Table of Contents

Go Back

On Positioning System Dynamics as an Applied Science of Strategy

Or: SD is Scientific. We Haven't Said So Explicitly, and We Should. *

Alan K. Graham
PA Consulting Group, Inc.

One Memorial Drive, Cambridge, Massachusetts 02172 USA
Voice: 617-864-8880, Fax: 617-864-8884
Alan.Graham@PAConsulting.com

Abstract

The field of System Dynamics (SD) and strategy practitioners share a common aspiration: to understand and control the success and failure of companies through appropriately rigorous methods. SD practitioners know that it has all the elements of proper scientific method, but generally do not teach or explain the method with that framework and vocabulary. Outside the field, SD is generally perceived as lacking rigor. This state of affairs has come to be through a kind of "Newton's law", where SD started off distant from both academic and corporate strategy, and no outside forces have motivated a rapprochement. The founding of SD in fact predates both the use of the strategy concept in business, and the knowledge of how to do practical quantitative hypothesis testing for dynamic models with messy data. In the four decades since the founding of SD, its practitioners have been confronted with a "relevance gap" between themselves and academic strategy approaches, and a "rigor gap" between themselves and corporate practice of strategy "analysis". These gaps represent lost opportunity for all involved.

Positioning System Dynamics as an applied science of strategy would mitigate several of the difficulties that SD practitioners face, and create benefits both personal and societal. This repositioning could be accomplished through 1) teaching and publishing in the explicit language of iterative hypothesis testing and explicit choice of testing criteria, 2) teaching iteration and testing as the fundamental method, with knowledge of formulation and dynamics as refinements, rather than vice-versa, and 3) explicitly claiming a scientific approach to strategy as an area of personal or organizational competence.

Keywords: Scientific method, applied science, strategy, hypothesis, testing, validation, systems thinking, phased modeling.

---

* The author acknowledges PA Consulting Group and Ken Cooper in particular for the time and resources for this research. The corporations with whom the author has worked on strategic issues have also contributed to this work with their blood, sweat and tears. Finally, the colleagues who have done so much to push forward the effective use of System Dynamics as a strategy tool, including Carlos Ariza, Adolfo Canovi, Bill Dalton, Sharon Els, Peter Genta, Tamara Kett, Donna Mayo, Evonne Meranus, Kim Reichelt, Craig Stephens and Tameez Sunderji.

1. Introduction: A Good Field without Good Public Relations

As a discipline, System Dynamics is healthy and growing. Starting from one MIT professor and a handful of graduate students, it went public in 1961 with *Industrial Dynamics* (Forrester 1961). Today, full courses in System Dynamics (SD) are taught at over 80 universities around the world (Stephens, Lyneis and Graham 2002). There is a professional society with a refereed journal and multiple chapters and conferences. Practitioners and users applaud its rigorous yet effective approach to high-stakes strategic issues.

Inside the System Dynamics community, its practitioners, if asked, will freely characterize System Dynamics as following the scientific method, although in all likelihood few have formally examined the scientific method. The authors have usually found quick acceptance to this definition of the field, which, broadly interpreted encompasses not only corporate strategy but ecology and economics, large projects and public sector work:

> *System Dynamics uses the scientific method to construct simulation models that address dynamically complex, high-stakes strategic issues*.

(Oliva 2002), for example, extensively recasts system dynamics testing, specifically calibration, in terms of more widely-recognized processes of scientific hypothesis testing. Finally, a sister paper to this shows how System Dynamics meets a newly-articulated legal definition of "scientific method" (Stephens, Lyneis and Graham 2002).

System Dynamics falls squarely within the Karl Popper's (1990) refutationist school of scientific method of the 1930s. As Sterman (2000, Ch. 21) summarizes, "Popper famously argued that it is possible to show a theory to be false…. Once a theory is falsified empirically it has to be discarded and replaced by some new, more accurate theory. Only those theories that have not yet been refuted should be accepted and that acceptance is always conditional." Sterman also notes the added fillip that since all models are inaccurate in some respect or another, the choice of tests aimed at falsifying are appropriately chosen by the modeler to align with the purpose or use of the model (Forrester and Senge 1980). (We will return to this variability of testing methods near the end of this paper, in Section 4.3.)

But things are much less uniform outside the SD community. Being taught in 80 universities around the world means that the vast majority of universities are not teaching SD as a distinct course. By the author's informal sampling in both academia and the corporate world, the majority do not know of the existence of System Dynamics. Of those that do, it's an even split between "yes, we did a unit on that in school" and "wasn't that the '*Limits To Growth'* methodology that was discredited?" To quote Ali Mashayekhi (2001), current President of the System Dynamics Society, "we are far from the real potential of our field. Today not many institutions in the world teach system dynamics, not many decision makers are aware of system dynamics or adopt systems thinking manage their organizations." So for the overwhelming majority of potential teachers and users of SD, it has little legitimacy of any kind, let alone having its true nature as a scientific endeavor.

What causes this cognitive dissonance? It is simplistic to explain this situation of many decade's standing on two publications, *Urban Dynamics* and *Limits to Growth*. After all, many individuals over the years have produced dissertations and publications in highly respected academic journals outside the field. Dozen if not hundreds of person-years have been spent doing rigorous, peer-reviewed scientific work in the field. To understand why this body of work hasn't created the public image its practitioners believe it should have, we must look more carefully at the field, how it characterizes itself, and how it interacts with other constituencies. What will be proposed here is a kind of "Newton's law" for disciplines: Disciplines in a particular positioning will stay in that positioning until acted on by an outside force (which has thus far been mostly absent). "Positioning" here is used very much in the marketing sense of defining the messages we send within the profession and outside the profession about who we are, what we do, and what our relationship is to other groups: Is SD scientific, or something else? Is SD related to some other fields and activities, or irreconcilably different?

Section 2 begins by describing the initial "rest" position at the creation of the discipline. Section 3 then examines the (absence of) forces that might have changed the positioning of the field since that time. Section 4 enumerates several benefits that would accrue with a change in positioning of the field, and Section 5 outlines the steps needed to realize such a repositioning. Section 6 concludes with a call to action.

2. SD originated with very little impetus to align with "science" or "strategy"

"Historical reasons" play a surprisingly large role in many human endeavors. Something is a certain way because it's always been that way. Professional disciplines collectively hold some set of beliefs, purposes, questions, and approaches important. (That may be a definition of what a discipline is, along with a system for propagating those beliefs.) And what gets propagated is in large measure what was propagated earlier. This is what makes scientific paradigms persist, as (Kuhn 1996) discusses, and (Sterman 1985) explicitly models.

So in thinking about why SD has not laid claim to its scientific nature, one place to start is origins. This paper suggests that there are characteristics of the origins of SD that have de-emphasized the importance or usefulness of cloaking itself in the mantle of either "science" or "strategy".

2.1. SD originated as a very applied science

There is a continuum in the practice of science, from the "pure" to the "applied". "Pure" sciences have the understanding that the science is being performed to obtain knowledge for its own sake, or at least for the purpose of a very general good. Cell biology, physics, physical chemistry, linguistics, and marine biology can fall mostly in this category. By contrast, "applied" sciences employ the scientific method, along with the considerable body of scientifically-obtained knowledge, to accomplish some specific purpose. And specifically, applied sciences often draw on the body of pure science results.

For example, medicine has the purpose of postponing death and improving well-being in humans, and it draws on pure sciences of cell biology, organic and inorganic chemistry, physics and other fields for useful scientific results. Indeed, the transformation of medicine into a highly professional discipline in the late nineteenth and early twentieth century quite deliberately resulted from the addition of academic scientific training to the medical education (Ludmerer 1985, Ch. 11).

Engineering is an applied field has the purpose of creating useful devices and artifacts, and it draws on the pure sciences of physics, chemistry, materials science, and sometimes biology for useful scientific results. It also uses mathematics which, although sometimes called "the queen of sciences" is not a science. (Mathematics conducts no investigations, nor makes any claims about, what is true about the real world.) Even engineering itself has a continuum. In US technical education, four-year "Engineering" degrees have clear scientific, theoretical, and methodological content. Thesis research in engineering is expected to expand the body of knowledge about engineering through rigorous scientific method. By contrast, four-year "Technology" degrees focus on how to do things, without the scientific, theoretical, and methodological content. There are no thesis requirements as such, and no PhD programs. In such established "applied science" fields as medicine and engineering, then, there is a mixture of some scientific method to discover new knowledge, and considerable amounts of previously-created knowledge. While academic physicians and engineers do scientific research, the focus of these applied sciences in teaching and in use is dominated by use of existing knowledge for practical application.

In parallel fashion, management has the purpose of effectively directing activities of groups of people, and it draws on the pure (but sometimes dismal) sciences of microeconomics, macroeconomics, marketing science, psychology, and so on. It can be argued that management in general is a field evolving into an applied science. Indeed Forrester (1961, Introduction) makes this argument, positioning System Dynamics (at the time, Industrial Dynamics), as part of the evolution from "art" to "science".

But we must note that System Dynamics is an applied science, which has the purpose of effectively directing action, typically in a corporate setting, but elsewhere as well. Forrester (1961, pg. 14) describes SD as drawing on four bodies of prior work: 1. Feedback control theory, 2. Studies of management decision-making as exemplified by the "bounded rationality" school of Herb Simon (Cyert and March 1963, Simon 1979, Morecroft 1985), 3. Experimental approach to policy design and 4. Computer simulation. More detailed analysis in (Richardson 1991) reaches much the same conclusions.

The bodies of knowledge from which SD evolved, then, consist of one very applied science (management decision-making), and three areas of applied mathematics. This is not to say SD modelers have excluded other sources for prior theory to work only with primary research on the corporation at hand. Indeed, there is extensive but usually informal borrowing from microeconomics, and many, more specialized fields, e.g. (Lane 2002a-b). Scholarly borrowing, and even formal databases of previous findings (Graham 2000) occur within the SD field. Nonetheless, even the manner of borrowing again underscores the applied character of the SD field.

So there is a seemingly paradoxical cultural tension built in to the System Dynamics tradition from its beginnings. On one hand, to build a model is to build a theory of an individual firm or market with the scientific method. On the other hand, there is a traditional of ruthless rejection of many of the trappings of scientific method if there is danger of conflict with practical utility.

"Academic" in management and even engineering circles is pejorative – it connotes excessive concern over rigor and precision, which gets in the way of making things better. This perhaps explains the character of a method that arguably falls completely within the scientific framework, and yet whose seminal work does not discuss the scientific method as such: Neither "scientific" nor "scientific method" appear in the index of (Forrester 1961), and "hypothesis" is discussed only in one appendix (O) of many appendices as a starting point, not as something to be refined and ultimately used to guide action.

But the applied nature of System Dynamics is far from the only reason that it started off describing itself so differently from being a "science of strategy", for in fact SD predates important elements of both "science" and "strategy".

<u>2.2 SD originated before the knowledge of how to test hypotheses in complex dynamic systems</u>

The theoretical framework for the rigorous hypothesis-testing of models that was being practiced in the 1950s and 1960s, econometric modeling, was impractical and flawed for dynamic models. In an era of computerized payrolls and accounting, and manual recording for everything else, using the standard econometric approach wasn't going to happen. Most key variables were not routinely measured. Even today, many variables key to the dynamics still aren't measured. And all data is flawed, and standard econometric model-testing is not robust with respect to such flaws (Senge 1974a-b). So the purely statistical hypothesis testing, using high-quality and complete data, nearly mandatory in many branches of science, wasn't going to happen.

Moreover, Forrester (1961, Chapter 13 and Appendix K) noted that simple reproduction of history via simulation wasn't a suitable test either, if the real system and the model were driven by different unknown random noise sequences.

Only in the late 1960s did Astrom and colleagues in the control theory community develop a feasible framework for hypothesis-testing and parameter identification in dynamic systems where data were missing and corrupted, and where the system could be perturbed by random noise (Astrom 1970). This framework is Full-Information Maximum Likelihood through Optimal Filtering (FIMLOF), outlined in the Appendix. Indeed, the mathematical prerequisite of optimal filtering, the now well-known Kalman-Bucy filtering, wasn't published until 1960 (Kalman 1960), which was then linked to the likelihood function by MIT's Fred Schweppe (1965, 1973) and Peterson and Schweppe (1974).

So when SD was first being created and taught in the early 1960s, quantitative hypothesis testing, and the whole aura of academic rigor that came with it, were correctly discouraged in favor of tests less quantitative, less reproducible, but drawing on far more extensive information in a far more robust manner (Forrester 1961, Ch. 13, Forrester and Senge 1980).

Perhaps as a small example of inertia of a discipline once it has gone down a certain path, Peterson (1975, 1980) explicitly described to the System Dynamics community the technique of parameter identification through FIMLOF, with full knowledge of its much greater suitability for SD than standard econometric techniques. But apart from being a seldom-used feature of the Vensim simulation package, the field has by and large ignored it. Even in a prominent discussion of statistical validation, it appears in one footnote (Sterman 1984).

Now it is true that use of time series data has gradually increased over the years, lead in corporate practice by the author's company (the former Pugh-Roberts Associates), and academically by Sterman and several others (e.g., Repenning and Kofman 1997). And it can be argued that the iterative refinement of data, model structure, and parameters using both qualitative tests and a quantitative fit criterion in the end approximation the logic of the rigorous framework, FIMLOF, even for many oscillatory systems (see the Appendix). But by and large, such testing is still not described in terms of the scientific method or hypothesis testing and refinement, and the exceptions notwithstanding, the distrust and neglect of significant use of quantitative data has persisted in the SD to this day.

In terms of the "Newton's law" for disciplines, a practical orientation that predated appropriate tools for hypothesis testing quite rightly lead to rejecting use of inappropriate tools, and positioning at the beginning of the field with less than full embrace of the framework and vocabulary of the scientific method. But not only did SD predate appropriate tools for hypothesis testing, but it also predates the concept of corporate strategy itself.

2.3. SD originated before the concept of "corporate strategy"

System Dynamics modeling was being created in the late 1950s (Forrester 1961, Preface). It may surprise contemporary management scholars that "strategy" seems to have emerged as a distinct business issue and function later than that, specifically in the mid-1960s. Prior to that period, according to (Andrews 1971), the integration of various bits of management and business knowledge went under the heading of "business policy", as in "foreign policy" or "public policy", which didn't carry the connotation of giving a top-down and structured direction to a company, as the concept of "strategy" does. Dating the rise of the concept of strategy during the 1960s is triangulated from four sources:

1. A survey of business research (Sibbet 1977) dates the introduction with Chandler's (1962) *Strategy and Structure* at MIT.

2. In the most-cited strategy journal, the *Strategy Management Journal* (Tahai and Meyer 1999), in the entire 1980 volume, its first year, no article refers to any publication prior to 1962 with "strategy" in the title. Most strategy references were to works from the 1970s. References prior to 1962 were to microeconomics and finance journals.

3. Andrews (1971) dates the beginning of the teaching of strategy as such in the mid-1960s with (Andrews et al. 1965). He introduced the concept into Harvard's MBA curriculum.

4. The first firm to prominently position itself as a "strategy boutique", Boston Consulting Group, apparently practiced its learning-curve and growth-share matrix methodologies as proprietary methodologies during the 1970s and only began to publish widely during toward the end of that decade, e.g. (Henderson 1979). See also (Seeger 1984).

So the seminal system dynamics modeling in the 1960s and 1970s stressed the high-level operational policies (Forrester 1961, Part III) that would be more appropriate to business units than to a modern, multi-divisional corporation. But paradoxically, Forrester's philosophy stressed an approach of organizational design (Forrester 1975), an attitude and viewpoint applicable at both the corporate and business unit levels, squarely overlapping the domain of what is now called "strategy". Regardless, even though Forrester consistently urges examination of the issues that determine corporate success or failure (1961, Introduction, Section 5), the word "strategy" doesn't even appear in the index. The field by and large has not attached primary significance to the concept since then.

In fact, the phrase "picking a strategy" will often cause System Dynamics modelers to deliver lectures on the importance of designing rules for feedback response—a policy—rather than making decisions. While "picking a strategy" is in a sense a decision, designing and implementing a strategy means re-doing goals and responses to deviations from them (for example via Balanced Scorecard as in (Kaplan and Norton 1996, 2000)), which fits the SD definition of policy.

System Dynamics, then, started out close in principle but distant in rhetoric from both focus on strategy as a field and explicitly espousing the scientific method. The next question to explore is why this state of affairs has persisted.

3. Characteristics of real-world strategy practice and academic scientific analysis of strategy do not motivate embrace by System Dynamicists

The second part of the argument is that there are causes – perceptions and motivations for personal choices of SD practitioners – that have maintained that distance in the decades following the founding of SD. This paper will suggest below that there are three main approaches to strategic analysis: academic cross-sectional analysis, corporate statistical studies, and strategic frameworks. Each of these, it will be suggested, has prominent differences and flaws in relation to System Dynamics which have made them unattractive for System Dynamics research and practice.

3.1. Academic strategy research appears mostly impractical and simplistic

The vast majority of academic research on strategy is cross-sectional analysis of many companies, done from mostly publicly-available information. Extensively time-based or internally oriented research is rare. For example, in the twelve 1999 issues of the *Strategy Management Journal,* only five or six articles qualified, depending on precise definitions, or about one every other issue. The rest were cross-sectional studies. As Michael Porter (1991) observes in his "Towards a Dynamic Theory of Strategy",

> …one might approach the task of developing a theory of strategy by creating a wide range of situation-specific but rigorous (read mathematical) models of limited complexity. … This approach to theory building has been characteristic of economics in the last few decades. … These models provide clear conclusions, but it is well-known that they are highly sensitive to the assumptions underlying them… it is hard to integrate the many models into a general framework for approaching any situation, or even to make the findings of the various models consistent. …
>
> … In my own research, I pursued cross-sectional econometric studies in the 1970s but ultimately gave up as the complexity of the frameworks I was developing ran ahead of the available cross-sectional data. I was forced to turn to large numbers of in-depth case studies to identify signifi-

> cant variables, explore the relationships among them, and cope with industry and firm specificity in strategy choices.
>
> The need for more and better empirical testing will be a chronic issue in dealing with this subject. *Academic journals have traditionally not accepted or encouraged the deep examination of case studies, but the nature of strategy requires it.* [emphasis AKG]

To cite a relatively visible example of the difficulties of cross-sectional studies, consider four well-known and reasonably thorough empirical examinations of outstanding companies. They come up with four very different conclusions, depending on the precise definition of "outstanding", whether long-lasting and "visionary" (Collins and Porras 1994), simply long-lasting (Foster and Kaplan, Ch. 1), harnessing Schumpeterian "creative destruction" (Foster and Kaplan, Chs. 2-12), or having gone from "good to great" (Collins, 2001).

To characterize the situation somewhat less charitably than Porter himself might, he found rigorous academic cross-sectional studies from available numerical data less than helpful in creating managerially useful insights and conclusions. Porter found that even merely conceptual case frameworks drawn from extensive case studies with mixed quantitative and qualitative information to be more effective in practice. System Dynamics as a field often finds itself much in the same position of examining cases in depth one at a time. Unlike typical case studies of course, SD modelers are able to quantify the framework, albeit for just one company and a handful of specific issues. (Graham and Ariza 2002 illustrate how taking such a sharp focus dramatically reduces time and effort needed to do the hypothesis-testing needed for a suitably well-validated model.)

In addition to the (cross-sectional) hypotheses being tested, the character of hypothesis-testing as customarily taught is palpably irrelevant to the normal System Dynamics modeling of corporate strategy. The idea that one can test hypotheses (including model structures) *qualitatively*—against anecdotal evidence—is very rarely taught at all, and *quantitative* hypothesis testing is usually taught in a very simplified context (single equation, perfect information). Material beyond that is usually taught only to PhDs. The problem isn't that the mathematics of ordinary least squares is in any way less true than those of FIMLOF. The problem is that an applied science like SD has to live with the characteristics of its universe of discourse (corporate strategy) and select methods consonant with those characteristics. FIMLOF and the heuristic approximations to it (see the Appendix) that form the core of the system dynamics methods to validate behavior are useful in the corporate strategy setting. So the explicit role of hypothesis testing in SD is often limited to Forrester's original concept of the initial dynamic hypothesis (1961, Appendix O). The model is hardly ever explicitly characterized as the hypothesis for the next iteration of modeling.

So between standard academic strategy practice and system dynamics there is a "relevance gap"—by SD standards, the practice of strategy analysis in academic fashion is far removed from the concerns of an applied science, and therefore not particularly useful, appealing or practical as a focus of attention or source of knowledge. Doing System Dynamics modeling is largely inconsistent with conducting what by SD standards are superficial surveys of large numbers of companies. (One exception is benchmarking to determine the maximum or minimum realistic variation in performance parameters like quality, productivity, or costs of specific activities.) An analogous gap exists between SD and typical corporate strategy analysis.

3.2. Strategy in corporate practice appears to be more adversarial discussion than scientific process

Before further characterizing the methods of strategy in practice, it behooves the author writing for a System Dynamics audience to identify two noteworthy exceptions to the usual complete disconnects between academic strategy research, popularized strategy frameworks, and System Dynamics practice. The exceptions go by different names within the different communities, but they are nonetheless at least loosely aligned and mutually supporting. There is an academic school of microeconomic / strategy analysis, the "resource-based view of strategy" (Wernefelt 1984, Peteraf 1993), which emphasizes looking at the underlying natural advantages (human, technological, geographical, reputational, etc.) possessed by corporations rather than more superficial manifestations like price or market share. This analytical framework seems to have inspired the popular "core competency" framework of strategy analysis (Prahalad and Hamel 1990, Hamel and Prahalad 1994). Morecroft and Warren have also used it as the basis for a modestly systematic approach to creating simple dynamic models for strategy analysis (Warren 1999a-b, 2000).

The other exception is the typical alignment between Porter's (1980) five fields framework on the one hand, and on the other, the architecture of major strategy models by the former Pugh-Roberts Associates, now part of PA Consulting (Graham and Ariza 2001, Lyneis 2000).

Those two modest exceptions aside, strategy in practice is dominated by the methods and concepts such as those shown in Figure 2. The Figure plots a variety of methods (which are described and referenced individually below in the text) according to two characteristics. First, the vertical axis characterizes the rigor of use of facts and the extent of tests for hypotheses, ranging from no quantification, through fact-based (but not scientific) to the scientific method (hypothesize, predict quantitatively, test, and refine). Second, the horizontal axis characterizes the scope of each method: roughly how many variables or issues it incorporates.

The upper left corner of Figure 1 contains academic strategy studies, and their commercial cousins such as price/demand studies, conjoint analysis of customer preferences, or segmentation studies. The work can be rigorous, but very narrow in scope, as discussed previously. The Figure shows the "relevance gap" as the significant horizontal distance between academic studies and System Dynamics modeling.

The "regression line" on Figure 1 labeled "strategic frameworks" surrounds the methods such as value maps (or value chains, c.f. Porter 1985) for two reasons. a) These are where the most frequently used and discussed frameworks lie, and b) they symptomatize the tradeoff to which Porter referred in the previous quotation. If you start by characterizing many companies, then the more issues you consider, the less rigorously you can say things about them. The "regression line" is downward-sloping, such that the most comprehensive frameworks are the least rigorous. In fact they are therefore the least predictive or prescriptive. Growth-share matrices (Henderson 1979) and value maps are potentially predictive, but generally used as basis for discussion and action directly, not for prediction, iteration, and extension or refinement. Learning curves are theories that in fact explain knowledge about the past, although again they are used as a tool for discussion and action (Henderson 1979). The limited scope of issues being characterized relegates it to that use. In effect, the "regression line" represents a kind of "production-possibility frontier" (or efficient frontier from portfolio analysis), where for general theories, the broader their scope, the less rigorous they can be.

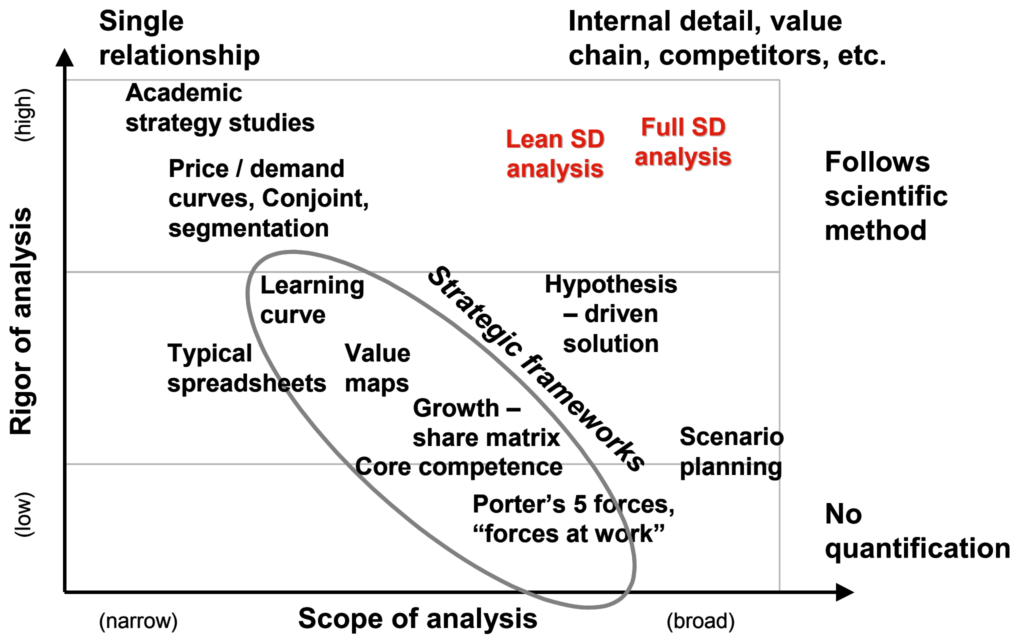

Figure 1. Various strategy tools, methods, and frameworks: How do they incorporate facts well (rigor) and in adequate number (scope)? There is a horizontal "gap of relevance" (scope) between SD and academic and corporate statistical analyses. There is a vertical "gap of rigor" between SD and corporate strategy frameworks and analysis processes.

There are three approaches that focus on specific companies and issues for specific purposes, which thereby can rise above the trend line. The lowest and broadest-scoped is scenario planning, as in (Schwarz 1991). The next method up is hypothesis-driven solution, made prominent by McKinsey and Co. (Rasiel 1999, Ch. 1). It uses a tree structure of hypothetical strategic actions ("if we do this, that will happen"), and reality-checks the feasibility and impact of each action through fact-gathering. This approach is placed on the Figure as being somewhat scientific, as there is some checking of facts against the hypothesis, but there is generally no specific prediction, and if there is checking of the entirety of the hypothesis tree against facts, it is intuitive only; the conceptual form of the hypothesis tree does not allow any quantitative test of whether it has captured the essential drivers of behavior. And the tree structure creates difficulties in considering, e.g. price changes which change market penetration, share, and margins simultaneously. All the same, the hypothesis-driven solution approach is the high end of scientific rigor within mainstream strategy practice, and even so is not terribly prominent.

Of course the limitations of the strategic frameworks are well-known at least in theory, so their use in practice is often buried in ad-hoc meetings, and thoroughly intermixed with an unstructured bottom-up approach (Mintzberg and Waters 1985, Mintzberg 1994). All of this is to underscore that, for a body of practitioners accustomed and indeed drawn to the peculiar rigor and breadth of SD modeling, there is a "rigor gap" between SD and the customary process of strategy-making, which consequently has little visible appeal. (or even sufficient tangibility to become involved with as such.)

And of course in the upper right is the (potentially) broadest-scoped and most rigorous method, System Dynamics. (It is worth noting that recent research (Graham, Choi and Mullen 2002, Graham, Moore and Choi 2002) has defined a reasonably rigorous and practical approach to confidence-bounding and hypothesis testing dynamic simulation models tuned to data, such that a scientific process can terminate with a quantification of confidence in model results—like econometrics and all good engineering science.

3.3. Corporate statistical analyses create obvious dynamic shortcomings

Corporations do conduct fact-based investigations, performing primary research with the intent of guiding strategic decision-making. Market segmentation studies, price-demand regressions and conjoint analysis (quantifying the tradeoffs consumers choose) quantify some aspects of customer behavior. Product-by-product (or business-by-business) analysis of risk-adjusted return on capital can say a lot about which businesses may be appropriate targets for divestiture. Forecasts give some clues about future events.

But for dynamic modelers, these are all partial approaches, with obvious flaws. The strategic implications of customer behavior aren't clear until the analysis considers the interaction with the company's internal capabilities and the responses of its competitors. Business-by-business analysis neglects synergies and economies of scale created by interactions among the business units (Graham and Ariza 2001). Forecasts, when acted upon, alter the basis of the forecast. By and large, dynamic modelers avoid methods based on such systematic data gathering and analysis, preferring to articulate (model-based) theory first, and determine whether such empirical research is needed or practical. In addition to theoretical objections, there are practical reasons that few dynamic modelers engage in primary research. Few modelers have the personal skills, client buy-in, demonstrated need for such methods, and schedule slack needed to do primary research. It perforce ends up being declared impractical.

There are occasional exceptions. Modelers can and do use statistical studies when already available, such as conjoint analysis (Graham and Walker 1998). Modelers do sometimes conduct their own primary research, as in segmentation (Graham and Walker 1998) or conjoint analysis (Schmidt and Gary 2002). And dynamic models, when suitably validated against quantitative data, can provide superior forecasts, especially in extreme conditions (Lyneis 2000). But by and large, dynamic modelers continue to calibrate to minimal data, often qualitative behavior modes, and rely on modest model purpose, sensitivity analysis, or robust policy design to obviate the need for more aggressive numerical inquiry.

This completes the discussion of a "Newton's law" for the discipline of System Dynamics. System Dynamics started off with an explicit orientation away from both "scientific" hypothesis testing and strategy. And since that time, just as with academic cross-sectional analysis and corporate strategy frameworks, there is little about corporate statistical investigations that complements the skill set, philosophy, and purposes of a dynamic modeler. The "relevance gap" and "rigor gap" have remained in force from the founding of the field to this day, and would seem to have deterred most practitioners from positioning SD as an applied science of strategy.

The next questions that arise are whether it is desirable that this state of affairs persist, and if not, what can be done to alter it?

4. Practitioners and students would benefit if SD were positioned as an applied science of strategy

Specifically, the next discussion examines three current difficulties with System Dynamic (in teaching, understanding modeling choices, and relating to other strategy activities), and suggests that positioning SD as an applied science of strategy substantially lightens these difficulties.

4.1. SD is more clearly communicated and taught as a scientific method.

There seems little doubt that System Dynamics would be in greater demand, done more effectively, and taught more efficiently if the theory and practice of the discipline were clearer and more easily communicable. Positioning SD as an applied science of strategy seems to be a key to clarifying a great deal about the field.

Fundamentally, no amount of description of system dynamics at the *operational level* will ever convincingly communicate to a normal person what SD is and what differentiates it from conventional strategy approaches. The more one describes multiple validation tests, iteration, inappropriateness of point predictions, feedback loops, etc. etc., the less comprehensible the explanation becomes, regardless of whether the listener is a full professor (outside SD) or a middle manager. The mass of SD techniques at the operational level is simply too large to comprehend in a realistic amount of time and effort. By contrast, the author is finding that explaining the *overall process* of SD modeling as the scientific method is easily and quickly understood.

But the problems with the “not a science” posture aren’t just in communication or conceptual understanding, they are in the practice of SD also. SD modeling is customarily taught in bottom-up fashion, starting from levels and rates and feedback loops, going through basic formulation, and building up to the construction of entire models. Iteration and refinement are taught almost as an afterthought. As a result, SD beginners almost always spend most of the available time formulating a model, and attempt to fix flaws at the last minute. The usefulness of beginning modeler’s work is thereby almost always uncertain. Most SD instruction (to which the author has been exposed) has this imbalance, and experience with spreadsheet models teaches people the same lesson: iteration—“fixing it”—is an afterthought, something to do at the last minute. As a result, even moderately-experienced dynamic modelers still don’t allow enough time for iterative refinement of their hypothesis (model). And more perniciously, corporate customers of modelers are often suspicious of modelers that come back several times asking for more information, and become especially suspicious if apparent conclusions change between an interim model and a well-validated model.

Such misperceptions would be substantially reduced if all concerned understand from the beginning that iterative hypothesis-testing is a legitimate, critical and substantial facet of model development through the scientific method.

But in the practice of SD, there are not only misperceptions about the process of the scientific method, but also midperceptions that arise from confusion about the appropriateness of specific techniques of hypothesis testing, as will be seen next.

4.2. Several confusing issues in SD become clearer in the framework of scientific method.

There are several issues which seem to become clearer in the framework of scientific method, especially given that within the SD method, validation tests and extend of iteration and refinement are chosen to fit the model's purpose, which can vary widely from one model to another (Sterman 2000, Ch. 21).

As one example, Systems Thinking (Senge, 1990) is often looked askance at by even the System Dynamics modelers who are accustomed to more rigorous model development. Within that framework, it lacks a certain legitimacy. But Systems Thinking can also be understood as scientific method whose hypothesis tests (comparison with anecdotal data only) are chosen to fit the purpose of being developable and useful to managers without extensive SD training. So understood within the broader framework of scientific method, Systems Thinking is theory-building that stops at the qualitative stage.

Lyneis' (1999) (and PA's) "phased approach" just becomes recognition that theories of socio-economic systems start small and qualitative, and iterate to a quantitative theory of a complexity depending on the problem and the stakes. The first three "phases" of Lyneis's phased projects are completions of the cycle of scientific method at three successive levels of testing criteria: first, versus anecdotal information alone (like Systems Thinking), second, that plus comparison to roughly quantitative behavior modes, and third, that plus clearly quantitative calibration testing. (See Graham and Ariza 2002 for further discussion.)

Another example: A discussion recently re-emerged on whether some qualitative models are best left unquantified (Coyle 2001, Homer and Oliva 2001). In the framework of iterative refinement of theories, the issues seem clear-cut. A qualitatively validated, qualitative model (causal diagram) has usefulness and known limitations. A quantitatively-validated quantitative model almost always has more usefulness and known limitations. But a quantitative model that will not be validated, for whatever reason, has little usefulness because its limitations are unknown. One reason a model may not be quantitatively validated, that discussed by Coyle, Homer, and Oliva, is that there may be too many soft variables for even SD techniques to deal with in a feasible amount of time and effort. But there are other possible reasons for not quantitatively validating as well. Most of the world limps along with mental and quantitative models only superficially validated, because no one with the right skills is available at the right place and the right time. And even corporate strategy modeling sometimes must yield to the "gut call" method if time is extremely short. So the "apostasy" that one may choose not to quantify, when considered in the right framework, becomes the modeler's equivalent of "don't drive into places you can't get out of": Don't start to quantify a theory that in practical terms you won't be able to validate adequately, because then you can't and you shouldn't trust the result.

A third example: exploratory models like *Urban Dynamics* (Forrester 1969) and *Limits to Growth* (Meadows, *et al.* 1974) have the purpose of raising issues and looking at possibilities in an unexplored area, at least by dynamic modeling. They should and must have very different validation criteria than well-studied fields. This is a trivial observation within the framework of scientific method, and there is a precise language for explaining it. But when this issue was described without using the common language of scientific method in the above-mentioned books, the point became extraordinarily difficult to communicate and was widely misunderstood. As a result, these works were widely pilloried within the larger academic community, and few would argue against the proposition that, notwithstanding the publicity gained, the field of SD suffered because of those misperceptions, and that the field would be better off had the nature of the validation criteria been recognized as appropriate.

But to focus too narrowly on avoiding misunderstanding may underestimate the true potential gains from positioning SD as an applied science of strategy.

4.3. SD has potential to create synergies in the currently fragmented field of strategy.

Figure 1 paints a pessimistic picture, with the "relevance gap" between SD and rigorous statistical work, and the "rigor gap" between SD and commonly-used corporate strategic frameworks. Figure 2 presents a related view, diagrammatically suggesting how at present, the worlds of academic strategy, corporate strategy, and System Dynamics hardly overlap, and the work within the overlaps is often distant from the mainstream. *The widely-separated mainstreams are foregoing enormous benefit by their distance from the other mainstreams.*

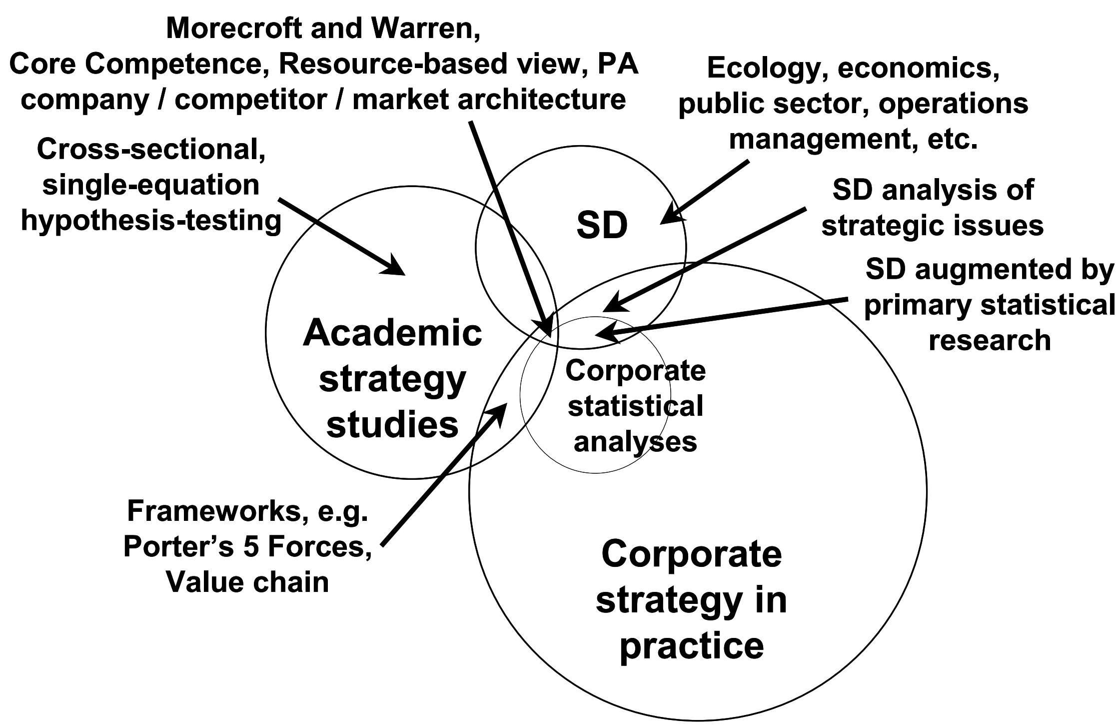

Figure 2. (Minimal) overlaps and synergies between academic strategy studies, corporate strategy in practice, and System Dynamics suggest major potential gains when fields take more advantage of one another.

(Graham 1982) and (Senge 1990) discuss the idea of an ensemble of technologies, in which there are several technologies, each of which produces a slight improvement in performance over the status quo, but which together can produce a quantum improvement. Both have argued that SD is likely one component in such an ensemble. To put a sharper point on the analogy, one common behavior pattern for technology ensembles is that there can be a "holding pattern" of many years duration, that lasts until all of the necessary technologies are used together. It is certainly arguable that there has been a holding pattern, with little fundamental innovation or change in performance over the past 15 years or more:

- in corporate strategy practice after (Porter 1985) (admittedly with considerable intensification of infrastructure and practice),
- in System Dynamics (in the post-*Limits to* Growth era),
- in academic strategy (since the founding of the *Journal of Strategic Management* in 1980), and
- in corporate statistical analysis (since conjoint analysis, Green and Wind 1975).

Certainly the consequences of lack of synergy are evident today. Corporations do far less statistical research than seems appropriate. All too often, corporations will consider and decide on strategies for which in fact the obvious data (for example, market segments or customer preferences) have not been collected. Of course, hesitation in spending money on primary research, or for that matter on consulting and strategic frameworks is understandable, for there is still an intuitive "leap of faith" from any such information to strategic conclusions. At the same time, System Dynamics studies (which *would* offer a rigorous "audit trail" from information to strategic conclusions) are often declined in part because potential corporate clients don't feel that there is enough information from which to draw reliable conclusions. (This is an incorrect perception (see Graham, Moore and Choi 2002) but it is a real and common perception.) There are also lingering doubts or uncertainties about the legitimacy of the process.

Although this author lacks direct data, it would be surprising indeed if the more rigorous academic strategy researchers did not experience more difficulty in working with corporate clients than practitioners of strategy frameworks because of a perceived "relevance gap" on the part of corporate clients. And of course, System Dynamics offers an escape from rigorous cross-sectional studies, into the kind of rigorous longitudinal studies that would be much more relevant and useful to corporate clients, as well as more directly useful in teaching strategy (as in Graham et al. 1996).

So there are clearly personal and societal benefits to be had if the gaps of relevance and rigor can be bridged. The next question is how this might be done.

5. The means to reposition SD as an applied science of strategy are straightforward

There are at least three types of action that seem necessary to reposition System Dynamics as an applied science of strategy.

First, consistently communicate in terms of the scientific method. Students need a clarifying framework upon which to hang new knowledge as it is acquired. Corporate clients need the "elevator story" that allows an abstract understanding (and obviates the compulsion to wallow in the details until either it's understood or they give up). Academics outside SD need a shared framework to understand how their work relates to SD. (Ironically, in almost all cases, other academics' work complements SD, but without communications in the common language of applied scientific method, this is extremely difficult for even highly educated people to discern.)

Use terms such as purpose, hypothesis, prediction, test, test criteria, and refinement, and iteration. Use these terms as precise technical terms when teaching, publishing, and even speaking with clients. Use near-synonyms only when necessary (For example, corporate clients will likely be uncomfortable if the word "theory" pops up too often, although many have become accustomed to "hypothesis".). Given that SD has not taken the medical route of creating a separate Latinate vocabulary, be careful to distinguish common usage from technical usage. (Look at the confusion arising from not all "models" being quantitative!) Likewise, when discussing borrowings from other fields, be clear on what part of the process or specific model features they pertain to. Avoid publishing articles that merely say "this seems relevant".

Second, teach iteration and testing as the fundamental modeling paradigm of SD, with knowledge of formulation tactics and structure-behavior relations added as a refinement, rather than the reverse.

The bad news here is that many tried-and-true lesson plans have the opposite orientation: The start is a bit of conceptual introduction about the unique concepts (feedback, stocks and flows), and then on into formulation exercises, then conceptualization exercises, with more realistic modeling exercises only at the end of the course. For example, the extensive discussion of hypothesis testing occurs in Sterman's *Business Dynamics* in the twenty-first chapter. There is considerable danger in a "bottom-up" approach that equation-writing will have been exercised many times, and model refinement only a few times, if at all.

A curriculum that focuses on hypothesis and testing will probably have to start with problems representable with a simple algebraic model, and deal with issues of how to refine such simple models to the point of usefulness. Then the problems and models get more complex, and more of the tactical information about SD comes into play. By the end of the course, the students will have practiced problem-solving through the scientific method many times, and testing a larger model only once, but with principles clearly understood. My hypothesis is that the students will learn much more of permanent value through such an approach, about both problem-solving and corporate life in general and about SD in particular. The current experience in teaching public school subjects through SD models instead of rote learning seems to confirm this hypothesis.

An important component of teaching iteration is teaching calibration to quantitative time series, in classrooms, as a legitimate technique within SD.

The third and most arduous thing to be done, to position SD as a science of strategy is for individual academics and consultants to claim expertise as such. I am urging a small number of self-selected academics to choose their path of research, publication and tenure to focus on questions of strategic dynamics (most likely for cases of individual corporations, for all the reasons Porter (1991) discusses). It is likely that the academic publications will need to be as much about defending an unconventional research methodology as about strategic issues of a particular company. (Here I am thinking of the extraordinary effort Chris Argyris and colleagues put into articulating and defending his methodology of active intervention in dysfunctional management teams (Argyris, *et al.* 1985).) But clearly, SD as an applied science of strategy represents a relevant and accessible research area with which academics can create careers.

In parallel fashion, SD consultants seemingly should offer their services as a scientific approach to strategy, with distinct value to corporations over and above that obtainable from "experts" with strategic frameworks.

6. Conclusion: Shall we do it?

In a way, the current situation of SD relative the rest of the world is like the very old joke about the two shoe salespeople who come to nineteenth-century China. One salesperson telegraphs back to the home office "The situation is hopeless. No one here wears shoes". The other salesperson telegraphs back to the home office "The situation has unlimited potential. No one here wears shoes".

Who must consider whether to take this banner forward? Regarding the third means, (individuals claiming expertise as such), it is clearly a matter for an individual person or an individual organization to decide to make this claim (and then back it up with deeds). Regarding the second means (giving primacy to scientific method over specific modeling technique), this is an empirical question, for which evidence from individual's experimentation may already be available. I call for discussion and publication in this area among professional academic teachers of SD. Regarding the first means (describing model-building in explicit terms of scientific method), why wouldn't we, each of us, start immediately? Only those practicing inappropriate process need have concerns. We should start immediately to describe what we do in terms the rest of the world can relate to.

References

Andrews, Kenneth 1971. *The Concept of Corporate Strategy.* Homewood, Illinois: Dow-Jones Irwin.

Andrews, Kenneth, Edmund P. Learned, C. Roland Christenson and William D. Guth 1965. *Business Policy: Text and Cases.* Homewood, Illinois: Dow-Jones Irwin.

Ariza, Carlos A. and Alan K. Graham 2002. Quick and Rigorous, Strategic and Participative: 10 ways to improve on the expected tradeoffs. *Proceedings of the 2002 International System Dynamics Conference.* Palermo, Italy. Forthcoming.

Argyris, Chris, Diana M. Smith and Robert Putnam 1985. *Action Science.* New York: Jossey-Bass.

Astrom, K. 1970. *Introduction to Stochastic Control Theory.* New York: Academic Press, 1970.

Chandler, Alfred. D., Jr. 1962. *Strategy and Structure: Chapters in the History of American Industrial Enterprise*, Cambridge Mass: MIT Press.

Collins, James C. 2001. *Good to Great: Why Some Companies Make the Leap and Others Don't.* New York: HarperCollins.

Collins, James C. and Jerry I. Porras 1994. *Built to Last: Successful Habits of Visionary Companies.* New York: HarperCollins.

Coyle, Geoff 2001. Rejoinder to Homer and Oliva. *System Dynamics Review* **17**(4) 357-363.

Cyert, R. M. and J. G. March 1963. A Behavioral Theory of the Firm (Engelwood Cliffs NJ: Prentice-Hall.

Forrester, Jay W. 1961. *Industrial Dynamics*. Waltham, Mass.: Pegasus Communications.

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

Forrester, Jay W. 1975. Toward a New Corporate Design. In Forrester, J. W.. *Collected Papers of Jay W. Forrester*. Portland Oregon: Productivity Press, 93-110.

Forrester, Jay W. and Peter M. Senge 1980. Tests for Building Confidence in System Dynamics Models. *TIMS Studies in the Management Sciences* **14**, 209-228.

Graham, Alan K. 1977. *Principles on the Relationship between Structure and Behavior of Dynamic Systems.* Cambridge, Massachusetts: Massachusetts Institute of Technology Ph.D. dissertation.

Graham, Alan K. 1982. Software Design: Breaking the Bottleneck. *IEEE Spectrum* (March), 43-50.

Graham, Alan K. 2000. Beyond PM101: 20 Lessons for Managing Large Development Programs. *Project Management Journal* **31**(4) 7-18.

Graham, Alan K. and Carlos A. Ariza 2001. Dynamic, hard and strategic questions: Using optimization to answer a marketing resource allocation question. *Proceedings of the 2001 International System Dynamics Conference.* Atlanta, Georgia.

Graham, Alan K. and Robert J. Walker 1998. Strategy Modeling for Top Management: Going Beyond Modeling Orthodoxy at Bell Canada. *Proceedings of the 16th International Conference of the System Dynamics Society, Quebec, Canada, July 20-23.*

Graham, Alan K., John D. W. Morecroft, Peter M. Senge and John D. Sterman 1996. Model-Supported Case Studies for Management Education, in Morecroft, John D. W. and John D. Sterman, eds. 1996. *Modeling for Learning Organizations.* Portland, Oregon: Productivity Press.

Graham, Alan K., Jonathan Moore and Carol Y. 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 (forthcoming).

Graham, Alan K., Carol Y. Choi and Thomas W. Mullen 2002. Using Fit-Constrained Monte Carlo Trials to Quantify Confidence in Simulation Model Outcomes. *Proceedings of the 2002 Hawaii Systems Conference*.

Green, Paul E. and Y. Wind 1975. A New Way to Measure Consumer's Judgments. *Harvard Business Review* **53** (July-August), 107-117.

Hamel, Gary and C. K. Prahalad 1994. *Competing for the Future,* Boston Mass.: Harvard Business School Press.

Henderson, Bruce D. 1979. *Henderson on Corporate Strategy,* Cambridge Mass: Abt Books.

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

Kalman, R. 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering,* Series D, **82** (March), 35-45.

Kaplan, Robert S. and David P. Norton 1996. *The Balanced Scorecard: Translating strategy into action.* Boston, Mass.: Harvard Business School Publishing.

Kaplan, Robert S. and David P. Norton 2000a. *Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment.* Boston, Mass.: Harvard Business School Publishing.

Kaplan, Robert S. and David P. Norton 2000b, "Having Trouble with Your Strategy? Then Map It. *Harvard Business Review* (September-October), 170-171.

Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lane, David C. 2001a. *Rerum cognoscere causas:* Part I – How do the ideas of system dynamics relate to traditional social theories and the voluntarism / determinism debate?. *System Dynamics Review.* **17**(2) 97-118.

Lane, David C. 2001b. *Rerum cognoscere causas:* Part II – Opportunities generated by the agency / structure debate and suggestions for clarifying the social theoretic position of system dynamics. *System Dynamics Review.* **17**(4) 293-309.

Lumerer, Kenneth M. 1985. *Learning to Heal: The Development of American Medical Education.* New York: Basic Books.

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

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

Mashayekhi, Ali N. 2001. President's Address: A Vision for the Field of System Dynamics. *System Dynamics Newsletter* **14**(2) December 2001, 1, 9-13.

Meadows, Donnela H., William W. Behrens, Jorgen Randers and Dennis L. Meadows 1974. *Limits to Growth.* New York: Universe Publishing.

Mintzberg, Henry 1994. The Fall and Rise of Strategic Planning. *Harvard Business Review*, (January-February), 107-114.

Mintzberg, Henry and James A. Waters 1985. Of Strategies Delivered and Emergent. *Strategic Management Journal* **6**(3): 257-272

Morecroft, John D. W. 1985. Rationality in the Analysis of Behavior Simulation Models. *Management Science,* **31**(7), 900-916.

Oliva, Rogelio 2002. Model Calibration as a Testing Strategy for Dynamic Hypotheses. *European Journal of Operational Research* (forthcoming).

Peteraf, Margaret A. 1993. The Cornerstones of Competitive Advantage: A Resource-Based View. *Strategic Management Journal,* **14**(3) 179-192.

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.

Peterson, David W. and Fred C. Schweppe 1974. Code for a General Purpose System Identifier and Evaluator: GPSIE. *IEEE Transactions on Automatic Control* (December).

Popper, Karl R. 1990. *The Logic of Scientific Discovery.* New York: HarperCollins.

Porter, Michael E. 1980. *Competitive Strategy*. New York: The Free Press.

Porter, Michael E. 1985. *Competitive Advantage.* New York: The Free Press.

Porter, Michael E. 1991. Towards a Dynamic Theory of Strategy. *Strategic Management Journal* **12**, 95-117.

Prahalad, C. K. and Gary Hamel 1990. The Core Competence of the Corporation", *Harvard Business Review* (May-June), 71-91.

Randers, Jorgen, ed. 1980. *Elements of the System Dynamics Method*. Portland, Oregon: Productivity Press.

Rasiel, Ethan M. 1999. *The McKinsey Way.* New York: McGraw-Hill.

Repenning, Nelson and Fred Kofman 1997. Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement. *Management Science* **43**(4), (April).

Richardson, George P. 1991. *Feedback Thought in Social Science and Systems Theory*. University of Pennsylvania Press.

Schmidt, Markus J. and M. Shayne Gary 2002. Guiding New Product Development and Pricing in an Automotive High Tech SME. *Proceedings of the 2002 International System Dynamics Conference.* Palermo, Italy (forthcoming).

Schweppe, Fred C. 1965. Evaluation of Likelihood Functions for Gaussian Signals. *IEEE Transactions of Information Theory*. **IT-11**(1) (January), 61-70.

Schweppe, Fred C. 1973. *Uncertain Dynamic Systems.* Engelwood Cliffs, NJ: Prentice-Hall.

Seeger, John A. 1984. Reversing the Images of BCG's Growth/Share Matrix. *Strategic Management Journal.* **5**: 93-97.

Senge, Peter M. 1974a. Evaluating the Validity of Econometric Methods for Estimation and Testing of Dynamic Systems. Cambridge, Mass.: Massachusetts Institute of Technology Sloan School of Management System Dynamics Group Working Paper D-1944-2.

Senge, Peter M. 1974b. An Experimental Evaluation of Generalized Least Squares Estimation. Cambridge, Mass.: Massachusetts Institute of Technology Sloan School of Management System Dynamics Group Working Paper D-1944-6.

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

Sibbet, David 1977. 75 Years of Management Ideas and Practice: 1922-1997. *Harvard Business Review.* Sept-Oct 1997, supplement.

Simon Herbert A. 1979. Rational Decision-Making in Business Organizations. *American Economic Review* **69**(4).

Stephens, Craig A., Alan K. Graham, and James M. Lyneis 2002. System Dynamics Modeling in the Legal Arena: Special Challenges of the Expert Witness Role. Palermo, Sicily: *Proceedings of the 2002 International System Dynamics Conference.* (forthcoming)

Sterman, John D 1984. Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models. *Dynamica.* **10**(2) 51-66.

Sterman, John D. 1985. The Growth of Knowledge: Testing a Theory of Scientific Revolutions with a Formal Model. *Technological Forecasting and Social Change.* **28**(2) 93-122.

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

Tahai, Alireza and Michael J. Meyer 1999. A Revealed Preference Study of Management Journal's Direct Influences. *Strategy Management Journal* **20**(3), 279-296.

Warren, Kim 1999a. The Dynamics of Strategy. *Business Strategy Review* **10**(3) 1-16.

Warren, Kim 1999b. The Dynamics of Rivalry. *Business Strategy Review* **10**(4) 41-54.

Warren, Kim 2000. The Softer Side of Strategy Dynamics. *Business Strategy Review* **11**(1) 45-58.

Wernefelt, Birger 1984. A Resource-Based View of the Firm. *Strategic Management Journal,* **5** 171-180.

Appendix: How the Iterative Validation Process for High-End Modeling Approximates System and Parameter Identification through Full-Information Maximum Likelihood via Optimal Filtering (FIMLOF)

This paper can give no more than a cursory description of FIMLOF works; the interested reader will find a very approachable exegesis in (Peterson 1980). However, a simplified outline of the algorithm in SD terms will suffice to understand the nature of the approximation.

The heart of the FIMLOF algorithm is the "predict-correct" cycle: starting from estimated values of the level variables, use the model equations to simulate forward to the next time for which real data are available. The "predict" part is then using model equations to predict what the observed data "should" be *a priori* for that time, i.e. the estimated observation given the estimated levels from the previous time. Of course, the real data will differ from the estimate, because of random noise driving the dynamics and random noise corrupting the data. Those differences are called the *residuals*. Standard Bayesian estimation can use model equations and the residuals for that time to calculate an *a posteriori* estimate of the level variables for that point in time. Those estimates are the starting point for the next predict-correct cycle. So the algorithm described thus far takes a stream of observations, and produces a stream of estimated level variables, and a stream of residuals. This is the Optimal Filtering (OF) part of FIMLOF.

Full-information maximum likelihood estimation backs into parameter values by calculating, for a given sample set of real observations and their residuals, the parameters that maximize the probability density of that aggregate sample. It turns out that the logarithm of the probability density, the likelihood, is a function of a quadratic weighting of the residuals. Therefore, the Full-Information Maximum Likelihood (FIML) estimate of the parameter values is a particular Weighted Least Squares (WLS) parameter estimate, which can be found by standard search methods. In essence, then, FIMLOF estimates parameters by finding a best WLS fit for a predict-correct cycle.

Predict and correct solves a problem with calibration via pure simulation, which Forrester (1961, Chapter 13 and Appendix K) identified, as noted earlier: if small amounts of unknown, random noise are driving an oscillatory system, a simulation of a completely accurate model and the real data can be out of phase. One runs the risk of using parameter changes to compensate for the accumulated consequences of random noise. Standard high-end SD handles this problem in a different way.

But in a real application, if a large perturbation is throwing model behavior off, the investigator can find out what was going on, and incorporate it into the model. The intention is that iterating in this fashion whittles away at the drivers not modeled explicitly to the point where the major drivers are known, and the modeler will not be biasing the parameter estimation to compensate for unknown exogenous drivers. Analysis of entrainment (Graham 1977, Ch. 5) implies that even for the case of moderately-damped oscillatory systems, once the major exogenous driving forces are known, differences simulated behavior and real behavior tend to manifest in phase differences, which are a smooth and continuous function of differences in resonant period over a broad range. So adjusting parameters to improve fit will in fact be moving them toward their "real" values. Seemingly, the only situation where this approach would fail is where a moderately-damped or lightly-damped system existed in virtual isolation in unchanging fashion, such that there are no time-varying structural changes and no major exogenous events, ever. At least in commercial practice, this hardly ever happens.

If major flaws in the data exist, the investigator must of course dig out the story. But then there are options: like the econometrician, the SD modeler may undertake to correct the data. (The econometrician needs to do so, lest corrupt data corrupt the results, cf Peterson 1975) But if the flaws have knowable causes, an SD modeler has the option of simply ignoring flaws and discrepancies. If anything, the model and modeler gain credibility by pointing to a discrepancy between data and model behavior and describing a very credible explanation.

Few modelers formally derive the WLS weightings that formally implement maximum likelihood estimation. However, engineering experience is that plausible *ad hoc* weightings often perform satisfactorily (Schweppe.1973). Carefully chosen weightings protect one in the event that errors in one data series can only be decreased at the expense of increasing errors for some other series. However, high-stakes dynamic models often seem to be over-identified (Graham, 2000, Graham, Moore and Choi 2002); at the end of an iterative modeling process, such tradeoffs seldom happen in any major way.

As another format of comparison of iterative manual hypothesis testing and calibration versus FIMLOF, Peterson (1980) lists eight characteristics that are difficult for standard econometric techniques to deal with, which FIMLOF handles rigorously. These are listed in Table 1 (next page), with corresponding entries describing how a high-end model validation process in effect arrives at the same point: a model structure and parameter consistent with known facts about the system, tuned to a suitable objective function.

In the end, FIMLOF and iterative deal with life's imperfections in different ways. FIMLOF provides a way for objective maximization to encounter the vicissitudes of life then recover robustly. Iterative hypothesis testing and calibration might be said to go around those vicissitudes.

Condition of data and system	**How iterative modeling and hand-calibration deal with the condition**
Nonlinearities in model dynamics	Simulation of models in levels and rates format in the "predict" phase
Nonlinear measurement functions	Non-reduced form equation formulation that creates variables in exact form of measurements, even if nonlinear (e.g., ratios)
Measurement error (errors in the variables)	Large residuals: investigate and correct the data or dynamics through the iterative process Minor residual error: Use a formal fit criterion like WLS to guide parameter estimation, but do not set parameters outside their a priori limits without empirical investigation
Mixed sampling intervals (e.g., estimating a weekly model, using monthly and yearly data)	Not a problem for visual calibration. For tuning to a fit criterion, use a fit criterion that weights according to quantity of information
Models with unmeasured endogenous variables	Incorporate the variables into the model and calibrate to measurements on driving and driven variables, direct and indirect, to approximate the Bayesian inference of the "correct" portion of the predict–correct cycle. In addition, part of the iterative scientific method here is to go get more data. If something was truly unmeasured, approximate information often can be obtained through interviews.
Cross-sectional, time series mixed data	Usually, a major strategy model is disaggregated somewhat, so that cross-sectional data is also (possibly incomplete) time series data at a more granular level
Unknown characteristics of equation errors and measurement noise	As above, investigate and incorporate findings to the point where unknown equation errors (e.g. exogenous drivers) and measurement errors no longer dominate the fit between data and simulation.

Table 1. SD used as an iterative scientific method approximates the results of Full-Information Maximum-Likelihood (estimation) via Optimal Filtering.

Back to the Top