The Agent-Based Modeling Canvas: A Modeling *Lingua Franca* for Computational Social Science

Ivan Garibay, Chathika Gunaratne, Niloofar Yousefi, and Steve Scheinert

Complex Adaptive Systems Lab, University of Central Florida

February 24, 2018

1 Introduction

When Grimm et al. (2006) first published their Overview, Design concepts, and Details (ODD) framework, they recognized both the growing use of computational models in the social sciences, particularly for individual- (IBMs) and agent-based models (ABMs), and the difficulty in describing agent-based models with sufficient clarity for understanding and replication. When Ostrom (2005) published the Institutional Analysis and Development (IAD) framework, she recognized the difficulty in describing complex institutions and the spaces in which they operate. The framework provides a tools for organizing inquiry and description of institutions by providing researchers with a set of questions to answer which, when all are answered, provide a thorough description and analysis of that institution. Both identified a common challenge for complex computational models: describing models is itself a difficult and complicated task. Both responded to this challenge by providing a framework to guide researchers' efforts in meeting this challenge. Ostrom's framework guides the design of conceptual models while Grimm et al's framework guides the description of computational models. Neither framework guides the translation of theories and conceptual models into computational models. Like the specification of statistical or econometric models (Kennedy, 2003), this process is left largely for the researcher to determine and apply. An additional framework, one that can help bridge the gap between conceptual, modeling, and technical languages used in different fields, is now needed to support the growing use of agent-based and individual-based models in the social sciences.

Computational modeling and simulation is now playing a fundamental role in promoting data-driven investigation of human behaviors and understanding of social systems. Social science researchers, in a wide array of social scientific fields, have long used a variety of computational methods. Econometrics was sufficiently developed to be recognized and its meaning debated more than half a century ago (Tintner, 1953). Mathematically-intensive game theory ungirded the United States' strategies during the Cold War (Schelling, 1960, 1966). Network analytic techniques are now regularly used in a variety of fields, particularly including sociology, political science, and policy science (Borgatti, Mehra, Brass, & Labianca, 2009; Comfort, Wukich, Scheinert, & Huggins, 2011; Koliba, Meek, & Zia, 2011). The growing use of network analysis and simulation, along with a clear articulation of a epistemology for applying agent-based models to the social sciences (Epstein, 2006), is helping to drive the emergence of a new field, computational social science (CSS), that brings together researchers with computational backgrounds and researchers with social scientific backgrounds to apply ever more sophisticated computational analytic techniques to social scientific research questions.

It is the interdisciplinary nature of computational social science that enables and necessitates collaboration between social and computational scientists, which are often now being referred to as "data scientists." In utilizing modeling and simulation approaches this collaboration has the potential to make monumental changes in addressing society's most challenging problems in ways not previously possible. Opportunities resulting from these collaborative efforts are many (Epstein, 2008). Social scientists provide substantive theories on whether, when, and under what conditions certain behaviors occur. Computational scientists and statisticians provide advanced techniques analyzing social scientific data to identify what and when certain social scientific phenomena will take place. Although this collaboration can lead to breakthrough innovations that no discipline can single-handedly produce, just as with any interdisciplinary team project, a primary obstacle is establishing a common language for effective communication among researchers with diverse backgrounds. Without such language, the collaboration is unlikely to succeed.

1.1 The Stakeholders

Three stakeholder groups are important for the success of a computational social science research effort:

Modeling and Simulation Specialists: Modeling and simulation specialists have the necessary knowledge of constructing computer models and simulation that are able replicate complex system phenomena. They understand the capabilities and limitations of modeling and simulation tools and which tools are appropriate under the given problem context. Importantly, they recognize the requirement for a model to have enough detail to be computationally implemented and understand how to analyze simulation output. For example, *starting from a generic ABM simulation experts are able to put forth questions that help define how a model must be constructed so that it can provide an answer to the research question.*

Social Scientists: Social scientists in this context are the domain experts. Despite the comprehensiveness of the agent-based model implementation tools and even the availability of design documentation (ODD/ODD+D) domain experts will often have a difficult time producing ABMs to suit their work. Without the input from social scientists, the models built by simulation experts and data scientists might fail to leverage the existing body of knowledge from the social sciences. As a result, these models, in turn, might fail to answer questions grounded in social science theory. Further, social scientists, without understanding the level of detail required of variables and relationships, could fail to provide a description with enough information to implement the ABM as a computer program. For example, rules for movement of agents are commonly overlooked: *what does a movement in the horizontal direction mean in relation to a movement in the vertical direction in a social context?*

Data Scientists/Machine Learning Experts: Data scientists and machine learning experts specialize in the knowledge and skills in using advanced analytic techniques that can be used to test the theories that social scientists propose. Though many social scientists regularly perform their own analyses, they often cannot stay up to date with the plethora of methodological developments or the scripting languages upon which the newest tools and techniques rely. For example, *in a model of social media, the domain expert may realize the importance of influence, yet not have the expertise to extract influence from the big data resulting from social media.*

In turn, the data scientist may not know which variables to look at in their search for user influence. This is where the communication between the data expert and the domain expert is crucial.

1.2 Need for a *"Lingua Franca"*

Furthering the computational advancement in the social sciences now requires the collaborative effort of experts from multiple domains to form inter-disciplinary research teams. The answers to today's most exciting research questions lie buried in large data sets. These data sets are often unstructured and have complex relationships, requiring specialized, cutting-edge computational analyses. As will be fully developed in the next section, the different stakeholder groups rely on diverging sets of concepts, methods, and lexicons. In many cases, the same specific term may have vastly different meanings in each stakeholder group. Building effective inter-disciplinary research teams means bridging this language gap between social scientists, machine learning experts and data scientists, and modeling and simulation experts. The absence of a common language or a *lingua franca*, spoken by all three types experts, is then a key concern and key requirement for effective computational social scientific research.

1.3 The Agent-Based Modeling Canvas

We propose such a *lingua franca*, the Agent-Based Modeling Canvas (ABMC). The ABMC aims to bring the collective effort of modeling and simulation experts, social scientists, and data scientists/machine learning experts towards the successful completion of a computational social sciences research collaboration. The ABMC is inspired by the Business Model Canvas (Osterwalder, Pigneur, Bernarda, & Smith, 2014). using a similar approach the ABMC visualizes all key computational social science project building blocks into a single paged canvas. This representation allows the multi-disciplinary discussion, design, and iterative refinement of the CSS project. A completed ABMC is able to completely define the requirements of a CSS research effort, including the guiding theories; the concepts, variables, and relationships; the data required for hypothesis building; the construction of the ABM itself; the expected target output; and the data and methods for the validation and output analysis. The canvas consist of 9 building blocks that a CSS research team must define in order to fully specify an ABM guided CSS project. It acts as a framework, similar to ODD and IAD for guiding the construction of agent-based models using steps drawn from the scientific method that is common to all scientists.

2 The Language Gap

Languages, by which we mean the varying tools that represent concepts and variables for analysis, are the key tools for scientific research. They are the mediums through which theories and concepts are expressed, data are structured and analyzed, and results interpreted, all in the service or testing hypotheses and providing evidence for and against theories. Social scientists, modeling and simulation experts, data scientists/machine learning experts have different languages for communication within their respective fields. These languages help with knowledge representation and dissemination by building a common lexicon with highly-refined definitions that allow quick, easy, and succinct expression of complicated ideas. However, in the case where the research outcomes are the result of inter-disciplinary teams working towards trans-disciplinary outcomes, these multiple communication streams render uninterpretable and communication breaks down. Jargon from one camp is misunderstood by the other and the shared mental model is never truly communicated across the various experts. In this section we explore some the existing efforts of research communication in the three disciplines considered in the CSS efforts and emphasize the disparity between them.

2.1 The Modelers

Agent(or individual)-based modeling has become a common approach to answering social science research questions through simulation. Agent-based models deal with large orders of agents, each activated at every discrete time step. Within a given time step an agent must plan and perform actions that can range in computational complexity. Due to the massive amount of computations, agent-based simulation experiments must make use of the computing power of modern computers. Agent-based modeling toolkits encapsulate and implement functions common across agent-based models such as the scheduler which determines the order in which agents act per time step. In doing so, they provide a further level of abstraction upon which model developers can skip the detail necessary to implement such a project directly on a lower level language such as Java or C++ by providing a scripting interface. For example: Repast Simphony written in Java and RepastHPC written in C++ is abstracted through Relogo, NetLogo which provides NetLogo scripting language, is an abstraction of Java and Scala implementations, and AnyLogic, again an implementation in Java abstracted as a visual programming language allowing the developer to organize blocks of encapsulated agent behaviors.

Thus, the entirety of the agent-based model is captured through the programming language used to implement it (provided the model is complete as was intended during its design). Yet is usually unreadable to domain experts. Further, concepts and key relationships between variables are now expressed in deep logic that is not easily interpretable through the eyes of social scientists or data scientists/machine learning experts. To address this issue, ABM documentation standards have emerged such as ODD and the human decision making specific version of ODD, ODD+D (ODD + Decision) (Grimm et al., 2006, 2010). The ODD standards help frame the agent-based model itself highlighting its:

> *"Purpose, Entities, state variables, and scales, Process overview and scheduling, Design Concepts (Basic principles, Emergence, Adaptation, Objectives, Learning, Prediction, Sensing, Interaction, Stochasticity, Collectives, Observation), Initialization, Input data, and sub models"*

The ODD standard itself is regularly maintained and updated by the authors and has been adapted by NetLogo as well where users are able to describe their model in the 'info' tab, which follows an ODD like structure. Together, NetLogo and ODD have standardized the agent-based model implementation and design documentation process. The penetration of NetLogo can be seen in 1, where we compare the ABM implementation software used by models submitted to the OpenABM (now CoMSES net) model library [1]. This comparison gives us some insight into the ABM tool requirements of the computational social sciences community. Many of the other platforms listed in 1 are carefully designed, robust, intuitive toolkits. They range from using visual programming (Anylogic) to Object Oriented Programming Languages like Java (Repast and Mason). However, the community has clearly chosen NetLogo as itself language of implementation. A few reasons to explain this adoption: 1) NetLogo is regularly maintained by CCL, 2) NetLogo provides a simple, Logo-like scripting language (with functional programming type constructs) , 3) it has many read-to-use functions important to ABMs in general including the scheduler, random number generators, and agent movement functions, 4) Its visual display is wrapped in

[1]https://www.comses.net

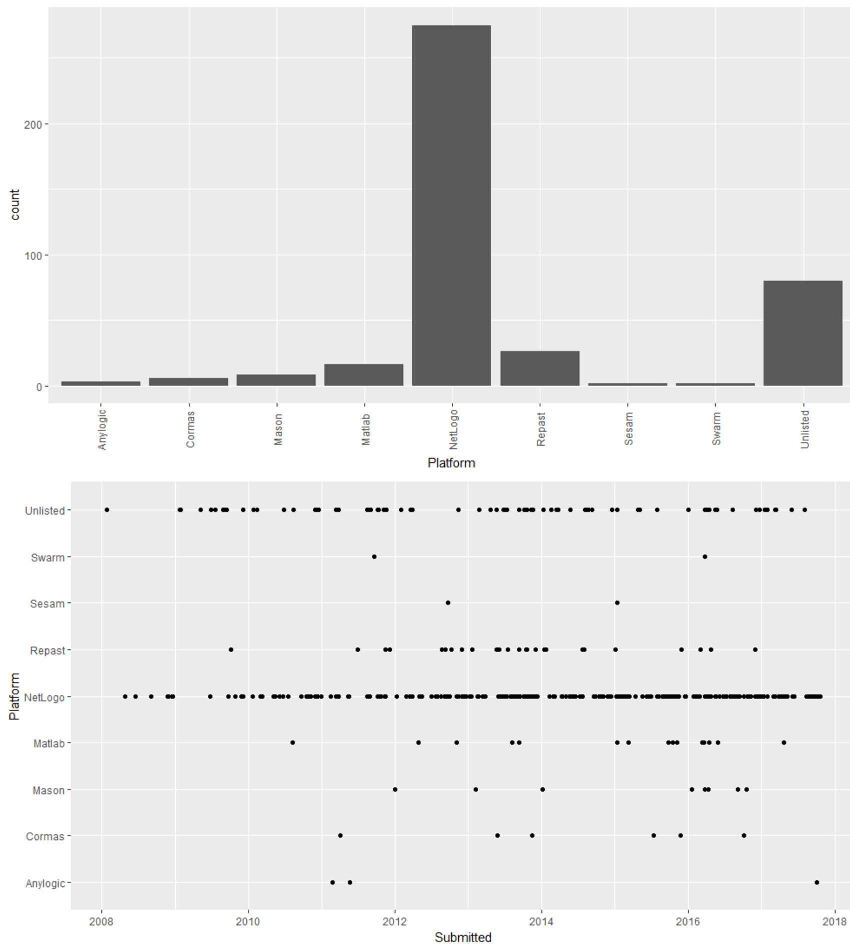

Figure 1: Platform usage of agent-based models on the OpenABM library from 2008 to November 2018. NetLogo has persisted throughout the decade as the most common platform for ABM development.

with the scripting interface allowing for quick transitions between coding and observing results (without having to compile and initialize runs separately), 5) its graphical components are drag-and-droppable, and 6) it comes ready with tools for parameter exploration (BehaviorSpace), parameter calibration (BehaviorSearch) (Stonedahl & Wilensky, 2010a), multi-scale modeling (LevelSpace (Hjorth, Head, & Wilensky, 2015)).

2.2 Social Scientific Languages in CSS

Certain communication standards are prevalent in the social science literature. Concept maps are common artifacts embodying social science concepts and research outcomes (Novak, 2010). However, diverging efforts at standardizing concept maps has meant that no standard does exist(Rafiei & Kardan, 2015; Kinchin, Hay, & Adams, 2000; Haymovitz, Houseal-Allport, Lee, & Svistova, 2017; Dipeolu et al., 2016; Schwendimann & Linn, 2016; Roth & Roychoudhury, 1992; Kamsu-Foguem, Tchuenté-Foguem, & Foguem, 2014; Fu, Huang, Ren, Weng, & Wang, 2017).

Frames (Minsky, 1975) and factor trees (Davis & O'Mahony, 2013) offer social scientific tools that are closer in their execution to the tools used by modelers and data scientists. Frames have helped the artificial intelligence community define knowledge representation and flow. Knowledge representations are common to both the social sciences and artificial intelligence. Good knowledge representations can be directly converted into models that can be programmed into a computer to simulate reality and may have symbolic representations or non-symbolic representations (Ramirez & Valdes, 2012). Factor trees have be introduced specifically for the specification of the factors that influence specific human behaviors (Davis & O'Mahony, 2013). A full descriptive factor tree can give a sufficiently complete definition of variables and their relationships to be fed into an ABM. Davis et al. have been able to directly convert factor trees into machine-runnable code.

Still, natural language narrative is quite common in describing social theory (Czarniawska, 2004) and has been shown lead to ambiguous and abstract theory definitions due to lack of formalism (Casas, 2013). Instead, several researchers have promoted the use of mathematics-friendly representations such as Petri-Nets (Casas, 2013; Köhler et al., 2007), Discrete Event System Specifications (DEVS) (Zeigler, 1989) or SysML (Weilkiens, 2011). Nevertheless, even methodological rigorous case study techniques, which are ubiquitous throughout the social sciences (George & Bennett, 2005; King, Keohane, & Verba, 1994), including case studies conducted with complexity-friendly frameworks like IAD, are not designed and rarely executed with computational or simulation modeling in mind. This manner of usage presents difficulties for researchers trying to translate case studies into computational models, adding an additional analytic step before machine-readable modeling codes can be constructed.

2.3 Data Analysis Languages

Most data analysis or machine learning efforts happen within the disciplines of Computer Science, Statistics, or Mathematics. Much of the recent data analysis and machine learning work has been conducted and packaged with high-level scripting programming languages such as Matlab/Octave, R, or Python for direct use. Methods that are established are algorithms that are commonly communicated in the form of pseudo-code (Horowitz, Sahni, & Rajasekaran, 1997) or flowcharts. These definitions are typically complete and ready to implement due to the nature of the purpose: to be run on computers. Large-scale software designed for such research are described in UML (Fowler & Scott, 2000). All of these methods have been established with the direct conversion to computer programming languages in mind and are easily convertible to code. Discrepancies in pseudo-code may exist, that mostly resulting to follow the evolution of programming languages and to reflect differences in different programming styles (object-oriented programming, functional programming, etc). Formal grammars and their many variants (Chomsky, 1956) can also be used to specify a complete functional set used for computer programming, but is probably less common for higher-level machine learning or data analysis type definitions.

2.4 A Comparison of Existing Languages

In Table 1, we compare and contrast some of the qualities of the more commonly used of these research languages to emphasize that there is a requirement for a *lingua franca* across the three disciplines. Here, languages that are adaptable to a discipline but require extra effort on the expert's part are awarded a ?, and languages that are actively used to cater to a certain discipline are marked with a ✓.

Comparing across the three columns, what we notice is that computational modeling shares languages with social sciences due to its penetration as a methodology for the social sciences, and with data analysis due to its roots as a computer science domain. However, there is a clear gap between

Table 1: A comparison of existing languages for knowledge and flow representation by the three stakeholder types of CSS. ?: adaptable to the domain, ✓: actively used by the domain.

Language	Social Science Experts	Data Analysis	Computational Modeling
NetLogo (and related software)	?		✓
ODD/ODD+D	✓		✓
Concept maps	✓		?
Factor Trees	✓		✓
Frames	✓		
Natural language narratives	✓		
UML	?	✓	?
High-level scripting languages (MatLab,R,Python)		✓	✓
Psuedocode, Flowcharts, Formal Grammars		✓	✓

the social sciences and machine learning/data scientists in the languages they use to frame and execute their research.

3 The Agent-Based Modeling Canvas

In this section we introduce our *lingua franca* towards the constructive collaboration between the three stakeholder types involved in a large scale CSS research project. Our approach, the Agent-Based Modeling Canvas (ABMC), shown in Figure 2, is highly visual, making it easily accessible to researchers in all fields.

ABMC visualizes CSS research as the result of two, usually iterative, processes. First, during the hypothesis building process the input of domain and data experts are drawn upon to construct meaningful theory- and data-verified hypotheses that are complete and ready to be modeled. Next, the Model Construction and Experimentation process, uses the set of hypotheses as the ground for model construction. The Hypotheses Building process ensures theory grounded requirements of the ABM and the Model Construction and Experimentation process in turn ensures computational

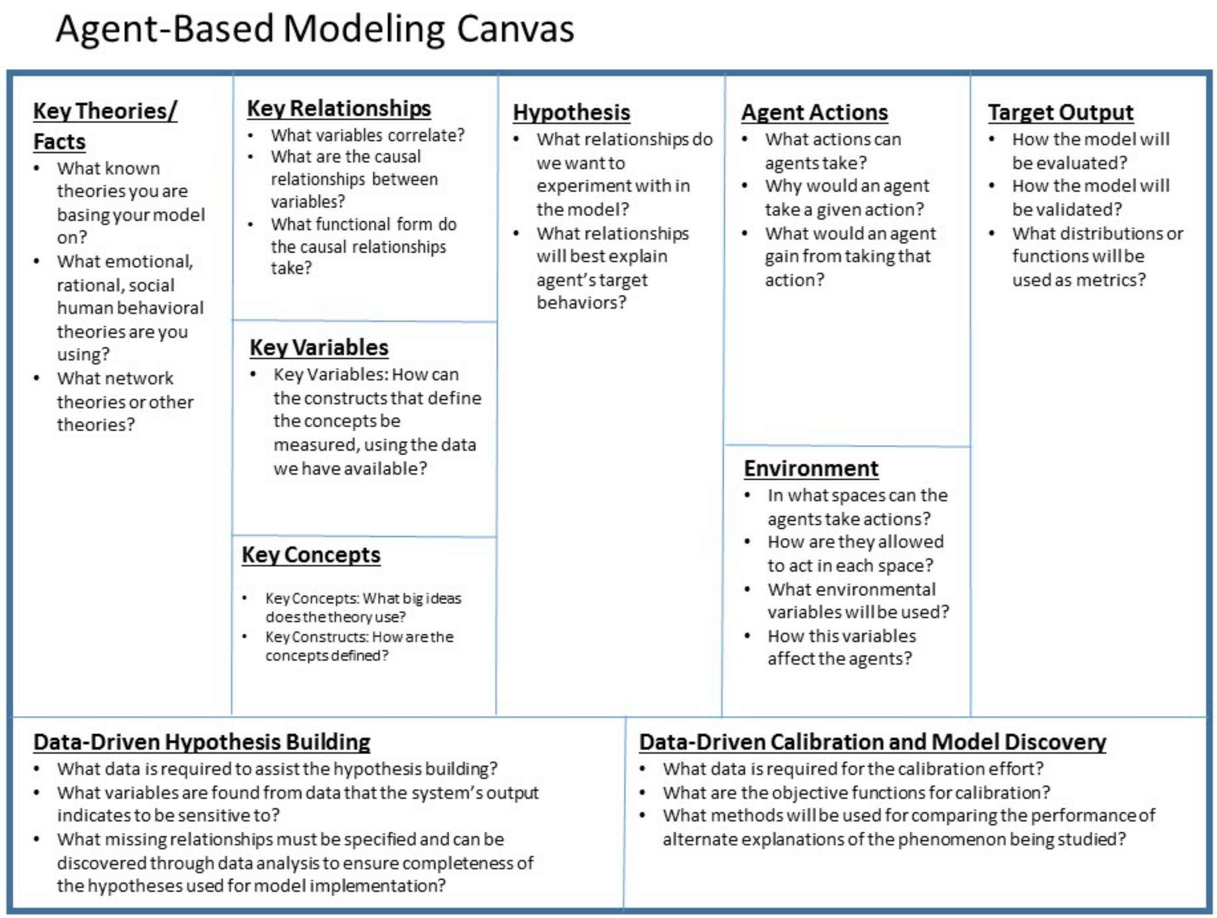

Figure 2: The Agent-Based Modeling Canvas

completeness of the experiments. Both processes require data for separate purposes. The first for filling in gaps of knowledge during hypotheses building and the second for validation and output exploration.

3.1 From Theory to Hypothesis: Human-aided Data-driven Hypothesis Building

Having a closer look at the hypothesis building process, one can realize that several steps are needed to get from a guiding theory to a testable hypothesis. Starting with the ***key theories***, for the purpose of the canvas, we follow the definition by (Shaughnessy & Zechmeister, 1985), where scientific theory is a coherent explanation of one or more phenomena observed reliably in systematic empirical research. The following questions might help fill this building block: *What known theories you are basing your model on? What emotional, rational, social human behavioral theories are you using? What network theories or other theories?*

Note that for a theory to be tested and successfully form more precise research questions, it must be narrow in scope and focus on specific ***key concepts***. Questions such as *what big ideas does the theory use? or how are the concepts defined?* can help fill this block. Besides concepts, what is common among all scientific theories is that they involve a set of ***key variables*** along with ***key relationships***, interrelating those variables. For example, we should ask *how can the constructs that define the concepts be measured, using the data we have available?*. Also, *how variables correlate? what are the casual relationships between variables? what functional form do the casual relationship take?*

Identifying context, variables and the way they interact helps narrow down a social theory to more specific and concrete instances that we call ***hypothesis***. These hypotheses will drive agent behavior which in turn, will determine the simulation output needed to test the hypothesis. In this block, we try answer questions such as *what relationship do we want to experiment within the model? or what relationships will best explain agent's target behaviors?*

Therefore, the three important blocks in building a hypothesis from a theory are (1) identifying the appropriate context (testbed) for replicating the theory, (2) discovering the relevant variables underlying that theory; and (3) measuring how those variables are interacting. The first task requires specialized knowledge and experience in the field of study. For example to study social phenomena, the knowledge of the social scientists is required to define assumptions, concepts, and processes to present a systematic view of social phenomena by specifying relations among variables (Kerlinger & Lee, 2000). The second task can be effectively and efficiently carried out by computational techniques and methodologies, which make use of data to create an algorithm to discover how variables interrelate. This theory-augmented machine learning approach can be quite powerful in establishing scientific hypotheses and relating them more fully to the relevant theories in a fast-paced, iterative fashion. Hence the importance of data for the hypothesis building process.

3.2 From Hypothesis to Model: Data-Driven Calibration and Model Discovery

As mentioned earlier, a theory helps investigating a phenomenon starting from the "big picture" and then narrowing it down to more specific hypotheses that can be tested. It is worth pointing out that there is often more than one plausible theory explaining a phenomenon. No matter if the existing theories are complementary or competing, it is essential to identify and consider all plausible theories implying the phenomena of interest. In this case, the ABM needs to consider not one but multiple social theories that are competing or cooperating to drive ***agent actions***. In order to fill this block we need to identify *what actions can agents take? why would an agent take a given action? what would an agent gain from taking that action?*

An evolutionary model discovery approach coupled with agent-based models can be used to explore multiple candidate models embodying existing theories of micro-behavior rules of agents (Gunaratne & Garibay, 2017, 2018). Although the need for this type of rule-space exploration has been emphasized for a while (Epstein, 1999, 2006), this has not been sufficiently investigated by the modeling and simulation community until now.

Furthermore, for an ABM to work, other entities called ***environment*** are also needed to drive the behavior and dynamics of the agents. To be able to identify these elements, we need to answer questions such as *in what spaces can the agents take actions? how are they allowed to act in each space? what environmental variables will be used? how this variables affect the agents?*

Another crucial step in developing any ABM is validation, which is typically performed by comparing the resulting output of the constructed model, with that of the real system. Note that a simulation resulting from an ABM is often characterized by a set of variables, which determine the:

1. Parameters of the model explaining the dynamics of simulation output

2. Initial conditions of the simulation run including, the characteristics of the agent populations and environmental state

3. The weights given to factors affecting behavior rules

Therefore, the validation process can be achieved by tuning of those associated variables for ***desired output***. This block of the canvas concerns *how the model will be evaluated? how the model will be validated? what distributions or functions will be used as metrics?* Finding the best set of parameters, typically, involves an optimization problem with respect to a (loss/objective) function, capturing the distance between the model's artificial output and the real experimental data. Parameter tuning, most often, requires exploring a gigantic parameter space, which might be a very tedious process. This motivates automatic machine learning-based or computational strategies, which facilitate faster and more efficient large scale explorations of the parameter-space. The result of this validation process will provide insight into the:

1. Basins of attraction of the phenomenon being simulated

2. Plausible archetypes/characteristics of entities in the real world

3. Importance of factors of behavior.

At the end, if the validation fails, one can conclude that the hypothesis being tested is invalid. In this case, the modeling team should consider alternative theories resulting in different testable hypotheses.

3.3 Two Application Examples

In this section, we demonstrate the usefulness of ABMC by example. We take two of the most well studied agent-based models and use the ABMC to fully describe them. The first, Schelling's seminal work on Segregation modeling (Schelling, 1969) is a good example of an abstract model of a social science concept that provides intriguing insights. The second, the Long House Valley project's Artificial Anasazi model (Dean et al., 2000), one of the first Computational Social Science models to be compared to data and have computational calibration efforts (Stonedahl & Wilensky, 2010b) and model discovery efforts (Gunaratne & Garibay, 2018) applied to. The purpose of choosing these examples is to allow future trans-disciplinary teams guidance toward the application of ABMC in their own CSS efforts.

3.3.1 Schelling's Segregation Model

The ABMC of the Schelling's Segregation Model (Schelling, 1969) is shown in Figure 3. The 9 building blocks are as follows.

Theory: The theory used is an individual-level description of discrimination. Motivation by statistics on the impact of racial segregation in the United States. (Pascal, 1967) states that segregation is driven by organized racial profiling based on the correlation of color with income (at the period of writing) and income with residential preference.

Key Concepts: Segregation is driven by individual-level neighbor preferences. Those of a similar race can require a minimum number of neighbors that are of the same race. Until this threshold is met, the individual is unsatisfied and seeks an alternate living situation.

Key Variables:

1. Parameter: Distribution of White threshold of tolerance of immediate neighbors of similar race
2. Parameter: Distribution of Black threshold of tolerance of immediate neighbors of similar race
3. Initial condition: White population density of the region
4. Initial condition: Black population density of the region

Key Relationships: $H_i = s_i > k_i$ where H_i is the happiness of the individual, s_i is the ratio of similar neighbors around the individual, and k_i is the desired minimum of neighbors of similar race by the individual.

If happiness is not met the agent should relocate to a new, free residence.

Hypothesis building data: Data on the correlation of race an economic status provided by (Pascal, 1967).

Hypothesis:

1. Segregation is the result of individual level need to be surrounded by neighbors of similar race.

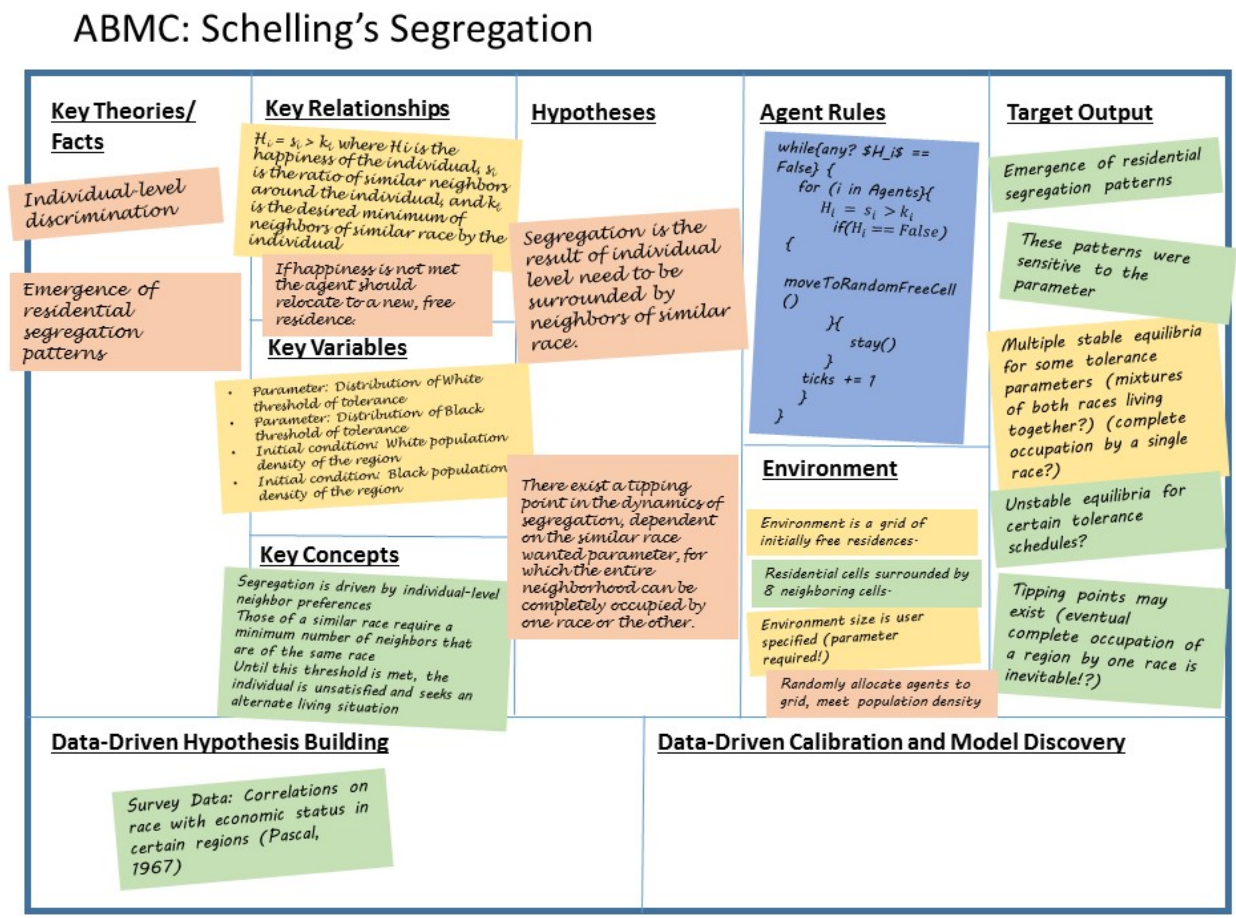

Figure 3: The Agent-Based Modeling Canvas of the Schelling's Segregation model.

2. There exist a tipping point in the dynamics of segregation, dependent on the similar race wanted parameter, for which the entire neighborhood can be completely occupied by one race or the other.

Agent Rules: Agent's behavior according to the following algorithm:

while *any?* $H_i == False$ **do**
  forall *i in Agents* **do**
    $H_i = s_i > k_i$ **if** $H_i == False$ **then**
      moveToRandomFreeCell()
    else
      Stay()
    end
    ticks += 1
  end
end

Environment:

1. Environment is a grid of initially free residences.
2. Each residential cell is surrounded by 8 neighboring cells.
3. Environment size is user specified (parameter required!)
4. Initialize grid at each simulation run start randomly allocating agents to the grid until the population density conditions are met.

Validation Data: None

Target Output:

1. The emergence of residential segregation patterns.
2. These patterns were sensitive to the parameter
3. There can exist multiple stable equilibria for some tolerance parameters and they may or may not allow mixtures of both races living together or complete occupation by a single race.
4. There may even exist unstable equilibria for certain tolerance schedules.

5. Depending on tolerance schedules tipping points may exist beyond which the eventual complete occupation of a region by one race is inevitable.

3.3.2 Artificial Anasazi

In figure 4 we consider ABMC applied to the Artificial Anasazi model (Dean et al., 2000) and some of the calibration (Stonedahl & Wilensky, 2010b) and model discovery efforts applied on it (Gunaratne & Garibay, 2017, 2018).

Theory:

1. Agriculture was a centerpiece of Pueblo cultures (Dean et al., 2000).
2. The typical household size at the period of study (800 AD to 1350 AD) consisted of five individuals on average (from archaeological ethnographic and demographic studies) (Dean et al., 2000).
3. Soil quality is determinant of farming potential and yield, with some zones of the Long House Valley being arable while others were not.

Key Concepts: The Kayenta Anasazi were an agriculture dependent civilization and fertility of land and availability of water would have affected the carrying capacity of the Valley.

Key Variables:

1. Initial condition: Map of the geography of the various fertility zones and water sources of the Valley along with annual water/drought estimates.
2. Initial condition: Variables of the land important to agriculture such as quality of soil and yield.
3. Initial condition: The annual nutritional need of households
4. Initial condition: Distribution of the threshold of the fertility of households
5. Initial condition: Distribution of the age of mortality of households
6. State variable: Quantity of food stored from previous harvest by household

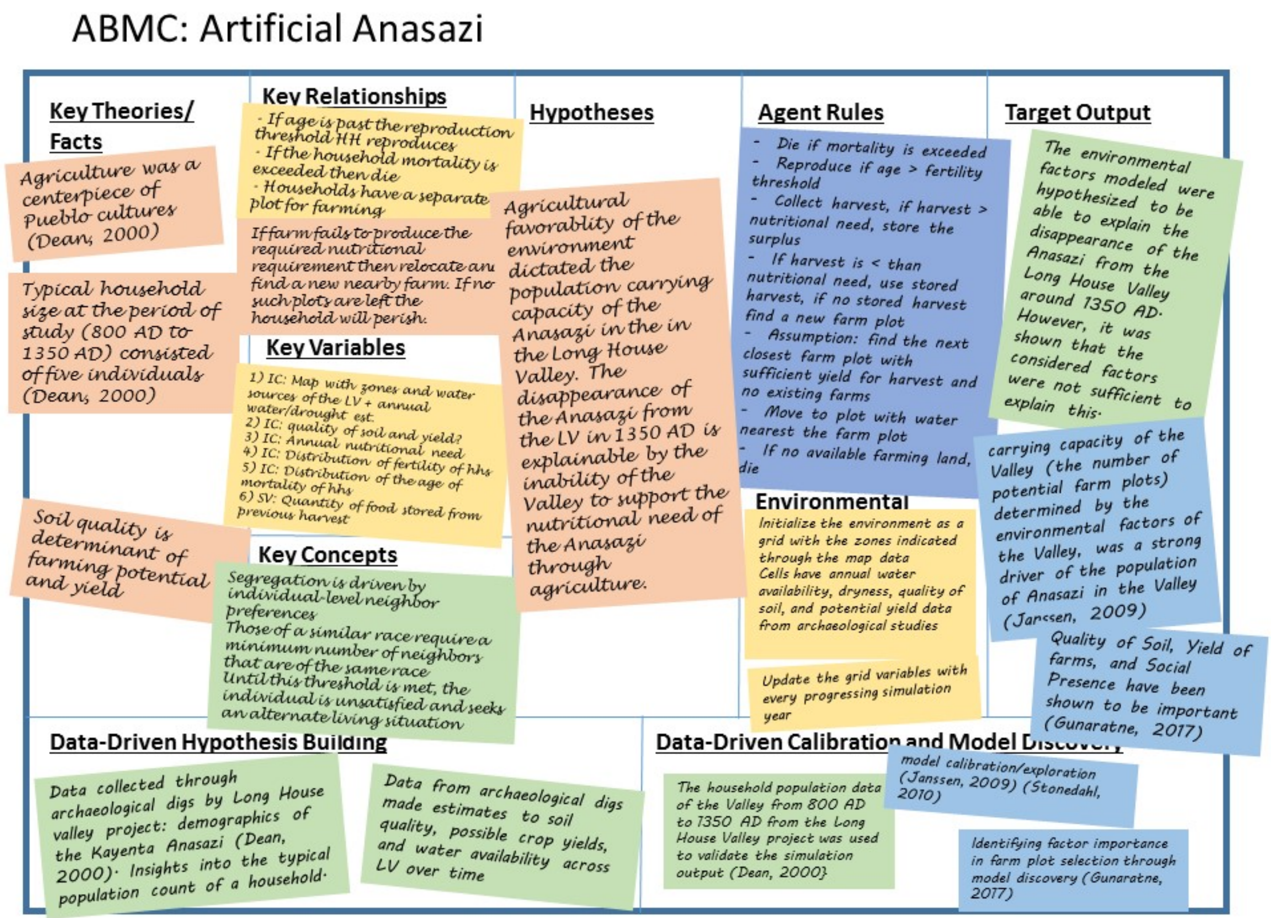

Figure 4: The Agent-Based Modeling Canvas applied to the Artificial Anasazi model, and later calibration and model discovery efforts performed on it.

Key Relationships:

1. If the household age is past the reproduction threshold of the agent it may reproduce.
2. If the household mortality is exceeded the household must die
3. Households have a separate plot for farming
4. If household's receive more harvest than their nutritional requirement the excess is stored
5. If the household's farm fails to produce the required nutritional requirement the household must relocate and find a new nearby farm. If no such plots are left the household will perish.

Hypothesis building data:

1. Data collected through archaeological digs during the course of the Long House valley project were used to determine the demographics of the Kayenta Anasazi (Dean et al., 2000). The data gave insights into the typical population count of a household (5 individuals).
2. Conditions of the Valley itself changed over the course of time. Data from archaeological digs made estimates to soil quality, possible crop yields, and water availability across the Valley.

Hypothesis: Agricultural favorablity of the environment dictated the population carrying capacity of the Anasazi in the in the Long House Valley. The disappearance of the Anasazi from the Valley in 1350 AD is explainable by the inability of the Valley to support the nutritional need of the Anasazi through agriculture.

Agent Rules:

1. Die if mortality is exceeded
2. Reproduce if age exceeds fertility threshold
3. Collect harvest, if harvest is greater than nutritional need, store the rest
4. if harvest is less than nutritional need, use stored harvest, if no stored harvest find a new farm plot

(a) (Assumption in original model) find the next closest farm plot with sufficient yield for harvest and no existing farms
(b) Move household to plot with water nearest the farm plot
(c) If no available farming land, die

Environment:

1. Initialize the environment as a grid with the zones indicated through the map data
2. Populate grid with annual water availability, dryness, quality of soil, and potential yield data from archaeological studies.
3. Update the grid variables with every progressing simulation year.

Validation Data: The household population data of the Valley from 800 AD to 1350 AD from the Long House Valley project was used to validate the simulation output (Dean et al., 2000). This data was used for model calibration (Janssen, 2009; Stonedahl & Wilensky, 2010b) and for identifying factor importance in farm plot selection through model discovery (Gunaratne & Garibay, 2018).

Target Output: The environmental factors modeled were hypothesized to be able to explain the disappearance of the Anasazi from the Long House Valley around 1350 AD. However, it was shown that the considered factors were not sufficient to explain this.

In further studies, (Janssen, 2009) demonstrates that the carrying capacity of the Valley (the number of potential farm plots) determined by the environmental factors of the Valley, was a strong driver of the population of Anasazi in the Valley, than in comparison to the agent parameters (fertility, mortality, etc). In (Gunaratne & Garibay, 2018) this result is strengthened by discovering and ranking the importance of factors of farm selection through data driven Evolutionary Model Discovery. It is shown that Quality of Soil, Yield of farms, and Social Presence (two factors affecting the carrying capacity and the last an emergent result of agent decisions) are actually stronger determinants of the population in the Valley than the other variables.

4 Conclusion

We are now witnessing a paradigm formation at the intersection of social, computer and data science. This interdisciplinary field needs a common language for effective communication between researchers who adopt very different frameworks and cultures of research within their own disciplines. Moving forward, it is therefore very crucial to adopt this common language in order to gain a mutual understating of computational social science research effort by these three types communities. The agent-based modeling canvas offers a *lingua franca* by creating a framework that enables social, computer, and data scientists to speak the same language. A completed canvas articulates research requirements derived from the guiding theories, the relevant concepts and variables, and the relationships between variables. In doing so, a completed canvas facilitates the formation of testable hypotheses. In this way, the proposed framework embodied by the canvas aids in constructing agent-based models by sketching the expected target output and employing data and methods for validations as well as analysis of the output. The result allows interdisciplinary model creation, description, validation, improvement and evolution by visually organizing the research project into the 9 model building blocks, and in our view, substantially contributing toward the success of any CSS project.

References

Becker, G. (n.d.). *S. 1964, 1993 . human capital: A theoretical and empirical analysis, with special reference to education.* Chicago, University of Chicago Press. ISBN.

Bersini, H. (2012). Uml for abm. *Journal of Artificial Societies and Social Simulation*, *15*(1), 9.

Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences*, *104*(17), 7301–7306.

Bezerra, R. F., Jalloh, S., & Stevenson, J. (1998). Formulating hypotheses graphically in social research. *Quality and Quantity*, *32*(4), 327–353.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892-895. doi: 10.1126/science.1165821

Calvez, B., & Hutzler, G. (2005). Parameter space exploration of agent-based models. In *International conference on knowledge-based and*

intelligent information and engineering systems (pp. 633–639).
Casas, P. F. I. (2013). *Formal languages for computer simulation: transdisciplinary models and applications.* Igi Global.
Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on information theory*, *2*(3), 113–124.
Cioffi-Revilla, C. (2014). Introduction to computational social science. *London and Heidelberg: Springer*.
Comfort, L. K., Wukich, C., Scheinert, S., & Huggins, L. J. (2011). Network theory and practice in public administration: Designing resilience for metropolitan regions. In D. Menzel & H. White (Eds.), *The state of public administration: Issues, challenges, and opportunities* (p. 257-271). M.E. Sharp Inc.
Czarniawska, B. (2004). *Narratives in social science research.* Sage.
Darity, W. (1975). Economic theory and racial economic inequality. *The Review of Black Political Economy*, *5*(3), 225–248.
Davis, P. K., & O'Mahony, A. (2013). *A computational model of public support for insurgency and terrorism: a prototype for more-general social-science modeling* (Tech. Rep.). RAND NATIONAL DEFENSE RESEARCH INST SANTA MONICA CA.
Dean, J. S., Gumerman, G. J., Epstein, J. M., Axtell, R. L., Swedlund, A. C., Parker, M. T., & McCarroll, S. (2000). Understanding anasazi culture change through agent-based modeling. *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes*, 179–205.
Dipeolu, A., Cook-Cottone, C., Lee, G. K., Donnelly, J. P., Janikowski, T. P., Reynolds, A. L., & Boling, T. (2016). A concept map of campers perceptions of camp experience: Implications for the practice of family counseling. *The Family Journal*, *24*(2), 182–189.
Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, *4*(5), 41–60.
Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling.* Princeton University Press.
Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12.
Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up.* Brookings Institution Press.
Feagin, J. (2013). *Systemic racism: A theory of oppression.* Routledge.
Fiore, S. M., Smith-Jentsch, K. A., Salas, E., Warner, N., & Letsky, M. (2010). Towards an understanding of macrocognition in teams: de-

veloping and defining complex collaborative processes and products. *Theoretical Issues in Ergonomics Science*, *11*(4), 250–271.

Fiore, S. M., Wiltshire, T. J., Oglesby, J. M., OKeefe, W. S., & Salas, E. (2014). Complex collaborative problem-solving processes in mission control. *Aviation, space, and environmental medicine*, *85*(4), 456–461.

Fowler, M., & Scott, K. (2000). Uml distilled second edition. *A brief guide to the standard object modelling language*.

Fu, Z., Huang, F., Ren, K., Weng, J., & Wang, C. (2017). Privacy-preserving smart semantic search based on conceptual graphs over encrypted outsourced data. *IEEE Transactions on Information Forensics and Security*, *12*(8), 1874–1884.

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences.* MIT Press.

González-Bailón, S. (2013). Social science in the era of big data. *Policy & Internet*, *5*(2), 147–160.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., & Giske, D. L., Jarl ... DeAngelis. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, *198*(1-2), 115–126.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The odd protocol: a review and first update. *Ecological modelling*, *221*(23), 2760–2768.

Gunaratne, C., & Garibay, I. (2017). Alternate social theory discovery using genetic programming: towards better understanding the artificial anasazi. In *Proceedings of the genetic and evolutionary computation conference* (pp. 115–122).

Gunaratne, C., & Garibay, I. (2018). Evolutionary model discovery of factors for farm selection by the artificial anasazi. *arXiv preprint arXiv:1802.00435*.

Havekes, E., Bader, M., & Krysan, M. (2016). Realizing racial and ethnic neighborhood preferences? exploring the mismatches between what people want, where they search, and where they live. *Population research and policy review*, *35*(1), 101–126.

Haymovitz, E., Houseal-Allport, P., Lee, R. S., & Svistova, J. (2017). Exploring the perceived benefits and limitations of a school-based social–emotional learning program: A concept map evaluation. *Children & Schools*, *40*(1), 45–54.

Hjorth, A., Head, B., & Wilensky, U. (2015). Levelspace netlogo extension. *Center for Connected Learning and Computer-Based Learning.*

Evanston, IL.

Horowitz, E., Sahni, S., & Rajasekaran, S. (1997). *Computer algorithms c++: C++ and pseudocode versions.* Macmillan.

Janssen, M. A. (2009). Understanding artificial anasazi. *Journal of Artificial Societies and Social Simulation*, *12*(4), 13.

Kamsu-Foguem, B., Tchuenté-Foguem, G., & Foguem, C. (2014). Using conceptual graphs for clinical guidelines representation and knowledge visualization. *Information Systems Frontiers*, *16*(4), 571–589.

Kennedy, P. (2003). *A guide to econometrics: Fifth edition.* MIT Press.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research: Fourth edition.* Harcourt College Publishers.

Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational research*, *42*(1), 43–57.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference on qualitative research.* Princeton University Press.

Köhler, M., Langer, R., Von Lüde, R., Moldt, D., Rölke, H., & Valk, R. (2007). Socionic multi-agent systems based on reflexive petri nets and theories of social self-organisation. *Journal of artificial societies and social simulation*, *10*(1).

Koliba, C., Meek, J. W., & Zia, A. (2011). *Governance networks in public administration.* CRC Press.

Mason, W., Vaughan, J. W., & Wallach, H. (2014). *Computational social science and social computing.* Springer.

McFarland, D. A., Lewis, K., & Goldberg, A. (2016). Sociology in the era of big data: The ascent of forensic social science. *The American Sociologist*, *47*(1), 12–35.

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? a field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, *100*(6), 1678.

Minsky, M. (1975). Minskys frame system theory. In *Tinlap75: Proceedings of the 1975 workshop on theoretical issues in natural language processing* (pp. 104–116).

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., … Schwarz, N. (2013). Describing human decisions in agent-based models–odd+ d, an extension of the odd protocol. *Environmental Modelling & Software*, *48*, 37–48.

Novak, J. (2010). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations.* Routledge.

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value proposition design: How to create products and services customers want.* John Wiley & Sons.

Ostrom, E. (2005). *Understanding institutional diversity.* Princeton University Press.

Pascal, A. H. (1967). *The economics of housing segregation* (Unpublished doctoral dissertation). Elsevier.

Rafiei, M., & Kardan, A. A. (2015). A novel method for expert finding in online communities based on concept map and pagerank. *Human-centric computing and information sciences*, *5*(1), 10.

Ramirez, C., & Valdes, B. (2012). A general knowledge representation model of concepts. In *Advances in knowledge representation.* InTech.

Roth, W.-M., & Roychoudhury, A. (1992). The social construction of scientific concepts or the concept map as device and tool thinking in high conscription for social school science. *Science education*, *76*(5), 531–557.

Schelling, T. C. (1960). *The strategy of conflict.* Harvard University Press.

Schelling, T. C. (1966). *Arms and influence.* Yale University Press.

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

Schwendimann, B. A., & Linn, M. C. (2016). Comparing two forms of concept map critique activities to facilitate knowledge integration processes in evolution education. *Journal of Research in Science Teaching*, *53*(1), 70–94.

Shaughnessy, J. J., & Zechmeister, E. B. (1985). *Research methods in psychology.* Alfred A. Knopf.

Sowa, J. F. (1983). Conceptual structures: information processing in mind and machine.

Stonedahl, F., & Wilensky, U. (2010a). Behaviorsearch [computer software]. *Center for Connected Learning and Computer Based Modeling, Northwestern University, Evanston, IL. Available online: http://www. behaviorsearch. org*.

Stonedahl, F., & Wilensky, U. (2010b). Evolutionary robustness checking in the artificial anasazi model. In *Aaai fall symposium: Complex adaptive systems* (pp. 120–129).

Tintner, G. (1953). The definition of econometrics. *Econometrica*, *21*(1), 31-40. doi: 10.2307/1906941

Weilkiens, T. (2011). *Systems engineering with sysml/uml: modeling, analysis, design*. Elsevier.

Zeigler, B. P. (1989). Devs representation of dynamical systems: Event-based intelligent control. *Proceedings of the IEEE*, *77*(1), 72–80.