

# Varieties of System Dynamics Practice

Alan K. Graham, PhD

Greenwood Strategic Advisors

Zugerstrasse 40, CH-6314 Unterägeri, Switzerland

Tel. +1 781 862 0866, Fax +41 41 754 7448

[Alan.Graham@Greenwood-AG.com](mailto:Alan.Graham@Greenwood-AG.com); [Alan.K.Graham@alum.MIT.edu](mailto:Alan.K.Graham@alum.MIT.edu)

## Abstract

*What is system dynamics? Traditionally, the explanations have been complex and implicit, e.g. 10 steps, 25 types of validation, and 33 questions model users should ask. Such elaborate explanations can create hurdles in teaching SD, innovating within SD and communicating about it and its value to the wider world. This paper examines (and labels) several varieties of system dynamics practice, with “classic” SD more or less at the center: Qualitative systems thinking, and classic, industrial strength and legal strength system dynamics.*

*All of the variations of system dynamics can be defined as different elaborations and iterations of testing only three hypotheses: our understanding of 1) the **problem** to be addressed, 2) the **system** in which the problem is created, and 3) the **recommendations** that will improve the situation. We use diverse sources of knowledge for all three, especially general knowledge about how people interact, expert participant knowledge, and statistics. We use computer simulation to test the system and candidate actions. This comprehensive reality-checking is what makes system dynamics a trustworthy approach to complex situations.*

*This simple definition of system dynamics is a movement toward a simple and compelling “elevator explanation”, which should also be useful in structuring effective teaching and communicating about both the complexities and simplicities of SD within a comprehensible overarching structure. When expanded into specifics, it can give the system dynamics modeler a larger set of approaches, and presumably a better chance of making a difference in the real world.*

Key words: Aimless plateau, definition, hypothesis testing, reality check, problem

## What is System Dynamics?

Probably enough ink has already been spilled characterizing the “aimless plateau” that Jay Forrester suggests that the field of system dynamics occupies (Forrester 2007, pg. 360). I would

offer a more optimistic view, namely that considerable broadening and deepening of the discipline has occurred, but mostly outside of academia<sup>1</sup>, and that, paradoxically, those varieties<sup>2</sup> of system dynamics point to a view of what system dynamics is and how to do it that is simpler, easier to teach, and easier to communicate than the traditional descriptions.

Before describing these extensions, we should begin with a discussion of what we might consider the essence of system dynamics to be and why that is important. We can then arrive at a definition of system dynamics that is both simpler and more encompassing than many traditional definitions. With that framework, we can then discuss the varieties of practices encompassed by that definition of “system dynamics”. To begin with definitions of system dynamics, then:

*Is it defined by foundations?* The closest to a definition that Forrester gives is this passage from *Industrial Dynamics* (Forrester 1961, pg 14):

“Four foundations on which to construction an improved understanding of the dynamics of social organizations have been built in the United States since 1940, primarily as a by-product of military systems research. These four are:

- The theory of information-feedback systems
- A knowledge of decision-making processes
- The experimental model approach to complex systems
- The digital computer as a means to simulate realistic mathematical models

These describe where SD came from, but do not sharply define what it *is*, though. As a definition, its boundaries are overly wide: Bread comes from flour, yeast, water and heat, but so do bagels. The “foundations” definition of system dynamics would have it also include tailoring the response characteristics of a jet aircraft through flight simulator experiments with human pilots. We would like a definition with more specificity.

*Is it defined by our practices?* (Forrester 1961, pg. 13) lists ten fairly generic steps for dynamic modeling (under which the jet aircraft example would still qualify as system dynamics). He also lists four broad validation tests (Forrester 1961, Section 13.4-5):

“...proper selection of boundaries of the system to be represented.”

“...an effective choice of model variables, properly interconnected.”

“The third and least important aspect of a model to be considered...concerns the values for its parameters (constant coefficients). The system dynamics will be found relatively insensitive to many of them. They may be chosen anywhere within a plausible range.”

“A ... model should generate time patterns of behavior that do not differ in any significant way (judged within the framework of the objectives of the study) from the real system.”

Over the years, the various lists have grown—(Sterman 2000, Table 21-4) lists a dozen categories of validation tests, most of them still fairly broad, each with several associated questions that in effect comprise a total of 25 sub-categories. Those are tests the modeler should execute. For users of models, Table 21-1 lists 33 “Questions model users should ask—but usually don’t.” Moreover, all of these are each important—the nature of system dynamics is such that one “fly in the ointment” taints the whole effort. So we can’t make a useful definition of system dynamics that ignores these complexities.

One problem in defining and teaching our field as a static list of practices is that such a definition starts to close the field to innovation and change. Even “classic” SD has changed over the 50 years since the publication of *Industrial Dynamics*. For example: Causal diagrams, group model building, and graphic user interfaces for model development were all later additions to commonly-taught SD. Even comparison with time series has crept to the edge of the mainstream (Sterman 2000 pp. 874-880). We *really* don’t want to define ourselves so specifically that our self-definition inhibits exploration and innovation.

An equally significant drawback of a definition based on practice lists is that it’s difficult to explain to potential users of the modeling results and professional peers, in a way that both communicates well and makes the differences from other methods somewhat clear—and positions SD as clearly desirable. (For convenience, let us use “clients” to mean “users of the modeling results and other important stakeholders”, without making specific assumptions about who if anyone is paying for the modeling, who is responsible at higher and lower levels in an organization for using the results and achieving performance improvements, and so on.)

Perhaps the largest drawback of definition based on practice lists is that it’s difficult to teach and develop skills in a coherent way. It has traditionally taken many years of *ad hoc* learning for modelers to develop the entire ensemble of skills and knowledge needed to successfully make a difference in the organizations they work with.

Just as leaders become effective in part by articulating clear and compelling visions for their organizations, the field of system dynamics would seemingly become more effective by articulating a clear and compelling vision of what system dynamics is. The definitions discussed so far are particularly short of “compelling”.

*Have our definitions neglected model purpose and usefulness?* Probably. System dynamics texts are clear about the importance of purpose in many if not all phases of model-based system improvement, e.g.

“In designing a dynamic simulation model of a company or economy, the factors that must be included arise directly from the questions that are to be answered.”

(Forrester 1961, Section 5.1)

“**Validity without Goals.** Validity and significance are too often discussed outside the context of model purpose. Usefulness can be judged only in relation to a clear statement of purpose. The goals set the frame for deciding what a model must do.”

(Forrester 1961, pg 122)

We can distinguish two aspects having a “model purpose”. The first is having an explicitly stated purpose that is specific enough to guide subsequent modeling decisions. This is uncommon—too often a model purpose is “to understand” a behavior, period. The second aspect is showing how modeling decisions in fact derive from the model purpose. Readers of Forrester’s *Industrial Dynamics* will find Sections 15.1-4 generally exemplary,<sup>3</sup>—and seldom imitated since. Having only “understanding” as a modeling purpose provides no foundation for making decisions about model formulation, testing and use, and in particular, no criterion for when a modeling effort is “good enough” to fulfill the purpose.

Even experienced practitioners on occasion find themselves back-filling modeling work because of incomplete understanding of a model’s intended use, a context that could have been identified near the very beginning of a project. One cure is to use a one-page diagram or document that states a variety of specifics that together define what is meant by a model’s purpose, and review it quite carefully with the clients, until there is agreement that the diagram or document accurately captures the clients’ situation and intentions in engaging in modeling with the modelers. We’ll see an example shortly. Such diagrams are a “transitional object”<sup>4</sup>—something that a non-modeling client can look at, understand, and agree or disagree with, and yet which is translatable into precise and operational definitions of model purpose usable by the modelers.<sup>5</sup>

Treatment of purpose and usefulness is part of system dynamics in a more fundamental sense, because SD differs from many other academic fields as much as engineering differs from science. Engineering is about real-world results, and science seeks abstract knowledge. They are related, and frequently work hand-in-glove, but are profoundly different. System dynamics is not about discovering unequivocally true things in some particular application domain (“learning more and more about less and less”); it is rather about improving performance in at least one specific real-world system at a time. In marketing terms, purpose and usefulness are key differentiators. If we want a compelling elevator story, treatment of purpose usefulness is well worth emphasizing. If we want to quickly communicate how what we do in system dynamics is good and useful to many audiences (academic colleagues in other disciplines, students, clients, and our own family members) we can emphasize purpose and usefulness as fundamental elements of SD.

## Why start simple?

Before diving into an alternative definition of system dynamics, let me sketch out a hypothesis of why our definition and presentation of system dynamics matters. Far be it for one portion of one qualitative paper to offer a comprehensive diagnosis of how the field has evolved and could evolve further. Yet it can offer a dynamic hypothesis reasonably consistent with the history of the field and known cause and effect relationships, focusing on the role of how the field is defined and therefore how it is presented and taught.

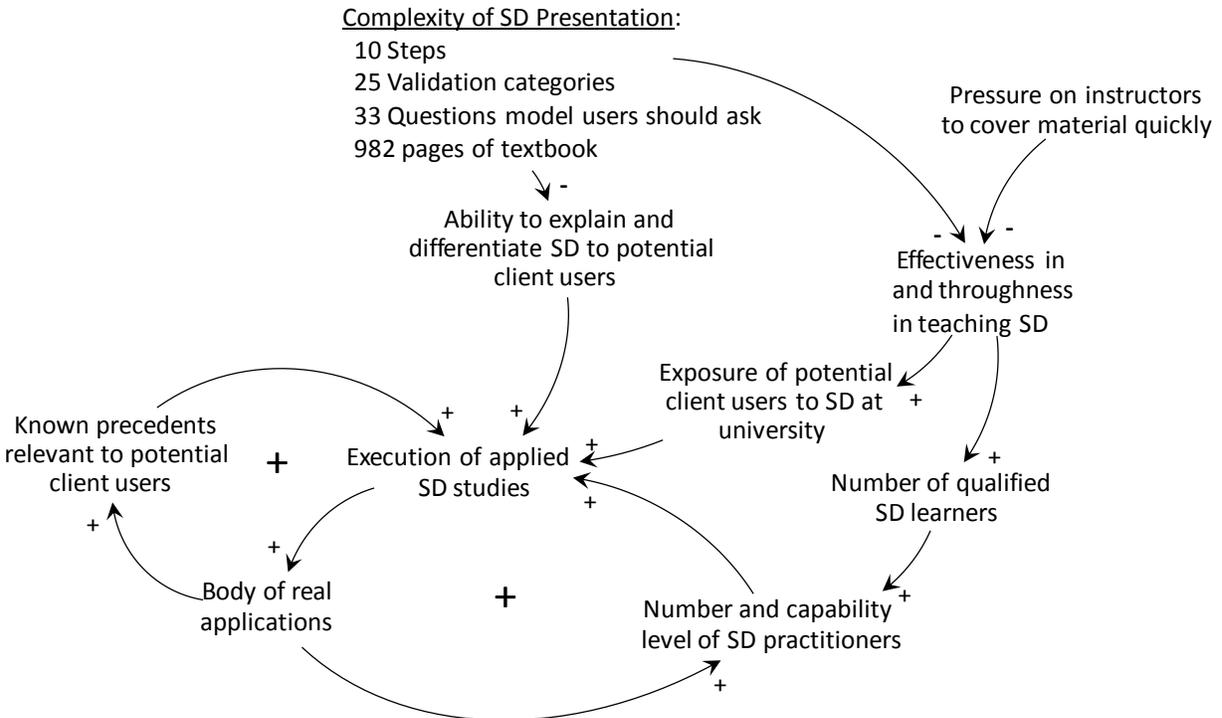
(Forrester 2007), (Homer 2007) and (Barlas 2007) all stress professional excellence as the path along which to move the system dynamics field upward, off of that “aimless plateau”. This is difficult to disagree with. But consider how to get more people moving along the path.

Specifically, let’s take a look at what happens when we start with a complex definition and presentation of system dynamics, as historically we have. The dynamic hypothesis shown in Figure 1 starts at the top from the current (implicit) definition of the field, characterized somewhat tongue in cheek by the enumerative multi-line variable name at the top of the diagram. The complexity with which the field is defined, presented and taught has two effects: SD is rendered harder to teach, and made harder to explain to potential clients (for convenience, the non-modelers who might hope to use SD or other strategic tools to make key choices).

Along the right side of Figure 1, the challenge of teaching system dynamics, especially within the time pressures of academic programs, is a brake upon reaching students who might eventually become practitioners. In many universities, it is difficult to regularly offer more than one or two semesters of system dynamics courses. Yet there is far more than two courses worth of content; coursework is therefore usually an insufficient foundation for professional practice of system dynamics.

The challenges of teaching system dynamics extensively and well are also a brake upon exposures that a wider university (or school, generally) audience that is exposed and familiarized with SD, which comes back later in their professional careers as being easier to engage to participate in or sponsor system dynamics work—“I saw it in school; I know it’s legitimate”.

These factors make it difficult to find the wherewithal to execute applied system dynamics work (in the middle of Figure 1). Only a handful of consulting companies have managed to survive for significant periods of time, and their size and scope of activities has not exhibited exponential growth. A low rate of execution of applied work implies slow growth in the extant body of real applications (bottom center of Figure 1).



**Figure 1. Dynamics of SD growth as influenced by complexity of definition of the field.**

The small body of applied system dynamics engagements in turn puts a severe limit on the number of truly experienced practitioners who can enter into a situation and address significant problems with a high degree of confidence. Too much of the work is doing things for the very first time. The rarity of truly experienced practitioners is in turn another factor making it difficult to engage real people and organizations in the use of system dynamics. This completes the positive loop at the bottom of Figure 1.

To add another positive loop to the situation, few real applications means that for any one organization, the number of known applications in organizations and situations similar to their own will be tiny. “If you haven’t done work like what I want done, I don’t want you to try.” This in turn makes it difficult to engage to do that SD work that would generate the needed precedents, completing the positive loop at the left side of Figure 1.

There are doubtless many more influences and loops. For example, pressure to cover material quickly means that what is taught becomes too narrow to support professional practice. Two examples: writing and presentation, or how to go about building consensus and implementing are of necessity excluded, and have become a real Achilles’ heel for the field.<sup>678</sup>

These issues are not particularly novel; doubtless many have identified the chicken and egg nature of the positive loops surrounding growth of this or any other field. But one thing seems

clear: As long as system dynamics practitioners continue to present and teach it as a highly complex, somewhat amorphous discipline, there will be neither chickens nor eggs on an industrial scale. Automobiles didn't really take off until they didn't require operation and servicing by someone both physically strong and mechanically inclined. Computers remained the domain of "eggheads" until packaged software allowed normal people to perform useful tasks with them. Bibliographic search remained the domain of professors, grad students and professional librarians until Google made it possible for anyone to search. And so on.

The field needs to be made accessible, and a good place to start is by offering an intelligible description of what we do that encompasses all of what we do. The raw material for such an offering is a systematic description of the varieties of system dynamics practice.

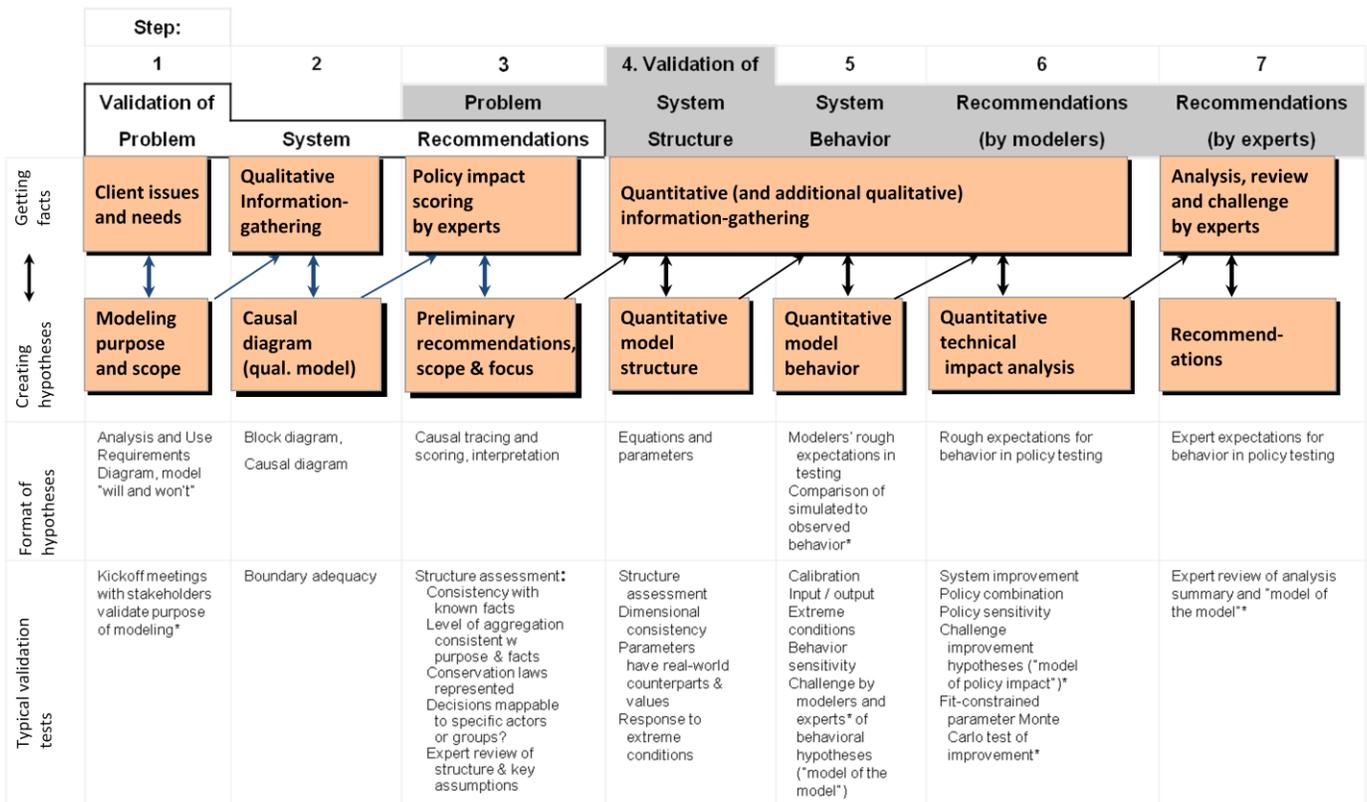
## **Toward a simpler framework**

By way of working *toward* a simpler view (but not yet having arrived at a simpler view), consider Figure 2, which is a diagram originally created (Graham 2010) to explain, in a compact way, how thoroughly validated a good system dynamics model really is. It is drawn from many dozens of modeling projects, at PA Consulting (the successor to Pugh Roberts Associates) and elsewhere.

Steps of modeling are along the top row. Any one step is hypothesis testing of a particular sort. Based on some facts we have gathered (the first pink row), we formulate a hypothesis (the second pink row) that we can test against those facts, or more facts as we gather them, in an iterative process. Of course, the format of the hypothesis (first uncolored row) and the specific tests one uses (bottom uncolored row) will vary, depending on which step is being executed. For example, at the first step, testing our understanding of what problems or issues are to be addressed, the hypothesis is formulated as a PowerPoint diagram, and the testing is done by reviewing by clients or other stakeholders or experts. When we have a simulator expressing a hypothesis about behavior (step 5), the format is the simulation model itself, the facts are qualitative and quantitative observations of system behavior, and the tests are the familiar sorts, e.g. behavior reproduction, behavior sensitivity, extreme conditions, etc.

Going horizontally across the diagram, you'll see a round of qualitative hypothesis testing with clients and experts, mostly before the usual quantitative modeling steps start. In this approach, the system is first "modeled" diagrammatically and reviewed by subject matter experts. This allows modelers to learn about the system in some depth before embarking on quantification, which are steps 4 through 7. The diagram shows along the bottom (at least implicitly) all of the tests in (Sternan 2000) and only a couple that aren't in that definitive compendium.<sup>9</sup> The diagram omits for clarity the occasional revisiting of earlier steps. Toward the end, the expert reviews of model and recommendations gradually become briefings to stakeholders, to socialize

the results of the modeling effort and take the first steps in translating the recommendations into actions.



**Figure 2. Generic steps in testing the three hypotheses: Format in which the hypothesis is expressed, facts against which to test them, and validation tests used to do the tests (After Graham 2010).**

This diagram is not in itself revolutionary. Alternating hypothesis and test is the essence of the modern understanding of the scientific method—(Graham 2002) explicitly argued for defining and teaching system dynamics as a scientific method. The idea of zigzagging back and forth between facts and successive, more and more specific hypotheses has long been part of quality management toolkit (Shiba, Graham and Walden 1993, Shiba and Walden 2001), as is the idea of associating particular tools with specific steps in a process. Within system dynamics, (Groesser and Schwaninger 2009) similarly associate validation tools with steps in the modeling process.

An important feature of this framework is that it is collapsible and expandable. Often, as (Lyneis 1999) recommends, quantitative modeling goes in several stages, starting with diagrammatic modeling, then building and validating simpler models before moving on to more complex, realistic and laborious models.<sup>10</sup> Diagrammatically, that additional iteration would just repeat steps 3 through 7.

Another development path is to use a larger model to validate and support a simpler, more approachable model (Saysel and Barlas 2006). Such processes add another, different, iteration to the quantitative portion (steps 3 through 7). One specific class of iterating into a simpler model is what might be called “message models”: Models used, not to investigate an issue for the first time, but rather models used to teach specific lessons about the real world to managers. The classic market growth model (Forrester 1968) is such a lesson model, based as it was on at least two previous full-blown, validated modeling efforts and numerous field cases.

In the other direction, contraction, some application domains are well-enough understood from previous case research that proceeding directly to the stocks-and-flows modeling stage is a reasonable decision,<sup>11</sup> e.g. in telecommunications markets (Graham and Walker 1998, Ariza and Graham 2002, Graham and Godfrey 2005) and management of complex projects (Cooper 1980, Graham 2000, Cooper and Lee 2009 and many others). Another example of bringing previous work to bear is Kim Warren’s (Warren 2008) use of the resource-based view of corporate strategy to create a situation-specific model in short order, for certain kinds of situations and strategic issues. (Sterman 2000, Ch. 5) seems to position causal diagramming as an *optional* tool on the way toward a quantitative simulator.

Another type of contraction occurs when some engagements stop after quantitative scoring based on causal diagrams. For example, (Mayo, *et al.* 2001) used a diagram and a scoring system in collaboration with experts to make recommendations about how to privatize the London Underground. A later, extensively validated, simulation analysis (undertaken to look at additional issues surrounding the privatization) showed the initial recommendations to be reasonably accurate. In another example, (Graham 2010) reports on an application in law enforcement where the recommendations were validated by a measurable improvement in the system performance. For convenience and consistency, nomenclature here will follow the earlier convention (Graham 2010) and call such engagements *quantitative systems thinking (QST)*. This distinguishes QST both from simulation modeling and from typical systems thinking exercises. QST will be discussed in more detail below.

## **Abstracting a simple definition of SD**

For the purpose of describing system dynamics both simply and generally, there are two remarkable conclusions to be drawn from Figure 2. First, *every step in the modeling process is the process of hypothesis testing in some form*. Every single one. That’s a powerful statement to make to students and potential clients alike.

Second, if we are willing to aggregate hypothesis testing of system *structure* with system *behavior*, and similarly aggregate *modeler* testing of recommendations with *expert* testing of recommendations, *there are only three types of hypothesis that all of the modeling activities are testing*, namely the hypotheses that:

1. We understand what **problem** that is to be addressed
2. We understand the (complex) **system** in which the problem is created<sup>12</sup>
3. We understand the **recommendations** needed to diminish the problem or improve system performance

Definition in terms of three hypotheses is not entirely unprecedented. Descriptions of validation testing (e.g. Sterman 2000 Ch. 21 and Forrester and Senge 1980, Richardson and Pugh 1981) usually distinguish validation tests that apply to system structure or behavior from those that apply to analytical results. In dispute resolution, (Stephens et al. 2005) distinguish between hypotheses of overall system structure and behavior on one hand and hypotheses on the specific analytical outcome on the other. (For dispute resolution modeling, the third hypothesis of the three, problem, is always clear, and needs little testing—it is the need to quantify the often indirect impacts of actions that are the subject of a dispute.)

Of course, the language in which these are expressed needs to change, depending on the listener, since, for example, the phrase “hypothesis testing” can be off-putting to non-academics. For a technical audience the three hypotheses may become three types of proposition we *validate*. For a nontechnical audience, they may become three things we must *reality-check*.<sup>13</sup> And so on.

(Between you and me, gentle reader, I prefer to speak of “hypothesis testing”, because it calls unequivocally for a testable hypothesis to be stated, including model purpose. But peer-to-peer jargon is often not the best way to communicate with the rest of the world.)

*Bookends.* Those three items describe in a very compact way *what* we are doing when we do system dynamics. For potential clients or other stakeholders concerned with the “*why*” and “*where*” of system dynamic, one or two “bookends” may be useful. We could preface the definition by describing the sorts of problems for which system dynamics is especially useful<sup>14</sup>:

System dynamics is especially useful in complex problems where causes and effects are intertwined, the implications of actions are unclear, and the concern is behavior over time, often aiming to improve future behavior.<sup>15</sup>

The other bookend is for those concerned with the “*how*” of SD—certainly the more technically inclined, but often managers who don’t believe that enough data exist to make modeling a useful exercise. For such people, we need a postscript that can lead into a discussion of data and information:

We usually use computer simulation to check the second and third hypotheses against both numerical data and expert knowledge of the system. Our hypotheses draw on a wide body of research about dynamic behavior (from feedback control theory) and about decision-making in many spheres of human activity<sup>16</sup>, and in particular, on the extant body of system dynamics research.

Even with bookends, this definition is simple: It avoids the complex “first we do this, then we do that, and we use this and that” *ad nauseum*. The definition is stated in terms of something that is obviously a good thing: explicit hypothesis testing or reality-checking (phraseology depending on the specific audience). It includes elements usually missing from descriptions of modeling methodologies: Listening to the customer and making sure of what they want and need (**problem**), and working with the customer’s experts to ensure a usable result (**recommendation**).

From that fundamental description, real modeling processes become more elaborate as needed for the problem, with tools drawn in at the appropriate stage. (And students can transfer the same validation skills to other forms of models, even spreadsheets.)

Note that this simple definition does not in any way call for a “dumbing down” of SD. Indeed as we will see shortly, a more professional approach to SD probably calls for use of a wider variety of processes and approaches than “classic” system dynamics, adapting the modeling approach to the unique features of each situation, which should increase the probability of success.<sup>17</sup> The simplicity and uniformity of the three types of hypothesis encompass considerable diversity. When one model purpose differs from another, myriad modeling choices of tools and approaches will be made differently.

For example, purposes of policy design, forecasting, and planning<sup>18</sup> all demand very different strategies for model and result testing. Even though *Industrial Dynamics* identifies problems with using forecasting as an interim purpose for policy design in one type of system (damped oscillatory, heavily driven by only by unknown noise)<sup>19</sup>, (Lyneis 2000) gives arguments, and an in-depth analysis of a real-world case, that system dynamics can be the most useful approach to medium-term forecasting. (In any event, the predictability of a system, and hence the usefulness of simulation-based predictions to a client, is an outcome of analysis, not a postulate.) For applications involving different purposes such as, for example, forecasting, the tests of recommendations would revolve around robustness of prediction, out-of-sample testing, and robust policy design within uncertainties created by imperfect forecasts, in addition to the more traditional tests of recommendations.

*Ubiquity of the hypothesis testing frame.* The hypothesis-testing framing also extends into the details of model development and testing. Traditionally, there is very little systematic instruction on how to start writing model equations and how to debug the results. However, starting to write equations for a piece of structure (from a causal diagram, for example) can be framed as creating a miniature dynamic hypothesis: “I think this kind of (equation) structure should produce this output in response to this input.” That immediately suggests how to structure test inputs and proceed with testing.

Likewise, validating a fully-written model against appropriate time series measures is traditionally intuitive only for very experienced modelers. Often, locating an effective point to

change the model is exceedingly difficult, especially in oscillatory or oscillation-driven systems. Tracing cause and effect around feedback loops simply reveals that everything is going up and down. Inexperienced modelers often start changing parameters and formulations almost at random. Experienced modelers do a scan of all of the points of discrepancy between simulation and real behavior, and use knowledge of feedback systems and the causal structure of the model in question to create a mini-dynamic hypothesis about what might correct the misbehavior.

Jay Forrester has always asserted that knowledge of how structure produces behavior is a necessary part of a modeler's toolkit. That is why (Forrester 1968) devotes considerable space exercising the properties of integrators, first- and higher-order delays. Similarly, (Graham 1977) collects generic principles relating cause to effect, mostly for oscillatory systems, that can serve as hypotheses in diagnosing causes and influences on oscillatory behavior. In the present discussion, we can now sharpen the justification for needing such knowledge: Professional modelers need knowledge of how structures create behavior as a source for hypotheses in creating system structure, and diagnosing system behavior and policy response.

The beauty of the hypothesis-testing framework and the framing definition of what system dynamics is about is that it is one simple idea, applied again and again, in describable and teachable ways. That simplicity has advantages in multiple venues:

First, even though the hypothesis-testing framework can be ramified into almost unbounded amounts of complexity, it can be taught, beginning to end, in the very first lesson. Teachers can start providing more structure and process for the discipline of system dynamics modeling, especially for previously-murky parts of the modeling process. In all elaborations of the framework after a simple beginning, the learners can have a clear picture of how whatever they are currently learning fits into the whole of system dynamics practice.

Second and perhaps most importantly, the hypothesis-testing framework (with language suitably adjusted to the listener) allows a brief and accurate “elevator version” that communicates the nature and value of system dynamics to non-system dynamics modelers—potential clients of system dynamics work—in a handful of sentences.

We can aspire to make the situation for system dynamics like that of chess, an activity that has endured for hundreds of years and spread around the globe, where the essential structure is simple and the variations are complex and seemingly endless. And yet it is teachable and usable (for our enjoyment) almost immediately.

Since the hypothesis-testing framework for system dynamics may be unfamiliar on the face of it, let me give some examples before discussing varieties of system dynamics practice.

## Three types of hypothesis

The first type of hypothesis is a characterization of the **problem** that is to be addressed, and the test is whether the recipients of the results see the characterization as an accurate statement of their needs or desires. Any number of formats will do for this, as long as they can be understood (in the same way) by modelers and clients. One format is what has been called a *model use diagram*, a kind of input-output diagram. It also gives a simple view of how the work will be done:



**Figure 3. Example hypothesis of understanding the client's problem—a model use diagram.**

This is a simple diagram, but it captures what's to be done at a very high level and in a way that's understandable by normal people. In the arrow on the left side, it establishes what policy actions are on the table (and off the table). It establishes important issues that we won't try to model dynamically—they're exogenous test or scenario inputs (bottom arrow). In the arrow on the right side, we've understood what constitutes a "good" outcome for a particular organization. (It's rarely just Net Present Value.) The lower right corner puts some explicit bounds on the study—time horizon and products. The lower left corner captures a puzzle of the type that often lies at the heart of an organization's perceived need to bring in dynamic modelers. That can be the kind of puzzle that if a modeling effort doesn't answer that question, it won't be seen as adequate. That corner is also the place to capture hoped-for and feared future behaviors, if these are at the forefront of the client's concerns. The diagram is stating the *modeler's* hypothesis on the problems to be addressed. It is tested by critical review with the client. Then it becomes a mutually understood purpose for the upcoming modeling.

A model of the **system** constitutes the second type of hypothesis. It codifies our understanding of the system in which the problems are (or will be) created—endo (within) genously (created). The hypothesis testing of simulation models uses the multiple tests familiar to system dynamics modelers, as enumerated in (Sterman 2000, Ch. 21), (Forrester 1959) and numerous other textbooks.

The hypothesis testing of the causal diagram model in qualitative systems thinking (QST) takes place in link-by-link discussion and review by subject matter experts—people who have been able to observe the real system in operation in depth and at length. Almost always, such hypothesis testing requires multiple experts and multiple sessions to review and ultimately agree on all aspects of the system being diagrammed. (Mayo, *et al.* 2001) describes the scope of one such effort. Mechanical details of this process will be discussed below.

The third type of hypothesis is our understanding of **recommendations** that will improve the performance of the system, and diminish or eliminate the problematic behavior. It might be argued that testing this hypothesis could be considered just one more type of system testing. It is separated here for four reasons, increasing order of importance

First, the hypothesis tests for policy or recommendations are uniformly quite different from those for system validation. For example, behavior sensitivity testing involves comparing two simulations, the base and the base with a parameter variation. Policy behavior sensitivity testing involves comparing two policy impacts and four simulations: the base and the base with a new policy, which defines one policy impact, and the base with a parameter variation, and the base with a parameter variation and a new policy, which defines another policy impact. Are the two policy impacts roughly comparable? If so, the policy sensitivity to that parameter is low. This can easily be true even if the behavior itself tests very sensitive to that same parameter.

Likewise for expert reviews, the foci are completely different between model reviews and recommendations reviews. Expert review of system behavior looks at a simulation in comparison to real data (using appropriate points of comparison, whether that is matching time series, or matching relative phase and amplitude of fluctuation). Expert review of a policy impact compares base case behavior with behavior under different policies, with a focus on examining relationships that the modelers have verified are important to creating the behavior change within the model.

The second reason for separating policy testing is a countermeasure to an extremely common problem in modeling engagements: There is almost always too little time devoted at the end of an engagement to policy testing by modelers and experts alike. In engagements of fixed duration, any slip along the way ends up shortening the calendar time devoted to policy testing. Too often, the last day of a modeling project finds the modelers making the last formulation fixes and running and reporting the policy results in the same day.

Third, policy testing is the capstone of the modeling effort. Until the modeler knows what conclusions are being generated, and the assumptions that are critical to those conclusions, it is impossible to even make final judgments about the previous step, whether validation of the underlying system is adequate to support the conclusions. Think of a hiker. Until the hiker knows the character of the destination and the terrain (and weather) in between, it is impossible to judge whether the backpack has appropriate equipment, or to even estimate what time of day to start out.

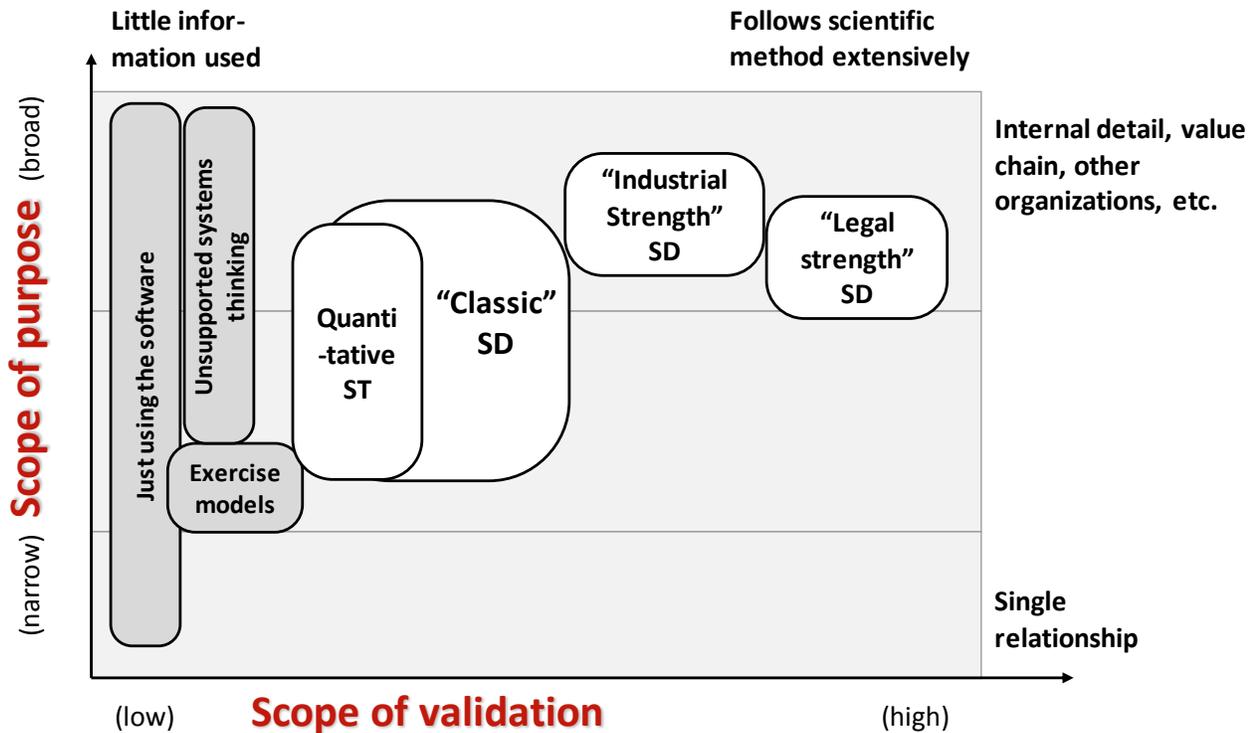
Fourth, the field of system dynamics has generally treated implementation as different and separate from system dynamics, and pretty much left students and future professors on their own.<sup>20</sup> Including an explicit step of testing hypotheses about recommendations against expert's knowledge and client perceptions is a bridge to implementation, using processes similar in form to other hypothesis-testing steps (that one hopes can therefore be understood and performed easily by modelers), that nonetheless is a start in the process of winning hearts and minds.

We have reached the point of having a reasonably encompassing definition of system dynamics, with three hypotheses and two bookends, which would seem about the right level of complexity. But the discussion has surfaced numerous variations in specific practices, and it should be clear that the discussion only started to identify potentially hundreds of variations. So, again in the spirit of simplification, let us now turn to typical varieties of practice.

## **Variations from the Classic**

Figure 2 is a graph using a pair of axes originally developed to position system dynamics in relation to common corporate strategy tools (Graham 2004), but they will serve here to discuss variations in system dynamics practice (and some practices that aren't system dynamics) in terms of broad clusters. The vertical axis is scope of analysis, which includes both diversity of organizations and decisions represented, and also the detail with which they are represented. The horizontal axis is rigor of analysis and use, meaning the intensity of information gathering and hypothesis testing.

Many of the labels are my own neologisms, which I hope convey the essence of differences between variations without prejudice. This is certainly not the only taxonomy possible, but it does serve to describe the observed varieties of system dynamics practice. Take it as a (hopefully) useful conceptual tool.



**Figure 4. Varieties of SD practice (white) and non-SD practice (grey).**

There are some variations that most system dynamicists can agree are better considered not to be part of system dynamics, due to lesser rigor than classic SD. Starting at the left of the Figure:

*“Just using the software”*. Probably most of us have run into people who claim to know about system dynamics, and it turns out that’s based on being able to use the software to produce simulations. Period. The stated purpose is often to “understand” the system. There is little effort to test the model against either numerical data or expert knowledge. Let’s exclude such “practices” from our definitions of system dynamics.

*“Exercise models”*. Their purpose is not improving performance of some system (real or imagined), but preparing modelers in specific technical skills.

Because model purpose and real use are so much a part of both the original conception of system dynamics and the simple definition above, I propose to exclude exercise models and their use from the definition of “doing” system dynamics—they’re the analogs of scales and arpeggios in musical performance: They support particular aspects of doing system dynamics, no more.

This is not to oppose use of exercise models; Rather, it is begging for “truth in labeling”—that when a given exercise model is presented, its explicit purpose is framed as technical exercise of a specific modeling skill. To the extent that we lump all simulators together, all equally “system dynamics models and modeling”, we run a risk of alienating the most critical thinkers, exactly

the people we'd like to attract into the field, as well as misleading students on the nature of professional practice in policy analysis.

For example, even papers published in the *System Dynamics Review* not infrequently derive recommendations by simply varying a policy parameter, with no reported attempt to understand why the results happen, nor with any reported testing of that understanding, having neither a stated hypothesis of what mechanisms are primary in producing the outcome, nor technical tests of such an hypothesis, nor expert review of the outcomes and the hypothesized mechanisms that produce them. Just vary some policies and state the recommendation.

This lack of hypothesis testing of recommendations would be roughly equivalent to the lack of hypothesis testing that would have occurred if one formulated and simulated a model, but then failed to looking at the simulation to see whether the behavior is realistic, or even whether it successfully implements the dynamic hypothesis. It would be roughly equivalent to the lack of hypothesis testing that would have occurred if one assumed a model purpose but never checked it at all with the clients. To avoid such gaps in professional practice, we need to show students an accurate picture of what professional practice is and isn't, which includes carefully distinguishing between exercise models and modeling exercises that cover a complete cycle of modeling and model use.

*“Unsupported systems thinking”* Systems thinking, with its use of archetypes and diagramming, is an attempt to bring some of the power of system dynamics to normal people without the extensive background and training that quantitative system dynamics modeling requires. Undoubtedly the results are less trustworthy than a well-done system dynamics effort. But the results seem like they would be somewhat *more* trustworthy than simply talking problems out, which in many cases is the alternative methodology.

“Unsupported systems thinking”, here denotes systems thinking exercises supported neither by the facilitation of an experienced system dynamics modeler, nor any effort to refine and sharpen the qualitative relationships on the diagram through quantification. Experienced system dynamics modelers will usually have dealt with similar or analogous problems, and verified through computer simulation how cause and effect relations produced the problem, and what kinds of actions remedy it. Such a modeler will also be adept at spotting compensating and amplifying feedback loops. In evolving a diagram without the support of a background of such rigor and experience, we should probably exclude such systems thinking exercises from our definition of system dynamics.

*“Quantitative systems thinking (QST)”*. This is a term used in an overview of economic modeling methods (Graham 2010) to describe the use of a causal diagram to do quantitative scoring of strength and timing of links with expert participants in the system, and based on those characterizations, scoring the effectiveness of alternative interventions (Mayo *et al.* 2001, Graham 2010, Pagani and Fine 2008). Imagine a causal loop diagram where each link between

variables subject matter experts facilitated by experienced SD modelers assign a strength score (high medium, low) and a delay score (long, medium, short), where these are defined suitable to the problem setting. Likewise each of the potential actions (among which will eventually be found the recommended actions or policies) is scored (high, medium, low) on variables they directly, causally impact. Measures of goodness of outcome will have been identified (for example, on a model use diagram) and diagrammed. So it becomes possible to trace all the (significant) paths from interventions to outcomes, and thus score the “bottom line” impacts of all of the candidate interventions. In its most intensive form, QST then examines (again, with subject matter expert review) interactions among actions and desirability scores of bundles of actions, and sensitivity testing to identify and revisit key quantitative assumptions.

On Figure 2, quantitative systems thinking consists of executing steps one through four. QST modeling engagements test all three of the fundamental hypotheses of system dynamics modeling—but using evaluation of algebraic relationships (scorings) instead of difference equations. There is the same stakeholder review of the problem to be solved, the same subject matter expert (SME) review of the diagram and its scorings, and the same SME and stakeholder review of recommendations. The knowledge of feedback dynamics does enter only qualitatively, through the experience of the (modeler) facilitator. There is the parallel extensive sensitivity testing to deal with the imprecise nature of the “data” and its translation into conclusions.

Quantitative scoring and analysis forces far more focused discussion on relative importances of links and impacts of actions than does simple examination of loops. As a result, the first two QST cases cited above were later vindicated. In the first case, a later and very well-validated quantitative model gave essentially the same results. In the second case, implementation of the recommendations created the predicted performance improvements. QST shares many of the process strengths of dynamic simulation-based system dynamics, including testing all three types of hypothesis. Moreover, in a practical setting, if a client’s time and budget situation do not permit any form of analysis based on dynamic simulation modeling, QST will often be the best alternative that a system dynamics modeler can offer. From published cases and my own (unpublished) case, it seems vastly superior to what would likely happen without an explicit modeling methodology: BOGSAT (Bunch of Guys Sitting Around Talking), or perhaps BOGLES (Bunch of Guys Lying with Excel Spreadsheets)<sup>21</sup>.

For all these reasons, quantitative systems thinking is here included within the broad definition of system dynamics.

Unfortunately, the published use of diagrammatic models has largely, if not entirely (exclusive of these three examples) remained at the level of simply examining diagrams, e.g. (Coyle 2000, Pruyt 2009) without any quantitative element. At the strictly qualitative, unscored level, diagrammatic models remain controversial within the field of system dynamics, e.g. (Richardson 1999, Homer and Oliva 2001). Rightly so.

Next come variations of fully quantitative system dynamics, foreshadowed above in the discussion of expansions and contractions of the modeling process in Figure 2. Differences among the variations are matters of degree, emphasis, and custom. Boundaries are not sharp and individual modeling efforts can easily be “in between”. Nonetheless, in practice we can distinguish three clusters of practices.

“*Classic*” system dynamics. This, in the center of the diagram, is the body of tools and processes first articulated in *Industrial Dynamics* and later codified in *Business Dynamics* and other textbooks. *Urban Dynamics* and *World Dynamics* (Forrester 1969 and 1971) are exemplars. Classic system dynamics usually represents the balance and preferences of tools and techniques of system dynamics as is usually articulated and taught.

- Model purpose is seldom stated explicitly or tested, but is nonetheless the implicit purpose is often used to make modeling choices by experienced practitioners.
- The starting point for modeling is usually considered to be a single dynamic hypothesis, which comes from the modeler, not the client or result user.
- Information for modeling tends to be dominated by qualitative knowledge about cause and effect
- Testing of model against that knowledge tends to be done within the modeler’s head.<sup>22</sup> (Group model building is a partial exception.)

*Group model building*. Some readers may have expected group model building to come next in the spectrum. However, “groupness” seems best considered as an additional dimension, orthogonal to those shown in Figure 4. Group model building has been done both in the classic system dynamics manner (e.g. how Jay Forrester interacted with a group of urban experts to create the *Urban Dynamics* model), and in the style of industrial strength system dynamics (e.g., Ariza and Graham 2002), which will be discussed next.<sup>23</sup>

“*Industrial strength*” system dynamics. This is a variety of system dynamics performed for paying customers in support of real decisions, as reported in (Roberts 1981) a handful of *System Dynamics Review* articles, and the System Dynamics Society’s Applications Award. It is the variety of system dynamics practiced by the enduring system dynamics consultancies, such as Ventana Systems or PA Consulting, and advocated by experienced practitioners such as (Homer 2007) and (Barlas 2007). Arguably, *Dynamics of Growth in a Finite World* (Meadows *et al.* 1974) is industrial-strength system dynamics.

This category is labeled “industrial strength” system dynamics to distinguish it from the weakly-validated published studies that are have not been initiated and designed to be used to guide real decisions in any meaningful way. This is not to say that models of the “classic” variety have not sometimes been enormously useful in real-world applications.<sup>24</sup> Nonetheless, there is a reasonably distinctive variety of system dynamics practice that is both quite common in the

setting of real-world application that has several characteristics quite different from “classic” system dynamics:

- Explicit (often contractual) statements of purpose, carefully checked with stakeholders. There is early and continued discussion about who wants the modeling results, why they’re wanted and what actions will be taken based (partially) upon them.
- Wide variation in the nature of purpose and “recommendations”, from classic SD policy design (algorithms for responding to system conditions), to single major decisions or results (e.g. “the damages claimed are 130 million person-hours at prevailing wage rates”, or “use a shared sales force”), to ongoing forecasts.
- There are usually multiple dynamic hypotheses to be dealt with, at least two, for example conflicting views within management of what’s important in the system and what’s implied for best strategy. Another source of multiple dynamic hypotheses is uncertainty, which gives rise to both hoped-for and feared scenarios, each with their own dynamic hypothesis.
- Explicit knowledge elicitation and Subject Matter Expert (SME) reviews<sup>25</sup> at multiple points in the modeling process. (Graham and Walker 1998) term the separate review by multiple diverse experts “distributed validation”. Of course, groups of SMEs are also often used, to get synergy of knowledge and viewpoints.
- Validation also uses available time series data as an additional source of model refinement, as well as an additional source of user confidence both directly (the model passes a validation test they can see and understand) and indirectly (for the modelers to understand what’s going on in the data creates much deeper understanding of the real world system). Usually, the characteristics of the system and problem allow simple comparison of simulation versus time series (more on this shortly).
- Choices of model format and level of detail, level of trustworthiness of results are all strongly influenced by schedule and budget, the objective being to produce the most trustworthy, useful results within time X and budget Y.
- Sensitivity testing and extreme conditions testing are performed according to expert judgment as to their potential contribution, weighed against schedule and labor cost.<sup>26</sup> Usually the judgment is to restrict sensitivity testing to a few key points in the system.

The use of falsification testing through comparison to time series bears elaboration, for it is probably the most prominent difference between “classic” system dynamics, and the “industrial strength” system dynamics that commercial practitioners have used for decades, and found to be useful and insightful. For example (Homer 2007, Barlas 2007) both stress use of time series as a practice that successful professionals use and benefit from.

The traditional reading of *Industrial Dynamics* is that it recommends against the use of time series data. This is inaccurate. Forrester *recommends* validating model behavior against *appropriate* properties of time series, e.g. (Forrester 1959 Section 13.5, pg. 121). For oscillatory

systems predominantly driven by unknown random noise, direct point-by-point comparison of simulation to data series can indeed be misleading. And both of the systems described in *Industrial Dynamics* were of that class of system. In such circumstances, one can either fall back on matching appropriate frequency domain characteristics such as relative phasing and relative amplitudes (Forrester 1959, pg. 120)<sup>27</sup>, or use mathematical validation appropriate to that type of system, which is optimal filtering, also known as Kalman filtering (Peterson 1975, 1980)<sup>28</sup>.

In fairness to Forrester, the mathematics of optimal filtering did not exist when *Industrial Dynamics* was written. But it has been available in the Vensim simulation software for decades. Arguably, the system dynamics field really has no excuse for avoiding the use of time series (often even to the point of neglecting to look for even behavior characteristics in data) merely as a matter of custom or ideology.

In practice, the need for full optimal filtering is relatively rare; most systems of interest are driven exogenously by large known events. Markets are typically driven exogenously by economic cycles. Complex projects are driven by their own milestones, and by prominent known setbacks along the way. When the largest of those events are represented, the numerous smaller variations typically average out, and comparison of simulation to time series is demonstrably a valid comparison to attempt.

One sometimes hears that there is little point to comparison of simulation to time series, because causing them to match is trivial curve-fitting. Indeed, (Forrester 1959, pg. 121) suggests this. In practice, “curve fitting” is anything but trivial. Any one parameter typically impacts behavior of multiple variables that must be matched to multiple time series, so the problem is finding whether there exists an ensemble of parameter values capable of causing behavior that matches all of the time series simultaneously. Technically, this is still curve-fitting of sorts, but where many parameters map into behavior of many time series, and the mapping isn’t decomposable into simpler curve-fitting problems. The situation often resembles a Rubik’s cube more than it does simple curve fitting.

Often, finding such an ensemble of parameter values proves to be simply impossible. In that case, a careful examination ensues of both the model formulations and the data sources and their interpretation. Detections of flaws (or flaws in understanding) occur in both data and model, in roughly equal proportions. Long experience has shown that comparison to time series is a fruitful means of refining model formulations, and deepening knowledge about the system being modeled. This is what’s supposed to happen in validation testing. It seems wasteful to abstain from such useful hypothesis tests, particularly in situations where the stakes are high.

Again in fairness to Forrester, his advice was written when simulations were still specified on punch cards and every single simulation experiment took some hours to turn around. For that matter, time series data were likely much less plentiful and easily obtainable at that time. Under the original circumstances, the (time) costs of experiments in parameter variation were several

orders of magnitude larger from what they are now. The cost-benefit calculus of testing against time series data has shifted dramatically in the half century since *Industrial Dynamics* was published.<sup>29</sup>

*“Legal strength” system dynamics.* This is system dynamics used for dispute resolution, either lawsuits or arbitration processes, typically concerning management of complex development and construction projects. Dispute resolutions are adversarial processes, with high stakes and long schedules, so there is not only more intensive validation testing, but also higher requirements for a documentary “audit trail” from expert interviews to model assumptions to model tests and analytical conclusions (Stephens *et al.* 2005). Regulatory disputes, such as in (Graham and Godfrey 2005) lie in between industrial strength and legal strength in terms of evidentiary requirements. Typical dimensions of legal strength system dynamics include:

- The purpose and fundamental analytical result is almost always the difference in a cost measure (dollars, labor hours, etc.) between an “as happened” simulation and a “but for” simulation, where the “but for” simulation takes away the specific actions and conditions that are the subject of disagreement, showing what would have happened “but for” those actions. For example, in the first arbitration to use system dynamics (Cooper 1980), the “as happened” simulation of a shipbuilding program included the customer making design changes, and delaying approval for other designs (which caused ripple effects in the design effort). The “but for” simulation simulated the shipbuilding without these elements, demonstrating that a significant portion of the project overrun was due to those customer actions and inactions.
- The majority of cases concern large development and construction projects, where the model structure is re-used from earlier applications, changing the number of project work phases to fit the project currently being modeled. This is both economical, and, by using a model structure already validated against other complex projects, increases the credibility of the analysis. There are usually no causal diagramming steps as such. Typically, diagramming is used to explain, rather than develop, the simulator.
- Expert interviews are more extensive and more structured than *de novo* industrial strength models for commercial or governmental strategy and policy.
- There is extensive matching to time series.
- Sources for time series and qualitative interview information, and their linkage to specific parts of the model, are completely documented.
- As the case progresses, the simulator is used to test alternative “theories of the case”—the opposition’s hypotheses why events happened the way that they did. The objective is to test the extent to which the alternative hypotheses are or are not consistent with the available facts of the case.
- The opposition hires its own system dynamics experts to review and critique the modeling, and the client often hires a third-party system dynamics expert reviewer to do “quality control” on the modeling effort.

- Systematic sensitivity testing tests both robustness of model structure, and robustness of analytical results. Recent practice in tested analytical results (the difference between “as happened” and “but-for” cost metric) is to use “fit-constrained Monte Carlo” analysis of outcome variability. This method finds confidence bounds for the outcome by looking at outcome variability from only those parameter changes that produce behavior consistent with the observed time-series behavior (Graham *et al.* 2002, Stephens *et al.* 2005).

To allow side-by-side comparison, Table 1 summarizes some of the features of qualitative systems thinking, classic SD, industrial-strength SD and legal-strength SD:

	<b>Problem</b>	<b>Dynamic Hypothesis</b>	<b>Information</b>	<b>System &amp; Recommendations Testing</b>
Quantitative systems thinking (QST)	Explicit	Many, with uncertainty	Expert cause & effect knowledge with scoring	Sensitivity testing, focused expert review
Classic SD	Implicit	One	Mostly cause & effect – little quantitative info.	Within modeler
Industrial strength SD	Explicit	Multiple competing	Quantitative & expert cause & effect knowledge	Focused, with experts
Legal strength SD	Impact quantification	Multiple (adversarial) theories of the case	Quantitative & expert cause & effect knowledge	Extensive, with experts. Confidence bounding. Third party review.

**Table 1. Selected features of varieties of SD.**

To summarize: Are all four of these practices “system dynamics”? They all test the same three fundamental hypotheses, use mostly same tools, and work from the same knowledge base, and differ more in matters of emphasis and custom than in fundamental concepts or skills. We can easily say yes, they are all system dynamics, and moreover, still other forms of system dynamics may lie in the future.

### Concluding Remarks

Three significant variations from “classic” system dynamics have emerged over the fifty years since *Industrial Dynamics* was published: Industrial- and Legal-strength system dynamics, and quantitative systems thinking. They all draw from the common toolkit of system dynamics, but with a different selection of tools, different choices of modeling steps, and different emphases.

It would be desirable for teachers and practitioners to incorporate all of these varieties of practice. But dealing with them one by one would be immensely time consuming and infeasible in practice; it is far more effective to teach and communicate the simple concepts that lie at the heart of all of them, and start from the simplest vocabulary to capture them:

System dynamics addresses complex, intertwined issues. To do that reliably, we reality-check that we understand three things:

- The **problem(s)** to be addressed
- The **system** they happen in
- The **recommendations** to address the problem(s)

We use computer simulation and lots of information and data about how people and organizations interact to check the system and the recommendations.

**Figure 6. “Elevator story” definition of system dynamics**

Following such a definition, the selling process or other conversations can go in any number of directions, from a solid foundation of shared understanding.

## **Acknowledgements**

The author is grateful to colleagues at Greenwood Strategic Advisors for time to set these thoughts on paper, and to colleagues at PA Consulting and to Jay Forrester and MIT, for the privilege of decades of experiencing the realities of system dynamics practice. Thanks to Jim Lyneis and George Richardson for thoughtful reviews of early drafts. Any unclarity or errors are of course my own.

## **References**

Andersen, David F. and George P. Richardson 1997. “Scripts for Group Model Building”. *System Dynamics Review* 13(2): pp. 107-129.

Ariza, Carlos A. and Alan K. Graham 2002. "Quick and Rigorous, Strategic and Participative: 12 Ways to Improve on the Expected Tradeoffs". *Proceedings of the 2002 International System Dynamics Conference*, Palermo, Italy.

Barlas, Yaman 2007. "Leverage Points to March 'Upward from the Aimless Plateau'". *System Dynamics Review* 23(4) 469-73.

Cooper, Kenneth G. and Gregory Lee 2009. "Managing the Dynamics of Projects and Changes at Fluor". *Proceedings of the 2009 International System Dynamics Conference*, Albuquerque, New Mexico, USA. This paper was the winner of the 2009 System Dynamics Society Application Award.

Cooper, Kenneth G. (1980). Naval Ship Production: A Claim Settled and a Framework Built. *Interfaces* 10(6), Dec. 1980, pp. 20-36.

Coyle, Geoffrey 2000. "Qualitative and Quantitative Modeling in System Dynamics: Some Research Questions". *System Dynamics Review* 16 (3), 225–244.

Forrester, Jay W. 1961. *Industrial Dynamics*. Cambridge, Mass.: MIT Press.

Forrester, Jay W. 1968. *Principles of Systems*. Waltham, Mass: Pegasus Communications.

Forrester, Jay W. 1968. "Market Growth as Influenced by Capital Investment". Originally published in *Industrial Management Review*, Cambridge, MA: MIT Sloan School, 9(2). Reprinted in: Forrester, Jay W., 1975. *Collected Papers of Jay W. Forrester*. Waltham, MA: Pegasus Communications.

Forrester, Jay W. 1969. *Urban Dynamics*. Waltham, Mass: Pegasus Communications.

Forrester, Jay W. 1971. *World Dynamics*. Waltham, Mass: Pegasus Communications.

Forrester, Jay W. 2007. "System Dynamics--the Next Fifty Years". *System Dynamics Review* 23(2/3) Summer-Fall, pp. 359-370.

Forrester, Jay W. and Peter M. Senge 1980. "Tests for building confidence in system dynamics models". In Legasto, Augusto, Jay W. Forrester and James M. Lyneis, ed.s, *System Dynamics*. TIMS Studies in the Management Sciences 14. New York: North Holland, 209-228.

Forrester, Nathan B. 1982. *A Dynamic Synthesis of Basic Macroeconomic Theory: Implications for Stabilization Policy Analysis* (PhD thesis). Cambridge, MA: Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

Graham, Alan K. 1977. *Principles on the Relationship between Structure and Behavior of Dynamic Systems*. Cambridge, Mass.: Massachusetts Institute of Technology Ph.D. dissertation, available at <http://libraries.mit.edu/docs/theses.html>.

Graham, Alan K. 2000. "Beyond PM 101: Lessons for Managing Large Development Programs". *Project Management Journal* 31(4), pp. 7-18.

Graham, Alan K. 2002. "On Positioning System Dynamics as an Applied Science of Strategy, or SD is Scientific. We Haven't Said So Explicitly and We Should". In *Proceedings of the 2002 International System Dynamics Conference*, Palermo, Italy. Available at <http://www.systemdynamics.org/publications.htm>

Graham, Alan K. 2004. "Pandemic Strategy Disconnects and How to Bridge Them". *Proceedings of the 2004 Strategic Management Society Annual Conference*, San Juan, Puerto Rico, USA

Graham, Alan K. 2010. "Economics and Markets" in Kott, Alexander and Gary Citrenbaum, eds., *Estimating Impact: A Handbook of Computational Methods and Models for Anticipating Economic, Social, Political and Security Effects in International Interventions*. New York: Springer.

Graham, Alan K. and Jeremy Godfrey 2005. "Achieving Win-Win in a Regulatory Dispute: Managing 3G Competition". *Proceedings of the 2005 International System Dynamics Conference*, Boston Massachusetts. Available from <http://www.systemdynamics.org/>.

Graham, Alan K., Jonathan Moore and Carol Choi 2002. "How Robust Are Conclusions from a Complex Calibrated Model, Really? A Project Management Model Benchmark Using Fit-Constrained Monte Carlo Analysis". *Proceedings of the 2002 International System Dynamics Conference*, Palermo, Italy. Available from <http://www.systemdynamics.org/>.

Graham, Alan K. and Robert J. Walker 1998. "Strategy Modeling for Top Management: Going beyond Modeling Orthodoxy at Bell Canada". *Proceedings of the 1998 International System Dynamics Conference*, Quebec Canada.

Groesser, Stefan and Markus Schwaninger 2009. "A Validation Methodology for System Dynamics Models". Paper presented at the 27th International System Dynamics Conference, Albuquerque, NM.

Homer, Jack and Rogelio Oliva 2001. "Maps and Models in System Dynamics: A Response to Coyle". *System Dynamics Review* 17 (4), 347-355.

Homer, Jack 2007. "Reply to Jay Forrester's 'System Dynamics--The Next Fifty Years'". *System Dynamics Review* 23(4), 465-7.

Lyneis, James M. 1999. "System Dynamics for Strategy: A Phased Approach". *System Dynamics Review* 15(1), pp. 37-70.

Lyneis, James M. 2000. "System Dynamics for Market Forecasting and Structural Analysis". *System Dynamics Review* 16(1), pp. 3-25.

Mayo, Donna, Michael Callaghan and William J. Dalton 2001. "Aiming for restructuring success at London Underground". *System Dynamics Review*, 17(3), pp. 261-289.

Meadows, Dennis L., William W. Behrens III, Donella H. Meadows, Roger F. Naill, Joergen Randers and Erich K. O. Zahn 1974. *Dynamics of Growth in a Finite World*. Waltham, MA: Pegasus Communications.

Pagani, Margherita and Charles H. Fine 2008. "Value network dynamics in 3G-4g wireless communications: A systems thinking approach to strategic value assessment". *Journal of Business Research* 61: 1102-1112.

Peterson, David W. 1975. *Hypothesis, Estimation, and Validation of Dynamic Social Models*. Cambridge, Massachusetts: Massachusetts Institute of Technology Ph.D. thesis.

Peterson, David W. 1980. "Statistical Tools for System Dynamics". In Randers Jorgen, ed. 1980. *Elements of the System Dynamics Method*. Portland, Oregon: Productivity Press.

Pruyt, Erik 2009. "Making System Dynamics Cool? Using Hot Testing & Teaching Cases". In *Proceedings of the 27th International System Dynamics Conference*, Albuquerque NM, USA.

Richardson, George P. and Alexander L. Pugh III 1981. *Introduction to System Dynamics Modeling Using DYNAMO*. Portland, Oregon: Productivity Press.

Richardson, George P. 1999. "Reflections for the future of system dynamics". *Journal of the Operational Research Society* 50 (4), 440-449.

Richardson, George P. 2011. "Reflections on the Foundations of System Dynamics". *System Dynamics Review*, forthcoming, published online 2011.

Richmond, Barry 1997. "The Strategic Forum: Aligning Objectives, Strategy and Process". *System Dynamics Review* 13(2), pp. 131-148.

Roberts, Edward B. 1981. *Managerial Applications of System Dynamics*. New York: Productivity Press.

Saeed, Khalid and Oleg V. Pavlov 2008. "Dynastic Cycle: A Generic Structure Describing resource allocation in political economies, markets and firms". *Journal of the Operational Research Society* 59(10): 1289-98.

Saysel A.K. and Yaman Barlas 2006. "Model Simplification and Validation with Indirect Structure Validity Tests". *System Dynamics Review* 22(3) 241-62.

Senge, Peter M. 1990. *The Fifth Discipline*. New York, NY: Doubleday.

Senge, Peter M., Art Kleiner, Bryan J. Smith, Charlotte Roberts and Richard B. Ross 1994. *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization*. New York: Crown Publishing Group

Senge, Peter M., Richard B. Ross, Art Kleiner, Charlotte Roberts and George Roth 1999. *The Dance of Change: The Challenges of Sustaining Momentum in a Learning Organization*. New York: Crown Publishing Group

Shiba, Shoji, Alan K. Graham and David Walden 1993. *A New American TQM: Four Practical Revolutions in Management*. Portland, Oregon: Productivity Press.

Shiba, Shoji and David Walden 2001. *Four Practical Revolutions in Management: Systems for Creating Unique Organizational Capability*. Portland, Oregon: Productivity Press.

Stephens, Craig A., Alan K. Graham and James M. Lyneis 2005. "System dynamics modeling in the legal arena: Meeting the challenges of expert witness admissibility". *System Dynamics Review* 21(2), pp. 95-122.

Sterman, John D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York, NY: McGraw-Hill/Irwin

Vennix, Jac A.M., David F. Andersen and George P. Richardson, eds. 1997. Special Issue on Group Model Building, *System Dynamics Review* 13(2).

Warren, Kim 2008. *Strategic Management Dynamics*. New York: John Wiley & Sons.

Warren, Kim 2011. "Challenging our slogans". Discussion Paper for the 2011 European System Dynamics Conference.

## Notes

---

<sup>1</sup> In commercial practice experience can be institutionalized and retained (rather than throwing someone into solo practice at a university after the first sizable research effort). Consider this: If PhD thesis research represents a person-year's worth of effort, a consulting organization of 15 co-located modelers is turning out the research equivalent of 15 PhD theses every year. And most modelers will be intimately involved in 3-4 projects per year, which means that, e.g. five years of experience translates to perhaps 12-15 research projects (some projects go longer than the typical 8-12 months). And since modeling is easier to sell if a client can see that similar work has succeeded, many projects will be in the same industry, or tackle related problems. So experience can deepen as well as broaden. And opportunities to innovate in method are similarly legion. Also, it is quite true that necessity is the mother of invention: Many of the variations in SD practice discussed here were necessitated by the client situation.

<sup>2</sup> English-speaking readers may find the title of this paper familiar-sounding. It is an allusion to William James' *Varieties of Religious Experience, a Study in Human Nature*. It is even more an homage to James' *Pragmatism*, a 1906 work that foreshadows much of current thinking about scientific method, as well as the primacy of purpose in analysis: "The pragmatic method...is to try to interpret each notion by tracing its respective practical consequences. What difference would it practically make to anyone if this notion rather than that notion were true? If no practical difference whatever can be traced, then the alternatives mean practically the same thing, and all dispute is idle." (Lecture 2) "...all the branches of science that investigators have become accustomed to the notion that no theory is absolutely a transcript of reality, but that any one of them may from some point of view be useful." (Lecture 6)

---

<sup>3</sup> The reason for the qualifier “generally” is that the stated purpose is merely “to explore” a specific set of questions. Forrester did carry the exploration far enough that managerial consequences became obvious. In general, for a well-focused modeling effort, if better decisions in some specific organization are the real objective, it is best to state that, and indeed explore questions about the limits of decisions and actions that are “on the table” for exploration. It is a waste of time and resources to explore options over which the particular organization truly has no control.

<sup>4</sup> Pierre Wack, originator of Shell Oil’s scenario planning process, borrowed the term “transitional object” from developmental psychology, where it denoted an object used to transition an infant from constant nurturing by a mother to independent existence. “Security blankets” are one example. Since Wack’s borrowing, the term has come to denote conceptual transitions as well.

<sup>5</sup> Proposals from experienced consultancies also are often accompanied by a block diagram showing endogenous versus exogenous elements, some important dimensions of level of detail, and primary cause and effect relationships. A block diagram tends to anchor expectations about the scope of detailing in the future model. This is another transitional object.

<sup>6</sup> Jay Forrester was unusual in this regard. During the period of Urban Dynamics, World Dynamics and the SD National Economic model, he regularly hired staff to help and teach modelers to write, as well as offering a writing seminar just for system dynamics topics.

<sup>7</sup> One gratifying feature of the sequels to *Fifth Discipline* (Senge 1990) such as (Senge *et al* 1994, 1999) is the renewed emphasis on how to engage organizations and translate analytical findings into decisions and action. These books have in effect been bringing many of the learnings of the field of Organizational Development to a new generation of aspiring practitioners.

<sup>8</sup> Diagnoses and remedies are numerous. This paper offers one simplification as a (still partial) remedy. Kim Warren’s work (Warren 2008) offers simplification in an entirely different dimension, by beginning from a general theory of corporate strategy and performance. Along the way, he has discovered several of system dynamic’s “folk sayings” get in the way of effective modeling and performance improvement (Warren 2011), including “model the problem not the system”, “do not make forecasts”, “close the loop” and “the mental database is the most important source of information”.

<sup>9</sup> Notably tests of the “model of the model”—a simple diagrammatic theory of why a problem arises and why a given set of actions can improve performance. This is very useful in communicating results. In addition, by virtue of being a compact (if diagrammatic) hypothesis of both why a problem arises and what actions are effective and ineffective in improving performance, the model of the model immediately suggests tests of the hypothesis that are considerably more focused than general system testing and policy testing.

<sup>10</sup> Lyneis also recommends a fourth stage, which is ongoing strategic use of the simulator and its descendants. There are a handful of examples known to the author, but overall, it is very rare to reach and stay at this stage. The business model of consultancy encourages client organizations to stop after the first Big Questions (often a crisis) have been addressed.

<sup>11</sup> A method that predates causal diagramming, and which is often espoused by Jay Forrester, is to state the problem and the dynamic hypothesis, and then (primarily from personal knowledge about the system) write down all the levels that comprise the “memory” of the system state. Connecting the levels then gives a stock and flow diagram, then a simulation model.

<sup>12</sup> There is a common admonition among system dynamicists to “model the problem, not the system” See (Warren 2011) for a discussion of why taking this at face value can be damaging, if for example, the modeler draws boundaries inappropriately narrowly. I sometimes jokingly say “every model represents the entire universe. It’s just that we choose *how* to represent it based on the problem at hand. We hardly ever have to model the availability of oxygen on our planet, for example. For most purposes, we can assume that oxygen is implicitly represented in our formulations, without calling it out. Similarly, the world economy will often be an exogenous input for scenario testing. And so on.”

---

<sup>13</sup> Not to be confused with Reality Check®, a functionality of Vensim that automatically checks whether pieces of a model behave within specified conditions.

<sup>14</sup> By no means is usefulness sharply restricted to what system dynamicists would call problems embedded in complex dynamic feedback systems. Some engagements end up with results that, in retrospect, could have been obtained with econometric modeling or database manipulations. But such a conclusions are after the fact, and weren't knowable at the beginning. In a field where most problems can't be addressed solely by more standard methodologies, choosing in advance to use only econometrics or some other non-SD method would be a very poor modeling strategy.

<sup>15</sup> (Richardson 2011) is the inspiration for this bookend: “the endogenous point of view is the *sine qua non* of systems approaches. What expert systems teachers and practitioners have to offer their students and the world is a set of tools, habits of thought, and skills enabling the discovery and understanding of endogenous sources of complex system behavior. Of course, in a definition aimed at the man in the street (with responsibilities or other stake in complex problems), one doesn't use phrases like “endogenous point of view”.

<sup>16</sup> There are a couple issues lying underneath the phrase “how people interact”. The first is that most system dynamics modelers would probably exclude purely engineering systems from our field; our expertise lies in large part in disentangling the complexities of organizations, markets, economies involving decision-making by humans. By this standard, the “avionics design” example would not be considered within the realm of system dynamics. The other issue is that system dynamics modelers have had reasonable success modeling interactions among natural resources, and plant and animal populations also. Ecology is an exception to the “people interacting” clause, but one which is obviously workable within the practices of system dynamics.

<sup>17</sup> The traditional approach has been rather Procrustean. More than once, when discussing the role of time series data, an academic has asserted to me that I should be educating the clients that time series data are not necessary. Even if I believed that to be true (which in general I do not, and which is in any event a proposition to be tested, not a received doctrine) the notion of educating clients just to avoid this bit of work is in most situations ludicrous. The only thing most clients want is to be educated about the nature of their situation and how to perform well in it. Time and resources are too short to add anything at all to that goal.

<sup>18</sup> Use of simulators for planning would seem to have purpose, model structure, and recommendations that lie in some intermediate range between pure forecasting and pure (feedback) policy design. There is a need for outputs like budgets, spending and performance that are likely to be achieved, at least in the short and medium term. Yet the most robust process for generating consistent, achievable plans would be those generated in the simulator by policies that respond to the state of the system. Non-modelers, too, are aware of the pitfalls inherent in what system dynamicists would call open-loop, short-term optimization—robustness and adaptation are desirable qualities in a planning process.

<sup>19</sup> Orally, Jay Forrester often summarizes the problem as “the time horizon over which you can forecast well is smaller than the time horizon of being able to do something about the system behavior”. Moreover, he shows that attempting prediction as the basis for countercyclical policies just shifts the cyclicity to a different frequency. More recently, Nathan Forrester's (1982) work shows the same phenomenon for countercyclical economic policy—stabilizing at one periodicity destabilizes another periodicity. That being said, not many real-world applications, at least outside of the economic realm, seem to revolve around designing countercyclical policies for endogenously-created oscillations. Even that minority of situations that do involve oscillation more typically revolve around predicting market cyclicity, and designing policies for individual organizations to predict or cope with that externally-imposed instability.

<sup>20</sup> By contrast, implementation is taught in internal courses of large consulting companies, and in process management (Shiba, Graham and Walden 1993), and is becoming part of the corpus of systems thinking (e.g. Senge *et al.* 1994) even though it was largely absent from the “founding document” (Senge 1990). So, more specifically, teaching hypothesis testing of recommendations against expert knowledge is an entrée into these bodies of practice.

---

<sup>21</sup> Thanks to George Richardson for inventing this sorely-needed acronym. And, as with many acronyms, there is some distortion—“Guys” does not imply that females are excluded from this activity.)

<sup>22</sup> In rereading *Industrial Dynamics*, it was surprising how ubiquitous was the assumption that the modeler would be the one to know about the real system and how it works. For example, page 118: “For now, it is best to acknowledge that ***the most useful models will be constructed by those who know the actual system*** [emphasis added] and who at the same time have a background in dynamic system analysis.”

<sup>23</sup> “Groupness” can vary by modeling stage. In the System Dynamics Review special issue on group modeling (Vennix, *et al.* 1997), (Andersen and Richardson 1997) describe what might be characterized as an industrial-strength front end (model purpose and conceptualization) with a classical back end (model formulation and policy testing). Inversely, (Richmond 1997) describes a classical one-modeler front end (model purpose, conceptualization, formulation and testing) with an industrial strength back end (policy evaluation and expert review). So to my mind, “groupness” is a subvariation, and to characterize a modeling effort, one needs to ask how and where the group is being used, and what else is going on in the modeling process.

<sup>24</sup> The recipients of the System Dynamics Society’s Applications Award span a spectrum from fairly “classical” approaches (Mark Paich and the OnStar case), to an industrial-strength approach (Homer, Hirsch, Milstein, *et al.* in the CDC public health strategy modeling) and in between, an adaptation of industrial-strength modeling to a lighter-weight use (Ken Cooper’s work with Fluor-Daniel’s engineering project change management).

<sup>25</sup> Group model building (Vennix *et al.* 1997), another recent addition to “classic” system dynamics, goes a considerable distance in obtaining expert review. Industrial-strength modeling structures and narrows the focus of meetings with experts to be explicitly reviewing a particular modeling deliverable (model purpose diagram, causal diagram, calibration, analytical results), excluding almost all discussion about how a given phenomenon is to be represented (which is much more a discussion topic for modelers only).

<sup>26</sup> The process of creating model structure, even when reusing large portions of a simulator, is highly iterative, with early iterations marked by, in Forrester’s words (2007, pg. 360), “obviously implausible behavior”. One silver lining in these clouds is that by the time a model reaches the point of fine calibration, it has already been subjected to numerous extreme conditions and parameter sensitivity tests. This is one reason that formal sensitivity testing on a “completed” model and its analytical results is usually focused rather than comprehensive in industrial strength system dynamics.

<sup>27</sup> (Saeed and Pavlov 2008) identify a state-space validation criterion, coordinated movement through a state space, which removes speed of movement from consideration. They apply develop this in a generic structure describing dynastic change in ancient China, with applications to political change in developing economies, and the emergence of the Mafia in 19<sup>th</sup> century Sicily—situations often not easily amenable to more traditional forms of time-series validation.

<sup>28</sup> Peterson points out that optimal filtering is only part of the mathematics needed to, e.g. estimate parameters. He has called the method Full-Information Maximum Likelihood via Optimal Filtering, or FIMLOF.

<sup>29</sup> In addition, Forrester presents validation of behavior as going through several stages, from the first stage of “elimination of ‘obvious’ implausibilities” to a second stage of “all of the model behavior that can be compared with the real system” to a third stage of “an objective set of quantitative criteria... [after deciding] what significance to attach to differences in the results of applying the criteria” (pp. 120-121). Arguably, the experience in Industrial-strength modeling (garnered over the five decades following the publication of *Industrial Dynamics*, and using far more abundant time-series data and far easier model experimentation) suggests that usually, going to the third stage of fitting models of important issues to time series data is leads to better models and still more trustworthy, useful results.