[Joshua M. Epstein](#) (2008)

Why Model?

For information about citing this article, click [here](#)

Abstract

This lecture treats some enduring misconceptions about modeling. One of these is that the goal is always prediction. The lecture distinguishes between explanation and prediction as modeling goals, and offers sixteen reasons other than prediction to build a model. It also challenges the common assumption that scientific theories arise from and 'summarize' data, when often, theories precede and guide data collection; without theory, in other words, it is not clear what data to collect. Among other things, it also argues that the modeling enterprise enforces habits of mind essential to freedom. It is based on the author's 2008 Bastille Day keynote address to the Second World Congress on Social Simulation, George Mason University, and earlier addresses at the Institute of Medicine, the University of Michigan, and the Santa Fe Institute.

1.1

The modeling enterprise extends as far back as Archimedes; and so does its misunderstanding. I have been invited to share my thoughts on some enduring misconceptions about modeling. I hope that by doing so, I will give heart to aspiring modelers, and give pause to misguided critics.

Why Model?

1.2

The first question that arises frequently—sometimes innocently and sometimes not—is simply, "Why model?" Imagining a rhetorical (non-innocent) inquisitor, my favorite retort is, "You *are* a modeler." Anyone who ventures a projection, or imagines how a social dynamic—an epidemic, war, or migration—would unfold is running *some* model.

1.3

But typically, it is an *implicit* model in which the assumptions are hidden, their internal consistency is untested, their logical consequences are unknown, and their relation to data is unknown. But, when you close your eyes and imagine an epidemic spreading, or any other social dynamic, you are running *some* model or other. It is just an implicit model that you haven't written

down (see Epstein 2007).

1.4

This being the case, I am always amused when these same people challenge me with the question, "Can you validate your model?" The appropriate retort, of course, is, "Can you validate yours?" At least I can write mine down so that it can, in principle, be calibrated to data, if that is what you mean by "validate," a term I assiduously avoid (good Popperian that I am).

1.5

The choice, then, is not whether to build models; it's whether to build *explicit* ones. In *explicit* models, assumptions are laid out in detail, so we can study exactly what they entail. On these assumptions, *this* sort of thing happens. When you alter the assumptions *that* is what happens. By writing explicit models, you let others replicate your results.

1.6

You *can* in fact calibrate to historical cases if there are data, and can test against current data to the extent that exists. And, importantly, you can incorporate the best domain (e.g., biomedical, ethnographic) expertise in a rigorous way. Indeed, models can be the focal points of teams involving experts from many disciplines.

1.7

Another advantage of explicit models is the feasibility of sensitivity analysis. One can sweep a huge range of parameters over a vast range of possible scenarios to identify the most salient uncertainties, regions of robustness, and important thresholds. I don't see how to do that with an implicit mental model. It is important to note that in the policy sphere (if not in particle physics) models do *not* obviate the need for judgment. However, by revealing tradeoffs, uncertainties, and sensitivities, models can *discipline the dialogue* about options and make unavoidable judgments more considered.

Can You Predict?

1.8

No sooner are these points granted than the next question inevitably arises: "But can you predict?" For some reason, the moment you posit a model, prediction—as in a crystal ball that can tell the future—is reflexively presumed to be your goal. Of course, prediction might be a goal, and it might well be feasible, particularly if one admits statistical prediction in which stationary distributions (of wealth or epidemic sizes, for instance) are the regularities of interest. I'm sure that before Newton, people would have said "the orbits of the planets will never be predicted." I don't see how macroscopic prediction—*pacem* Heisenberg—can be definitively and eternally precluded.

Sixteen Reasons Other Than Prediction to Build Models

1.9

But, more to the point, I can quickly think of *16 reasons other than prediction* (at least in this bald sense) to build a model. In the space afforded, I cannot discuss all of these, and some have been treated *en passant* above. But, off the top of my head, and in no particular order, such modeling goals include:

1. Explain (very distinct from predict)
2. Guide data collection
3. Illuminate core dynamics
4. Suggest dynamical analogies
5. Discover new questions
6. Promote a scientific habit of mind

7. Bound (bracket) outcomes to plausible ranges
8. Illuminate core uncertainties.
9. Offer crisis options in near-real time
10. Demonstrate tradeoffs / suggest efficiencies
11. Challenge the robustness of prevailing theory through perturbations
12. Expose prevailing wisdom as incompatible with available data
13. Train practitioners
14. Discipline the policy dialogue
15. Educate the general public
16. Reveal the apparently simple (complex) to be complex (simple)

Explanation Does Not Imply Prediction

1.10

One crucial distinction is between explain and predict. Plate tectonics surely *explains* earthquakes, but does not permit us to predict the time and place of their occurrence. Electrostatics explains lightning, but we cannot predict when or where the next bolt will strike. In all but certain (regrettably consequential) quarters, evolution is accepted as explaining speciation, but we cannot even predict next year's flu strain. In the social sciences, I have tried to articulate and to demonstrate an approach I call *generative* explanation, in which macroscopic *explananda*—large scale regularities such as wealth distributions, spatial settlement patterns, or epidemic dynamics—emerge in populations of heterogeneous software individuals (agents) interacting locally under plausible behavioral rules (Epstein 2006; Ball 2007). For example, the computational reconstruction of an ancient civilization (the Anasazi) has been accomplished by this agent-based approach (Axtell et al. 2002; Diamond 2002) I consider this model to be explanatory, but I would not insist that it is predictive on that account. This work was data-driven. But I don't think that is necessary.

To Guide Data Collection

1.11

On this point, many non-modelers, and indeed many modelers, harbor a naïve inductivism that might be paraphrased as follows: 'Science proceeds from observation, and then models are constructed to 'account for' the data.' The social science rendition— with which I am most familiar —would be that one first collects lots of data and then runs regressions on it. This can be very productive, but it is not the rule in science, where theory often precedes data collection. Maxwell's electromagnetic theory is a prime example. From his equations the existence of radio waves was deduced. Only then were they sought … and found! General relativity predicted the deflection of light by gravity, which was only later confirmed by experiment. Without models, in other words, it is not always clear what data to collect!

Illuminate Core Dynamics: All the Best Models are Wrong

1.12

Simple models can be invaluable without being "right," in an engineering sense. Indeed, by such lights, all the best models are wrong. But they are fruitfully wrong. They are illuminating abstractions. I think it was Picasso who said, "Art is a lie that helps us see the truth." So it is with many simple beautiful models: the Lotka-Volterra ecosystem model, Hooke's Law, or the Kermack-McKendrick epidemic equations. They continue to form the conceptual foundations of their respective fields. They are universally taught: mature practitioners, knowing full-well the models' approximate nature, nonetheless entrust to them the formation of the student's most basic intuitions (see Epstein 1997). And this because they capture *qualitative behaviors of overarching interest*, such as predator-prey cycles, or the nonlinear threshold nature of epidemics and the

notion of herd immunity. Again, the issue isn't idealization—all models are idealizations. The issue is whether the model offers a *fertile* idealization. As George Box famously put it, "All models are wrong, but some are useful."

Suggest Analogies

1.13

It is a startling and wonderful fact that a huge variety of seemingly unrelated processes have formally identical models (i.e., they can all be seen as interpretations of the same underlying formalism). For example, electrostatic attraction under Coulomb's Law and gravitational attraction under Newton's Law have the same algebraic form. The physical diversity of diffusive processes satisfying the "heat" equation or of oscillatory processes satisfying the "wave" equation is virtually boundless. In his economics Nobel Lecture, Samuelson writes that, "if you look at the monopolistic firm as an example of a maximum system, you can connect up its structural relations with those that prevail for an entropy-maximizing thermodynamic system…absolute temperature and entropy have to each other the same conjugate or dual relation that the wage rate has to labor or the land rent has to acres of land." One diagram, in his words, does "double duty, depicting the economic relationships as well as the thermodynamic ones." (Samuelson 1972; see also Epstein 1997) In developing the Anasazi model noted earlier, my colleagues and I made a "computational analogy" between the well-known Sugarscape model (Epstein and Axtell 1996) and the actual MaiseScape on which the ancient Anasazi lived.

1.14

I am suggesting that analogies are more than beautiful testaments to the unifying power of models: they are headlights in dark unexplored territory. For instance, there is a powerful theory of infectious diseases. Do revolutions, or religions, or the adoption of innovations unfold like epidemics? Is it useful to think of these processes as formal analogues? If so, then a powerful pre-existing theory can be brought to bear on the unexplored field, perhaps leading to rapid advance.

Raise New Questions

1.15

Models can surprise us, make us curious, and lead to new questions. This is what I hate about exams. They only show that you can answer somebody else's question, when the most important thing is: Can you ask a new question? It's the new questions (e.g., Hilbert's Problems) that produce huge advances, and models can help us discover them.

From Ignorant Militance to Militant Ignorance

1.16

To me, however, the most important contribution of the modeling enterprise—as distinct from any particular model, or modeling technique—is that it enforces a scientific habit of mind, which I would characterize as one of *militant ignorance*—an iron commitment to "I don't know." That is, all scientific knowledge is uncertain, contingent, subject to revision, and falsifiable in principle. (This, of course, does not mean readily falsified. It means that one can in principle specify observations that, if made, would falsify it). One does not base beliefs on authority, but ultimately on evidence. This, of course, is a very dangerous idea. It levels the playing field, and permits the lowliest peasant to challenge the most exalted ruler—obviously an intolerable risk.

1.17

This is why science, *as a mode of inquiry*, is fundamentally antithetical to all monolithic intellectual systems. In a beautiful essay, Feynman (1999) talks about the hard-won "freedom to doubt." It was born of a long and brutal struggle, and is essential to a functioning democracy.

Intellectuals have a solemn duty to doubt, and to teach doubt. Education, in its truest sense, is not about "a saleable skill set." It's about freedom, from inherited prejudice and argument by authority. This is the deepest contribution of the modeling enterprise. It enforces habits of mind essential to freedom.

Acknowledgements

References

AXTELL, RL, JM Epstein, JS Dean, GJ Gumerman, AC Swedlund, JHarberger, S Chakravarty, R Hammond, J Parker, and M Parker, "Population Growth and Collapse in a Multi-Agent Model of the Kayenta Anasazi in Long House Valley". *Proceedings of the National Academy of Sciences, Colloquium* 99(3): 7275-79.

BALL, Philip (2007), "Social Science Goes Virtual" *Nature*, Vol 448/9 August .

DIAMOND, Jared M. , "Life with the Artificial Anasazi," *Nature* 419: 567-69.

EPSTEIN, Joshua M. and Robert Axtell (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press.

EPSTEIN, Joshua M. (1997). *Nonlinear Dynamics, Mathematical Biology, and Social Science*. Addison-Wesley Publishing Company, Inc.

EPSTEIN, Joshua M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press.

EPSTEIN, Joshua M. (2007). "Remarks on the Role of Modeling in Infectious Disease Mitigation and Containment". In Stanley M. Lemon, et al, Editors, *Ethical and Legal Considerations in Mitigating Pandemic Disease: Workshop Summary*. Forum on Microbial Threats, Institute of Medicine of the National Academies. National Academies Press.

FEYNMAN, Richard P. (1999) "The Value of Science." In Feynman, R. P. *The Pleasure of Finding Things Out*. Perseus Publishing.

SAMUELSON, Paul A. (1972). "Maximum Principles in Analytical Economics. In *The Collected Scientific Papers of Paul A. Samuelson*, edited by Robert Merton, Vol III, 8-9. Nobel Memorial Lecture, Dec. 11, 1970. MIT Press.