Building Confidence in Exploratory Models[1]

George P Richardson
Professor Emeritus, University at Albany
State University of New York

Beauty is truth, truth beauty.
That is all ye know, and all ye need to know.
(Keats, 1819)

Cherish those who seek the truth ,
but beware of those who find it.
(Voltaire, 1694-1778)

The truth is rarely pure and never simple.
(Oscar Wilde 1895)

Abstract
Some of the classic tests for building confidence in system dynamics models do not work well for models not rigorously calibrated to empirical data. Moreover, many of the classic tests do not go deep enough to build high confidence in a model-based investigation not using hard data for parameters or for fitting a model to reality. To instill as much rigor and quality as possible into such "exploratory models", this discussion identifies several categories of concerns about simulation models not rigorously fit to empirical data, drawing on examples of high quality exploratory modeling work, including Forrester's own classic works. It proposes guidelines and benchmarks for striving for the highest quality simulation work in the field, even for modeling work not extensively grounded in real-world data.

The paper focuses on simulation modeling. Some of it may have implications for studies and interventions employing only mapping techniques and qualitative analyses, but that is not its intent. Its special focus is on simulation modeling that strives for important contributions without the rigorous use of sophisticated analytic data-fitting tools. That focus probably characterizes the vast majority of the formal model-based work in the field of system dynamics.

It is divided into five sections: Motivation, Priors, Problems Inherent in Exploratory Modeling, Efforts to Build Confidence in Exploratory Modeling, and Conclusions.

Motivation
As recently as 2017, the Call for Papers for the the annual International Conference of the System Dynamics Society listed five requirements for submissions:
- Addresses an important issue,
- Takes a dynamic, endogenous perspective,
- Captures the behavioral decision making of the actors in the system,

[1] I have benefitted from advice from the 2023 ISDSM conference review process, the *System Dynamics Review* process, and a number of long-time friends and colleagues. I thank you all for your time and wisdom, and apologize if I haven't been able to meet all your requests.

- Is rigorously grounded in data,
- Is fully documented and replicable.

An admirable list. Yet many ostensibly important works in the field are not "rigorously grounded in data", including any of Forrester's classic works (*World Dynamics, Urban Dynamics,* and "Market Growth as Influenced by Capital Investment". None of these would have been eligible for presentation at the 2017 Conference.

Homer (1996, 2014) characterized the kinds of models not rigorously grounded in data as "exploratory" models[2]. Commonly included in this category are:

- Models that help support conversations. These often appear in participatory modeling workshops and may support the development of subsequent policy analyses.
- Modeling a "class" of problems rather than a specific case. These simulation models are commonly referred to as generic models. There is no single data set such a model can fit.
- Small models deliberately reduced to try to capture and communicate insights that apply to larger, more complex constructions .
- Models not based on real data but helping to form a foundation for subsequent data-based modeling.

The data distinctions Homer pointed out between "exploratory" models and what he called "scientific" models sound related to discussions of "theoretical" and "applied" work set out well by De Gooyert and Größler (2018). But they are not the same. Homer was talking about numerical data; De Gooyert and Größler focused on model context and purpose. Taken together, all these publications (including this one) are adding to what we understand about our diverse field and how we can do very high quality work in all our diversity.

It has been suggested that exploratory work is responsible for the impression that our field is on an "aimless plateau" (Forrester 2007). Yet it was Forrester himself who paved the way for such modeling efforts (Forrester 1964). One of the fourteen points he discussed in that classic article was labeled "Accuracy vs. Precision". Contrary to modeling traditions prevalent around 1960, Forrester asserted that it "is <u>not</u> true that Accuracy must be achieved before Precision is useful".

> The ability to precisely state a hypothesis and to examine its consequences can be tremendously revealing even though the accuracy of the statement is low. A precise and explicit statement with assumed numerical values will tell us the kinds of things which can happen. Should these things be important we can later devote attention to improving the accuracy of their statement. (Forrester 1960).

That old prescription is now in doubt in our field, with some suggesting, for example, that modeling without hard empirical data is unsuitable for publication (Sterman 2018). That is to say, exploratory modeling may not now be sufficiently rigorous to be "tremendously revealing" or acceptable for publication.

[2] Further described below. Homer's articles included a few other important characteristics of exploratory models, some of which are developed in this paper, but we begin with the single characteristic of exploratory models as simulations not based on, or calibrated with, empirically-derived numerical data.

I choose to take that conclusion as a challenge for us. To improve the perception of the quality of work in our field, we have to improve the rigor of practice that surrounds all forms of work in the field, including exploratory modeling.

Priors

Before proceeding we need to be clear about what we mean here by "data", "confidence" and the numerous subtle details that distinguish "exploratory modeling".

Data. All system dynamics simulation-based studies use qualitative data and quantitative data. I will use the adjective "qualitative" to refer to non-numeric information. I want to focus here on *quantitative* data – numbers. I may sometimes use the word "data" to mean just numerical data. We will distinguish between numbers many call "soft data" and "hard data" (Sterman 2000, 853). Soft data are numbers that people have in their heads informally or intuitively.

Hard data in this paper refers to numerical information that is "out there" in reality to uncover. It is the numbers that independent researchers would agree upon after detailed investigations. It is the stuff that "rigorously grounds models in data". More precisely, hard data is "empirically derived numerical data". But that's too long a phrase to use repeatedly, so I will substitute the term "hard data" throughout.[3]

Hard data is used for model calibration, which includes formal estimations of parameters and fitting models to quantifiable reality.[4] Soft data doesn't have the authority of hard data. Exploratory modeling often uses some hard data, but the models rely more on soft data.

Exploratory modeling. Following Homer's intent, we will call a model "exploratory" if it is not rigorously grounded in hard data. An exploratory simulation model is a tool to support investigations and explanations. Model parameters tend to be based on soft data – memory, intuitive estimates, and plausibility – or a mixture of soft and hard data. Formal data-estimating or data-fitting tools tend not to be employed. Homer envisioned exploratory modeling as a preliminary stage in building rigorous models, before they are carefully calibrated to hard data from a particular case. Yet here we will treat exploratory modeling as an important kind of system dynamics modeling that stands on its own. In fact, it is quite likely that most simulation-based work in the field is exploratory modeling (but I do not have the hard data to support that thought).

Confidence. In this article we take "confidence" in a model and its implications to be a kind of multifaceted trust that grows with the research and modeling efforts. It is not all or nothing, but rather a judgment that is constantly being re-evaluated by the reader.

---

[3] I tend to use "hard data" and "real data" interchangeably, as many do. To avoid confusions with "real qualitative data", the paper sticks to "hard" data exclusively to represent empirically derived numerical data.

[4] See Rahmandad, Oliva, and Osgood (2015) for an excellent treatment of analytic methods relevant to system dynamics modeling.

> “Confidence in a system dynamics model accumulates gradually as the model passes more tests and as new points of correspondence between the model and empirical reality are identified.” (Forrester & Senge 1979).

We do not “validate” any model-based work; we work to build confidence in all aspects of it, as it grows.

> “No model has ever been or ever will be thoroughly validated. ... ‘Useful’, ‘illuminating’, or ‘inspiring confidence’ are more apt descriptors applying to models than ‘valid’.” (Greenberger et al., (1976)).

Purpose of this paper. Many may assume that only models grounded in hard data can support high confidence in the details of their design, formulation, testing, and implications. Here we seek to push back on that assumption and to develop ways to grow the greatest possible confidence in studies based on exploratory simulation models.

Problems inherent in Exploratory Modeling

No real data, no real focus?

It is deceptively easy to believe that a specific case with hard data is necessary in order to know what we are talking about. Urban problems: Which ones? Which city? Central city decay? Crime? Traffic congestion? Quality of life? Slums? Segregation?

But we can know the focus without hard data. And almost always, the focus has deeply important qualitative aspects, which can be explored in detail without numerical data fitting specific cases.

Figure 1: A picture (a kind of qualitative data) of polar mother and cub on shrinking ice flows in search of food (Source: uk.news.yahoo.com)

Often no obvious 'problem owner'?
Exploratory work is often undertaken by a single system dynamics practitioner, or a small group working on a problem they find so important they want to work on it even though they have not been asked to do that. There is often not a client group (yet?) or other expert colleagues to talk with. Lacking a 'problem owner', the modelers have to deal with problems on their own, or at best by email with someone who might be able to provide guidance or answers. That can mean that the exploratory work rests heavily on the modelers' limited knowledge of the problem and consequently on the modelers' own mental model(s). Without extreme care, professional expertise, and serious research, the work may be perceived as superficial and not well-grounded.

Model purpose: Foggy without data?
The purposes of an exploratory study are just as important to detail as the purposes of a study based on hard data. It is presumably set by the modeler(s) and their clients or advisors. Evidence from hard urban data may reveal that unemployment rises in a mature city. But we don't need the data of a particular city to sharpen the focus, just the knowledge of the pattern in the data. Figure 2 shows the example of five urban population patterns in the U.S. as the cities reach their land limits.

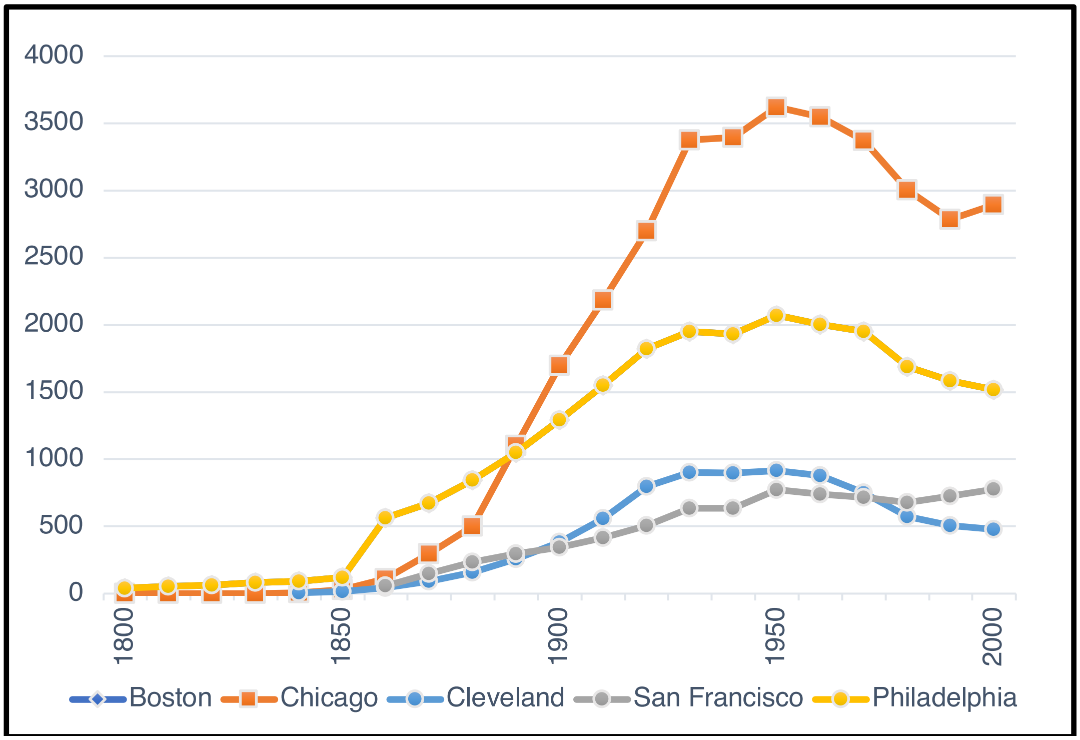

Figure 2: Populations in five large United States cities, 1800 to 2000, showing a generic urban population pattern over time (Sources: Web searches on individual cities)

The natural question posed by Figure 2 is why do five U.S. cities show the same pattern of growth, stagnation, and decline in population? For that matter, why does it also happen in twenty other U.S. cities, and the Netherlands[5] (and presumably elsewhere)? That was, of course, the question Forrester (1969) addressed in *Urban Dynamics*.

An instructive example about model purpose is Forrester (1968), which goes by the title "Market Growth as Influenced by Capital Investment". Yet Forrester's real purpose was to reveal the power of an endogenous point of view of corporate dynamics. To do that he formulated the model with a completely unrealistic assumption: the model contains *no constraint* on how large the company can grow – no limit to the market, no responses from rival corporations that could steal market share from the model company. That model purpose required an exploratory approach: there could be no actual case, no actual data, in which the corporation could grow without bound.

Quantification: Choosing data without data?

To be simulated, any model needs initial values and parameters, but we know that for an exploratory model it is likely we aren't using hard data from which to draw the numbers.

---

[5] Personal communication

Consider the simplified urban model in Figure 2.

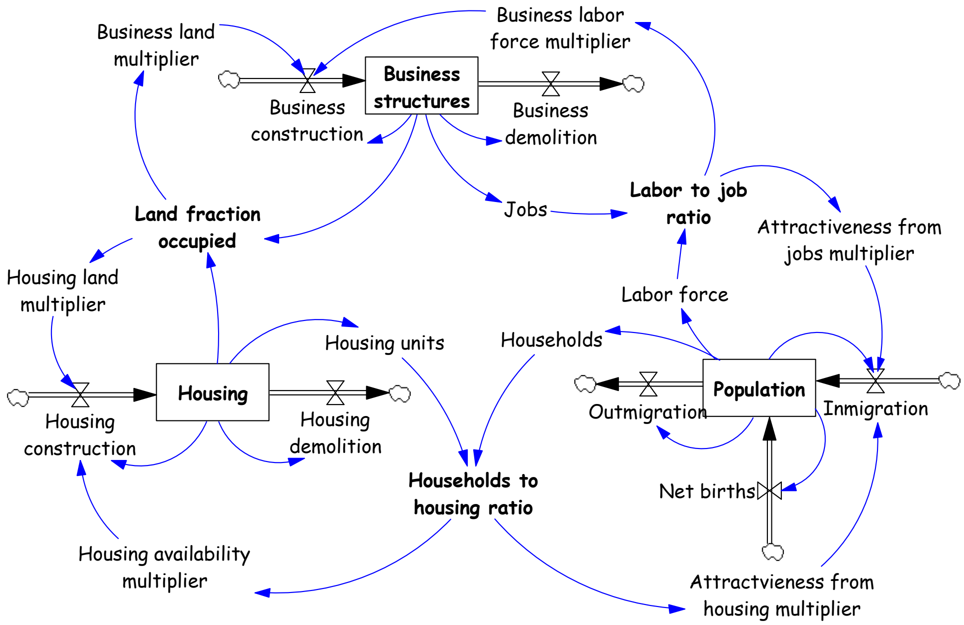

Figure 2: Simplified urban model (Alfeld and Graham 1976)

To make it easy to look at, the parameters here are not shown.[6] We need to set values for all of them. But the really important numbers in the model are not the individual made-up parameters, but the ratios:

Land Fraction Occupied
Labor-to-Job Ratio
Households-to-Housing Ratio

No matter how the generic parameters in the model are set, those ratios have to be realistic. LFO must be between 0 and 1; LJR must hover around 1 (above 1 means unemployment, below 1 a demand for labor); HHR must also hover around 1 (above 1 means crowded housing, below 1 empty apartments). Readers will ignore the little made-up parameters in the model, but they would insist that the ratios are realistic. For example, the initial value of Business Structures must generate a number of jobs that comes close to matching the size of the workforce, and so on.

It is probably a principle of exploratory modeling that parameterization requires numbers that work together to make observable *ratios* realistic.

Model boundary concerns, with only intuitive data to guide us?

[6] There are a lot of them: land per business structure, land per housing structure, jobs per business structure, number of people in a household, number of housing units in a housing structure, and so on.

What's in, and what's out, what's detailed and explicit, what's aggregated and implicit? Without hard data and efforts to fit a model to it, it can seem that the study is not adequately constrained to a well-defined problem. Where should one draw the boundary for exploratory studies of urban problems, or global problems, or declining market share for a generic corporation?

For *Urban Dynamics* Forrester did have a small advisory group of highly experienced urban professionals (all from Boston). Presumably with their help, but perhaps on his own, Forrester chose to draw the boundary around the generic city itself, and not to include suburbs. They were interested in the endogenous dynamics over which the city itself could have control.

But additionally there was a serious scientific purpose to leaving out suburbs. If an urban model without suburbs could <u>not</u> show urban decay, no matter how carefully constructed were the experiments with the model, then Forrester and his colleagues would have to conclude that suburbs are crucial to urban decay. But if the model <u>could</u> show urban decay without explicit involvement of suburbs, then great potential insights about urban dynamics would be suggested. So in the belief that suburbs were not the causes of urban decay, or maybe in the hope that they were, suburbs were left out. That decision ultimately produced the important insights of *Urban Dynamics.*

For *World Dynamics* he had the principals in the Club of Rome, but they left all the details to the modeler. All the individual regions of the world had to be inside the model boundary, but should each one be explicitly represented, or should they be lumped together (aggregated) somehow? Forrester chose one aggregate world, not disaggregated by north, south, east, or west, or any other disaggregation. He was criticized for that decision, but it makes sense to think that Forrester knew that was how to begin. Insights derived in that aggregate world could be explored in subsequent disaggregations by others, and they were.

For the Market Growth model Forrester turned a conversation into an instructive model by himself. The focus of the conversation was that a particular company was losing market share during economic downturns, while other similar companies were not. Forrester was able to create an exploratory model-based explanation for that phenomenon from assumptions about the inner workings of the companies, an inherently endogenous perspective.

While it may not be obvious from these summaries, in each of these cases the model boundary emerged from reflections on audience – who should want to know? – and model purpose – what questions are they seeking to answer?. The first – audience – should be identified very early on: Who should want to read this? The second – what will they gain from the study? – is likely to emerge as the model-based insights begin to emerge.[7]

---

[7] For example, Forrester once told me of his experiences exploring the favorite urban renewal policy of the 1960s – building lost-cost housing. He found it tended to make conditions in the *Urban Dynamics* city better in the short term, but either didn't work in the long run or actually made conditions worse. He kept at it for two weeks before he explored reducing housing (knocking down the worst of slum housing). I don't think he had those "better before worse" and "worse before better" insights when he started.

Endogeneity without data?: To what extent should an exploratory model include adaptive societal responses, and to what extent should that be left to users of the model? For example, in *World Dynamics* there are no endogenous societal responses, no built-in policies for adapting to resource shortages, pollution threats, declining ability to feed the world's billions, rampant population growth, and so on. All of these were left to people using the model to explore global futures by changing parameters. The model could have been written with at least some of those formulated explicitly with endogenous adaptive policy controls, but the modeler chose not to.

It seems reasonable that this question is handled the same way for all system dynamics studies. Like model boundary questions, it appears to come down to model purpose. How will the model be used? And that does not appear to rest on data from a particular case.

We have identified six potential problem areas in exploratory modeling:

- *No real data, no real focus?*
- *Often no obvious 'problem owner?*
- *Model purpose: Foggy without data?*
- *Quantification: Choosing data without data?!*
- *Model boundary concerns: What's in, what's out, with no data to guide us?*
- *Endogeneity without data?*

How can such problems be dealt with in exploratory models? To what extent does the lack of hard data for a specific case make these problem areas more difficult? And how do we deal with them? The following section tries to provide some answers.

Efforts to Build Confidence in Exploratory Models

The subsections named here and expanded below come from years of reflections on our best exploratory work. They describe considerations that apply to all system dynamics work, but are especially vital in building confidence in exploratory models.

- *Vivid dynamic problem*
  - *Perhaps in words, but much better in graphs over time tied to descriptive text*
- *Vividly understandable model structure*
  - *Not just adequate figures. Communicate as well as you possibly can.*
- *Tight links between structure and behavior*
  - *Convincing stories of how particular loops are involved in particular behavior.*
- *Alternative dynamic hypotheses*
  - *Not just one! Explore, present, and discuss as many as you can think of.*
- *Many simulation runs*
  - *The more runs showing realistic model behavior, the more confidence.*
- *Real world meanings of simulations*
  - *Describe every simulation in real-world terms, to keep the focus there.*

They emerged for me in much the same way as the thoughts presented in Richardson (2022).

While many could argue these six categories are important for building confidence in any model-based study, I must emphasize that they are especially important for exploratory models. In models with a heavy emphasis on hard data, various quantitative tools and fits-to-data may provide enough confidence in the work. But for exploratory simulation studies these six areas are crucial. They have to carry the day.

Vividly understandable model structure
Since the rise of "influence diagrams" in the UK about 1985 (Wolstenholme and Coyle 1983), it has become common to reveal model structure in causal-loop diagrams and then perhaps to move on to diagrams that also show stocks and flows. Sadly, there are awful flaws in CLDs in some published work that reduce or destroy confidence in the work. You know what they are: Putting everything in one diagram, simplifying so much that the algebra (if any) and causal structure is impossible to figure out, overwhelming with detail, not identifying accumulations as stocks, or leaving out accumulations entirely, drawing the links in ways that make feedback loops hard to see (e.g., straight lines), and so on. We will not show any of that.

The diagraming principles great modelers appear to use to build confidence involve these habits:

- *Build up to complexity.* Not one picture, but a sequence of pictures unfolding model structure step by step. Either show different parts of the model in small separate structures, or add structure to the growing diagram, a step at a time.
- *Respect cognitive limits.* The reason unfolding is necessary. There is a limit on how much new material can be absorbed. Keep the number of new entries at each build-up step small, say seven plus-or-minus two (Miller 1956).
- *Explain..* After each view, explain in satisfying detail the new entries. If possible, preview how a loop may be important in some dynamic patterns to be encountered.
- *Beauty is truth, truth beauty* (Keats 1819). Make each successive figure handsome – I mean that literally – elegant, easy to read and follow, a pleasure to look at, reasonably easy to guess equations (not necessary, but a great boost for audience confidence). You are using the elegant clarity of your pictures to move your audience toward greater complexity than they may have ever thought they could understand about their problem. Well-structured diagrams help your audience to feel that they "see" what you are saying.[8]
- *Omit link polarities.* Usually too much tiny detail (remember cognitive limits). But name the major loops and include their polarities; they provide part of your explanations. For audiences unfamiliar with reinforcing and balancing structures, spend a few words (repeatedly?) to make the ideas clear.
- *Distinguish pipes and causal links.* Always use pipes between stocks in conserved-flow chains (for clarity about flows), but use causal links everywhere else (for visual simplicity).
- *Use flags.* If it helps, use a special symbol to draw attention to key parts of model structure – e.g., the circled items in Figures 4-7.

---

[8] But contrary to Keats, beauty in modeling is NOT all they need to know! Pretty pictures by themselves cannot make a model excellent, or even adequate; they can't substitute for insightful modeling. What great model pictures can do is be *welcoming.* It is like the door to your home. If the paint on the door is peeling, the doorbell doesn't work and no one answers it, and the welcome mat is filthy, the entry does not build your guest's confidence. But if everything about the doorway is lovely and pleasing to your visitors, it adds to their comfort entering your home and probably shapes their expectations. Construct *welcoming* diagrams (of excellent models!) to help build confidence.

- *The goal.* You are building understandings of the structure of the emerging model, working to fit the model to the mental models of your readers, and so building *confidence* in the study. Act accordingly.

It is worth noting that these principles were implicitly given to us as early as 1972 in *World Dynamics* (Forrester 1972) and *Limits to Growth* (Meadows et al. 1972). Figure 3 shows an example from *Limits to Growth* that shows most of these principles.

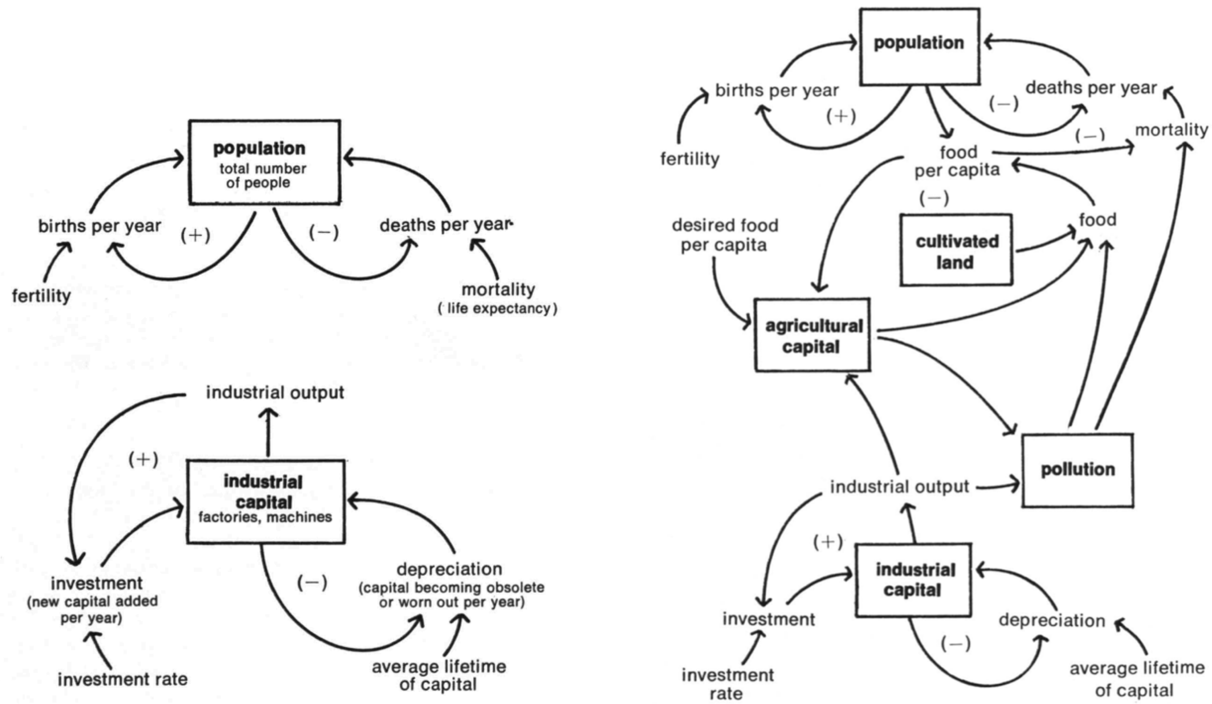

Figure 3: Diagrams from *Limits to Growth* (1972, pp. 95 & 97)
illustrating many of our diagramming principles.[9]

Unquestionably, the clearest way to present complex structure is to *unfold the model* step by step. To illustrate, I will show an unpublished piece of my own work. It's not perfect, but it tries to exemplify all the things we are talking about here. Figures 4 through 7 show a build up of an model from Richardson (2014), an exploratory model focused on growth dynamics of a management science field, to which we will return to a bit later.

Other examples are in Richardson (2021), Morrison et al. (2022), and scores of others. See Sterman (2000, Figs. 5.21-5.24) unfolding a complex CLD, and Chapter 20 for a superb build up of a large model).

***Vividly understandable model structure**:* Unfolding "Drawing Insights from a Small Model of the Growth of a Management Science Field" (Richardson 2014).

[9] A diagramming form we have largely forgotten, or never noticed, to our detriment.

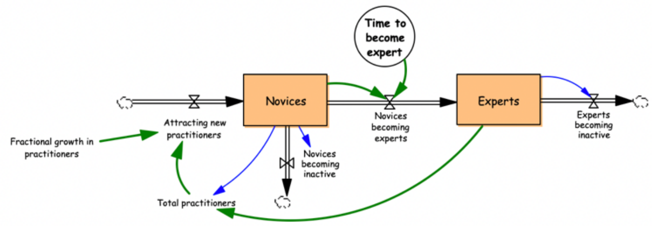

Figure 4: Reinforcing loops in the growth of practitioners.
(Note that almost all the equations are intuitively obvious.)

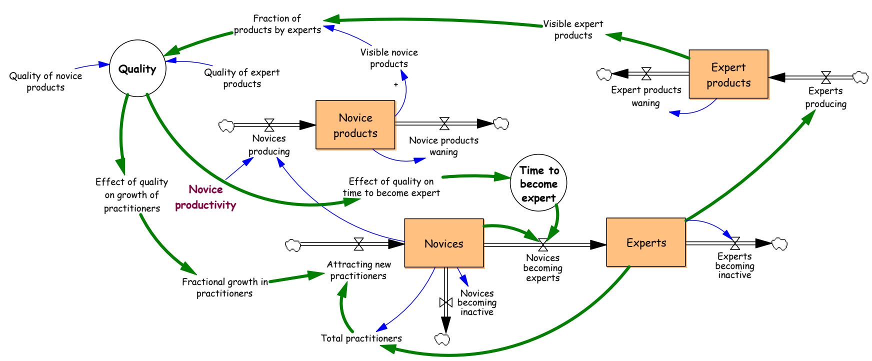

Figure 5: Feedback loops associated with promulgation of work in the field and the resulting perceptions of Quality.

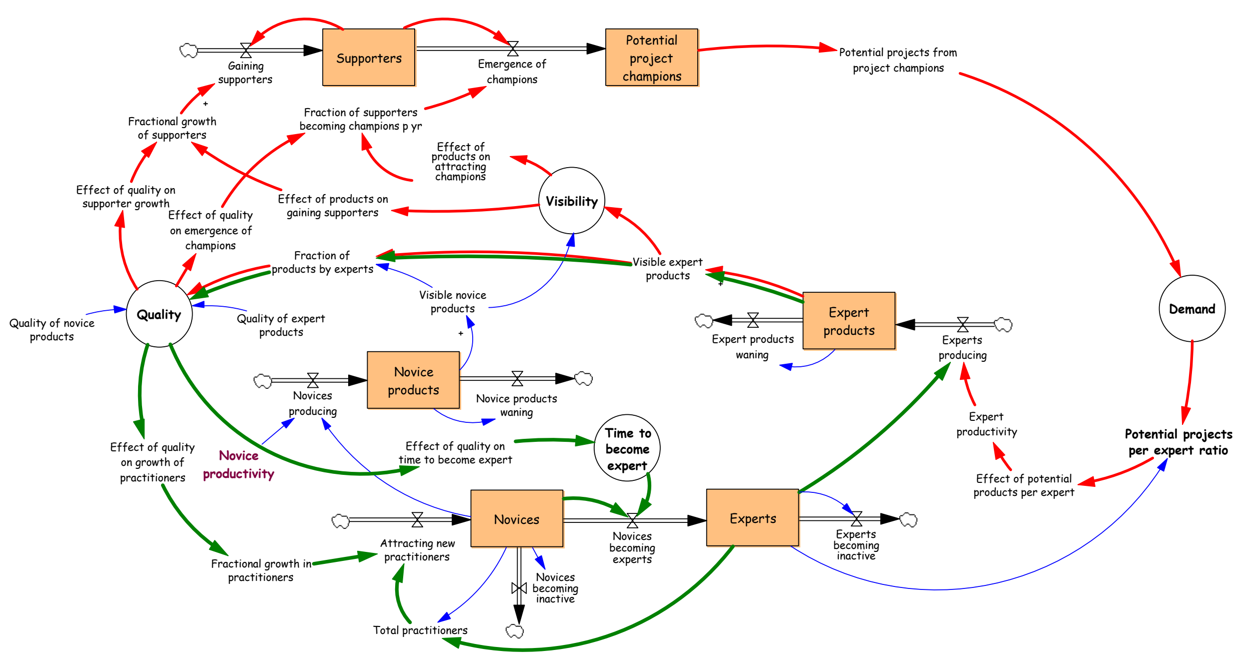

Figure 6: Adding stocks of supporters and potential project champions, together with the structure surrounding visibility. Red feedback loops are associated with growing demand for work in the field; green loops are associated with the supply of practitioners and their products. (Note that your eye can easily see most of the loops.)

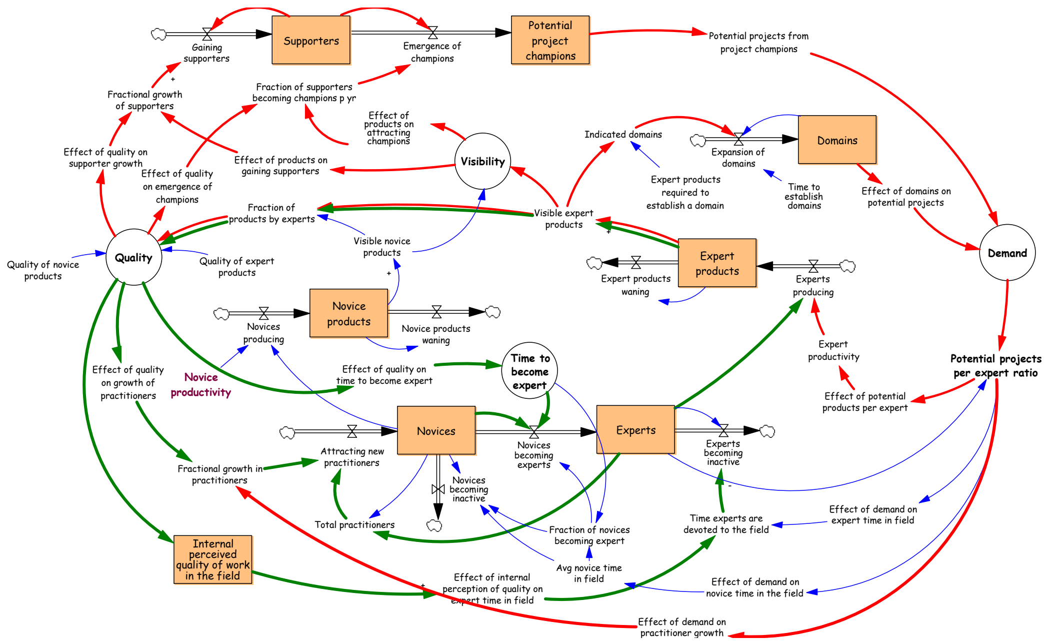

Figure 7: Adding the insightful stock of Domains[10] and closing loops back to the dynamics of practitioners completes the stock-and-flow / feedback structure of the model.

We've now reached a level of complexity that is about the most we can hope readers can understand in one figure. Note, of course, there would be no hope of anyone making sense of Figure 7 by itself, without the unfolding build-up in Figures 4, 5, and 6.

We see it takes a lot of space to unfold a model. But you can afford to take the space. That build-up is doing two things for us and the reader: It is presenting model structure in manageable steps as understandably as possible, and it is part of our substitute for fitting the model to data. The sequence of diagrams is striving to fit the model's structure to the *mental models* of system structure that readers bring to the problem. A great presentation of model structure can serve as a test of model fit to mental models that can build confidence in an exploratory model.

Furthermore, we need vividly understandable structure for the next step.

Linking structure to behavior

In fully quantified data-based modeling, there are always model runs exploring the goodness-of-fit of simulated output to real data. We now have increasingly sophisticated tools for doing that (Rahmandad et al., 2015). In exploratory modeling we are not using case-based data and so no tests of goodness-of-fit to real data. So what do we do to substitute for those fit-to-data tests?

---

[10] "Domains" are highly developed, familiar, even famous areas of application and expertise within the field. System dynamics domains thus so far appear to include urban dynamics, project management, litigation in project disputes, medicine and health, climate dynamics, and largely hidden military and government studies.

There are three issues here: *discovering* structure that influences a behavior mode, *explaining* such a structure/behavior connection to your audience, and *justifying* it.

Discovering structure/behavior connections probably involves things you already know. Experimenting with up-and-down changes in a parameter helps to see loops that make the behavior increase and/or decrease or not move at all. Setting a multiplicative graphical function to all 1's disables it, and essentially eliminates the loops that go through it. Behavior changes that result should begin to make your structure/behavior stories clear. Explaining structure/behavior links should be easy once you understand them. Justifying for your audience that some aspect of system structure influences some piece of system behavior requires showing the runs that convinced you of the connection. Depending on your purposes and your audience, you may be able to skip showing actual simulations to them and leave the justification to your description of the structure/behavior link.

Now comes building confidence in a model from *how it behaves.* In the exploratory works I have looked at there seems to be an almost universal practice: Lots of simulations!

Many (many!) simulation runs

This is probably the most important advice. Experienced modelers working with exploratory models run *many* simulations, covering an extensive range of scenarios. More simulations and more scenarios can help readers to assess more thoroughly the quality of the work. Not only do such tests show readers that your model can stand up to demanding tests; they also illustrate the range of insights readers may learn from the model. Both build confidence.

In *World Dynamics* there are about 20 different simulations. In *Urban* Dynamics there are 14. In Richardson (2014), a short 13-page paper, there are 11 simulations exploring different scenarios in the growth dynamics of a management science field (discussed in more detail below). And those shown are undoubtedly not all of the runs explored. Those add to your own understandings and help answer questions in presentations.

In exploratory modeling, different scenarios are something like goodness-of-fit tests of simulated model behavior with *mental model behavior*. Simulating many different scenarios adds confidence in the study with every successful "fit" to our mental models. The more runs and the greater diversity of scenarios, the greater the chance that any existing model flaws would appear. When flaws don't appear, when the model repeatedly behaves as our mental models say it should, confidence in the model has to grow.

Alternative dynamic hypotheses

All good model-based analyses explore more than one dynamic hypothesis to account for the behavior(s) of interest in the study. Great exploratory models should be no exception. Does a peak and decline in global population come from declining availability of natural resources?, changes in the roles of women?, changes in desired family size?, increases in the death rate from lack of food or growing pollution?, and so on. A good exploratory model can address many such questions, often simply using tests of alternative parameter values representing different policies or scenarios and analyzing them carefully.

Real world meanings of simulations
Changing a parameter value and simulating a model can be looked at two ways by the modeler and the audience. It can be thought of and analyzed as an *abstract sensitivity test*, and it can be thought and analyzed as a *potential real-world scenario.* Both are valuable for building confidence in the model.

At the highest level of quantitative sophistication, supported by advanced mathematical tools, practitioners explore parameter sensitivities at some length, comparing the graphs to real data, gaining insights about structure and behavior. They use the results to strengthen (or weaken) confidence in the model and to flag excessively sensitive parameters for possible reformulations.

In exploratory modeling our interest is most effectively placed on the real-world. Parameter changes are best phrased as changes in a real *scenario* or changes in actual *policies*. They should never be phrased as "Let's raise that parameter 20% and see what happens", but rather "Let's raise the time it takes to establish a well-known new domain of system dynamics work from five years to twenty years" (see Figure 6 above). The change should represent a real aspect of the system we are trying to learn about. If the model behaves in some unexpected way – that is, fails to align with our mental models – we learn we have work to do to fix some errors, or revise our thinking. But if the model behaves as we thought it would, we can explore further whether we think that really makes sense to our mental models and teaches us something potentially important.

The point here is important, so I would like to present an example in more detail. In the short paper presenting the model pictured in figures 4 through 7 (Richardson 2014), there were eleven simulations exploring different scenarios in the growth dynamics of a management science field. All eleven of those runs are described and interpreted in Table 1 in the appendix.[11]

Parameter Changes in model runs	**Potential Real World Implications from the simulations described on the left**
Setting the parameter for the normal fractional growth of practitioners to 20% per year, up from 15% in the base run, results in *less* growth of the field:	Trying to grow too fast can actually slow the spread of a field, just as it can a company or a political movement.
No growth in Potential Project Champions. The field fails to grow beyond 1988.	Failure to attract supporters and champions forces the field to try to grow without them: consultants would find it difficult to find clients, and the growth of the field would probably be left to academics and retirees, people who are freer to pursue their interests even if the marketplace isn't particularly interested. Growth is severely constrained, if not eliminated completely.

Figure 8: The first two rows of Table I in the Appendix, for simplicity here.

The left-hand column in that table contains the detailed descriptions of the parameter changes – they were the figure captions in the original paper. The right-hand column contains discussions

[11] The paper and the simulations were used as the basis of a special workshop of the Policy Council at the International Conference in 2014 focusing on potentials for growth of the field.

of the real-world implications of those parameter changes, which appeared in a grand summary at the end of the paper.

The first thing to notice is that on the left we are talking about numbers, but on the right we are talking about interpretations of the simulations, phrased as potential real world implications. For the first row, the normal fractional growth of practitioners is increased from 15% to 20% per year. The result was actually *fewer* practitioners! The paper explains why that oddity could really happen and how the structure of the model produces that behavior for very plausible reasons.

Because you are stating everything in real-world terms, your audience will be thinking about real-world implications. The implicit message, emphasized over and over, is that *the model and the parameter changes are tools for thinking about the real world.* In addition, the results of the parameter changes help your audience to evaluate the model, and we hope, gain confidence in the model itself and its potential implications.

Conclusions

In most of our applications, fitting a model to an individual case and its particular data is the strongest way to proceed. It is the foundation for what Homer called "scientific modeling" (see Rahmandad et al. (2015) for tools and techniques). It is the coin of the management science realm.

But there are also times when an exploratory approach is more appropriate for the purposes of the study and the needs of the audience. Forrester (1969) wrote *Urban Dynamics* focusing on a generic city so that implications of the study would not be seen as applying only to a particular place or a particular data set. I chose to design the model shown in Figure 3-6 and Table 1 as an exploratory model rather than fit it to data of the system dynamics community so that in the workshop discussion we could avoid focusing on data fit and focus instead on healthy growth strategies for a management science field.

In this paper I have focused on the problems and potential of building confidence in exploratory models and the studies they support. Confidence appears to grow from

- *Vividly communicated dynamic problem*
- *Vividly understandable model structure*
- *Tight links between structure and behavior*
- *Many simulation runs*
- *Alternative dynamic hypotheses*
- *Real world meanings of simulations*

But in another sense, confidence in our exploratory modeling work will grow if we are focusing on building confidence throughout the entire process. For example, surface things we can do that we know are important but I have omitted. Consider whether all the parts of our approach support each other. Does this picture build confidence? Does this discussion of structure and behavior build confidence? Do our conclusions support or improve insights? As far as we can test and tell, is it realistic, that is, does it conform to our mental models of the ways we think the real world is structured and behaves over time? If not completely, then as we reflect upon it does most of it make sense? Can we build on it?

This paper starts to point a way for us. It suggests a number of guidelines to follow. Follow as many of them as you can. Improve them. Develop more. If we do exploratory studies and modeling work as well as we possibly can, there is no reason exploratory work can not be a strong, insightful, often quoted part of the excellent reputation of the system dynamics field.

References

Alfeld, LE and A K Graham (1976). *Introduction to Urban Dynamics*. Available from the System Dynamics Society.

De Gooyert, V and A Größler (2018). "On the differences between Theoretical and Applied System Dynamics Modeling. *System Dynamics Review* **34**(4): 575-583.

Forrester, JW (1964). "Common Foundations Underlying Engineering and Management."*IEEE Spectrum* **1**(9): 66-67.

_______(1968). "Market Growth as Influenced by Capital Investment." *Industrial Management Rev. (MIT)* **9**(2): 83-105.

_______(1969). *Urban Dynamics*. Cambridge MA, Productivity Press."

_______(1971). *World Dynamics*. Tokyo, Nippon Keiei Shuppankei.

_______(2007). "System dynamics – the next fifty years." *System Dynamics Review* **23**(2-3): 359-370.

Forrester, JW and P Senge (1979). Tests for Building Confidence in System Dynamics Models. Reprinted in Chapter 4 of *Social Dynamics,* online curated collection of Forrester's selected works: <https://collections.systemdynamics.org/jwf/social-dynamics>

Greenberger, M, MA Crenson, BL Crissey (1976). *Models in the Policy Process: Public decision making in the computer era.* New York: Russell Sage Foundation.

Homer, JB (1996). "Why We Iterate: Scientific Modeling in Theory and Practice." *System Dynamics Review* **12**(1): 1-19.

_______(2013). The aimless plateau, revisited: why the field of system dynamics needs to establish a more coherent identity. *System Dynamics Review* **29** (2): 124–127.

_______(2014). "Levels of evidence in system dynamics modeling." *System Dynamics Review* **30** (1-2): 75-80.

Homer, JB and GP Richardson (2018). Growth of the system dynamics field. *System Dynamics Review,* 33, 3-4 (2018).

Keats, J. (1810). "Ode to a Grecian Urn. <https://en.wikipedia.org/wiki/Ode_on_a_Grecian_Urn>

Meadows, DH, DL Meadows, J Randers, WW Behrins (1972). *Limits to Growth.* (New York: Universe Books, A Potomac Associates Book).

Miller, GA (1956). “The Magic number Seven, Plus or Minus Two”, *Psychological Review.* **63** (2), 81-97.

Morrison, JB, L Berkowitz & R Minehart (2022). “Residency Program Cultures: How Dynamic Interactions Shape Learning Outcomes”, *2022 International System Dynamics Conference*. Frankfurt, Germany.

Rahmandad, H, R Oliva, & ND Osgood, eds. (2015). *Analytic Methods for Dynamic Modelers* (Cambridge, MA: The MIT Press).

Richardson, GP (2014). “Drawing Insights from a Small Model of the Growth of a Management Science Field”. Unpublished paper used to structure a workshop at the 2014 International System Dynamics Conference.

_______ (2021). Can Systems Thinking be an Antidote to Extensive Evil? *Systems Research & Behavioral Science*. 2021; 38: 401– 412. https://doi.org/10.1002/sres.2677

_______(2022). “What Jay Didn’t Tell Us: Hidden Gems in the System Dynamics Practice of Jay W. Forrester”. *System Dynamics* Review (2022.

Sterman J 2018. System dynamics at sixty: The path forward. *System Dynamics Review* **34**(1–2): 5–47.

Voltaire (1694-1778). Repeatedly attributed to Voltaire (and other later authors perhaps borrowing the idea) but at this writing my source is unknown.

Wilde, Oscar (1895), “The Importance of Being Earnest, A Trivial Comedy for Serious People”

Wolstenholme, EF and RG Coyle (1983). “The Development of System Dynamics as a Methodology for System Description and Qualitative Analysis.” *Journal of the Operational Research Society* **34**(7): 569-581.

Appendix
Table 1: Simulation runs in Richardson (2014) building confidence in the model

Parameter Changes in model runs	**Potential Real World Implications from the simulations described on the left**
Setting the parameter for the normal fractional growth of practitioners to 20% per year, up from 15% in the base run, results in *less* growth of the field:	Trying to grow too fast can actually slow the spread of a field, just as it can a company or a political movement.
No growth in Potential Project Champions. The field fails to grow beyond 1988.	Failure to attract supporters and champions forces the field to try to grow without them: consultants would find it difficult to find clients, and the growth of the field would probably be left to academics and retirees, people who are freer to pursue their interests even if the marketplace isn't particularly interested. Growth is severely constrained, if not eliminated completely.
More difficult to enter new domains for system dynamics research and applications. (The number of products required to establish a domain is raised from 25 to 50 in this run.)	Expanding into new domains of research and application is key to the growth of a field like system dynamics.
The time it takes to establish a new domain of work is increased from 5 years to 20 years.	The more effort and time it takes, the slower or more constrained is the growth potential of the field.
More visible products from novice practitioners. beginning in year 2000 the fraction of novice products visible (published, available on the web, or otherwise readable) rises from 70% to 90%.	The result is less growth of the field -- fewer practitioners, fewer supporters and potential project champions, lower overall quality, and fewer visible products in spite of the greater visibility of novice products.
The visible life (availability) of novice products is raised beginning in 2000 from five years to the life of expert products, here assumed to be eight years.	Managing the visibility of work in a field is undoubtedly difficult. A field can monitor its own conferences and journals and perhaps prevent all but high quality work from being promulgated, but it would have trouble managing all the possible outlets for publication or dissemination. However, the power of limiting the visibility of the work of novices in the field is so great for the health and growth of a field that much thought should go into how best to accomplish it.
Less visibility of novice products. Beginning in 2000, the fraction of novice products visible is reduced from 70% to 30%.	It is fair to say, the field of system dynamics has erred (if it is erring) on the side of inclusiveness, welcoming early practitioners to its conferences and publishing their work along with the work of our most expert practitioners. Limiting the exposure of their work in our conference proceedings would be distressing for them, and for those of us who care about them, until we all realize that such limits grow the field dramatically and actually shorten the time to become expert. Visibility means making our good work known widely. As disturbing as it sounds, doing that means that academics should be publishing much of their best system dynamics-based work in peer-reviewed journals other than the *System Dynamics Review.*
Market the field (by unspecified mechanisms). Great growth in supporters (grey and green curves on the right), but not a significantly greater growth in domains or practitioners.	Visibility also means finding ways to make widely known the work of consultants and practitioners in the public and private sectors. They may lack the time and incentives necessary to publish, but perhaps we can link advanced PhD students with practitioners and get out publications that benefit both. Practitioners would get their work out with appropriate control and minimal effort; PhD students would learn about state-of-the art applications and processes; and PhD students would get publications as second-authors reporting on great work.
Spread conferences around the world to market the field more widely (not simulated, inferred from above runs)	Expanding conference opportunities with such initiatives as the recent Asia-Pacific Conference in Tokyo creates growth potential for the field, but carries with it the dilemma about promulgating only high quality work. That is not to say practitioners in the Asia-Pacific area are less expert than anywhere else, but rather to say that all conferences inevitably have to address the contradictions inherent in welcoming beginning or less expert practitioners while wanting to make only expert work visible.
Mentoring, to reduce the time it takes to become expert by 40%.	Mentoring is another enormously high-leverage policy. Consulting firms naturally take a mentoring approach, much like the internship process for growing the expertise of doctors.
Mentoring for quality. (The quality of novice products is assumed to rise from 1 to 4 beginning in the year 2000.)	Mentoring goes beyond coursework. In its most powerful forms in a field like system dynamics, it involves skilled experts (academics, consultants, practitioners in business and the public sector) working every week with people who have graduated from university coursework and want to become truly expert. As a field we are probably missing much of the growth potential of mentoring.