Mental Models and Dynamic Decision Making in a Simulation of Welfare Reform[1]

David F Andersen, Terrence A Maxwell, George P Richardson, and Thomas R Stewart[2]
Center for Policy Research
Rockefeller College of Public Affairs and Policy
University at Albany, State University of New York
Albany, New York 12222 USA
Tel: 518 442 3850
Fax: 518 442 3398
E-mail: fadum@csc.albany.edu

Abstract

This paper is the second in a pair presented in this volume. The first paper presents a theoretical view of mental models appropriate for carrying out empirically-based research on system dynamics modeling interventions. Mental models consist of three types of measurable sub-models--ends models, means models, and means-ends models. The means-ends models may be thought of as containing either detailed "design" logic or much more simple "operator" logic. This paper presents an empirical test of the impact of interventions intended to improve design versus operator logic for 53 participants in a dynamic learning laboratory with a task centering on implementing welfare reform over a simulated twenty year period. Results suggest that providing managers with high level heuristic results from modeling interventions is a necessary condition for achieving improvement in system performance. Focusing on operator logic is key to improving managerial performance of dynamic tasks.

1 Preparation of this paper was supported in part by the National Science Foundation (Decision, Risk, and Management Science Program) under grant no. SES-9211521.

2 This paper represents a fully collaborative effort. The authors are listed in alphabetical order.

Mental Models and Dynamic Decision Making in a Simulation of Welfare Reform

System dynamics literature has paid a great deal of attention over the last decade to the importance o aligning mental models with the systems people are trying to control or change. This focus on alignmen is particularly critical in training programs which employ systems thinking and seek to impar understanding of the structure of a system to mangers who are undertaking simulation-based training ii management laboratories or computer "microworlds."

Several questions about the efficacy and design of such workshops need to be explored. Does detaile understanding of structure and the dynamic behavior resulting from that structure help managers t perform dynamic tasks better in uncertain environments? Is structural understanding the most importan understanding to impart or should trainers place greater emphasis on the development of key heuristic which managers can use to operate the system in question? And finally, what is the generalizability o such training? This paper presents a preliminary empirical investigation of these questions.

I. Theoretical Foundations

The concept of managers' mental models as the key intervening concept between system dynamic modeling interventions and managerial behavior designed to improve system functioning was introduce at an early date by Forrester (1961) and remains a powerful organizing concept even in recent literatur (Senge, 1990). However useful this concept may be for guiding the practice of system dynamic modeling, we believe that the construct of mental models as used in the present systems interventio literature is often still discussed as a pre-scientific construct. Within much of the tradition of the systen dynamics intervention literature, the concept of mental models has defied careful and precise description has not yielded a consistent set of measurable attributes from one study to the next, and for the most par remains undescribed and unmeasured. However, other recent work more carefully grounded in th psychological literature has made better progress in terms of defining and measuring what one means b mental models (e.g. Sterman 1989,1994, Brehmer 1988, Vennix 1990, and Diehl 1992).

In work related to this paper, Andersen, Maxwell, Richardson, and Stewart (1994) have proposed marriage of a cybernetic view of decision making with the Brunswickean lense model (Brunswik 1956 Hammond 1955, Hammond, Stewart, Brehmer and Steinman 1975, Brehmer and Joyce 1988) to produc a more conceptually complete view of what one might mean by a mental model when used to describ system dynamics modeling interventions. We posit that mental models must be viewed as a specific an well-defined subset of all types of mental activity used to support managerial decision making. Mor specifically, three types of constructs combine to create useful and measurable mental models--mean models, ends models, and means-to-ends models.

This expanded theoretical framework was used to derive competing hypotheses concerning what i important in mental model research. Four such competing hypotheses were sketched--(1) the outcom feedback hypothesis, (2) the cue selection hypothesis, (3) the design logic hypothesis, and (4) th operator logic hypothesis. Each of these four views highlights different aspects of the overall decision making process, predicts that differing managerial interventions will be effective in improving manager ability to manage complex situations, and yields empirically verifiable propositions that may be used t test these sometimes competing predictions.

Our related theoretical work also appearing in this volume (Andersen, Maxwell, Richardson, and Stewa 1994) suggests that the design logic versus operator logic hypotheses yield quite different prediction concerning how a systems intervention will act to improve managerial decision processes. The design logic proposition emphasizes the need to create more elaborate, causally sophisticated, and feedback-sensitive cognitive models of means-ends effects. Only by understanding the complexity of the systems that they are managing will managers be better able to improve system performance.

On the other hand, the operator logic perspective predicts that causally sophisticated and feedback oriented cognitive understanding of means-ends effects are not effective ways to intervene to improve performance. The key to improving performance is to support management's strategy selection process directly by giving management key strategic insights in the form of highly "chunked" heuristics.

II. An Empirical Test of Using Design Versus Operator Logic to Improve System Performance

Between September 1992 and December 1993, we have been engaged in an empirical exercise designed to test the relative power of the designer logic versus operator logic hypotheses in terms of their ability to predict the performance of human decision makers in a complex and dynamic decision-making situation.

The task that these decision makers must face is the management of welfare reform in a large county over a twenty year period. The simulation model upon which this task is based was first published in 1990 (Andersen, Richardson, Lurie, and Ratanawijitrasin 1990) and has subsequently been converted to a four person management laboratory for training MPA students (Maxwell, Andersen, Richardson, and Stewart 1991). The present version of the exercise is a one person lab that runs for a full day. The first half of the day consists of orientation to and training in the simulation and the task. As a part of the first half day, participants are given a training module representing one of three treatments as described below. The second half of the day requires participants to actually manage the system over a twenty year period (after training, this exercise usually takes between one and two hours) and then to complete an extensive set of debriefing exercises designed to measure participants understanding of many aspects of the lab. The details of the design, instruments, and overall approach and results--sketched briefly below--are discussed in more detail elsewhere (Maxwell, Stewart, Richardson, and Andersen 1994).

Research Design. 58 students, including mostly masters candidates in public administration and social welfare at the Rockefeller College, participated in the lab. Five cases were removed because of technical difficulties in model simulation and debriefing, leaving 53 cases for analysis. All participants were paid $40 for their participation and a bonus of an additional $40 was offered for best performance associated with each treatment group. Approximately half of the students (we refer to these as "experts") had been through a two-day management lab focusing on the four player version of the JOBS simulation. They had completed previous readings on the simulation, had worked in groups before coming to the lab, and spent two full days working with the four person model with about half of the two days spent in debriefing discussions. The other half of the students came to the one day lab "cold" and were classified as "novices" with respect to this task.

The novice and expert groups were randomly assigned to one of three treatments. The first treatment, **causal loop training,** gave students detailed information about the causal structures underpinning the simulation. Causal structures were explained in terms of key feedback loops with increasingly complex views of system structure being provided on a HyperCard interface. Students were given complete hard copies of all key system structure during and after their training. They had access to the hard copy during the graded performance of the exercise. This treatment was intended to provide detailed design logic-type of information to participants.

Participants receiving the second treatment were given a summary analysis of five strategies that might be used to manage the system. For each strategy, participants were given full information concerning how the whole system behaved over time in response to the five strategies. Well-crafted graphs of system performance over time were explained to participants through a hypercard interface. In addition, participants were given hard copy of these training modules and allowed to keep and refer to them during practice runs as well as the final graded run. This treatment was referred to as the **strategic time plots training.**

Finally, in the third treatment, participants were given the same five strategies with a summary table of overall system performance for each of the five strategies. In addition, simple rules of thumb for interpreting how each strategy would affect the system were explicitly given out to participants. These participants did not get any detailed causal information nor any information concerning how the whole system behaved over time in response to the five policy packages. We refer to this treatment as the **strategic heuristic treatment.**

After receiving training appropriate to their treatment, all participants were allowed to practice their strategies for 90 minutes. During the practice runs they were free to stop the simulation and restart it at will to practice various strategies. After the practice runs and a lunch break, participants returned to a single graded run of twenty simulated years where they were not allowed to restart. A data collection debriefing then followed.

Performance Measures. The overall goal of the simulation was to minimize the costs of welfare reform. The cost score included actual costs of providing Aid for Families with Dependent Children (AFDC) plus the costs associated with job training minus a bonus for each welfare client taken off the welfare rolls. All scores were normalized so that a normalized score of 100 meant that overall welfare costs were the same as in the initial condition.

A good final score at time period 20 could be as low as 50, indicating that the total costs of welfare had been cut in half by the reform after 20 years. In addition to the final score at time period 20, participants were given information on their 20 year average costs. In general, at the end of 20 years these costs were greater than 100 since the system exhibited "worse before better" behavior requiring large expenditures in the early years in order to achieve substantial savings in the out years.

The performance score that was used to award the prize was a cumulative score (CUMMSCOR) -- a weighted average of the final period score and the 20 year average costs.

Measurements. During the graded run of the task, unobtrusive measures of participant approaches to the task were measured. For example, the number of clicks requesting additional information from other cards in the hypertext stack was recorded along with measures of amount of time spent on task. After the run, a first debriefing module asked a number of open-ended questions to which the participants typed answers directly into a recording system. Next participants returned to the main screen for the game and were asked to click on all fields that they found to be important in playing the game. For each field so highlighted, participants were asked to explain why that field was important. Participants were then led through 56 forced response scales that measured their causal knowledge, knowledge of appropriate strategies, knowledge of system goals (these were given at the start), perception of model training clarity, self-report of management learning, and self-report of previous task-related experience. Finally, participants were administered a standard short form of the Meyers-Briggs personality test to be used in subsequent dissertation research associated with the overall projects. Full details of the data collection procedures and results are given elsewhere (Maxwell, Stewart, Richardson, and Andersen 1994).

III. Preliminary Analysis of the Data and Results

Performance vs. Training and experience with JOBS simulation. For preliminary analysis, the measure of performance described above (CUMMSCOR) was used as the dependent variable in a 3 (training condition) X 2 (level of experience with the JOBS simulation) analysis of variance. The main effect for treatments was statistically significant ($F=3.69$; $df=2,46$; $p < .05$). Post hoc comparisons revealed a statistically significant difference between treatment 3 and the other two treatments, but not between treatments 1 and 2. The main effect for experience and the treatment X experience interaction were not significant. The means and standard errors for the treatment main effect are given in Table 1.
Table 1. Means and standard errors for performance (CUMMSCOR) by training condition

Training condition	Mean	Standard error	N
Causal loop training	105.3	2.96	20
Strategic time plots	104.3	3.12	18
Strategic heuristics	94.9	3.54	14

Performance vs. Knowledge. Three knowledge tests were used to measure goal knowledge (6 items), causal knowledge (18 items) and strategy knowledge (10 items). The score on each test was simply the number of items correctly answered. There were no significant correlations between scores on the knowledge tests, indicating that they did in fact measure different types of knowledge.

Table 2 presents correlations between three measures of performance and scores on the knowledge tests. These correlations indicate that strategy knowledge correlates with the final performance score but not with the average score. Since lower scores indicate better performance, the negative correlations indicate that a higher strategy knowledge score was associated with better performance. Goal and causal knowledge were not related to performance.

Table 2. Correlations between knowledge scores and performance measures

	Performance measure		
Knowledge test score	Average score	Current (final) score	Cumulative (combined) score
Goal Knowledge	.07	-.10	-.03
Causal Knowledge	.18	.14	.18
Strategy Knowledge	-.07	-.36 **	-.28 *

** $p < .01$ * $p < .05$

These scores indicate that only strategy knowledge correlates with performance, but only with the final end point reached, not with the average score. Strategy knowledge is also not correlated with the other knowledge test scores.

Training vs. Knowledge. The mean percent correct for each knowledge test by training conditions is presented in Table 3. Goal knowledge differs little across training conditions (goal training did not differ across conditions). Causal knowledge score is highest for training conditions 1, which focused on causal knowledge. Strategy knowledge was highest in condition 3, which consisted of strategy training. However, condition 2, which also included strategy training, has the lowest strategy score. With the exception of the low strategy knowledge for condition 2, the mean knowledge scores conform to expectations, but none of the differences between treatments is statistically significant.

Table 3. Mean knowledge scores by treatment

	Knowledge scores		
Training Conditions	Goal Knowledge	Causal Knowledge	Strategy Knowledge
Causal loop training	83.3	74.4	58.0
Strategy time plots	80.5	67.6	545.4
Strategy heuristics	86.6	63.9	61.4

Decision process, training, and performance. Two measures related to decision process were measured: the number of buttons pushed, which indicates how much information was requested during the test run and the time (in seconds) taken to complete the test run. Both can be considered indicators related to the use of analytic thought processes which typically take longer and consume more information. The two variables were significantly, but weakly correlated ($r = .32$, $p < .05$), indicating that they do not measure the same construct. It may be that they measure different aspects of analytic/intuitive processes. Table 4 shows the correlations between decision process variables and performance.

The means of the decision process variables by training condition are shown in Table 5. Subjects in condition 1 selected, on average, nearly twice as many buttons as those in condition 3 and took an average of nearly 10 minutes longer to complete the test run. These differences were not statistically significant, however.

Table 4. Correlations between decision process variables and performance.

	Performance measure		
Decision process measure	Average score	Current (final) score	Cumulative (combined) score
Number of buttons selected	.27	.17	.24
Time to complete test run (seconds)	.30*	.14	.23

*$p < .05$

Table 5. Means of decision process measures by training condition

Training condition	Number of buttons selected	Time to complete test run (seconds)
Causal loop training	41.7	2681
Strategy time plots	28.9	2235
Strategy heuristics	20.9	2089

IV. Conclusions

These results are preliminary. We have not yet had time to determine the sensitivity of the results to distributions of the variables and outliers. We plan a detailed structural analysis to examine alternative models to explain the covariance structure of the variables.

The preliminary results are both tantalizing and frustrating. A number of intriguing, non-intuitive relations were found, but all are weak and on the border of statistical significance. Our tentative conclusions are:

1. Experience with the JOBS simulation was not related to performance.

2. Causal knowledge was not related to performance.

3. Greater strategy knowledge was related to improved performance.

4. In comparison with the other groups, the group with causal training (Condition 1) had greater causal knowledge, took longer to complete the test run, requested more information, and did not perform as well.

5. In comparison with the other groups, the group with strategy training (Condition 3) had less causal knowledge, greater strategy knowledge, took less time, requested less information, and performed better.

If these results are confirmed and generalize to other situations, they will indicate a clear superiority of strategy knowledge over causal knowledge for both efficiency and effectiveness in managing dynamic systems.

References

Andersen, D.F., T. Maxwell, G.P. Richardson, and T. Stewart. 1994. Foundations of mental mode research. *Proceedings of the 1994 International System Dynamics Conference.* July 1994, Stirling Scotland.

Andersen, D.F., G.P. Richardson, I. Lurie, and S. Ratanawijitrasin. 1990. Simulation Modeling t Enhance Information System Design: The Case of Welfare Reform. ORSA/TIMS Joint Nationa Meeting. Philadelphia, PA.

Bakken, B. E. 1993. Learning and Transfer of Understanding in Dynamic Decision Environments Ph.D. dissertation, M.I.T., Cambridge, MA.

Brehmer, B. and Joyce, C.R.B., Eds. 1988. *Human Judgment: The Social Judgment Theory View* Amsterdam: North-Holland.

Brehmer, B. 1988. In one word: Not from experience. In B. Brehmer and C.R. B. Joyce (Eds.), *Huma Judgment: The Social Judgment Theory View.* Amsterdam, North-Holland.

Brunswik, E. 1956. *Perception and the Representative Design of Psychological Experiments* (2n edition). Berkeley: University of California Press.

Diehl, E.W. 1992. Effects of Feedback Structure on Dynamic Decision Making. Ph.D. dissertatior M.I.T., Cambridge, MA.

Forrester, Jay W. 1961. *Industrial Dynamics.* M.I.T. Press

Hammond, K.R. 1955. Probabilistic functioning and the clinical method. *Psychological Review* 62: 255 262.

Hammond, K.R., T.R. Stewart, B. Brehmer and D.O. Steinman. 1975. Social Judgment theory. In M.F Kaplan and S. Schwartz (Eds.), *Human Judgment and Decision Processes.* New York: Academi Press.

Maxwell, T.A., D.F. Andersen, G.P. Richardson, and T.R. Stewart. 1991. Learning Through Simulate Experience: Building a Computer-Based Learning Environment to Help Public Managers Implemer the JOBS Program. National Public Management Research Conference. Syracuse Universit Syracuse, NY.

Maxwell, T.A., T.R. Stewart, G.P. Richardson, and D.F. Andersen. 1994. Jobs Management Laborator Technical Report. Unpublished working paper, Center for Policy Research, University at Albany- SUNY. Albany, NY.

Senge, P. 1990. *The Fifth Discipline: The Art & Practice of the Learning Organization.* NY Doubleday.

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

Sterman, J.D. 1994. Learning in and about complex systems. *System Dynamics Review* 10(2-3).

Vennix, J. 1990. *Mental Models and Computer Models: Design and Evaluation of a Computer-Base Learning Environment for Policy-Making.* University of Nijmegen, Netherlands.