September 5, 2003

# Validating Agent –Based Models: From Metaphysics to Applications

Darren Schreiber, J.D., C.Phil.

Prepared for the Midwestern Political Science Association's Annual Conference in Chicago, April 2002.

This is a work in progress. Please contact Darren Schreiber (dschreib@ucla.edu) with comments and critiques. You can find the most current version of this paper and updates on his research agenda at http://www.bol.ucla.edu/~dschreib.

In this paper, I will contend that multi-mode validation of the implications of agent-based models can support the quest for social science knowledge in a postmodern era. I will first layout a brief description of some of the epistemological problems that contemporary social science faces. Next, I will describe an epistemology that addresses these problems in light of the postmodern critique. The claim is that a process of validating models with a variety of tools can accomplish knowledge goals while responding to many criticisms coming from postmodernism. The next section makes the epistemological claim a bit more specific by defining more of the terms. Then, I proceed to specify a number of validation tools. Finally, I show how I have applied those tools to a few agent-based models that I have developed.

Rene Descarte (Descartes 1956 (original 1637)) embarked on a journey with lofty ambitions to be sure. He wanted to find some truth that he could not doubt. This intellectual walk was tremendously productive. His "cogito ergo sum" and his ontological proof for the existence of God, were both fruits of this trip. Even more important, was the path paved for future scientists and philosophers. With his example, many found an increasing belief that solutions to the really hard problems were just down the road a bit.

This hopeful enthusiasm was typified by the publication of <u>Principia Mathematica</u> (Whitehead and Russell 1997 (original work 1910)) by Alfred North Whitehead and Bertrand Russell, which endeavored to root mathematics in a logically consistent set of axioms. Unfortunately, a number of results in the early twentieth century severely undermined the optimism that had hit a crescendo at the end of the nineteenth. Kurt Gödel (Godel 1992 (original work 1931)) proved that some propositions would be inherently undecidable regardless of the axioms, undermining the Cartesian faith in deduction. Heissenberg showed that some information about physical systems had to be uncertain, undermining the belief in empirical certainty. Wittgenstein's private language argument undermined the definitiveness of language. And, Einstein's relativity work undermined the certainty of time and space, bending even the geometry that bears Descartes' name.

Richard Bernstein (Bernstein 1988) labels this upended desire for certainty the "Cartesian Anxiety." As relativity spread from physics to morality, the anxiety heightened. Postmodernism brought a critique to most facets of the academy, following Descartes' passion for doubting, but finding nothing undoubtable. "The Method" so beloved by Cartesian heirs became a target of scorn as did "Truth," "Law," "Cause and Effect," and many other comforting ideas. Bernstein did a great job articulating the need to get Beyond Objectivism and Relativism as the debate was plaguing postmodern science, but did not provide clear road map.

An empirical study of epistemology, however did provide a vision of what post-postmodern knowledge might look like. In Women's Ways of Knowing (Belenky et al. 1986), a group of feminist psychologists and philosophers went out to look at how a diverse group of women made knowledge claims. They discovered four broad groupings: silent knowers, women (typically abused) who made no knowledge claims at all; received knowers, who relied on the truth claims of others; subjective knowers, women who relied on intuition or other subjective forms of knowledge; procedural knowers, who had a method for discerning; and constructed knowers, who integrated a full variety of epistemological tools and interacted with others in their truth-seeking.

My first contention in this paper is that a successful social science in the wake of postmodernism needs to function like a community of "constructed knowers." Without the consensus on method or truth that Cartesians shared, we should embrace the natural diversity of methods that contemporary social science reflects. We should endeavor to make our received ideas, subjective insights, and procedures commensurable or risk becoming "silent knowers" like so many academics who fear making knowledge claims in the wake of postmodernism.

The constructed knowing epistemology makes even more sense when we consider an appropriate ontological framework. Immanuel Kant (Kant 1990 (original work 1781)) distinguished between the "noumena" and the "phenomena." The "phenomena" is things as they appear, our world of ready sense impressions. But the "noumena" is things as they actually are. Kant's critique of Descarte is that we simply cannot see things as they actually are. We may postulate a correspondence between noumena and phenomena (Kant's "Postulate of Rational Cosmology"), but this is merely a postulate, not certain truth.

I will contend that we should follow Kant in hoping (i.e. postulating) that the phenomena we observe correspond with the noumena, but retaining the humility that this may not be the case. I also concur with McKelvey (Henrickson and McKelvey 2001) that a "model -centered science" is the right approach to the postmodern epistemological problem. In my version of McKelvey's approach, scientists have theories (i.e. beliefs about the world) that are formalized as models, which are tested against phenomena, which we postulate correspond to noumena.

Freidrich Hayek (Hayek 1967) supports this model-centered view by arguing that the world has no closed systems. While the concepts of laws and cause and effect maybe appropriate for simple phenomena, we must add "ceteras paribus" assumptions one after another when we deal with the kind of complex things that social scientists endeavor to talk about. In such a complex world, we may need to abandoned simple notions of cause and effect and instead describe "transition functions" (Holland 1998) that can connect time T with time T+1. Models, as operationalized thought experiments, can serve this role of transition function.

Rom Harré (Harré 1970) suggested a useful taxonomy of models based upon the ways a model can relate to its subject and source. He begins with three general classes of models: homeomorphs, protomorphs, and paramorphs. The class of protomorphs is comprised of logical icons and geometrisations. Harré describes these protomorphs in relationship to homeomorphs and paramorphs, as lamprey are to fish … with more development protomorphs may become proper models. In a homeomorph, the source and the subject of the model are the same. For instance, "a toy car is an example of a homeomorph, as the subject of the toy car is a real car, and a real car is also the source of the model. The use of a computer (source) as a model of human cognitive processes (subject) is an example of a paramorph (Azevedo 1997): 126."

As suggested by Azevedo, computer models like Agent-Based Models, fall into the class of paramorphs. In the subclass of partial paramorphic analogues, the input and output of the subject and the model are exactly alike, but the causal processes of the model and the subject differ. A computer simulation of a person's purchasing decision may have the exact same inputs and outputs, but their method of processing the information are distinct.

However, most agent-based models do not produce the exact same outputs as their target. Rather "the model process and model input and output are analogues of the subject process and the subject input and output respectively (Azevedo 1997): 128." This makes most agent-based models complete paramorphic analogues, they work similarly to the real world and produced output that is similar, but they not exactly the same.

Some experimental economists have taken agent-based modeling in a slightly different direction by using real humans as their agents. In the California Social Sciences Experimental Laboratory (CASSEL -- http://www.cassel.ucla.edu/), many economists have been running models where undergraduates are given the information that agent-based modelers typically give their computerized agents. Harré would classify these models as paramorphic homologues because the process for decision making (the undergraduate buying stocks) is the same in the source problem and in the model.

Harre's taxonomy makes a clear distinction between exact equivalence and analogous equivalence. However, I would contend that equivalence is much more complicated and nuanced. Is an undergraduate playing with fake credits really using the same processes as a sophisticated investor planning their retirement? How close does a model have to be to the source to be equivalent, if the tenth decimal place is the accurate? If the second decimal place is the accurate? While the taxonomy is a useful conceptual organizing tool, its actual application may be less clear. Where on the continuum does a particular results fall in terms of equivalence?

First, we might expect that the direction of the relationships between two variables is consistent. We might expect if the correlation between $x_{source}$ and $y_{source}$ is positive, then the relationship between $x_{model}$ and $y_{model}$ should also be positive. Or, we might expect that the magnitude of changes is similar. For instance, if the source shows that large changes in a particular x variable have only negligible impact on a particular y variable, then we should expect the model to reflect such a relationship. Furthermore, we might be content with our model even if it got the direction of the relationship wrong, given that it correctly predicted a small magnitude.

Beyond the simple kinds of equivalence in the first model, we might think about the derivatives of the relationships between an x and a y. While both the source and the model agree that increases in x lead to increases in y, we may require that if the source has a geometric increase in y that the model have a geometric increase in y as well. Or, we may find that the source relationship between x and y is more complex with y increasing across some range of x and decreasing in another range. I would contend that equivalence of this type is a more stringent standard than the first kind.

Another kind of equivalence would be to compare the digits in the y for the source and the model. If we find that a ten-unit increase in x will cause a 250-unit increase in y for the source, we might be content with the model predicting a 270-unit increase in y. Or, we could imagine tightening the standards for more and more accuracy, depending on the state of the science. John Zaller has often joked that while you could have gotten tenure for getting showing the sign of a relationship in the data in the early days of political science, the discipline is now debating the second decimal place of some coefficients. In contrast, the recent search for the Higg's particle failed because the results could not be verified at physics' benchmark of $p<0.0001$ (http://spot.colorado.edu/~vstenger/Briefs/Higgs.html)

To this point, I have been describing the kinds of equivalence between only a single x and a single y in the model and source. However, one would hardly bother making an agent based model if the problem could be described so simply. More typical are models with a string of inputs and a string of outputs to be compared in a multi-dimensional space. Getting similar results for the combined effects of two or more x variables on two or more y variables between the source and model is much more impressive than mere one to one correspondences. And, this is more impressive still when the relationships are non-linear.

This description of some of the issues in thinking about equivalence of results is really just a precursor to the discussion about tests. Now that we have some sense of the kinds of equivalence in results we might look for, what kinds of test might we run with the model? Because most agent-based models are partial paramorphic analogues, we are not merely concerned with looking at the inputs and outputs. Rather, we are also concerned with evaluating the process as well. This makes the study of agent-based models much more complex than something like the typical regression analysis, but it also has the potential to give us more insight into causal inference as well.

In the context of postmodernism viewing model testing as "proof" would be highly problematic. We are not going to find truth in the absolute noumenal sense. Rather, we are going to validate a model for a particular purpose.

Putting validation in the context of purpose becomes very important if cannot assume ontological "Truth." While the video game agent-based model "The Sims" may be valid as a model of human behavior for the purpose of entertainment, it would be lacking if we tried to organize our families around principles we learn from playing the game. Some models are "operationalized thought experiments" (Dennet 1998) only designed to answer speculative what-ifs. Other models seek to integrate data and formal methods. And, still others are used to guide policy planning and implementation. The contention here is that when we validation a model, we are validating it for a particular purpose.

This further means that model validation, like model construction, should be a process. After only a few lines of code are written, we should validate that these parts function as our theory said they would. This is not only good programming practice, it is good modeling in general. Our theories about the distributions of variables in a survey should be validated, before we go off running hypothesis tests in the form of regression analysis.

Kathleen Carley has aptly described this process as Model-Test-Model. The result of the modeling is not "proof" or "truth." Instead, we get a model that has been validated in a particular context and revised in light of the validation process. The model always awaits further validation and further tests.

The great benefit of this process-centered epistemology is that our models should not be simply shelved after success, but should be part of a vital and cumulative intellectual life. A well-constructed model should be designed with this process epistemology in mind. Too many statistical models illustrate a very small point in the literature and fail to engage any of the many potential implications. If our statistical validation threshold is 0.05 and we have tested only one implication then we have a relatively high chance of false positives. In contrast, John Zaller has advocated that the solution to the data-mining problem is to test your model against multiple implications. In one of his papers (Zaller and Feldman 1992), he outlined sixteen implications from his theory and demonstrated all but one of those at the 0.05 level. The joint probability of a spurious correlation for all those implications simultaneously is staggeringly small. Perhaps we can claim to be a political "science" after all.

As mentioned above, well-constructed models can be part of a cumulative knowledge process as well. The "constructed knower" of Belenky et al's work does not rely merely upon her own insights. Rather, she is connected to authority, intuition, history, reason, and her community of other knowers. As part of an intellectual community, we can validate our model against the claims made by others' models and increase the credibility of our beliefs.

Furthermore, this process-centered epistemology highlights the intellectual role of "levels" (Holland 1995) and "emergence" (Holland 1998). As we validation our assumptions of the micromotives of our agents we simultaneously validate our understanding of their macrobehaviors (Schelling 1978), and vice versa. If valid behaviors emerge from models in ways we did not anticipate (Johnson 2001), this can be much more compelling than merely constructing a model to do something and having it do that. All models have multiple implications and the practice of exploring only one implication per publication is extremely limiting in terms of validation testing.

Well this metaphysical discussion of validation is nice, but what is to be done (as Lenin would ask)? Fortunately we political scientists are not alone on our quest for validity. Most of engineer's work is done in modeling. Their blue prints or scale models or CAD/CAM designs are not just fanciful thought experiments, however. They are actually built and lived with in our daily lives. As a result, validation is a critical issue, perhaps <u>the</u> critical issue, for engineers. And, they have entire conferences on the subject of model validation and verification. Robert Sargent is a leader in the field of validation in engineering. Much of what I present in this next section is based on his thinking.

To connect the list of validation techniques I present below to the ontological framework described above, I have categorized the tests by their ontological footing. The first category I describe is "Theory – Model" tests. In this category, we are essentially looking to see if the conceptions that we have in our mind are met by their operationalization in the model. The second category of "Model – Model" tests connects the model we have developed to other pertinent models. The third category of "Model – Phenomena" tests connects the model we have operationalized to the phenomena that we observe. And, the final category of "Theory – Model –

Phenomena" tests examines the model in the context of theory and phenomena simultaneously. Because models, theories, and phenomena often overlap, these categories are more constructed conveniences than concrete, noumenal truth. But, that probably goes without saying in the context of this paper.

In "Theory—Model" tests, we are asking whether the model matches the theory we had. As operationalized thought experiments, models can have a transparency (when code is clearly written and assumptions well laid out) that our raw theories may not enjoy. In the "Face Validity" test, we present the model to persons who are knowledgeable about the source problem and ask whether this model is reasonably compatible with their knowledge and experience (Sargent 1987). Certainly, a first step in this process would be to ask whether the model results fit with the sensibilities of the substantive expert in the modeling team. Presenting the model at conferences and in publications is another way of getting the kind of feedback needed to appraise the facial validity of the model. In the face validity test, the source of the data is the broad knowledge and experience of the substantive experts.

We might also compare the model with results from narrative descriptions of the source. This "Narrative Validity" test is similar to face validity, but relying on published accounts about the process or the input and output of the source is a more formal step. Furthermore, the narrative validity test is more amenable to consensus in that a team of scholars will be likely to have more disagreement about whether a model fits their experience than whether a model fits a particular narrative description. Sargent's formulation of this test is similar to the "Theoretical Validity" test proposed by Kathleen Carley (1999).

In the "Turing Test" (Sargent 1987), named after mathematician Alan Turing, the test is whether a group of experts could tell the difference between data generated by the model and data generated by the real world. Turing's original application was to use this as a test of artificial intelligence. If a machine can fool another human into believing that its communications come from a person rather than a machine, it must have achieved a high level of artificial intelligence. Similarly, if the output of our computational modeling is indistinguishable from real events, we have achieved a substantial level of validation.

A final type of test I am tentatively labeling as the "Surprise Test." Sometimes our agent-based models do something well beyond what we designed them to do. If these behaviors match our theories about the world, but emerged from our models as unanticipated implications, then we can have a very interesting kind of validity.

The "Model—Model" tests emphasize the connectedness of the epistemological framework I am articulating. "Docking" is a classic test of validity. In this case, the source data comes from another model of the problem. This method has also been called "Cross-Model Validity" (Carley 1999). The techniques for docking vary wildly. We may compare our model with another model that has been validated to some level. Or, we might see what changes in the parameters or algorithm are required in our model to allow it to replicate results from another. Or, we might even have a separate team re-code our model.

Similar to docking is "Analytical Validity." In this test, we can use the results from analytic methods or formal proofs. In many social sciences, formal methods have been used to make predictions for a specific set of circumstances. We may set our model parameters to the assumptions of the formal model and compare the results.

Akin to "Analytic Validity Tests" are "Fixed Values Tests." Sargent suggests that in cases where results can be hand calculated we use a set of fixed values and compare the results of hand calculation with the results of the model. For the domains that most agent-based models cover this is probably not appropriate, but there may be occasions when one is able to compare model results against easily calculable results. For instance, Schelling (Schelling 1978) initially worked out the implications of his theory of housing segregation by testing it with a chessboard and some spare change. He later compared his manual manipulations with the computer output.

Comparing our models with real world data is the domain of the "Model—Phenomena" tests. For instance, we might use "Historical Data Validity" tests to compare our model results with the results of previously collected data. Similarly, we might use "Predictive Data Validity" tests and have our model forecast results and compare our forecast with actual outcomes (Sargent 1987). Another related concept that mixes historical data and prediction is "Out of Sample Forecast" tests. Familiar in the statistical literature, we can use a portion of our historical data to tune the model and another portion of our data to test the predictive outcomes of our model.

Since many fields of social science are not as practiced in running experiments, it is also worthwhile to remember that we might conduct "Experimental Data Validity" tests when we can create good experiments for testing our hypotheses. In particular, we should consider experimental data validity on the micromotives parts of our model since those are usually more tractably testable than the macrobehaviors. The final "Model—Phenomena" category is "Event Validity Tests" which compare the occurrence of particular events in the model with the occurrence of the events in the source data.

Many of the tests that I am putting into the "Theory – Model – Phenomena" test category are useful for understanding the "phenomenal" robustness of the theoretical implications of the model. In this category, we are looking at the theory and the phenomena at the same time we tweak the model to see how well it bridges the divide.

"Extreme-Bounds Analysis" or "Extreme Condition Testing" is one method for robustness testing a model. In these tests, the experimenter uses very high or very low values for the inputs and/or parameters of the model to test whether the model continues to make sense at the margins (Leamer 1985). For instance, we should be surprised if our model eliminates all the agents and yet we still have trade. While a model that generates absurd results for extreme values may not need to be rejected purely on that ground, we should at least bracket any results we claim with a warning about the model failures. Extreme bounds analysis can be particularly useful when dealing with models where the parameter values are arbitrary. If we set up a utility function for some of our agents and the average utility is 50, this value may not have any real world value to correspond with. Thus, we can use extreme bounds analysis to test the implications of various extreme values for the model parameters. This helps us to validate that the parameters are relatively sensible in relation to each other even if we do not have a real world value to compare with.

A related method is "Global Sensitivity Analysis", proposed by Ed Leamer (Leamer 1985). In this technique, we adjust the parameter settings of the model and the inputs to the model and look to see how sturdy or fragile the results are. If particular results evaporate when the model is tuned to slightly different parameters then we should be concerned about making robust claims.

However, running a good set of global sensitivity analysis tests is not a trivial issue. If we imagine a model that has a mere five parameter settings, sweeping those parameters with 100 distinct values for each parameter requires 500 runs of the model.[1] However, if we believe that there might be any interactions among parameters, we should really sweep them in combination. If we sweep the space of pairs of those parameters for 100 distinct values each, then we require 10,000 runs per pair, which would require 150,000 runs to get all the possible pairs of the five parameters. We would need 10 billion runs to sweep all five parameters simultaneously with 100 distinct values per parameter. Even at one second per run, this would still take 317 years of serial processing time

Since sweeping the entire model space is such a problem, John Miller proposed the Automated Non-Linear Testing System (ANTS) as a tool for exploring the model space more efficiently. Miller's method involves running the target agent-based model inside of a model where the agents try to break the target model. The ANTS are given parameter settings to use in the model and their fitness is determined by how far from the expected model results they can go with minimal changes to the parameter settings. By using a genetic algorithm for such a purpose, a very complex parameter topology can be explored relatively quickly. For non-robust models, ANTS can identify a set of parameter values that give inconsistent results but are very similar to the default parameters of the model.

In addition to using the ANTS genetic algorithm to break the model, one might also use ANTS to tune the model. Instead of finding parameter values that diverge from expected results, we could use source data and tune the model parameter settings to the known output from the source data. By partitioning a data set and using ANTS with half the data to tune the model and ANTS with the other half to break the model, we could implement the "out of sample" validation test in combination with the robustness "ANTS" test.

---

[1] I usually use the extreme-bounds analysis first to determine an appropriate range for performing a global sensitivity analysis. This reduces the one-dimensional space that we will search for each

To this point we have been discussing testing the entire model as a single entity. We may also test robustness and validity by looking at the components. "Validating substructures" allows us to look at the performance of individual components of the model. For instance, we may have individual level decision data that we can compare to the agents in our model. This allows a modeler to study both the micro-motives as well as the macro-behaviors of the model. Similarly, we might use "Degenerate Tests" of interrupting some components of the model and noting the impact on overall results Or, we may uses "Traces" testing to look at individual agents as they work through the modeling environment (Sargent 1987).

As a related issue, we could use “Animation Validity” tests to compare the visually displayed qualities of the model with the qualities observed in source data. As the other members in this category, such test combines our theoretical expectations of the model and our observations about the model and real world phenomena. The category of “Theory—Model – Phenomena” tests are thus extremely important, both as theory building tools and as insures of robustness.

In my own work, I have used a number of the tests described above to validate the agent-based models that I have developed. Below, I review three models I have worked on and the validation tools I used. The first model is a simple model of a cocktail party, designed mostly as a programming exercise. The second model is much more sophisticated and illustrates the formation of political parties. The third model comes out of the Schelling tradition and is a sophisticated framework for studying housing segregation.

In the cocktail party simulation, Troy Tessier and I wrote a very simple model where agents start off at a random location in a square room. They look for interesting

---

parameter.

agents talk with and their interest in each of these people decays as they talk with them. When the they are bored enough, they go looking for someone new to talk with.

This very stylized model was designed, coded, debugged and validated in a single evening. We were not making major sociological claims, just showing a proof of concept. We validated the model with the facial validity test; making sure that they dynamics of the agents looked right. You could image this simulation showing the movement of real people at a party. It would not pass the Turing Test since people ended up a bit too tightly clumped. And, when we ran a tracing test by following a single agent through out the simulation, her movements looked sensible.

The political party formation model in contrast is one of my major research projects. It has been designed and refined over a number of years and I have visions for extensions of the model that will take many more years to fully implement. The model makes some very standard formal theory assumptions about the preferences and behaviors of political actors and the agents actions are based on only five rules.

This model connects to the two party theory of Duverger's law, the minimum winning coalitions result of Riker (Riker 1962), the median voter theorem of Black (Black 1958), and the realignment work of Sundquist (Sundquist 1983). In the process of validating this model, I started with face validity tests to see if the basic results looked right. I validated the substructures, often using fixed values tests to make sure that the agents were doing what I thought they should do. I had surprising results, such as getting the median voter theorem result when my model was much simpler than I thought it would need to be.

As the modeling process developed, I looked to make sure that my results match formal theory predictions (analytical validity), empirical data (historical data validity), and the work of some other agent-based models (docking.) These tests were done with varying levels of sophistication. In some cases, I was merely looking for simple equivalence. In other cases, I ran the model hundreds of times to ensure that the results were robust across a variety of parameter settings.

The housing segregation model that I have developed with Rick Sander has gone through even more rigorous testing. After designing the model, I spent a full summer testing its implications under a variety of conditions. We validated substructures with survey data where citizens of five different cities were asked to rank their preferences in a manner very similar to the task we have our agents perform. We presented our results to subject area experts who looked at the facial validity of the work and confirmed similarity of its output to their perceptions of how actual cities developed. We tested it against multiple implications by listing the implications of the theory, getting real data from other sources and comparing our results with that data.

Later, we looked at the robustness of those findings. I ran extreme bounds analysis, by tweaking the parameters of the model to the extremes and checking what happened. Having determined the logical boundaries of our parameter settings, I ran a global sensitivity analysis by looking at model results across a wide range of parameter settings. This allowed us to understand how movement in each of our parameters would affect the various dependent variables. Furthermore, we developed response surface maps where we would vary multiple parameters simultaneously and study their effects.

As is evident from this brief overview of validation practices, agent-based modelers have a large toolbox to draw from in evaluating their models. Agent based modeling has great potential as a tool that can implement our theories, connect to our data, and experiment with an world that we could not otherwise alter. ABM's are also a tool perfectly suited to address the Cartesian Anxiety remaining after postmodernism.

**Works Cited**


Azevedo, Jane 1997. *Mapping Reality: An Evolutionary Realist Methodology for the Natural and Social Sciences*. Albany, NY: State University of New York.

Belenky, Mary Field, et al. 1986. *Women's Ways of Knowing: The Development of Self, Voice, and Mind*. New York: Basic Books.

Bernstein, Richard J. 1988. *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*. Philadelphia: University of Pennsylvania Press.

Black, Duncan 1958. *The Theory of Committees and Elections*. New York: Cambridge University Press.

Dennet, Daniel C. 1998. *Brainchildren: Essays on Designing Minds*. Cambridge, Massachusetts: MIT Press.

Descartes, René 1956 (original 1637). *Discourse on Method*. New Jersey: Prentice-Hall.

Godel, Kurt 1992 (original work 1931). *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*: Dover.

Harré, Rom 1970. *The Principles of Scientific Thinking*. London: Macmillan.

Hayek, F.A. 1967. "The Theory of Complex Phenomena." *Studies in Philosophy, Politics, and Economics*. Chicago: University of Chicago Press: 22-42.

Henrickson, Leslie and Bill McKelvey 2001. *Foundations of "New" Social Science: Institutional Legitimacy from Philosophy, Complexity Science, Postmodernism, and Agent-Based Modeling*. "Paper prepared for presentation at the Annual Meeting of the " Sackler Colloquium on Agents, Intelligence, and Emergent Human Organization, Irvine, California.

Holland, John H. 1995. *Hidden Order: How Adaptation Builds Complexity*. New York: Addison-Wesley Publishing.

Holland, John H. 1998. *Emergence: From Chaos to Order*. Reading, Massachusetts: Perseus Books.

Johnson, Steven 2001. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Scribner.

Kant, Immanuel 1990 (original work 1781). *Critique of Pure Reason*. London: MacMillan.

Leamer, Edward E. 1985. "Sensitivity Analyses Would Help." *American Economic Review* 75 : 308-313.

Riker, William 1962. *The Theory of Political Coalitions*. New Haven: Yale University Press.

Sargent, Robert 1987. *An Overview of Verification and Validation of Simulation Models*. "Paper prepared for presentation at the Annual Meeting of the " Winter Simulation Conference.

Schelling, Thomas C. 1978. *Micromotives and Macrobehaviors*. New York: W.W. Norton & Company.

Sundquist, James L. 1983. *Dynamics of the Party System: Alignment and Realignment of Political Parties in the United States*. Washington, D.C.: The Brookings Institution.

Whitehead, Alfred North and Bertrand Russell 1997 (original work 1910). *Principia Mathematica*: Cambridge University Press.

Zaller, John R. and Stanley Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences." *American Journal of Political Science* 36 : 579-616.