Table of Contents

Go Back

Might Twenty Models Cover Ninety Percent of All Situations Managers Encounter?

Scott F. Rockart, Fuqua School of Business

Duke University, Durham, N.C. 90120

email: srockart@duke.edu - phone: 919-660-7998

Original Draft: 3 January, 2004 Current Revision: 27 August 2004

Abstract

Jay Forrester's call for a set of 'general, transferable computerized cases' to cover most managerial situations is one of the great tasks standing before the field of system dynamics. A library of widely applicable cases, if accompanied by reliable guidelines for when to apply them, would be a boon to research, teaching, and management. Researchers could use these cases as strong null models when evaluating new situations and new theories (Bell 2001). Teachers could use the most widely applicable cases for a general management course or could use a thematic subset of the cases for specialized courses. Managers could approach a new situation by selecting then tailoring the model or models most likely to shed light on that situation. This paper discusses progress that has been made toward this library of general and transferable models, and describes a research project under way to both evaluate models for the library and produce the application guidelines needed for the library to fulfill its promise.

(Learning, Competitive Dynamics, Capabilities, Professional Services)

“System dynamics models usually represent a family of social systems … *when possible* a model should be a general model of a class of system to which belongs the particular member of interest … An important step in validation is to show that the model takes on the characteristics of different members of the class when policies are altered in accordance with the known decision making differences between the members.” (Forrester and Senge 1980) (italics in original)

“…20 such general, transferable, computerized cases would cover perhaps 90 percent of the situations that managers ordinarily encounter." (Forrester 1991)

Jay Forrester’s call for a set of ‘general, transferable computerized cases’ to cover most managerial situations is one of the great tasks standing before the field of system dynamics. A library of widely applicable cases, if accompanied by reliable guidelines for when to apply them, would be a boon to research, teaching, and management. Researchers could use these cases as strong null models when evaluating new situations and new theories (Bell 2001). Teachers could use the most widely applicable cases for a general management course or could use a thematic subset of the cases for specialized courses. Managers could approach a new situation by selecting then tailoring the model or models most likely to shed light on that situation. This paper discusses progress that has been made toward this library of general and transferable models, and describes a research project under way to both evaluate models for the library and produce the application guidelines needed for the library to fulfill its promise.

What is a general transferable case?

A lot of attention and effort in the SD field has focused on generic feedback structures (AKA archetypes) and model components (AKA molecules). However, by ‘general, transferable computerized cases’ Jay appears to be talking about a description of a situation coupled with a fully-specified mathematical model tied to a specific kind of situation that can be simulated to understand that kind of situation. As Lane and Smart (1996) discuss in detail, a computerized case of this kind is a far cry in substance and purpose from the small sets of coupled feedback loops that have occupied a great deal of attention in the SD field. What is the appropriate level of model upon which to build a library?

If we pursue Jay's vision, we can be fairly certain that the models in the library will seem quite redundant to the expert modeler. The models will share not only the most basic mathematical structures (stocks and flows) but also more complex component mathematical structures (e.g., aging chains, perception delays, and co-flows). Many of the models may even be indistinguishable at the level of feedback structure.

However, the hunt for ever greater generality can be taken too far. At the extreme, the systems field could be reduced to general statements about reinforcing and balancing feedback loops, the role of delays, and the potential for shifts in loop dominance. Managers (and policy makers) and most academics desire more guidance than to look for 'limits to growth' or 'overshoot and decline' no matter how useful such general guidance may be.

We already have good candidates for library models of the kind Jay is advocating. Reflecting on one of these illustrates the value of less general but still powerful models. The Bass Diffusion Model may be considered a leading candidate for inclusion in the library. The model sheds light on an issue that managers frequently encounter; the diffusion of a new product or service. The model comes complete with setting, content, and a mathematical specification. The model is easily coupled with one of many cases used in business schools for teaching about diffusion. A manager or researcher encountering a diffusion setting will usually find that the model predicts the behavior in that setting well, and she can quickly adopt and adapt the model even without great modeling expertise. This is true in no small part because the model is very clear about the content of the stocks (people) and dimensions of the one flow in the model (people per unit of time). In addition, the Bass Model's generalizability has been reasonably well established. It can be and has been parameterized to fit a wide range of diffusion processes. When the situation requires more than parameter changes to the Bass Model, managers and researchers are likely to find that it still serves as a good starting point or core model (Homer 1987).

One could argue that the Bass Diffusion Model is simply one example of the ‘limits to growth’ archetype. It is. But so is Jay’s Market Growth Model (Forrester 1968), John Sterman’s model behind the People Express flight simulator, the World models behind the *World Dynamics* and *Limits to Growth* books (Meadows 1992), and many other models that shed light on new situations. We could argue that we need only one of these models in the library or merely a description of how one reinforcing and one balancing loop can interact to create growth and then saturation. Applying the same logic, we could argue that the library needs only one model of cyclical behavior or merely a discussion of a balancing feedback loop with delays. If so, should we include one of Meadow’s commodity models (Meadows 1970), a predator-prey model, Jay’s model of inventory and human resources from Industrial Dynamics, or perhaps a model of the economic long wave?

If we reduced our model library to archetypes in this way, we would lose a great deal of the settings and constructs that make models compelling to managers. More importantly, a manager may be able to see the connections between basic feedback structures and the situations she faces, but still want evidence that these structures are both critical to the situation she is facing and adequate to guide decisions. To satisfy the real needs of managers, the model library has to accept some redundancy for the sake of detail.

This is not to say that many kinds of system descriptions (archetypes, molecules, general cases) cannot coexist. However, if the library of models Jay envisions is to become a reality, we have a great challenge ahead of us. Models in the library need to be coupled to evidence that they are in fact both general and powerful when applied to some clearly defined range of issues (i.e., context). Gathering such evidence will require a substantial shift in the emphasis our field has placed on research methods.

How do we know that our models are general and transferable?

Most of the data, data collection methods, and data analysis techniques that we can use to determine the extent and limits of a model’s explanatory power (i.e., whether a model belongs in the library) differ from those that we commonly use to develop and evaluate our models. Surveys and broad archival analyses are the norm for investigating descriptive questions and questions about generalizability such as *how often* and *how many* as well as who, what, where, which, how much (Yin 1994). These methods, unfortunately, rarely provide the level of detailed insights into process that we need to develop explanations and thus are rarely used by our modeling community. Experiments, interviews, and the careful historical analysis of individual cases that we use to collect detailed explanatory information tell us a great deal about a model's fit with a specific case, but little about the generalizability that we need to identify library models.

This is not to say that the field has ignored generalizability entirely. Researchers have relied on a number of cues to generalizability in order to leverage case findings beyond the original case or cases studied. The primary means of leverage have been: directed case selection and structural comparisons, testing propositions, replication, transfer of generic structures, and enumeration of patterns.

The selection of cases appears to be the most common way that researchers bridge the gap between descriptive breadth and explanatory depth. Cases are chosen as exemplars of (or as revealing exceptions to) a phenomenon known to be widespread but poorly understood. Sterman, Kaufman, and Repenning (1997) investigated the case of Analog Devices because it had struggled to make its improvement efforts benefit the company. Others had already documented that this was a widespread phenomenon; many firms experience trouble translating operational improvement efforts into financial improvement. The case was intriguing and was likely to be particularly illuminating; the firm had actually succeeded in making substantial operational improvements yet the firm's financial results worsened compared to those of its peers.

Similarly, Hall (1976) describes the rapid growth then collapse of several leading magazine companies. He studied one company in depth to understand why it – and hopefully by extension all three – ran into problems. He extends the analysis beyond the original case by presenting evidence that all three firms shared key policies implicated in the downfall of the Saturday Evening Post.

A classic way to test a model broadly is to reduce it to conform to the data available on a larger set of cases. The reduction is often down to a set of propositions deduced from the model. Ideally these propositions discriminate not only between the model and 'no effect' null hypothesis, but also between the model and predominant (and presumably plausible) rival hypotheses. SD studies of behavioral decision making have made extensive use of proposition testing through experimental methods (Sterman 1989; Sterman 1989; Kampmann 1992; Moxnes 1998). Experimental settings are particularly amenable to this kind of inquiry since conditions can be manipulated to provide the data needed to discriminate among rival propositions.

Models of cyclical behavior (Antonelli 1989), forecasting and the Bass diffusion model have also been subjected to tests of propositions with non-experimental data. Much of this testing has occurred because these models have a much wider following and originated outside of the SD tradition. This wider interest group has provided both skilled practitioners and an audience for testing propositions. Large-sample tests of propositions derived from SD models are far from the norm. The field has long emphasized the value of including non-numerical data (and the foolishness of excluding it) in understanding a situation, making it seem odd to reduce a SD model for the sake of accommodating limits on numerical data (Forrester 1961). SD practitioners, and many outside of SD, were for a long time wary of available empirical methods for studying non-experimental data on systems (Senge 1977; Randers 1980; Leamer 1983). When there were few trained SD practitioners it seems sensible that they would have spent their time building SD models rather than traditional empirical testing. With the great richness of the models they created it is hard (philosophically if not practically) to reduce them to component insights that would make good fodder for testing.

The family-member test is the SD world's terminology for replication testing (Forrester and Senge 1980; Sterman 2000). Replication tests a model by seeing if it can explain situations where similar or different behavior has been observed (Yin 1994). There is a widely noted and oft bemoaned dearth of replication research across the social sciences (Madden et al.). While there are exceptions (Meadows 1970; Hall 1976) the same is generally true of SD studies of the social sciences. Most replications are relatively casual, limited in scope, and with little rigorous basis for the selection of replication settings.

There are a lot of reasons why replication research gets little attention. Successful discovery of new ideas provides avenues for subsequent research and with it acclaim and citations. Successful replication efforts, in contrast, are unlikely to be published. Unsuccessful replication is more likely to be published, but only if it directly challenges the original research and this is a hard thing to accomplish outside of experimentally controlled settings. Less stark replication failures are likely to promote enmity rather than acclaim, and social connections are critical to success in academia. Replicating one's own research is costly, risky, and generally not required to obtain credit for the original research. As a result, it is not surprising that published replication research derives substantial merit from having extended the original model considerably rather than merely replicated it in new settings (Homer 1987).

Establishing that twenty models cover ninety percent of situations would require extensive replication research but is conceptually feasible (Lane 1988). How do we encourage more replication research and should we encourage replication research? I think one of the most valuable things we can do is increase our expectations of replication and thus increase the benefit to the researcher for performing replication research. What do I mean? Currently, replication research serves two primary purposes. The first purpose replication research serves is as a test of the original finding's internal validity. Where controlled experimentation is possible, replication is a stark test of the internal validity original experiment. An experiment that cannot be replicated is fundamentally discredited. Where controlled experimentation is not possible, as it is not in case study research, we are much slower to reject the original finding because it is not repeated in additional settings. The reduced potential in non-experimental settings makes replication research less attractive. The second purpose replication research serves is as a test of the external validity, or generalizability, of the finding. In addition questioning the validity of the insights for the original case setting, replication research can help to establish the boundaries of the findings.

Transfer of generic structures (AKA archetypes) across settings is replication's pedagogical sibling and the enumeration of archetypes has been a particularly popular way that SD researchers have promoted transfer. SD archetypes are fundamentally abstract mathematical equations presented graphically in the form of feedback loops. They are made useful by coupling them with examples of how the feedback loops generate behavior and how they can be populated with many different constructs to show how behavior may be generated in a wide variety of settings. Generic structures arguably help people to grasp new situations more quickly and manage them more effectively. As researchers and managers identify and document additional situations where the generic structures are applicable, we can consider these to be casual replications. However, such casual replications do not get us closer to evaluating Jay's conjecture about how many models it would take to cover a wide range of managerial situations.

The use of generic structures raises an important question. What are we replicating? The generic structures strip past theories of their specific constructs and mechanisms leaving only the most general statement of mathematical structure (causal loops) tied to the very loosely defined constructs and mechanisms and a set of examples. The generic structures lack the specificity that many see as the fundamentals of a theory. While SD overall may constitute a 'theory of structure' (Forrester 1968) we should not sell the specific theories short by abstracting them into oblivion.

A great deal of the vibrancy of academic debate is about the relative importance of specific constructs and mechanisms that determine organizational behavior including: institutions, firm networks, inertia, information asymmetries, agency and governance, routines, transaction costs, social capital, reputation, and resources. The debate is vibrant because knowledge of specific constructs and mechanisms helps policy makers and managers; particularly so when and where specific predictions about what is important can be linked to easily observed characteristics of situations. Most SD models have such specificity but need a solid empirical foundation suggesting their generality as well as their internal validity. I think it would be a mistake to be satisfied with transfer of generic structures in place of replication of specific theories and we are shy on true replication research.

A worthwhile precursor to investigating the empirical generalizability of models would be to look at the empirical data on the generalizability of patterns. Forrester (1964) proposed four basic reference modes for company growth (exponential growth up to a healthy plateau, continued growth with oscillation, growth to a plateau with oscillation, and growth and collapse) and spoke about their relative frequency based on his own experience. These reference modes are powerful guides for framing our questions. They have reappeared frequently within SD discussions and are carefully replicated with credit our field's most recent text (Sterman 2000, pp.99). However, it does not appear that we have attempted an empirical study to check Jay's intuition about the relative frequency of these patterns and to see how frequently other patterns occur.

Such a study would honor Jay's original work by treating it as important enough to investigate and provide a sound basis for establishing the extent and limitations of our knowledge. Even if the four reference modes turn out to cover the vast majority of all situations, and their relative frequency across the population proves to be similar to Jay's experience, we still would benefit from knowing how the time frame differs for the reference modes across settings.

Similarly, Sterman provides six modes of behavior in his text that he asserts are the most common (2000, pp.99). These include three of Forrester's four reference modes (exponential growth up to a healthy plateau, growth to a plateau with oscillation/overshoot, and growth and collapse) along with three more modes that are described as the 'most fundamental' and include goal seeking behavior along with two more modes (pure oscillation and exponential growth) that are components of Forrester's three. These modes have great pedagogical value and Sterman provides a wealth of examples illustrating how often and how widely they reappear.

In both cases, the authors use the modes for pedagogical purposes and rely on their experience to make judgments about frequency and importance. Given their importance to the field, the prevalence of modes deserves further research along with the generalizability of models. Such research might shift our beliefs about the importance of certain modes and could likely reveal more subtle combinations of modes that have received inadequate attention. Behavior modes such as punctuated change (Sastry 1997) and the fall-and-then-rise mode often termed worse-before-better behavior (Forrester

1969) may deserve further attention and eventual inclusion in a short list of fundamental behavior modes.

Generalization of theories to nested populations

Selective replication research builds greater confidence in the soundness of the original research and broad study of patterns helps to focus research and pedagogical effort. However, it represents only a small step toward the goal of twenty models for ninety percent of all situations. Insight about prevalence of phenomena requires a sampling logic; we must first define a population to which we believe (hope) a model can be generalized and then collect data on the census or a sample of that population.

With several research assistants, I am currently working to test the feasibility of generalizing slightly simplified versions of several classic models to nested (increasingly broad) populations. One of these models is Roger Hall's model of the Saturday Evening Post that appears to do a very good job of helping us to understand other magazines, notably the recent issues facing Martha Stewart Living (Hall 1976). A second model is a simplified version of John Sterman's People Express model. In each case the work involves first defining a narrow population of firms (magazines and point-to-point airlines respectively) and seeing how easily the models can be adapted to fit the behavior of members of the population. Based on the initial results, a wider population will be defined and the tests repeated until the models cease to be good guides to understanding the new situations.

It is worth saying that this endeavor is not without big questions. Yin (1994, p. 48) rejects the use of sampling logic with case studies arguing that case studies "should not generally be used to assess the incidence of phenomena" and that "a case study would have to cover both the phenomenon of interest and its context, yielding a large number of potentially relevant variables …require an impossibly large number of cases…" However, I do not think we should be deterred by these arguments. The goal of testing is not to extend the model until it fits all cases but rather to see how many and which cases the model fits with little or no modification. At a minimum, we would expect to have a few details that we would need to know for each case (the parameters that our base model is sensitive to) after which we see how well the model lines up with the family of cases without further modification.

Another issue that must be confronted is that thehe SD field has long advocated modeling problems and warned against modeling systems. This style of research seems like developing models of systems and in fact it is. However, once we have a model motivated by a problem it is worth asking how common that problem is among firms with similar systems, and whether the system components included in the model to explain that problem in one case are able to produce behavior shown by other members of the family (nested population) to which the system belongs. If the population is structurally similar enough, and the original model captures key parts of the structure, then additional problems (reference modes) might all be generated by small variants on the interactions of the constructs and mechanisms underlying the original problem that motivated the modeling effort. That, in effect, is the hope behind Jay's conjecture about twenty models.

This work does have successful precedents with simple models. Sterman shows how simple models of expectation formation mimic the inflation forecasts found in a survey of business, academic, and government economists and the energy use forecasts collected by the Department of Energy (Sterman 1987) (Sterman 1988).

Ultimately, what I am proposing is the development of exhaustive but not exclusive categorizations of firms based on a variety of factors including: industry characteristics, firm age, firm size, extent and logic of diversification, firm strategy, or behavior. For each category, a model will be adapted or developed and tested. Making these groupings exhaustive will reveal the prevalence of managerial situations where our models fail to provide insight. Allowing groupings to be non-exclusive simply recognizes that more than one model might need to be investigated for any given setting to be reasonably certain that key issues have been identified. At the end of the process I hope we will be able to say "Here is a model that will apply to most situations where …" and "any argument that purports to explain this situation should be compared to that model."

References

(1998). Building Edison's Bulb. Technology Review. **May/June:** 96.

Antonelli, C. (1989). "The role of technological expectations in a mixed model of international diffusion of process innovations: the case of open-end spinning rotors." Research Policy **18**.

Bell, G. (2001). "Neutral Macroecology." Science **293**: 2413-2418.

Forrester, J. W. (1961). Industrial Dynamics. Cambridge MA, Productivity Press.

Forrester, J. W. (1964). Modeling the Dynamic Processes of Corporate Growth. IBM Scientific Computing Symposium on Simulation Models and Gaming, Thomas J. Watson Research Center, Yorktown Heights, NY.

Forrester, J. W. (1968). "Market growth as influenced by capital investment." Industrial Management Review **9**(2): 83-106.

Forrester, J. W. (1968). Principles of systems. Cambridge, MA, Wright-Allen Press Inc.

Forrester, J. W. (1969). Urban Dynamics. Cambridge MA, Productivity Press.

Forrester, J. W. (1991). From the Ranch to System Dynamics: An Autobiography, M.I.T., System Dynamics Group.

Forrester, J. W. and P. M. Senge (1980). Tests for Building Confidence in System Dynamics Models. System Dynamics. A. A. Legasto, Jr. and e. al. New York, North-Holland. **14:** 209-228.

Hall, R. I. (1976). "A system pathology of an organization: The rise and fall of the old Saturday Evening Post." Administrative Science Quarterly **21**(2): 185-211.

Homer, J. B. (1987). "A Diffusion Model with Application to Evolving Medical Technologies." Technological Forecasting and Social Change **31**(3): 197-218.

Kampmann, C. P. E. (1992). Feedback complexity and market adjustment: An experimental approach. Cambridge, MA, Massachusetts Institute of Technology.

Leamer, E. E. (1983). "Let's take the con out of econometrics." The American Economic Review **73**(1): 31-43.

Lane, D. and C. Smart (1996). "Reinterpreting 'generic structure': evolution, application, and limitations of a concept." System Dynamics Review **12**(2): 87-120.

Lane, D. (1998). "Can we have confidence in generic structures?" Journal of the Operational Research Society **49**: 936-947.

Madden, C. S., R. W. Easley, et al. (1995). "How journal editors view replication research." Journal of Advertising **24**(4): 77.

Meadows, D. H. e. a., Ed. (1992). Beyond the Limits. Post Mills, VT, Chelsea Green Publishing.

Meadows, D. L. (1970). Dynamics of Commodity Production Cycles. Cambridge, MA, Wright-Allen Press.

Moxnes, E. (1998). "Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics." Management Science **44**(9): 1234-1248.

Randers, J., Ed. (1980). Elements of the System Dynamics Method. Cambridge, MA, Productivity Press.

Sastry, A. M. (1997). "Problems and paradoxes in a theoretical model of punctuated organizational change." Administrative Science Quarterly **42**: 237-275.

Senge, P. (1977). "Statistical estimation of feedback models." Simulation(June): 177-184.

Sterman, J. D. (1987). "Expectation Formation in Behavioral Simulation Models." Behavioral Science **32**: 190-211.

Sterman, J. D. (1988). "Modeling the Formation of Expectations: The History of Energy Demand Forecasts." International Journal of Forecasting **4**: 243-259.

Sterman, J. D. (1989). "Misperceptions of Feedback in Dynamic Decision Making." Organizational Behavior and Human Decision Processes **43**(3): 301-335.

Sterman, J. D. (1989). "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." Management Science **35**(3): 321-339.

Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston, McGraw-Hill Companies.

Sterman, J. D., N. P. Repenning, et al. (1997). "Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement." Management Science **43**(4): 503-521.

Yin, R. K. (1994). Case Study Research. London, SAGE Publications.