIS specialists, operations researchers, systems analysts, and engineers from various subfields (e.g., systems engineering or transportation) is not uncommon. (The grand total of disciplines in large projects—as in large archaeological field teams, or in space exploration programs—can sometimes approximate the size and complexity of a small college). This multicultural group must learn to communicate efficiently and effectively, which requires a great deal of translation, patience, and capacity for teamwork and related skills—both professionally and interpersonally—if the project is to succeed. Without minimizing the significance of this sociological aspect, the focus of this paper is methodological, so this sociological issue is not discussed as much as it would deserve, due to space limitations.

**2.4** These three defining features—which fundamentally stem from the very first, having to model a *complex socio-natural target system*—require a development program or methodology that aims toward the completion of a final model as the end result, but which cannot begin by immediately taking on the full complexity of the target system. Rather, the development program must start from some simpler model or models (building blocs) and then systematically proceed through a sequence of models until the final model is reached. A critical question is therefore: How should the investigators design and schedule such a developmental sequence, such that the research program will yield increasingly realistic and insightful models approximating the final target

system? A first step towards answering this and related questions can benefit from ideas contained in the late Imre Lakatos' classic account of Isaac Newton's theoretical research program.

### **Building on Lakatos and Newton**

- 2.5** Interestingly, the developmental process implied by the above features of a complex social simulation is akin to the process of developing scientific research programs, as described by the classic work of Imre Lakatos (1970). Lakatos' framework remains inspiring because, unlike other philosophers of science, his focus on long-term research projects is both innovative and unique. Based mostly on his account of the Newtonian research program, arguably the most prominent theoretical modeling effort in the history of science, Lakatos described the brilliance of Newton's success in terms of a *progressive sequence of theoretically driven mathematical models*—from an initial simple, single, spherical, and moonless Earth model ( $M_0$ ), to the final planetary system model ( $M_F$ ) comprising the Sun, multiple planets, moons, and elliptical orbits that approximated the target system (real planetary system) to the desired degree of fidelity.
- 2.6** The genius of Newton, based on a study of Lakatos' account and his methodological interpretation, was twofold. First, he identified an initial simple model that provided a fertile start for a rich sequence consisting of increasingly complex models that eventually encompassed the entire solar system (target system). Secondly, Newton followed a specific sequence of models—not just an arbitrary path—from his initial model to the final one. The path was jointly determined by theoretical interest and mathematical tractability, which at times required Newton to change hats and become a mathematician to create the necessary formal calculus for modeling continuous change. A final and critical feature of the Newtonian program was that Newton did not test his theory until he had reached a sufficiently complete model approximating the target system—the system for which he had empirical data. Importantly, Newton did not succumb to the temptation of testing his models before they had attained sufficient complexity—he understood that the simpler initial models would fail empirical tests. Interestingly, Newton did utilize selected empirical features of the real world as he developed his model sequence (e.g., planets are nearly spherical, not otherwise; orbits are elliptical, not circular; and so on). So,  $M_0$  can and indeed should have some empirical features. The methodology of social simulation should build on these and other successful lessons taken from the history of science, with appropriate adaptations and additions. Other important lessons from the history of science cannot be included here for obvious space limitations. They would include—among my personal favorites—the development of relativity theory, quantum mechanics, the theory of evolution, the mathematical theory of arms races (see further below and Meeker 2002), systems reliability theory, and computational linguistics. Lakatos also discussed the quantum theory of light as a successful theoretical research program. I chose the Newtonian program because it is more broadly known.
- 2.7** Students of Lakatos and philosophers of science are of course familiar with numerous other important features of "the methodology of research programmes", such as *progressive problemshifts* (the quantum passage from simpler to more complex models in the main sequence), *negative heuristics* (model development paths to avoid), *positive heuristics* (directions in which models are developed), and several others; features that set Lakatos' account (theoretical methodology) apart from others', including his teacher Karl Popper and Thomas Kuhn's popular sociological paradigm. These notions are important for supporting the main idea

regarding the progressive nature of a successful research program, the key element being the unique sequence of models from simple to complex.

## **Complex Social Simulation Research Programmes**

**2.8** Given these premises concerning some modeling requirements and epistemological antecedents, I turn to a positive and normative statement of the proposed methodology of complex social simulations. I begin by summarizing several defining features of a social simulation, in general, and continue with the proposed methodology as a sequence of computational models from initial to final.

### ***Defining features of a social simulation***

**2.9** All social simulations have a set of defining features, including their formal nature, their scientific motivation based on substantive research questions, and their experimental character, in addition to other special features of the class of complex social simulations discussed earlier.

**2.10** First, and perhaps most obvious, social simulations are *formal models* (Diesing 1971), except that they are expressed primarily in the language of computer code, not the mathematical calculus of Newton's theory (Taber and Timpone 1996), or other mathematical structures. Moreover, object-oriented models provide the most viable formal expression for social theories to date, because—unlike Newtonian mechanics—the main entities in the social world consist of individuals, thoughts, groups, and social relations, not variables or equations (Cioffi-Revilla 2009). Even game theory falls short of providing adequate formal support for modeling realistic multi-actor social dynamics, as witnessed by the incompleteness of  $N$ -person game theory and related features. Since social simulations are formal models, then their development is susceptible to similar considerations, rules, and challenges as those found in the more traditional domains of mathematical models in the social sciences (e.g., dynamical systems, game-theoretic models), with the additional requirements of the simulation form.

**2.11** Second, in addition to being a formal theory, a simulation model is also constructed for the purpose of answering a set of well-defined *core research question(s)* (Gilbert and Troitzsch 2005). There are obviously as many instances of research questions as there are target systems and interesting issues to investigate, but a clear and well-defined set of questions (which in some instances can be just a single question, not several) is always necessary. The core research questions of a social simulation also define the final model.

**2.12** By contrast, it is worth noting that the a priori identification of a specific research question in the case of a purely traditional mathematical model is often not an essential preliminary, because the nature of the mathematical formalism itself can often define such questions. For example, in the case of a social system formalized through differential equations (e.g., dynamical systems, a la Lotka-Volterra or comparable models), some natural questions include stability conditions, phase portraits, sensitivity to initial conditions, and other features for which the mathematical methodology is already established. Similarly, in the case of a social system formalized by a graph-theoretic model (e.g., models from social network analysis), naturally interesting questions include the formal properties of structural features (e.g., the distribution of degree and other network statistics).

**2.13** The situation is rather different for social simulation systems, such as agent-based models and

object-oriented simulations in general, because the formalism of objects and code offer a far greater spectrum for investigation and interesting research questions; so the computational formalism is by itself insufficient to define the most interesting or worthwhile research questions. Granted, some a priori research questions are (or should be!) common in many social simulations: Are the simulation results sensitive to initial conditions? What is the structure of the parameter space? Does interaction topology (e.g., von Neumann or Moore neighborhoods) make a difference? Does cell or patch geometry (triangular, square, hexagonal) matter? What is the long-term asymptotic equilibrium of the system, assuming there is one? However, these and similar questions are rather more generic and technically related to verification and validation, rather than being inspired by domain-specific and theoretically driven or substantive research questions. (For a discussion of these and other related methodological questions in computational social science and social simulation, see Cioffi-Revilla (2002).)

- 2.14** The core domain-specific questions addressed by complex social simulation models must be substantive in nature, not only technical or methodological. Substantive questions include, for instance, questions concerning the effect of reproductive rate of agents on overall population sizes; the effects of terrain or technology on social interactions; or the existence of special agent attributes (for instance, leadership skills or memory) on social organization and environmental impacts. Regardless of the specifics, every simulation model is constructed to answer one or more set of questions. And, as noted earlier, the formalization in computational objects and code does not always suggest such questions—beyond the generic technical questions already noted.
- 2.15** A third defining feature of complex social simulations is their *experimental* capacity, or at least their experimental potential given by their ability to answer "what if ... " questions *in silico*—especially concerning experiments that are not feasible in the real world, whether for ethical or other reasons. Accordingly, social simulations must be supported by a viable infrastructure for record-keeping for experiments, including not just results but also the exact procedures followed to obtain results. This is a demanding feature, as anyone with experience knows, because it can often require resources that are many times the resources needed to conduct the actual simulation runs. The design and maintenance of proper archival records is often uninteresting or unnecessary in traditional mathematical modeling projects, except perhaps for historical purposes. By contrast, it is essential for social simulations, even in the case of simple ones.
- 2.16** Based on these and related features, a viable methodology for developing a complex simulation model for addressing well-defined research questions is to begin with some simple initial model  $M_0$  and end with a final model  $M_F$ , with an appropriate (i.e., efficient) sequence of intermediate models connecting  $M_0$  to  $M_F$ . As mentioned earlier, the three critical issues are:
1. Which should be the starting simple model  $M_0$ ? This question concerns the subset of the target system to be chosen as the initial model's basic ontology.
  2. Which should be the final complex model  $M_F$ ? This question concerns the definition of the last model as it approximates the target system, given that a model of any kind (simulation, mathematical, or statistical) is always a simplification  $S$  of empirical reality  $R$ , such that  $S$  is a subset of  $R$ . The question is, which of the many possible kinds of representations should be chosen as  $M_F$ .
  3. Which should be the developmental sequence from initial model to final model? This question concerns all the intermediate steps along the modeling path from initial simplicity ( $M_0$ ) to greater complexity ( $M_F$ ).

The endpoints of the desired sequence are discussed next (2.17 and 2.20), followed by the process in between (2.24).

### ***The initial model $M_0$ : What are the core questions being answered?***

- 2.17** Every long journey begins with the first step, which must be taken in some direction—even if later such a direction requires correction through learning. Although first steps can be difficult, for many reasons, some viable heuristics for formulating an initial model include the following:
- Specify the basic questions being asked about the target system. These are the main questions that must be answered by the final model  $M_F$ . Complex social simulations rarely ask a single question, unlike simple social simulations; rather, they ask a set of related questions linked by relations and composing an overall unifying theme.
  - Identify a minimal set of fundamental or basic social entities (actors) and elementary environmental components in the target system and represent these as an initial simple world. The object-oriented framework of agent-based models facilitates this and other modeling steps.
  - Define an initial set of relations among model entities, such as agent-environment interactions and interactions among the agents themselves.
  - For each object in the simple world (agents and environment) select a small (minimal) set of meaningful attributes based on relevant social theory.
  - Similarly, select a minimal set of dynamics (operations, methods) that affect the various entity attributes, again based on theoretical considerations.
  - Develop new theory as necessary; object-oriented modeling often reveals gaps in the necessary social theory, especially for the specification of dynamics.
  - Refrain from testing this initial model against real world data, since it will obviously fail.
  - Decide if users other than researchers will use the model (e.g., policy analysts), and if so decide if the initial model needs to be designed accordingly or if end-user considerations can wait. End-user participation in model development is highly recommended (Moss 2008).
- 2.18** Selection of an initial model cannot be arbitrary, but neither can it be allowed to stifle initial model development by excessive discussions on finding "the right" initial model. Better select an initial simple model that is reasonable, tractable, and susceptible of further development than trying to find a perfect solution. From a social perspective, leadership on the part of the principal investigator(s) is almost always necessary to overcome disagreements, centrifugal discussions, and other inconclusive exchanges that consume valuable time and valuable resources (tangible and intangible).
- 2.19** Realism or high empirical fidelity is not a goal in an initial model  $M_0$ , but rather capturing the fundamental elementary structure and dynamics of the target system. Although this initial model should not be tested against data, at the same time it makes no sense to begin with a distorted caricature of the target system instead of using more representative entities and relations. Some significant features of the target system should already be chosen for  $M_0$ , just like Newton began with spheres to represent planets (not point masses or cubes), and ellipses for orbits (not circles).

## ***The final model $M_F$ : What is the target system to be modeled?***

- 2.20** As pointed out earlier, the final model should be defined by the set of questions the simulation model is meant to address. In other words, it should be sufficiently rich or capable of answering the intended research questions, possibly with some additional margin to suggest further investigations that might prove potentially fruitful. Hence, the definition of domain specific research questions is clearly seen as the central and theoretically driven task, not instrumental or purely technical.
- 2.21** Some heuristics for defining a final model  $M_F$  include:
- Ask which of the central questions of the project are sufficiently well answered by the model in question.
  - Ask whether end-users other than researchers will be satisfied from their perspective.
  - Conduct sufficiently extensive tests of external validity (calibration and validation), such as testing for fit (qualitative and quantitative) between simulated data and target data.
  - Conduct sensitivity analysis to assess model robustness and other desirable properties, especially in the case of policy-related projects.
  - Compare the theoretical process explanation of the simulation to other extant explanations in relevant social theories.
  - Identify new features or discoveries uncovered by the simulation and conduct empirical tests.
  - Identify target phenomena omitted from the simulation and decide whether they should be included.
- 2.22** Focusing on purely technical questions to identify a final model is insufficient. Questions of sensitivity, interaction topology, asymptotic equilibria, and others cannot determine what should be included in a final model because such questions are unhelpful for bringing closure and focus—they leave too many options wide open. Instead, well-defined research questions, such as the core puzzles that motivate the project, go beyond technical questions and assist in the selection of the final model to approximate the desired target system.
- 2.23** I now turn to what lies between the endpoints of a theoretical research program: the progressive models that come after  $M_0$  and before  $M_F$ .

## ***The developmental sequence from $M_0$ to $M_F$***

- 2.24** The models that lie between initial and final simulations display features of the final model minus the initial model. Accordingly, such features should be well understood and their systematic introduction should occur in an order that maintains the final goal in sight but is also informed by the natural complexity of the social, technical, and/or natural phenomena being modeled. Given the nature of complex social simulations, the second model after the very first can add complexity along social, technical, or natural directions.
- 2.25** The second model in the sequence could be a further development of the natural environmental dynamics since, following Simon (1996) and others in the complex adaptive systems tradition, an important goal of social simulation is to understand how social agents display adaptive behavior when operating in their environment. For example, the second model might add additional

environmental dynamics, such as seasonality in the availability of resources or a somewhat more complex weather system (e.g., weather fronts that move over the landscape producing waves of rain) or short-term change. Alternatively, or as a complement, additional complexity can come from a technical system, such as a transportation system with a set of features (e.g., rate of technological innovation, structural features, operational characteristics).

- 2.26** Given a somewhat more complex natural or artificial environment, the third model in the sequence could consider more complex social dynamics able to cope with the increased environmental complexity introduced by the previous model. For example, whereas the first model contained only simple groups, such as households or other ascriptive groups (e.g., clans), the third model might introduce non-kin-based social relations, such as meritocratic or other associations that reflect a more complete representation of the social fabric in a given society of agents. Organizations are artifacts produced by human and social technology.
- 2.27** A fourth model might consider a larger territorial scale, such as a region that extends beyond the normal range of local agents. Eventually, the sequence of models should reach a stage when higher-order social relations are explicit—preferably as emergent structures—in the form of official or state-like governance institutions. Finally, at least in the case of final models where long-term change and societal adaptation is of interest, climate (not just weather as in earlier models) and other forms of diachronic or structural change is introduced.
- 2.28** Several technical features of the project can be used to define subsequent intermediate models. For example, automated information extraction tools can be used to build selected model components, in terms of agents, dynamics, or encapsulated features. Other methodologies, such as GIS or social network analysis (SNA), offer similar opportunities for developing intermediary models and features leading toward a final model.



## The Mason-Smithsonian Inner Asia Project

- 3.1** The Mason-Smithsonian Joint Project on Inner Asia (Cioffi-Revilla et al. 2007) is a complex social simulation project aimed at developing a better interdisciplinary scientific understanding of the rise and fall of polities—national territorial societies with a system of government—over a very long time period, sufficiently long to examine the social effects of climate and environmental change. The project began officially in January 2006, with initial discussions since 2004, and the present funding cycle ends in December 2009. Primary funding for the project is provided by the US National Science Foundation, under a three-year grant from the Human and Social Dynamics Program, with supplementary funding provided by the Center for Social Complexity of George Mason University and the National Museum of Natural History of the Smithsonian Institution in Washington DC.
- 3.2** Organizationally, the project includes a relative large collaborative community of investigators, including:
- *Numerous collaborators:* Senior and junior professors and scientific personnel, one postdoctoral researcher, graduate students, a high school teacher and his students, in addition to other unfunded but interested investigators.
  - *Several disciplines:* Computational social science and its specializations in agent-based modeling in political science, economic geography, social network analysis; computer science and its specializations in multi-agent systems and evolutionary computation;

anthropologists and specializations in archaeology and ethnography; and natural and environmental science.

- *Several institutions:* George Mason University (C. Cioffi, S. Luke, D. Parker, M. Tsvetovat), through the Center for Social Complexity at the Krasnow Institute for Advanced Study; the Smithsonian Institution (J.D. Rogers, W. Fitzhugh, B. Frohlich, W. Honeychurch), through the Department of Anthropology of the National Museum of Natural History; the Mongolian Academy of Sciences; the Thomas Jefferson High School of Science and Technology (R. Latimer), through its Department of Computer Science. (An earlier attempt to include additional international collaborators from Europe and Asia unfortunately failed, due to funding limitations and a reduced budget.)
- *Long time period:* Initial discussions and meetings to prepare the project proposal began as early as 2004. Project-related publications and other activities are expected to continue beyond 2009 (officially the last year of funding by the NSF grant), given the set of new models (and new data sets) generated by the project.

**3.3** What links together this community of scientists, disciplines, and institutions is a deeply shared interest in the investigation of long-term societal adaptations using comparative and computational methods—approaches that are now breaking new scientific ground that goes beyond earlier theories and methods. The Mason-Smithsonian Joint Project on Inner Asia includes many of the characteristic features found in other complex social simulation projects and is used here to illustrate the methodology outlined in the previous section.

### **Core questions and theory**

**3.4** The progression of theoretical models is driven by the core questions (positive heuristic) of the research program. In the case of the Mason-Smithsonian project, the questions are: How did polities first emerge in Inner Asia? What effect, if any, might neighboring polities (China) have had? How did Inner Asian polities evolve over periods of history that were sufficiently long to include climate variation? How did semi-nomadic (agropastoralist) societies eventually form territorial albeit non-sedentary polities? How can one explain the large expanse of the Mongol Empire, given its relatively simple form of government? From a comparative perspective, numerous smaller polities (e.g., Renaissance republics in Europe) have operated with vastly more complex systems of governance—how did the Mongols do it? These and other scientific puzzles constitute the core project questions of the Mason-Smithsonian Inner Asia Project.

**3.5** The main theoretical framework used in the project is the so-called "canonical theory" for explaining sociopolitical change (origins and development of polities) based on a process of variations around a recurring theme of societal challenges and collective action responses—with occasional failures (Cioffi-Revilla 2005). Over time, a society (social groups) succeeds or fails in addressing threats and exploiting opportunities, depending on their leadership and collective capacity to act; as well as on processes linked to the natural environment where they are situated. Degrees of success and failure produce various levels of development and decay, respectively. Collective action operates on a time scale akin to a "fast process" (synchronic change), whereas historical evolution and political change operate on a longer time-scale akin to a "slow process" (diachronic change). Pages from a historical atlas—a useful metaphor for this class of complex social simulation—mark the slow/diachronic process; pages from a historical memoir or chronicle (e.g., *The Secret History of the Mongols*; Weatherford, 2004) are replete with fast/synchronic processes.

- 3.6** The canonical theory comprises a set of related and interdisciplinary path-dependent explanations that account in an integrated way for variations in historical change, unlike earlier theories that rely mostly on implicitly deterministic mono-causal explanations and eschew the presence of multiple time-scales in sociopolitical evolution (synchronic and diachronic change). For example, the canonical theory includes original ideas as well as elements from earlier social theories by Carneiro (1970), Marcus (1993), Olson (1965), and Lichbach (1996), among others. Formally, the canonical theory is modeled as a probabilistic branching process (a forward logic sequential tree) with subtrees grafted onto the main nodes that consist of lotteries and decisions by Nature and human agents, respectively.
- 3.7** In terms of the earlier criteria and characterization of complex social simulations, the Mason-Smithsonian Inner Asia Project provides a nice fit in terms of investigating (1) coupled socio-natural systems (the canonical theory includes both social and natural dynamics), (2) ranging over multiple spatio-temporal scales (local household terrain to continental/international scale, over a period of approximately 3,000 years), and (3) requiring a multidisciplinary approach (integrating knowledge from the social, computational, and natural sciences).
- 3.8** The computational simulation models that formalize (instantiate) the canonical theory and related processes in the Mason-Smithsonian project are built with the MASON (Multi-Agent Simulator of Networks and Neighborhoods) toolkit (Luke et al. 2005). (Other agent-based simulation toolkits include Swarm, Netlogo, Repast, and Cormass. See Gilbert (2008) and Nikolai and Madey (2009) for recent surveys.) MASON is written in the Java programming language, but MASON models (such as those described in the next section and others) can be rendered in small .jar files (generally just a few MB) that run on a personal computer without a user having to know the Java language. However, knowledge of Java is necessary to build or modify (as opposed to experiment with) a given model.

### **Progressive computational models**

- 3.9** Inner Asia is a vast region of the world comprising several ecological zones (ecosystems) that range from desert, to steppe, to high mountain; and numerous and diverse societies that extend from modern-day Mongolia to its neighbors, especially to the west, towards Central Asia. During the first millennium BCE (and until recently) many of these societies were agro-pastoralist or semi-nomadic communities that relied on herding, trade, and limited forms of agriculture. Thus, the human and societal ontology of the target system for this region of the world includes numerous nomadic actors: migratory households with movable possessions, seasonal camps, "states on horseback", and other non-sedentary social entities that are relatively uncommon elsewhere in the world. In addition, sedentary polities also co-exist with nomads, as neighbors or mingled, depending on the time period and usually in strong competition.
- 3.10** The first computational simulation model developed by the Mason-Smithsonian Project (the project's initial "M<sub>0</sub>") was the so-called *MASON HouseholdsWorld model* (Cioffi-Revilla et al. 2010). In its earliest version (2008), HouseholdsWorld contained only a basic terrain with a distribution of food, comparably simple weather, and a set of nomadic households that followed herds on the landscape. This initial model provided a very simple abstraction of the target world but contained several key elements for understanding (and modeling) far more complex dynamics concerning the core questions of the project. For example, the Households model permitted the investigation of household migration and aggregation patterns, movement of herds, simple weather effects on both resources and households, and numerous other initial questions

that needed to be understood before addressing higher levels of social and natural complexity. Importantly, the Households model did not contain any politics, or political processes directly related to the emergence of polities or systems of governance.

- 3.11** Subsequent versions of the Household model included greater detail in terrain as well as weather, by introducing seasonal variation in precipitation and, therefore, annual fluctuations in resource stocks for the herds (Rogers et al. 2009)). None of this additional environmental and social development included any politics. The MASON Households model—especially later versions—has its own theoretical potential, besides being a stepping stone towards models with greater complexity. For example, Households can be used to replicate memory and sociality experiments first initiated with earlier versions of the MASON Wetlands model (Cioffi-Revilla et al. 2004), or to replicate and extend experiments from the MASON Sugarscape model (Bigbee et al. 2007), the latter replicating the model of Epstein and Axtell (1996).
- 3.12** The next milestone model of the project was the MASON Hierarchies model (Cioffi-Revilla et al. 2008), developed by adding social and natural features to the simulation (additional objects, attributes, operations). Unlike Households, Hierarchies takes on the challenge of modeling the explicit emergence of political entities in the socio-natural landscape. Thus far, the Hierarchies model has proven successful in generating one form of political entity: alliances of various sizes based on a canonical process of recurring endogenous leadership crises and revolts among clans. Although the model has some features that render it superficially similar to other simulations of politogenesis (e.g., Axelrod 2003; Turchin 2003), what makes the MASON Hierarchies model progressive with respect to earlier simulations is the reliance on explicitly political processes in the canonical theory and—consistent with the historical record—on the diversity of paths that produce growth and decay in sociopolitical complexity.
- 3.13** The MASON Hierarchies model is progressive with respect to the previous Households model, because it includes more elements of the target system (i.e., beyond those represented in Households) and also provides a new generative explanation for a greater variety of phenomena (namely the emergence of political alliances among clans). Hierarchies is among the few extant computational simulation models of politogenesis and the first to situate sociopolitical complexity within a changing natural environment; it is also compatible with the application of evolutionary computation (EC).
- 3.14** From a more technical standpoint, the Hierarchies model also implements several significant innovations in visualizations, including social network analysis graphics and multiple time series of key indicators of socio-natural evolution. For example, based on these facilities, investigators have been able to observe the emergence of distribution of capabilities and other features.
- 3.15** The next simulation models in the computational research program will increase the spatial and temporal scale of simulations. Already the Hierarchies model covers an area larger than the typical habitat of a household, but the spatial dimension of the simulation is only implicit and needs to be portrayed in relation to the known record of empirical sizes.
- 3.16** The final model is envisioned to have a set of features that address the main research questions, including:
1. Complete use of the multiple generative paths of the canonical theory, comprising both threats and opportunities, endogenous and exogenous situational changes, multiple modes of collective action, and other theoretically grounded mechanisms that drive the

slow/diachronic process of long-term change.

2. Extensive external validation through methods and data from multiple origins and domains, including different social and natural variables, distributions, longitudinal patterns, social network structures, and other empirical/operational links between the simulated world of  $M_F$  and the real target world of Inner Asia—and over the long time period of the project.
3. Sociopolitical cartographic visualization comparable to a "virtual historical atlas", demonstrating how the computational simulation model provides a credible replication of the known socio-natural history of the region.

**3.17** Models from the Mason-Smithsonian project on Inner Asia have also proved valuable for developing more advanced models in other complex social simulation projects, such as the Mason-HRAF Joint Project on Eastern Africa (Rouleau et al. 2009). This is because the preservation of legacy knowledge and code has proven essential for developing increasingly sophisticated models and maintaining continuing with personnel turnover. In sum, the Mason-Smithsonian Inner Asia Project has been well-served by the proposed methodology of complex social simulations, although ultimately the success of the project will depend on numerous other factors. Other similar projects on long-term evolution and adaptation in socio-natural systems could also be analyzed from a similar perspective to gain additional comparative insights.



## Conclusions

- 4.1** Social simulation is an emerging field of computational social science—especially agent-based modeling—and arguably a transforming trend in 21st century social science. In recent years social simulation has progressed from simple toy models such as Conway's or Schelling's (which still hold significant theoretical potential) to increasingly realistic models of complex social systems, such as agent-based models that include heterogeneous agents interacting in changing natural and/or artificial environments.
- 4.2** The class of these often large multidisciplinary computational projects requires a scientific research methodology that would seem to be qualitatively or organizationally distinct from, albeit grounded on, simpler social simulations that have more limited scope, largely intentionally minimal complexity, and typically evolve through the work of a single investigator. Large projects involving complex social simulations require more robust and resilient methodological features if they are to involve large numbers of participants, several disciplines, different institutions, and a long period of time.
- 4.3** This article proposed a methodology for theoretically-driven research programs involving complex social simulations—particularly inter- and multi-disciplinary socio-natural systems with multi-level architecture. The proposed methodology is based on a succession of models that is akin to but distinct from the late Imre Lakatos' notion of a "research programme". Lakatos' framework is inspiring because of its specific relevance for long-term research projects such as those that arise in complex social simulations. A major difference lies in the interdisciplinarity of the core research questions—that cannot be answered by knowledge from a single discipline—and by the computational (as opposed to mathematical) formalism, especially the object-orientation to modeling.
- 4.4** The proposed methodology was illustrated through examples from the Mason-Smithsonian project on developing agent-based models of the rise and fall of polities in Inner Asia. The initial model in the sequence of progressive problemshifts was the so-called MASON HouseholdsWorld

model, which contained a natural environment and households that form camps as an emergent phenomenon. The second major model, the so-called Hierarchies model, also in MASON, introduced politics into the virtual social world, resulting in a progressive problemshift where alliances of clans were generated as emergent phenomena. Additional models will follow, guided by the project's core questions and final goals.

- 4.5** The proposed methodology requires further development. Clearly, multidisciplinary and object-oriented modeling both provide promising, major, and unexplored potential contributions to computational social science and social simulation. So far this methodology of systematically progressive simulation models has proven valuable for advancing the scientific objectives of the project and avoiding some pitfalls. Importantly, it may also prove valuable for improving scientific communication and facilitating the growth of knowledge.



## Acknowledgements

An earlier version of this paper was presented at the III Edition of Epistemological Perspectives on Simulation (EPOS): A Cross-Disciplinary Workshop, October 2-3, 2008, Lisbon University Institute - ISCTE, Portugal. Thanks to conference participants, especially Nuno David and Luís Antunes, as well as two anonymous referees of this journal, for useful comments and criticisms. Funding for this study was provided by the Center for Social Complexity of George Mason University and by grant HSD-0527471 from the Human and Social Dynamics Program of the US National Science Foundation. The opinions, findings, and conclusions or recommendations expressed in this work are those of the author and do not necessarily reflect the views of the National Science Foundation. This paper is dedicated to Paul Diesing, who introduced me to the work of Imre Lakatos, and to all members of the Mason-Smithsonian Joint Project on Inner Asia.



## References

- AXELROD, Robert. (2003). Advancing the Art of Simulation in the Social Sciences. *Japanese Journal for Management Information Systems* 12 (3).
- BERGER, Joseph, Bernard P. COHEN, and Morris ZELDITCH. (1972). Status Characteristics and Social Interaction. *American Sociological Review* 37:241-255.
- BERGER, Joseph, and Morris ZELDITCH, eds. (1993). *Theoretical Research Programs: Studies in Theory Growth*. Stanford, CA: Stanford University Press.
- BIGBEE, Anthony, Claudio CIOFFI-REVILLA, and Sean LUKE. (2007). Replication of Sugarscape Using MASON. In *Agent-Based Approaches in Economic and Social Complex Systems IV: Post-Proceedings of The AESCS International Workshop 2005*, edited by T. Terano, H. Kita, H. Deguchi and K. Kijima. Tokyo: Springer.
- CARNEIRO, Robert. (1970). A Theory of the Origin of the State. *Science* 169:733-738.
- CIOFFI-REVILLA, Claudio. (1998). *Politics and Uncertainty: Theory, Models and Applications*. Cambridge and New York: Cambridge University Press.
- CIOFFI-REVILLA, Claudio. (2002). Invariance and universality in social agent-based simulations. *Proceedings of the National Academy of Science of the U.S.A.* 99 (Supp. 3) (14):7314-7316.

- CIOFFI-REVILLA, Claudio. (2005). A Canonical Theory of Origins and Development of Social Complexity. *Journal of Mathematical Sociology* 29 (April-June):133-153.
- CIOFFI-REVILLA, Claudio. (2009). Simplicity and Complexity in Computational Modeling of Politics. *Computational and Mathematical Organization Theory* 15(1): online.
- CIOFFI-REVILLA, Claudio, and Raymond DACEY. (1988). The probability of war in the  $N$ -crises problem: Modeling new alternative to Wright's solutions. *Synthese* 76 (2):285-306.
- CIOFFI-REVILLA, Claudio, Sean PAUS, Sean LUKE, James L. OLDS, and Jason THOMAS. (2004). Mnemonic Structure and Sociality: A Computational Agent-Based Simulation Model. In *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence*, edited by D. Sallach and C. Macal. Chicago, IL: Argonne National Laboratory and University of Chicago.
- CIOFFI-REVILLA, Claudio, Sean LUKE, Dawn C. PARKER, J. Daniel ROGERS, William W. FITZHUGH, William HONEYCHURCH, Bruno FROHLICH, Paula DEPRIEST, and Chunag AMARTUVSHIN. (2007). Agent-based Modeling Simulation of Social Adaptation and Long-Term Change in Inner Asia. In *Advancing Social Simulation: The First World Congress in Social Simulation*, edited by T. Terano and D. Sallach. Tokyo, New York, and Heidelberg: Springer Verlag.
- CIOFFI-REVILLA, Claudio, J. Daniel ROGERS, and Maciej LATEK. (2010). The MASON HouseholdsWorld Model of Pastoral Nomad Societies. In *The Science of Social Simulation: The Second World Congress in Social Simulation*, edited by Keiki Takadama, Claudio Cioffi-Revilla, and Guillaume Deffaut. Tokyo, New York, and Heidelberg: Springer Verlag.
- CIOFFI-REVILLA, Claudio, William HONEYCHURCH, Maciej LATEK, and Maksim TSVETOVAT. (2008). The MASON Hierarchies Model. Mason-Smithsonian Inner Asia Project, Working Paper.
- DIESING, Paul. (1971). *Patterns of Discovery in the Social Sciences*. Chicago: Aldine-Atherton, Inc.
- EPSTEIN, Joshua M., and Robert AXTELL. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, D.C.: The Brookings Institution and MIT Press.
- GARDNER, Martin. (1970). Mathematical Games: The fantastic combinations of John Conway's new solitaire game "Life". *Scientific American* October, 120-123.
- GILBERT, Nigel. (2008). *Agent-Based Models*. Thousand Okas, CA: Sage Publishers.
- GILBERT, Nigel, and Klaus TROITZSCH. (2005). *Simulation for the Social Scientist*. Second edition ed. Buckingham and Philadelphia: Open University Press.
- GILLESPIE, John V., and Dina A. ZINNES. (1975). Progressions in mathematical models of international conflict. *Synthese* 31 (2):289-321.
- HALES, David, Juliette ROUCHIER, and Bruce EDMONDS. (2003). Model-to-Model Analysis. *Journal of Artificial Societies and Social Simulation* 6(4)5. Available online: <http://jasss.soc.surrey.ac.uk/6/4/5.html>
- KOHLER, Timothy A., and Sander E. VAN DER LEEUW, eds. (2007). *The Model-Based*

*Archaeology of Socionatural Systems*. Santa Fe, NM: School for Advanced Research Press.

LAKATOS, Imre. (1970). Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave. London, UK: Cambridge University Press.

LAVE, Charles A., and James G. MARCH. (1993). *An Introduction to Models in the Social Sciences*. Lanham, MD: University Press of America.

LICHBACH, Mark I. (1996). *The Cooperator's Dilemma*. Ann Arbor, MI: University of Michigan Press.

LUKE, Sean, Claudio CIOFFI-REVILLA, Liviu PANAIT, and Keith SULLIVAN. (2005). MASON: A Java Multi-Agent Simulation Environment. *Simulation: Transactions of the Society for Modeling and Simulation International* 81 (7):517-527.

MARCUS, Joyce. (1993). Ancient Maya Political Organization. In *Lowland Maya Civilization in the Eighth Century A.D.*, edited by J. A. Sabloff and J. S. Henderson. Washington, D.C.: Dumbarton Oaks Research Library and Collection.

MEEKER, Barbara F. (2002). Some Philosophy of Science Issues in the Use of Complex Computer Simulation Theories. In *The Growth of Social Knowledge: Theory, Simulation, and Empirical Research in Group Processes*, edited by J. Szmataka, M. Lovaglia and K. Wysienska. Westport, CT and London: Praeger.

MOSS, Scott. (2008). Alternative Approaches to the Empirical Validation of Agent-Based Models. *Journal of Artificial Societies and Social Simulation* 11(1)5. Available online: <http://jasss.soc.surrey.ac.uk/11/1/5.html>

NIKOLAI, Cynthia, and Gregory MADEY. (2009). Tools of the Trade: A Survey of Various Agent-Based Modeling Platforms. *Journal of Artificial Societies and Social Simulations* 12(2)2. <http://jasss.soc.surrey.ac.uk/12/2/2.html>

OLSON Jr., Mancur. (1965). *The Logic of Collective Action: Public Goods and the Theory of Goods*. Cambridge, Mass.: Harvard University Press.

SHELLING, Thomas C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* 1:143-186.

ROGERS, J. Daniel, Maciek LATEK, Teresa NICHOLS, and Theresa EMMERICH. (2009). Weather, Scale, and Complexity in Inner Asian Pastoralist Adaptive Strategies. Washington, DC: Mason-Smithsonian Joint Project on Inner Asia.

ROUCHIER, Juliette, Claudio CIOFFI-REVILLA, J. Gary POLHILL and Keiki TAKADAMA. (2008). Progress in Model-to-Model Analysis. *Journal of Artificial Societies and Social Simulation* 11(2)8. <http://jasss.soc.surrey.ac.uk/11/2/8.html>

ROULEAU, Mark, Mark COLETTI, BASSETT, Jeffrey K., HEILEGIORGIS, Ates B., GULDEN, Tim, and KENNEDY, William G., . (2009). Conflict in Complex Socio-Natural Systems: Using Agent-Based Modeling to Understand the Behavioral Roots of Social Unrest within the Manderu Triangle. *Proceedings of the Human Behavior-Computational Modeling and Interoperability Conference 2009*. Oak Ridge, TN. 23-24 June 2009. Available from the author.

SIMON, Herbert A. (1996). *The Sciences of the Artificial*. 3rd ed. Cambridge, MA: MIT Press.

SZMATKA, Jacek, Michael LOVAGLIA, and Kinga WYSIENSKA, eds. (2002). *The Growth of Social Knowledge: Theory, Simulation, and Empirical Research in Group Processes*. Westport, CT and London: Praeger.

TABER, Charles S., and Richard J. TIMPONE. (1996). *Computational Modeling*. Thousand Oaks, London and New Dehli: Sage Publications.

TURCHIN, Peter. (2003). *Historical Dynamics: Why States Rise and Fall*. Princeton, NJ: Princeton University Press.

WEATHERFORD, Jack. (2004). *Genghis Kahn and the Making of the Modern World*. New York: Crown Publishers.