# Simulation-Oriented Cyberinfrastructure for Computational Social Science

Santiago Núñez-Corrales and Les Gasser
Informatics Institute and School of Information Sciences
University of Illinois at Urbana-Champaign

CSSSA Conference

November, 2016

**Abstract**

Computational social simulation (CSS) has proven its utility for basic research in social, behavioral, organizational, cognitive, linguistic, evolutionary, and biological sciences. It has been employed effectively in many application domains including government, policy, commerce, and industry. Along with basic science, applied problem-solving, and design, large-scale social simulations provide great educational opportunities for exploring and visualizing theories and models of complex social systems, and studying how they change. This paper reports on several facets of our approach to improving the realism, viability, and impact of simulation-oriented computational research in social sciences by i) raising the scale and complexity of simulation models; ii) grounding computational simulation research on properties of its fundamental constituents including social objects themselves, information, representational processes and their limits; iii) linking social simulations to very large data sets and to live streams of data and iv) binding and modernizing CSS tools into modern cyberinfrastructures compatible and interoperable with those of many other computational and data sciences.

## 1 Introduction

There is tremendous interest in computational modeling, simulation, and analysis of social, socio-technical, and socio-environmental systems, in many different application areas including basic science, government/policy, and commerce/industry. Computational social simulation (CSS) has proven its utility in social, behavioral, organizational, cognitive, linguistic, evolutionary, and biological sciences. Large-scale simulations provide tremendous educational opportunities to explore and visualize alternative theories and models of complex social systems, and to study how they change.

In our view, four avenues form the clearest path toward improving the realism, viability, and impact of simulation-oriented computational research in social sciences: i) raising the scale and complexity of models; ii) grounding computational simulation research on properties of its fundamental constituents including social objects themselves, information, representational processes and their limits; iii) linking social simulations to very large data sets and to live streams of data (e.g. driving CSS with live streams from large sensor networks); and iv) binding and modernizing CSS tools into modern cyberinfrastructures compatible and interoperable with those commonly used and under development in many other computational and data sciences (e.g. workflows; data repositories; analysis/visualization pipelines; traceability and provenance of results to/from data, theory, and models; reproducibility, etc.)

This paper reports on our thinking during the early stages of a new project called "Simulating Social Systems at Scale," hosted and supported by the National Center for Supercomputer Applications (NCSA) at the University of Illinois. The project's motivating vision is to create the basis for, and demonstrate the feasibility of effective, efficient, accessible, very large-scale simulations of social systems, as a central component of general scientific cyberinfrastructure and computational science practice.

### 1.1 Challenges of the scientific program in computational social science

The ultimate goal of a science is obtaining explanations that best approximate the measured state and structure of the world (Friedman, 1974), independent of domain. Observations and regularities are the raw materials from which theories and laws are obtained (Salmon et al., 1989). The particular procedural differences between scientific domains, however, depend on the structure of the relevant entities and how these are affected when measurements are performed upon them. Social science is not an exception in this sense.

Explanations need to correctly bridge different levels and scales, but their potentially large number leads to feasibility concerns. Reducing this complexity calls for macroscale descriptions from microscale states that are both accessible and deducible (Forge, 1980). A similar tension exists between individual and collective levels of analysis in society (Brewer, 1991). Deducibility is essential for theories to be considered candidate explanations, whether the particular problem at hand is one of prediction (gaining knowledge about the future), retrodiction (explaining the past) or sampling (accurately measuring reality). *Invariance* has become a keystone in the edifice of modern science because it leads often to fundamental explanations by removing complex arrays of particular interactions, simplifying the experimental and theoretical landscapes. Computational social science, as the discipline that makes use of an *in silico* empirical approach, is in the process of developing methods for finding the invariants and abstractions for making the understanding of society a more tractable task.

The methods of the physical sciences appear at first hand to be incompatible with some of the features of social research. One reason is due to the apparent absence of invariance in either the internal dynamics of social agents or the environment in which they are embedded, leading to large amount of possible hypotheses that can be formulated and tested without efficient ways to filter those which are correct (Kincaid, 2004). Biology, in contrast to physics, serves as a good example for the situation in social sciences: hypotheses about organisms need to be deduced both from limited measurements in a varying, complex environment and evidence from the past, whose quality decreases rapidly with temporal distance (Bernstein et al., 2000). Recent success of computational systems biology in providing sound explanations to diverse phenomena by integrating observations, theories and information contained both *in* and *about* the organism is undeniable (Zhao et al., 2013), where some invariants have been identified in high-level representation of many biological processes, and not in the experimental data itself.

Information, a central piece of the establishment of modern biology, appears also to be at the core of an effective computational social science. Most of the discoveries in organisms related to heredity pertain to the analysis of categorical, genomic or contextual information. In turn, information is then converted to models of the biological systems that greatly simplify the process of understanding relations between parts and their global effects. In these abstraction process, networks have become a *lingua franca* capable of revealing invariants that remained hidden at more grounded levels of representation (Cioffi-Revilla, 2002). The notion of agency, prevalent in the concepts behind ongoing research in computational social science, is an information-rich concept itself that has lead to significant advances (Cioffi-Revilla, 2014).

In view of the above discussion, we propose to address the following three central problems as the next frontier of (computational) social science. While these problems are embedded in a strong digital context, it must be stressed that good explanations of phenomena aiming at the development of solid theoretical arguments from an inherent richness of social structures (Flyvbjerg, 2001), and not software or hardware projects, are the relevant entities for the notion of any work pursuing the development of such science (Conte et al., 2012).

### 1.1.1 Moving from ontological debates to epistemic goals

Instead of presupposing levels of description and categories of objects (ontological), a more substantial approach to finding explanations attempts to determine the mechanics and the entities in a generalized way (epistemic) (Van Bouwel and Weber, 2008). By taking this road, four key tasks become soluble: (1) finding adequate levels of explanation from observable facts and approximating models, (2) determining the form of the explanations by integrating networks of hypotheses that may be falsified and are devised as to account for the complexity of social phenomena, (3) constructing structural and functional explanations that provide macroscale causal relations through contrasts of possible outcomes –for instance, T-contrast (Skocpol, 1969) - that match fine-scale causal explanations (Taylor, 1988) and (4) elucidating social mechanisms that involve both accumulation of facts and postulating general principles (i.e. covering laws) that may be partially adequate (Gross, 2009), but neither of which are equatable at all to social mechanisms on their own.

We consider that the latter view, common to discussions in social science in general, should be a structural guide for computational social science to realistically extend the knowledge domain it attempts to serve. While the nature of social systems apparently differs in many respects to the systems studied in the natural sciences, there seems to be no fundamental reason why an empirical study of systems of many agents may only be a theoretical device, but rather reach inner explanatory structures when properly constructed. Cyberinfrastructure and social modeling tools in this regard must be consistent from the conception up to the implementation, being sensitive to the fact that interacting with aspects of the social world may introduce unexpected effects (Gross, 2009), for instance, when simulations are interactive (Helbing and Balietti, 2013).

### 1.1.2 Moving from small worlds to large-scale experiments

Whether experimental, social or historical research (Cleland, 2002), reducing instance size is a commonplace heuristic for grappling with the complexity induced by systems that interact with the environment or other systems. For social systems as complex systems, the reduction is justified by assuming emergent behavior: large numbers of simple units, under similar forces and with similar behavior, may produce unexpected non-linear responses (Heylighen, 1989). Sometimes the inner structure of the responses is that of a small world (Kleinberg, 2000), which summarizes the fact that in many systems the entities have a statistically preferential and restricted set of interactions (Amaral et al., 2000). Small-world organization has been extensively observed and studied in areas such as cooperation (Challet and Zhang, 1997), human brain functional networks (Bassett and Bullmore, 2006), social and biological communities (Girvan and Newman, 2002) and human language research (i Cancho and Solé, 2001).

However, scaling limits have been identified for emergent behavior even in small problems (Barabási and Albert, 1999) and suggested in problems involved social systems (Robb, 1989). Some complex phenomena require sufficiently large problem sizes for emergence to occur. Moreover, even in that case, microscale description may not suffice to accurately compute expected outcomes of the whole system. Such failure in reductionistic explanations may be characteristic of the distinction between emergent and non-emergent systems (Wayne and Arciszewski, 2009) when either (1) there exist fundamental differences in the presupposition of the objects that interact at the different phenomenological levels, or (2) the description of the ensemble of the objects and their environment is incomplete (Batterman, 2001). One of those failures is the inability of models to account for the variations (either deterministic or random) of the world. Their origin may be intrinsic to the nature of the model of the world (i.e. agents only *sample* a flux of events) or dependent on the interactions with the agents (i.e. the model of the world is mutable and agents can modify it). Briefly, small-worlds are locally dependent models that do not often account for changes in the environment, crucial for unraveling the complexity of events in social systems; empirical embeddedness (Boero and Squazzoni, 2005) is a key property that cyberinfrastructure for computational social science ought to ensure when required.

### 1.1.3 Moving from information by-products to information-centric representations

For general theories and laws to be deduced from data, adequate information representation is fundamental (Van Fraassen, 2010). Representations frequently denote the entities, their properties and the mechanics of the system they compose; in the abstract, representations in agent-based systems are a special type of information that is often subtly embedded in the dynamics or the resulting patterns of emergence itself (Bickhard, 2000). The relative nature of information as an exchange between agents, as well as the intrinsic information content of the world in both static and dynamic contexts reveal a close relationship between information and circumstance (Barwise, 1986). For instance, a successful theory is capable of elucidating which constraints exist in a system, and what their applicability is such as in understanding how agents consult, cooperate and compete amongst themselves (Eaton et al., 1998). If the constraints change, the circumstances of the agents may lead them to new outcomes. Using small-world theories, when disconnected from the relational information contained both in the agents, the environment and their interrelation is a most probable cause for incompleteness in the specification of the rules of a system (e.g. partial representations, incomplete concept hierarchies) that leads to poor inferences and models (Li et al., 2013).

In a sense, gaining knowledge about a system may be equated to exhaustively finding the correct description for all the constraints that govern its dynamics. A theory then becomes a collection of related constraints capable of being instanced to yield specific predictions or explanations. But constraints are inferred from data as instances of known patterns. Hence, patterns (detailed prescriptions that match with particular instances) are high-level information tools that, coupled with adequate experimental design and data analysis tools, can lead to powerful insights in domains as complex as ecology (Grimm et al., 2005). Cyberinfrastructure that enables large-scale simulations in computational social science should provide means to instrument, capture and systematize information at different levels of representation, desirably with integrated automated pattern-matching and machine learning mechanisms to facilitate research.

## 2 Components of a large-scale simulation framework

In this work, we propose a general architecture for the development of computational social science experiments, rooted in the capabilities of existing and future cyberinfrastructure and compatible with the twin goals of i) achieving more complete and robust understanding of the useful range of social objects and processes (i.e., developing general social theory); and ii) achieving greater predictive and explanatory power in regard to specific social settings and cases (i.e. social problem solving and design). This section is concerned with its structural and functional description, as well as with some general properties and attributes expected of such infrastructure.

### 2.1 First-class entities: experiments as abstract contracts

Experiments are detailed specifications of realistic or abstract settings that contain the necessary and sufficient information constraints and processes for determining whether a particular set of assertions about the state of the world may be rejected, be subject to further enquiry or be considered as supported to a certain level of significance (Gooding, 2012). The specification contains the following general elements

- the pre-conditions of the experiment (i.e. state of the world prior to executing the actions, including variables out of the experimentalist's control),
- a detailed prescription of how processes are transformed into steps connected via intermediate inputs and outputs,
- a specification of the post-conditions after the processes are executed,
- a representation of the expected measurement outcome(s) and
- an analytic prescription for obtaining significance levels that would determine the success of the experiment.

Arguably, the main difficulty of performing experiments in the social sciences lies in the implications of (a) having agents that can interact with the world in complex, (b) the practical difficulties involved in finding or creating the appropriate conditions (c) and the existence of potential unethical consequences of enacting them in many situations (Benton and Craib, 2010). By replacing agents with any sufficiently complex entity, determining the practical possibilities of observational or experimental settings and bearing in mind the ethical consequences of any experiment, the situation is not so much different from the natural sciences in many contexts. In such cases, though experiments come to the rescue as as abstract specifications of possible worlds that strictly follow certain principles or laws (Gooding, 1992; Brown, 2011) that are not enacted in reality.

In the same way that a scientific instrument embodies the principles of a domain-specific context (i.e. an information background) (Baird, 2004), thought experiments contain the necessary and sufficient elements for theories to be tested. However, they are also present in the form of the *conceptualization* and design of experiments executed in real settings. In social sciences, thought experiments have been conceived as elucidating mechanisms in general social contexts for problems such as providing alternative explanations to the Social Contract as the governing element of society (Latour, 1998), causality of historical events (David, 2007) and comparative analysis of social scenarios (Schneider and Wagemann, 2012). In addition, thought experiments also define language that facilitates communication amongst peer researchers in expert audiences (Reiner, 1998), essential to creating community and critical mass in computational social science.

Therefore, experiments and their abstract specifications are first-class objects in our design of large-scale social simulation infrastructure. Their information representation is central for orchestrating the underlying computing infrastructure and subsequent data analysis.

### 2.2 Multi-agent systems: programs and constraints

The fundamental construct for general social simulation is the *Agent-Based Model* (ABM). An ABM most often comprises a set of *agents* which are the units of action, and some kind of *world* or environment in which the agents and their interactions are situated. There may be information flows between agents, and these may be represented as either coupled agent-environment interactions, or as explicit communications. This conceptualization of ABMs gives rise to three modeling *components*: 1) a physical world, 2) the agent's physical affordances (e.g. sensing/effecting) and decision procedures; and 3) an inter-agent communication infrastructure. Given the experimental philosophy and scientific-epistemic goals outlined above, these three layers provide serious constraints on the capabilities and scalability of ABMs. These issues are grounded in fundamentals of computational simulation and representation infrastructure, as constrained by the scientific goals such as reproducibility, and control over experimental conditions. Put briefly, on one extreme we could envision all agents acting independently with no interactions among each other. This frees us to execute agents as rapidly as computation resources permit, completely in parallel – an approach known as *divide-and-conquer*, widely used in extreme-scale data-processing. However, this approach obviates models in which agents interact, radically reducing the scope of possible science. It also eliminates the control needed for reproducibility, because there is no information about or possible control over the ordering of actions, which thus may turn out to be random and unrepeatable. Unfortunately, introducing interactions among the three model components (world, agents, and communication) puts extreme limitations on the scalability of models for fundamental complexity reasons. Thus the goal of achieving a robust and highly scalable cyberinfrastructure for CSS requires rethinking some simulation and experiment fundamentals and hence possibly rethinking the fundamentals of social theory that they represent. Possible refinements for scalability include probabilistic simulations, quality vs resource tradeoffs, and moving from deterministic to statistical representations of social phenomena.

### 2.3 Workflow-to-infrastructure as a goal satisfaction problem

Since experiments are first-class objects in our approach (above), it is possible to reason about dynamic matching between the emergent resource needs of a set of experiments and their scientific goals. Our idea here is to support dynamic provisioning of data sources, analysis tools, and simulation resources based on analyzing the goals of the experiment, exploring alternative provisionings. For example, a scheme like Decker and Lesser's TAEMS and PGP models for planning (Lesser et al., 2004) would allow this type of reasoning, based on satisfying *quality*, *cost*, and *time* tradeoffs for alternative experiment workflows, given available infrastructure components.

There exist similar attempts to synthesize execution-to-hardware mappings, most of which rely on using middleware stacks to make parallel access transparent. The Mesos platform is an example that in particular abstracts nodes that can execute MPI and Hadoop (Hindman et al., 2011). Another example is portfolio scheduling of scientific tasks for data centers in which a ledger of multiple infrastructure options are compared against economic criteria in terms of energy and estimated CPU time (Deng et al., 2013). Communication-aware schedulers optimized for expensive all-to-all operations exist (Subramoni et al., 2014) with analogues for mapping programs to cloud computing resources (Gupta and Kalé, 2013). These approaches are expensive and operate during the execution, not in general prior to it.

Another way to generate execution-to-infrastructure mappings is to interpret them as goal satisfaction problems with search and optimization-based satisfiers. Goal-driven workload assignment and redistribution is well a known strategy in clouds and supercomputers (Saifullah et al., 2013; Tang et al., 2013; Wang et al., 2013; Prabhakaran et al., 2014). Current work around execution-to-infrastructure mapping based on some preliminary form of goal satisfaction (Czarnul, 2013; Sarnowska-Upton, 2013) suggest this route is promising. In short, automated mechanisms that removes the complexity of infrastructure relates issues are central to increased supercomputer usage as well as the use of modeling and simulation across disciplines (Kelly et al., 2015). It is our view that an AI-oriented strategy will be beneficial beyond large-scale modeling in computational social science.

### 2.4 A proposed architecture for large-scale computational social science experiments

The general constructs for an integrated architecture aimed at enabling experimental work in the social sciences under the paradigm of very large scale multi-agent systems integrated into cyberinfrastructures are shown in Figure 1. Our design has been in part motivated by the ideas and goals discussed in Axelrod (1997), and extended to aim benefiting from existing state-of-the-art cyberinfrastructure.

With respect to actual usage of the system, we propose a series of reconfiguration steps that social science cyberinfrastructure would follow during execution of an experiment.

#### 2.4.1 Contract specification phase

Ontology development for representing scientific experiments is a long standing research area (Noy and Hafner, 1998). In conjunction with semantic provenance mechanisms (Sahoo et al., 2008), they are essential for making tractable the growing number of experimental explorations that are openly released into the public and research domains and being able to reason about the life cycle of scientific experiments (Mattoso et al., 2010). Ontologies for describing experiments exist in materials science (Cheung et al., 2008), bioinformatics (Usadel et al., 2006), biomedical research (Dumontier et al., 2014) and experimental microbiology (King et al., 2009), among others. Our current target ontology is EXPO (Soldatova and King, 2006) due to its generality in describing the context and contents of experiments.

A contract specification is composed of several elements. With respect to agents, both their communication model as well as their input data sources (i.e. archival, live) need to be specified along with their observables, measurements and ideally any associated hypotheses. In addition,

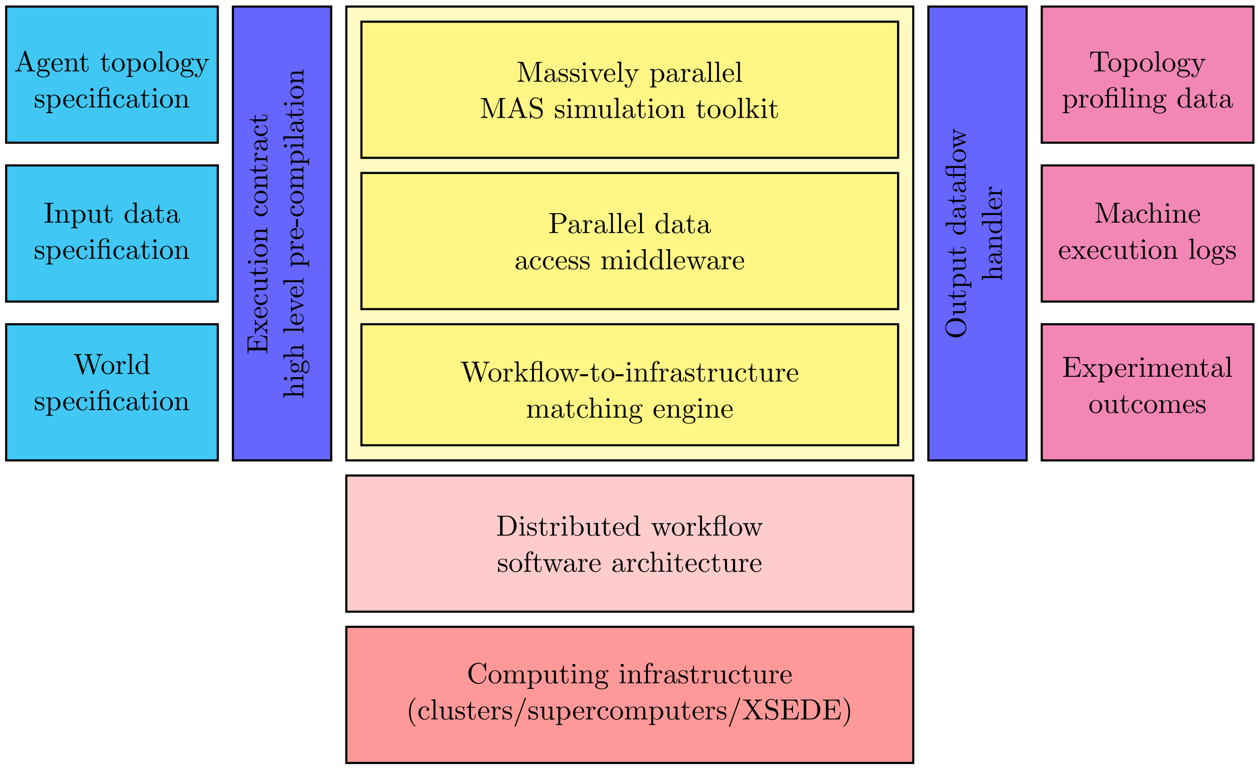


Figure 1: Abstract, high level block representation of a large-scale MAS experimental framework for enabling computational social science experiments. Clear separation of responsibilities at each stage in the execution of a computational social science experiment allows decoupling from particular details of the underlying computing infrastructure.

a description of the world in which agents interact and act is necessary, involving its distributed representation and its spatial features (e.g. GPS data, time constraints, update policies). The description is then compiled into a contract file as input to the next phase.

#### 2.4.2 Contract-to-execution phase

Once a contract is assembled, several software pieces are configured and instanced. a massively parallel MAS simulation toolkit will consume parts of the contract related to agent and world representations, constructing an instance of the specified model afterwards. Presently, the ROSS Time Warp System (Carothers et al., 2002) and (Collier and North, 2011) stand as examples of commonly used toolkits to be supported in such workflows. Our current efforts include the design and development of a new toolkit for agent-based models under uncertainty. In addition, other components for handling distributed representations of spatial (Brown et al., 2005) and temporal (Mattern, 1989; Nielsen et al., 2001) world models will be implemented and integrated.

One important consideration is the common execution background in systems that may be used by computational social scientists. The aim of this architecture is to be flexible enough as to operate in any provisioning-capable system (e.g. supercomputer, cloud) where middleware packages such as MPI, Hadoop and others may be installed and used by the different toolkits to make distribution transparent. The result of this phase is an executable specification in an intermediate language for later optimization to the particular infrastructure settings.

### 2.4.3 Model execution phase

As previously described, a specialized pre-execution step will transform the executable specification into a concrete execution schedule that matches the underlying infrastructure. The process needs to interact with workflow management systems and resource descriptions in order to gather sufficient information as to determine (possibly sub-) optimal resource allocation for a given simulation. The concrete executable schedule may be later hand tuned or through determination of simulation wide scaling laws.

### 2.4.4 Execution-to-outcomes phase

After the model is executed, outcomes are classified in three categories: profiling data in relation to communication topologies, human-readable machine execution logs and experimental results in structured form. In particular, standard scientific file formats are expected for later analysis and visualization. An important element of the latter process is determining whether experimental results come from a data flow in a continuous simulation (hence, derived from a sampling metaphor) or from a single execution unit. Compliance with standard scientific data formats such as NetCDF (Rew et al., 2006) and HDF5 (Yang et al., 2005) is to be implemented as well.

### 2.4.5 Post-experiment data management phase

After execution, a data consolidation toolset will provide visualization and analysis facilities after experiments are performed. Considering the dynamics of social science research and the scope of the types of research to which this architecture is aimed at, tools that allow interaction and have a low learning curve are ideal. In that sense, significant experience with large volumetric data analysis and visualization has been obtained through the yt Project (Turk et al., 2010), one of the initial platforms that will be evaluated for integration in our architecture. Guaranteeing attributes such as data curation, provenance and reusability are critical for enhancing the research experience in computational social science (Bechhofer et al., 2013) as to generate a fully digital context appropriate for community building and its broader scholarship.

## 3 Application targets

Social system are in general those in which a set of interacting entities share information towards explicit or implicit common goals such as lowering difficulty boundaries, survival, coordinating responses to threats or gaining knowledge through cooperation (Luhmann, 1995). Finding research domains where social systems are central is critical for assessing the utility of large-scale ABMs as research tools. In particular, we propose utilizing this infrastructure to model events in language evolution, soil microbiology, small interference RNA (siRNA) based therapies in neuromedicine and organizational evolution.

## 3.1 Soil ecology

The bacterial ecology of soils is a driving factor of fertility, productivity, and economic growth (Wardle et al., 2004). Function-wise, fungi and bacteria are dominant with respect to fertility because of their decomposition activity in the soil food web (Anderson, 2003). Aboveground and belowground microbiotas differ from one another and possess high degrees of specificity in their activities and molecular products; they tangle mutually through both negative and positive feedback loops (Ehrenfeld et al., 2005). In summary, the mechanics of soil microbiota is complex and driven by a soil food web that is subject to a variety of factors that work at the micro and macro levels.

The bioavailability of C, $N_2$, $CO_2$ and $SO_X$ depends on bacterial concentrations and species (Singh et al., 2004). Disease detection mechanisms in the form of excess of various molecular products trigger defense actions in communal plant bacteria, ultimately leading to collective responses against possibly pathogenic populations (Whipps, 2001). Resistance to environmental changes has been observed to occur thanks to dense mutualism networks in which several organisms have adapted to perform the same functions, to direct chemical signaling and to coordinate the expression of resistance mechanisms (Griffiths and Philippot, 2013). Individual cell-to-cell communication mechanisms exist and are modulated by several factors that affect a series of delicate molecular signalling pathways. Two examples are the signal-controlled biofilm development strategies (Kolenbrander et al., 2010) and the more indirect, asynchronous horizontal gene transfer mechanism mediated by plasmids and viruses (Aminov, 2011).

ABMs have been used for modelling ecological responses of collections of individual microorganisms. Hellweger and Bucci (2009) provides an extensive review of existing literature on the subject including techniques, theoretical aspects and challenges. Application examples include modelling host-pathogen immunology responses (Bauer et al., 2009), biofilm formation (Lardon et al., 2011), bacterial chemotaxis (Emonet et al., 2005) and selectivity responses to synergistic drug combinations (Lehár et al., 2009). Modelling molecular rich environments has also been performed through a molecules-as-agents approach, where reactions are interactions between agents with responses modulated by calculating kinetics from ODEs (Sneddon et al., 2011). However, existing applications of ABMs to understanding interactions soil ecology seem to be restricted to few examples such as the effect of earthworms in soil structure (Blanchart et al., 2009).

Successful application of ABMs may bring a new perspective into soil microbiota as long as a scalable approach is feasible (approximately $10^9$ ecologically diverse bacteria per gram of soil are present). Applied research targets include engineering the equilibrium of bacterial populations to specifically favourable to crops in sustainable ways (Sturz et al., 2000) and designing remediation strategies by transplantation of foreign soil bacteria accompanied with inorganic compounds (Nwachukwu, 2001).

## 3.2 siRNA-based therapies in neuromedicine

Small interference RNA (siRNA) is a mechanism of molecular pathway regulation in which precisely engineered small RNA strands (21-25 bps) bind selectively to mRNA required to produce a given protein, leading to its downregulation or complete inhibition (Pushparaj et al., 2008). This mechanism was first identified in viral defense and gene silencing mechanisms in plants (Ratcliff et al., 1997) and later experimentally demonstrated in *Caenorhabditis elegans* (Fire et al., 1998). Since then, siRNAs have become a major technology for both discovering new molecular pathways (Hood et al., 2012) and aspects of existing ones and for intervening in living systems with an emphasis in biomedical applications.

RNA interference occurs in mammalian neurons (Krichevsky and Kosik, 2002) and persist up to three weeks after therapeutic administration (Omi et al., 2004). Small RNA-mediated cell-to-cell communication is facilitated by exosomes (Frühbeis et al., 2012). The latter is critical in NMDA receptor-mediated excitotoxicity and ischemic neuronal death (Belousov and Fontes, 2016) and a strong connection of gap junctions and mRNA to gene expression patterns in human seizure disorder is known (Naus, 1991). There is direct evidence of sensorial reconfiguration dependency on mRNA and siRNA, with the olfactory system as a prime example (Juang et al., 2013). However, much ground needs to be covered in terms of mapping the diversity of regulatory functions that RNA interference may allow in conjunction with the growing body of knowledge about omics of the neuron (Elia and Finkbeiner, 2013).

Designing siRNA therapies against neurodegenerative diseases is of particular interest and difficulty because of the inherent complexity of the molecular pathways involved in brain activity and its impact towards finding molecular medicines for the brain (Davidson and Paulson, 2004). While significant progress has been made in the last ten years in molecular neurobiology (Forman et al., 2004), further integration into a whole brain picture remains as a grand challenge; siRNAs appear to be an important part of the puzzle.

There are no direct known applications of ABMs to model siRNA inter-cell communication despite the agent-like nature of neurons. Modeling of signal transduction in general (Chylek et al., 2014) had been attempted through rule-based systems, a technique that is becoming more frequent due to its representational convenience and analytic properties (Maus et al., 2011). All approaches attempting to overcome challenges of multi-scale modeling in biological systems (Qu et al., 2011) in the light of new perspectives made possible are immediate targets for large-scale ABM development. Open questions in siRNA cell-to-cell communication include how to develop a rational basis for improving safety (Boudreau et al., 2011) and estimating the potential effectiveness of new siRNA therapies (Shankar et al., 2005).

### 3.3 Emergence of organizations

Organization is a structural and functional feature of many abstract and complex systems. In human organizations, one perspective about its emergence is theoretically linked by the existence of multiple networks and transposition: when individuals from two different social networks share a new common environment, they bring knowledge of their previous context (patterns and rules) that serves as an *a priori* structuring element for new sets of behaviors, expectations and dynamics (Powell et al., 2012). Elements such as the notion of ownership, division of labor, fairness and others also allow moving beyond emergence to convergence (Fulmer and Ostroff, 2016). More over, the acquisition of individuality for an organization is an interesting phenomenon that is constructed, or rather *emerges* from other types of individuals (Scott, 2013).

Literature on agent-based models for the emergence of organization is extensive. Helbing (2012) provides an overview of many relevant application areas ranging from modeling of socio-economic systems to responding to systemic risks and managing internal complexity. Many of these models are based on the small-world network approach in which a large portion of the apparent complexity is reducible to few local interactions and few functional motifs (Amaral et al., 2000). Small-world networks are limited in various important situations, as when systems are coupled and clustering of entities in the organization is to be expected (Bruch and Atwell, 2013). The latter is critical when simulations are a vehicle to understand how human decisions are made in coupled human-natural systems (An, 2012).

At present, we are exploring possible applications of large-scale agent based models to areas tied to public policy such as innovation diffusion (Kiesling et al., 2012), design of macroeconomic policies (Fagiolo and Roventini, 2012), dynamics of health systems (Luke and Stamatakis, 2012) and modeling of consumer energy choices (Rai and Henry, 2016). As indicated in the discussion above, the potential of ABMs should not only predictive but also retrodictive. Current literature on reconstructing ancestral socio-ecological systems (Heckbert et al., 2013) is suggestive on that particular line of research.

## 4 Conclusion

The Computational Social Sciences can only benefit by the support of advanced cyberinfrastructure that integrates elements including streaming live data as both input and output, complex *world models* based on physical reality such as geographic information systems, computational and data infrastructure for running extreme-scale, multi-component and multi-level models; advanced communication models that support research on information-centric social phenomena, and integrative frameworks such as workflow systems, automated goal-to-workflow-to-infrastructure matching, and foundational scientific models. The project we report on is providing the seed and proof of concept of this integrated cyberinfrastructure at scale, along with identifying major bottlenecks and new theory needed to realize its vision.

## 5 Acknowledgements

We gratefully acknowledge the support of the National Center for Supercomputing Applications through the NCSA Fellows program in the project titled "Simulating Social Systems at Scale" (L. Gasser) and of the Illinois Informatics Institute through the Informatics PhD program (S. Núñez-Corrales).

## References

Amaral, L. A. N., Scala, A., Barthelemy, M., and Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the national academy of sciences*, 97(21):11149–11152.

Aminov, R. I. (2011). Horizontal gene exchange in environmental microbiota. *Frontiers in microbiology*, 2:158.

An, L. (2012). Modeling human decisions in coupled human and natural systems: review of agent-based models. *Ecological Modelling*, 229:25–36.

Anderson, T.-H. (2003). Microbial eco-physiological indicators to asses soil quality. *Agriculture, Ecosystems & Environment*, 98(1):285–293.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. In *Simulating social phenomena*, pages 21–40. Springer.

Baird, D. (2004). *Thing knowledge: A philosophy of scientific instruments*. Univ of California Press.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

Barwise, J. (1986). Information and circumstance. *Notre Dame Journal of Formal Logic*, 27(3):324–338.

Bassett, D. S. and Bullmore, E. (2006). Small-world brain networks. *The neuroscientist*, 12(6):512–523.

Batterman, R. W. (2001). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*. Oxford University Press.

Bauer, A. L., Beauchemin, C. A., and Perelson, A. S. (2009). Agent-based modeling of host–pathogen systems: The successes and challenges. *Information sciences*, 179(10):1379–1389.

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., et al. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2):599–611.

Belousov, A. B. and Fontes, J. D. (2016). Role of neuronal gap junctions in nmda receptor-mediated excitotoxicity and ischemic neuronal death. *Neural regeneration research*, 11(1):75.

Benton, T. and Craib, I. (2010). *Philosophy of social science: The philosophical foundations of social thought*. Palgrave Macmillan.

Bernstein, S., Lebow, R. N., Stein, J. G., and Weber, S. (2000). God gave physics the easy problems: adapting social science to an unpredictable world. *European Journal of International Relations*, 6(1):43–76.

Bickhard, M. H. (2000). Information and representation in autonomous agents. *Cognitive Systems Research*, 1(2):65–75.

Blanchart, E., Marilleau, N., Chotte, J.-L., Drogoul, A., Perrier, E., and Cambier, C. (2009). Sworm: an agent-based model to simulate the effect of earthworms on soil structure. *European Journal of Soil Science*, 60(1):13–21.

Boero, R. and Squazzoni, F. (2005). Does empirical embeddedness matter? methodological issues on agent-based models for analytical social science. *Journal of Artificial Societies and Social Simulation*, 8(4).

Boudreau, R. L., Spengler, R. M., and Davidson, B. L. (2011). Rational design of therapeutic sirnas: minimizing off-targeting potential to improve the safety of rnai therapy for huntington's disease. *Molecular Therapy*.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and social psychology bulletin*, 17(5):475–482.

Brown, D. G., Riolo, R., Robinson, D. T., North, M., and Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and gis. *Journal of Geographical Systems*, 7(1):25–47.

Brown, J. R. (2011). *The laboratory of the mind: Thought experiments in the natural sciences.* Routledge.

Bruch, E. and Atwell, J. (2013). Agent-based models in empirical social research. *Sociological methods & research*, page 0049124113506405.

Carothers, C. D., Bauer, D., and Pearce, S. (2002). Ross: A high-performance, low-memory, modular time warp system. *Journal of Parallel and Distributed Computing*, 62(11):1648–1669.

Challet, D. and Zhang, Y.-C. (1997). Emergence of cooperation and organization in an evolutionary game. *arXiv preprint adap-org/9708006*.

Cheung, K., Drennan, J., and Hunter, J. (2008). Towards an ontology for data-driven discovery of new materials. In *AAAI Spring Symposium: Semantic Scientific Knowledge Integration*, pages 9–14.

Chylek, L. A., Harris, L. A., Tung, C.-S., Faeder, J. R., Lopez, C. F., and Hlavacek, W. S. (2014). Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 6(1):13–36.

Cioffi-Revilla, C. (2002). Invariance and universality in social agent-based simulations. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7314–7316.

Cioffi-Revilla, C. (2014). Computation and social science. In *Introduction to Computational Social Science*, pages 23–66. Springer.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science*, 69(3):447–451.

Collier, N. and North, M. (2011). *Repast HPC: A platform for large-scale agentbased modeling.* Wiley.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., et al. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1):325–346.

Czarnul, P. (2013). A model, design, and implementation of an efficient multithreaded workflow execution engine with data streaming, caching, and storage constraints. *The Journal of Supercomputing*, 63(3):919–945.

David, P. A. (2007). Path dependence: a foundational concept for historical social science. *Cliometrica*, 1(2):91–114.

Davidson, B. L. and Paulson, H. L. (2004). Molecular medicine for the brain: silencing of disease genes with rna interference. *The Lancet Neurology*, 3(3):145–149.

Deng, K., Verboon, R., Ren, K., and Iosup, A. (2013). A periodic portfolio scheduler for scientific computing in the data center. In *Workshop on Job Scheduling Strategies for Parallel Processing*, pages 156–176. Springer.

Dumontier, M., Baker, C. J., Baran, J., Callahan, A., Chepelev, L., Cruz-Toledo, J., Del Rio, N. R., Duck, G., Furlong, L. I., Keath, N., et al. (2014). The semanticscience integrated ontology (sio) for biomedical research and knowledge discovery. *Journal of biomedical semantics*, 5(1):1.

Eaton, P. S., Freuder, E. C., and Wallace, R. J. (1998). Constraints and agents: Confronting ignorance. *AI magazine*, 19(2):51.

Ehrenfeld, J. G., Ravit, B., and Elgersma, K. (2005). Feedback in the plant-soil system. *Annu. Rev. Environ. Resour.*, 30:75–115.

Emonet, T., Macal, C. M., North, M. J., Wickersham, C. E., and Cluzel, P. (2005). Agentcell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics*, 21(11):2714–2721.

Fagiolo, G. and Roventini, A. (2012). Macroeconomic policy in dsge and agent-based models. *Revue de l'OFCE*, (5):67–116.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans. *Nature*, 391(6669):806–811.

Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge university press.

Forge, J. (1980). The structure of physical explanation. *Philosophy of Science*, pages 203–226.

Forman, M. S., Trojanowski, J. Q., and Lee, V. M. (2004). Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. *Nature medicine*, 10(10):1055–1063.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, 71(1):5–19.

Frühbeis, C., Fröhlich, D., and Krämer-Albers, E.-M. (2012). Emerging roles of exosomes in neuron–glia communication. *Frontiers in physiology*, 3:119.

Fulmer, C. A. and Ostroff, C. (2016). Convergence and emergence in organizations: An integrative framework and review. *Journal of Organizational Behavior*, 37(S1):S122–S145.

Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.

Gooding, D. C. (1992). What is experimental about thought experiments? In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 280–290. JSTOR.

Gooding, D. C. (2012). *Experiment and the making of meaning: Human agency in scientific observation and experiment*, volume 5. Springer Science & Business Media.

Griffiths, B. S. and Philippot, L. (2013). Insights into the resistance and resilience of the soil microbial community. *FEMS microbiology reviews*, 37(2):112–129.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., and DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *science*, 310(5750):987–991.

Gross, N. (2009). A pragmatist theory of social mechanisms. *American Sociological Review*, 74(3):358–379.

Gupta, A. and Kalé, L. V. (2013). Towards efficient mapping, scheduling, and execution of hpc applications on platforms in cloud. In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2013 IEEE 27th International*, pages 2294–2297. IEEE.

Heckbert, S., Isendahl, C., Gunn, J. D., Brewer, S., Scarborough, V. L., Chase, A. F., Chase, D. Z., Costanza, R., Dunning, N. P., Beach, T., et al. (2013). Growing the ancient maya social-ecological system from the bottom up. In *The Oxford Handbook of Historical Ecology and Applied Archaeology*.

Helbing, D. (2012). *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*. Springer.

Helbing, D. and Balietti, S. (2013). How to do agent-based simulations in the future: From modeling social mechanisms to emergent phenomena and interactive systems design (ssrn scholarly paper no. id 2339770). *Rochester, NY: Social Science Research Network. Retrieved from the Santa Fe Institute Archive: http://www. santafe. edu/research/workingpapers/abstract/51b331dfecab44d50dc35fed2c6bbd7b*.

Hellweger, F. L. and Bucci, V. (2009). A bunch of tiny individuals—individual-based modeling for microbes. *Ecological Modelling*, 220(1):8–22.

Heylighen, F. (1989). Self-organization, emergence and the architecture of complexity. In *Proceedings of the 1st European conference on System Science*, volume 18, pages 23–32. AFCET Paris.

Hindman, B., Konwinski, A., Zaharia, M., Ghodsi, A., Joseph, A. D., Katz, R. H., Shenker, S., and Stoica, I. (2011). Mesos: A platform for fine-grained resource sharing in the data center. In *NSDI*, volume 11, pages 22–22.

Hood, L. E., Omenn, G. S., Moritz, R. L., Aebersold, R., Yamamoto, K. R., Amos, M., Hunter-Cevera, J., and Locascio, L. (2012). New and improved proteomics technologies for understanding complex biological systems: addressing a grand challenge in the life sciences. *Proteomics*, 12(18):2773–2783.

i Cancho, R. F. and Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1482):2261–2265.

Kelly, R. C., Liu, S., Ghosh, S. S., Vento, D. D., Hart, D. L., Nagle, D., Smith, B., and Valent, R. A. (2015). Advanced user environment design and implementation on integrated multi-architecture supercomputers. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, page 33. ACM.

Kiesling, E., Günther, M., Stummer, C., and Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, 20(2):183–230.

Kincaid, H. (2004). Contextualism, explanation and the social sciences. *Philosophical Explorations*, 7(3):201–218.

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L. N., et al. (2009). The automation of science. *Science*, 324(5923):85–89.

Kleinberg, J. M. (2000). Navigation in a small world. *Nature*, 406(6798):845–845.

Kolenbrander, P. E., Palmer, R. J., Periasamy, S., and Jakubovics, N. S. (2010). Oral multispecies biofilm development and the key role of cell–cell distance. *Nature Reviews Microbiology*, 8(7):471–480.

Krichevsky, A. M. and Kosik, K. S. (2002). Rnai functions in cultured mammalian neurons. *Proceedings of the National Academy of Sciences*, 99(18):11926–11929.

Lardon, L. A., Merkey, B. V., Martins, S., Dötsch, A., Picioreanu, C., Kreft, J.-U., and Smets, B. F. (2011). idynomics: next-generation individual-based modelling of biofilms. *Environmental Microbiology*, 13(9):2416–2434.

Latour, B. (1998). Thought experiments in social science: from the social contract to virtual society. *1st virtual society? Annual public lecture*, 1.

Lehár, J., Krueger, A. S., Avery, W., Heilbut, A. M., Johansen, L. M., Price, E. R., Rickles, R. J., Short Iii, G. F., Staunton, J. E., Jin, X., et al. (2009). Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nature biotechnology*, 27(7):659–666.

Lesser, V., Decker, K., Wagner, T., Carver, N., Garvey, A., Horling, B., Neiman, D., Podorozhny, R., Prasad, M. N., Raja, A., et al. (2004). Evolution of the gpgp/taems domain-independent coordination framework. *Autonomous agents and multi-agent systems*, 9(1-2):87–143.

Li, J., Mei, C., and Lv, Y. (2013). Incomplete decision contexts: approximate concept construction, rule acquisition and knowledge reduction. *International Journal of Approximate Reasoning*, 54(1):149–165.

Luhmann, N. (1995). *Social systems*. Stanford University Press.

Luke, D. A. and Stamatakis, K. A. (2012). Systems science methods in public health: dynamics, networks, and agents. *Annual review of public health*, 33:357.

Mattern, F. (1989). Virtual time and global states of distributed systems. *Parallel and Distributed Algorithms*, 1(23):215–226.

Mattoso, M., Werner, C., Travassos, G. H., Braganholo, V., Ogasawara, E., Oliveira, D., Cruz, S., Martinho, W., and Murta, L. (2010). Towards supporting the life cycle of large scale scientific experiments. *International Journal of Business Process Integration and Management*, 5(1):79–92.

Maus, C., Rybacki, S., and Uhrmacher, A. M. (2011). Rule-based multi-level modeling of cell biological systems. *BMC Systems Biology*, 5(1):1.

Nielsen, M., Sassone, V., and Srba, J. (2001). Towards a notion of distributed time for petri nets. In *International Conference on Application and Theory of Petri Nets*, pages 23–31. Springer.

Noy, N. F. and Hafner, C. D. (1998). Representing scientific experiments: Implications for ontology design and knowledge sharing. In *AAAI/IAAI*, pages 615–622.

Nwachukwu, S. U. (2001). Bioremediation of sterile agricultural soils polluted with crude petroleum by application of the soil bacterium, pseudomonas putida, with inorganic nutrient supplementations. *Current Microbiology*, 42(4):231–236.

Omi, K., Tokunaga, K., and Hohjoh, H. (2004). Long-lasting rnai activity in mammalian neurons. *FEBS letters*, 558(1-3):89–95.

Powell, W. W., Packalen, K., and Whittington, K. (2012). Organizational and institutional genesis. *The emergence of organizations and markets*, 434.

Prabhakaran, S., Iqbal, M., Rinke, S., Windisch, C., and Wolf, F. (2014). A batch system with fair scheduling for evolving applications. In *2014 43rd International Conference on Parallel Processing*, pages 351–360. IEEE.

Pushparaj, P., Aarthi, J., Manikandan, J., and Kumar, S. (2008). sirna, mirna, and shrna: in vivo applications. *Journal of dental research*, 87(11):992–1003.

Qu, Z., Garfinkel, A., Weiss, J. N., and Nivala, M. (2011). Multi-scale modeling in biology: how to bridge the gaps between scales? *Progress in biophysics and molecular biology*, 107(1):21–31.

Rai, V. and Henry, A. D. (2016). Agent-based modelling of consumer energy choices. *Nature Climate Change*.

Ratcliff, F., Harrison, B. D., and Baulcombe, D. C. (1997). A similarity between viral defense and gene silencing in plants. *Science*, 276(5318):1558–1560.

Reiner, M. (1998). Thought experiments and collaborative learning in physics. *International Journal of Science Education*, 20(9):1043–1058.

Rew, R., Hartnett, E., Caron, J., et al. (2006). Netcdf-4: software implementing an enhanced data model for the geosciences. In *22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanograph, and Hydrology*.

Robb, F. F. (1989). The limits to human organisation: the emergence of autopoietic systems. In *Operational Research and the Social Sciences*, pages 247–251. Springer.

Sahoo, S. S., Sheth, A., and Henson, C. (2008). Semantic provenance for escience: Managing the deluge of scientific data. *IEEE Internet Computing*, 12(4):46–54.

Saifullah, A., Li, J., Agrawal, K., Lu, C., and Gill, C. (2013). Multi-core real-time scheduling for generalized parallel task models. *Real-Time Systems*, 49(4):404–435.

Salmon, W. C. et al. (1989). Four decades of scientific explanation. *Scientific explanation*, 13:3–219.

Sarnowska-Upton, K. (2013). *Methods Toward Automatic Configuration of Computing Environments for Application Execution*. PhD thesis, University of Virginia.

Schneider, C. Q. and Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.

Scott, W. R. (2013). *Institutions and organizations: Ideas, interests, and identities*. Sage Publications.

Shankar, P., Manjunath, N., and Lieberman, J. (2005). The prospect of silencing disease using rna interference. *Jama*, 293(11):1367–1373.

Singh, B. K., Millard, P., Whiteley, A. S., and Murrell, J. C. (2004). Unravelling rhizosphere–microbial interactions: opportunities and limitations. *Trends in microbiology*, 12(8):386–393.

Skocpol, T. (1969). *States and social revolutions*. Cambridge University Press.

Sneddon, M. W., Faeder, J. R., and Emonet, T. (2011). Efficient modeling, simulation and coarse-graining of biological complexity with nfsim. *Nature methods*, 8(2):177–183.

Soldatova, L. N. and King, R. D. (2006). An ontology of scientific experiments. *Journal of the Royal Society Interface*, 3(11):795–803.

Sturz, A., Christie, B., and Nowak, J. (2000). Bacterial endophytes: potential role in developing sustainable systems of crop production. *Critical Reviews in Plant Sciences*, 19(1):1–30.

Subramoni, H., Kandalla, K., Jose, J., Tomko, K., Schulz, K., Pekurovsky, D., and Panda, D. K. (2014). Designing topology-aware communication schedules for alltoall operations in large infiniband clusters. In *2014 43rd International Conference on Parallel Processing*, pages 231–240. IEEE.

Tang, W., Ren, D., Lan, Z., and Desai, N. (2013). Toward balanced and sustainable job scheduling for production supercomputers. *Parallel Computing*, 39(12):753–768.

Taylor, M. (1988). Rationality and revolutionary collective action. In Taylor, M., editor, *Rationality and revolution*, pages 63–97. Cambridge University Press, Cambridge.

Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T., and Norman, M. L. (2010). yt: A multi-code analysis toolkit for astrophysical simulation data. *The Astrophysical Journal Supplement Series*, 192(1):9.

Usadel, B., Nagel, A., Steinhauser, D., Gibon, Y., Bläsing, O. E., Redestig, H., Sreenivasulu, N., Krall, L., Hannah, M. A., Poree, F., et al. (2006). Pageman: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC bioinformatics*, 7(1):1.

Van Bouwel, J. and Weber, E. (2008). De-ontologizing the debate on social explanations: A pragmatic approach based on epistemic interests. *Human Studies*, 31(4):423–442.

Van Fraassen, B. C. (2010). Scientific representation: Paradoxes of perspective. *Analysis*, 70(3):511–514.

Wang, W.-J., Chang, Y.-S., Lo, W.-T., and Lee, Y.-K. (2013). Adaptive scheduling for parallel tasks with qos satisfaction for hybrid cloud environments. *The Journal of Supercomputing*, 66(2):783–811.

Wardle, D. A., Bardgett, R. D., Klironomos, J. N., Setälä, H., Van Der Putten, W. H., and Wall, D. H. (2004). Ecological linkages between aboveground and belowground biota. *Science*, 304(5677):1629–1633.

Wayne, A. and Arciszewski, M. (2009). Emergence in physics. *Philosophy Compass*, 4(5):846–858.

Whipps, J. M. (2001). Microbial interactions and biocontrol in the rhizosphere. *Journal of experimental Botany*, 52(suppl 1):487–511.

Yang, M., McGrath, R. E., and Folk, M. (2005). Hdf5-a high performance data format for earth science. In *Proceedings of the International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology*.

Zhao, X.-M., Tian, W., Jiang, R., and Wan, J. (2013). Computational systems biology. *The Scientific World Journal*, 2013.