SIMULATION GAMES. - DO PEOPLE REALLY LEARN ?

A case study with a System Dynamics approach.

Bent E Bakken
System Dynamics Group
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

ABSTRACT

The use of computers to improve general and task-specific mastery is the core of most computerized learning games. But they are also employed to communicate complex structures. Substantive theories explaining concept formation and learning has been developed by Piaget (1936) and used by Papert (1980) in a computer-specific sense. A recurring assumption in theories of learning is that people who learn from computer games learn in an abstract way, i.e. they learn more than to repeat efficient manipulations of the symbols on the screen.

This study develops a theory of the learning process involved when simulation games are used to reveal the structure of a complex system. Within the framework of a sample of student exposure to a computer simulation game, we find evidence for a theory that games do transmit knowledge of a complex system. In particular, the study indicates that participants Initial Experience, as well as its Relevance to the situation plays a major role in the mastery of a game. Another factor determining understanding of the structure is the use of gradually increased complexity. - A System Dynamics simulation model of the learning process is developed, giving new insights into how simulation models should be built to overcome the dilemmas of game transparency and real world complexity. A sample of 8 students show a close fit to the model predictions. A framework for further behavioral research and the state of the art implications for System Dynamics modelling practice is outlined.

INTRODUCTION

The advent of computers in the 1950's and 1960's and high-resolution color screens have led to a considerable development in computerized simulation games. - On the one hand are the amusement games. They purport to entertain players as well as to provide them with an interesting way to become familiar with computer screens and key-boards. A main marketing argument behind these games is the development of alertness and structural thinking necessary to be a successful player in the game - and in the real world.

On the other side are the learning laboratories where the use of games has a more clearly task of of either helping people to understand a specific issue or to develop intellectual or practical skills. This study is concerned with the latter class of games.

Educational simulators and games have a long and successful tradition, especially in areas where real-life training is inherently difficult, dangerous or expensive. Flight and weapon system simulators are good examples, and they have been in extensive use for the last twenty years. Such simulators are cheap and easy to use compared to the their real-life counterparts. - Not to mention the advantage for the absent passengers in the case of a simulated crash. Business simulators serve the same purpose as flight simulators: To provide an inexpensive and efficient learning environment. Most business simulations are however less task-specific than the typical flight simulator and more concerned with improvement of general concepts, although interactive video has been used for specific skill improvement to make salesmen's task-oriented training more efficient.

The organizationally inclined social psychology tradition, where people play different roles in specific situations and where debriefing is essential in helping the players to understand individual dynamics and their impact on group performance, is in many ways the father of the computerized business game. Instead of teaching a specific task, the game purports to use the specific situation to teach something of general value. This teaching is *assumed* to take place within the interaction between the individual and the game. If successful, the player gains insight both in his own behavior pattern as well as in how the rules of the game make some strategies more effective than others.

The inherent qualities of a business game then make the task of determining success or failure of such games a rather difficult one. Whereas task specific games can be measured along the task mastery dimension, business games should be measured in terms of improvements of the mental models of the players. As specific measures of the mental models are extremely difficult to get at, one must be content with indicative measures. In lack of good variable measurements, we have chosen to interpret improvement in game scores as evidence of a movement towards a more salient mental model of the system the game is portraying.

The concept of mental models in production planning is extensively discussed in Forrester (1961). His work indicates that unfit mental models are present in managers heads at all times and that computer simulations have the double purpose of making explicit these models as well as showing the consequence of their assumptions. Papert (1980) goes further in his work aiming at improving children's mental models. Very much concerned with computer aided learning, he concludes with saying that if given the option to actively *explore* physical assumption the students will unconsciously absorb the system more efficiently than if a student is given only questions and answers to a problem set. He is thus close to the Piagetian (Piaget 1936) development tradition (from which he also comes). This tradition sees childhood development very much as a process where abstraction makes children able to infer from their multidimensional environment into general concepts where (cause and effect) relationships are applicable across a lot of boundaries.[1]

This paper describes an experiment where individuals are exposed to a game portraying, in a non-transparent way, capital over-expansion due to the "multiplicator-accelerator" effect. There are two groups. In the first group players are exposed to a step increase in game complexity. The second group is exposed only to a difficult version of the same game. We also develop a System Dynamics simulation model to condense and simulate our findings. Later we discuss the results. We finally conclude in proposing improvements to the study, using the same game model as a starting point.

THE SIMULATION MODEL

[1]Piaget's findings might explain why traditional System Dynamics is not easily implemented. Human cognitive structures, upon which most knowledge must connect, is very much fixed at age 12. Since the experiences leading to these structures usually are both linear and closely related in time and space, the System Dynamics paradigm with emphasis on nonlinearity and long time lags between cause and behavioral effect, (and often conceived as counterintiutive) makes assimilation of System Dynamics ideas and structures bound to be problematic.

The game model has been developed by Sterman (1985) and is rooted in a phenomenon in economic activity first discussed by Kontradieff (1935). The economic long wave theory claims that economic activities is fluctuating with a period of 40 to 80 years. Sterman's theory describes the fact that capital self-ordering in itself is sufficient to cause such waves, and his game lets people be in charge of the capital ordering decision. The self-ordering mechanism is based upon the fact that to produce machinery necessary for finished goods, you also need *machinery to produce that machinery*. When building up capital, deliveries to the goods sector are restrained, so that a backlog of ordered but not delivered goods is piling up. Likewise, a pipeline of undelivered capital is created. If players are not able to detect this dynamic feature, production overcapacity builds up and cycles with a period of approximately 50 years develop.

Very much deviaung from the neo-classical assumptions of rational expectations, the model has been criticized for lack of realism by economists of that School. But by using concepts from Simon (1983) and describing the decision makers as boundedly rational, Sterman (1986) shows that players consistently crate economic long waves when they play the game and he asserts that people do not understand the underlying dynamics when exposed to the game. He further argues by saying that there is a parallel to the game in the real world, where decision makers' limited capacity for gathering and processing information result in a sustained long waves of economic expansion and retraction.

PROTOCOL

The protocol divided the players into two groups. All plays were individual with no communication within the groups. The first group was exposed to two different goods sector demand patterns, patterns that are exogenous variables in the game.The players did not know the demand pattern until post-game. The first pattern consists in a single, 11 % step increase in orders in year 6. Taught that "When you are comfortable with your understanding of how to play the first game version well, proceed to game version 3", they finally received a cyclical demand pattern. All players were told to write down scores, time used as well as any comments they had.The second group started with version 3, i.e. the cyclical demand pattern, otherwise their instructions were identical to the first group. The participants were graduate students in System Dynamics introductory courses at Sloan School of Management at MIT. They were between 23 and 38 years old, all male. They were randomly put into two groups, 5 i group 1 and 3 in group 2. Since the number of participants was very low the experiment at first might seem unreliable. But because the findings coincide with non-recorded impressions from earlier and later experiments with the game (Sterman 1986), we qualitatively assessed the results as representative for the individuals we have seen go through the test.

The experiment was voluntary, and held at the end of the spring semester. This made us believe that our participants had a more aggressive learning attitude than most MIT/Sloan School Students. That time of year is namely filled with final exams and term papers are due. Especially when considering the only incentives were improved understanding of an economic phenomenon (the Long Wave) as well as snacks, the group was in a sense self-select in its eagerness to learn. We do however expect that this fact was without importance for the results, since the random drawing s secured homogeneity of the 2 groups.

FINDINGS

The problems encountered by the players fell in two categories; The first was to understand the screen and the movements on the display; The second was to figure out what caused undesired behavior to take place, find the underlying causes and then take corrective action. We estimated that the first type of problems was trivial, since all participants had experience in using computers. The second category is what this study is concerned with; How do people form and change their mental models when exposed to a new situation ?

The findings indicates how learning takes place. It seems as if differences in time spent on each game only has insignificant impact on scores. This suggests that people follow their own speed, and that speed variations lead the players to spend as much time as they subjectively feel is necessary on the task. Individuals playing slowly are likely to use the time to actively explore what was happening. The scores, however, indicate that actions involving extensive intellectual effort might as well induce wrong moves as right ones.

A very strong correlation was found between the numbers of games played and the scores was found. This corresponds to the learning curve used in cost assessment of the manufacturing sector of the economy; Peoples' experience lead them to perform tasks more reliably and better as time progresses. Figure 1 shows the score pattern of each individual, and figure 2 portrays the average scores of the two groups as a function of previous game exposure.

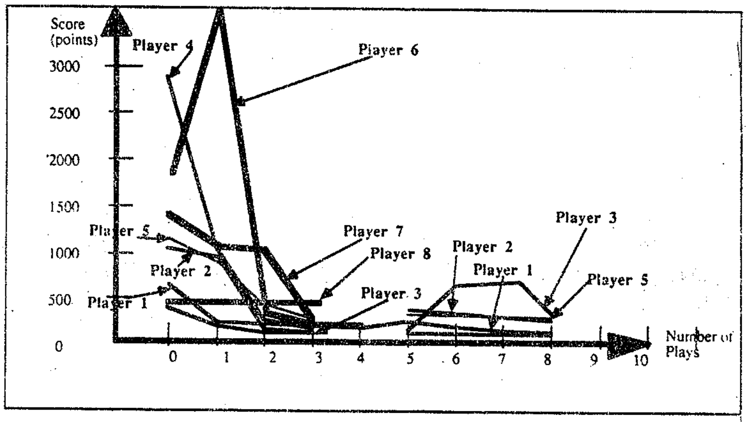


Figure 1: Individual scores

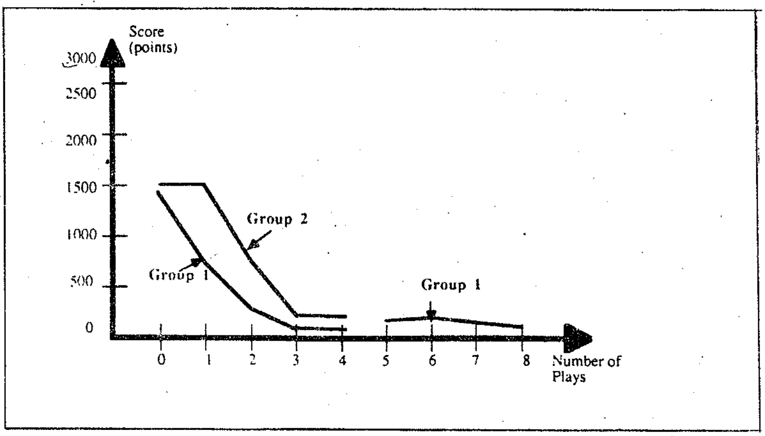


Figure 2: Average group scores

Figure 1 shows that among group 1 members (players 1, 2, 3, 4 and 5), some play the game more times than others before they feel comfortable with it. Because of the difference in numbers of games played, we have "transported" these players to game number 5 when they start up with a new and more complex input data series. So even if a player only played game numbers 0, 1, 2 and 3 he is moved to game number 5 when the second demand pattern occurs. This make evident the jump in scores from the easy to the difficult demand pattern.

Figure 2 averages the data from figure 1. Because of scale impreciseness, we need to clarify the fact that group 1 reaches a mean score of less than 100 after game number 3, whereas the average score for game number 5 jumps up to almost 200. Also worth noting is the fact that average score after game number 8 still is higher than 100 and never reaches its low average value after game number 4.

Also evident from the raw data but less clear on the graphs is that group 2 seems to flatten out with a value of over 200, whereas group 1 (with the same input data) starts game 5 with a score of under 200 and steadily improves (decreases) its score after that.

A MODEL OF THE LEARNING SITUATION

The experiment shows a characteristic learning curve pattern, but also evident is the jump in scores when a new goods order behavior is experienced. The interaction between mental and scoring processes is addressed in figure 3. The model is surely a mentalistic description and would be dismissed by behaviorally oriented psychologists (Skinner 1974), and even to cognitively inclined scholars (Newell and Simon, 1976) the model might seem a little broad and intuitive. The model does however fit quite nicely in the philosophical direction of idealism or mentalism (Sprigge 1984). Figure 4 shows the stock-and flow representation of the same model.

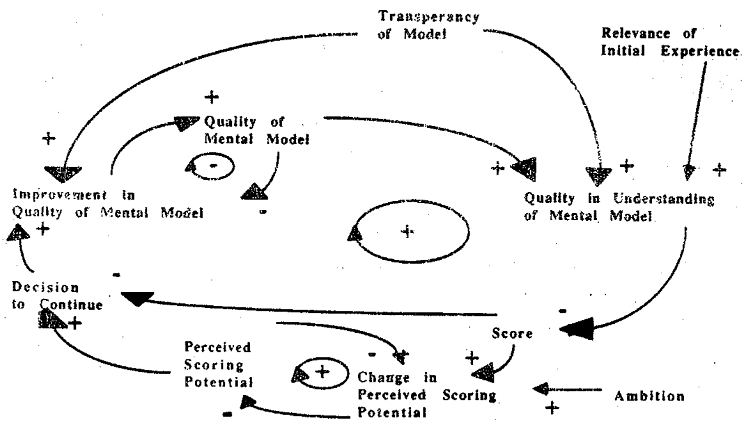
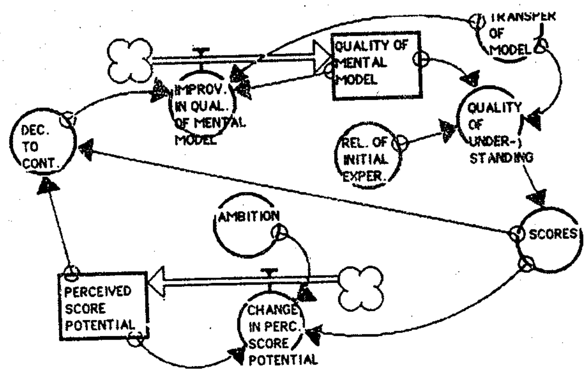


Figure 3: Causal loop diagram of the learning process involved when individuals repeatedly play the STRATEGEM 2 game.



<u>Figure 4: A Stock- and Flow description of the model.</u>

THE LEARNING MODEL'S VARIABLES

Quality of the Mental Model

The formulation of QMM captures the fact that mental models are task related. Here the task was the management of economic long wave dynamics. These models' are, in a Piagetian (1936) sense also built upon previous experiences and abstractive capabilities. As time goes by and a certain mental model is used and exposed to new information, the newfound knowledge is assimilated (Bakken 1987). Since our model only captures the 2 hours necessary to play the game, the decay of the mental model is not included.

The measure of the mental model is not subject to study. We have simply started with a value of 1 (dimensionless) and it increases as repeated experiences improves it. - Arbitrarely, its maximum value is set to 20 - to portray a "perfect" mental model of a given system.

Improvement in the Quality of the Mental Model

The improvement in the Mental Model describes the decreasing returns of the Mental Model on its own improvement. Initially, then, the existence of a mental model is crucial in its own improvement. After exposure to environment, however, the previously built mental model means less and less to its own improvement. We talk about satiation. The improvement is conditioned on the playing of the game. - No play, no improvement. So once the switch Decision to Continue is turned off, then this rate immediately takes the value of zero. The improvement of the Mental Model is also conditioned on the Transparency of Model. The more transparent the model, the faster the improvement of the Mental Model.

Decision To Continue

This switch is a simplified version of the withdrawal dynamics described in the game literature, particularly salient is the Milgram experiment (Richmond 1976). He describes that individuals withdraw from an experiment when they are content with their own performance. In the Milgram case, the crucial point was acceptable pain levels. In ours, the withdrawal depended on the relative difference between actual and perceived obtainable scores.

Scores

The Scores are measured in points, and low Scores indicates good performance. They are supposed to be a direct reflection of the Quality of Understanding of the game Model.

Quality of Understanding of the Model.

The Quality of Understanding of the game Model is a linear function of three variables;
a) The Mental Models's Quality, that is - the better the Mental Model quality, the better the understanding of the computer Model
b) The Relevance of Initial Experience, that is - the more relevant the background of a player, the better is the understanding of the simulation Model.
c) Transparency of simulation Model, since an opaque game is less likely to provide understanding than a transparent game.

Relevance of Initial Experience

This constant enables us to correct for the fact that individuals have different previous exposure to both computers, System Dynamics and long wave theory. - All factors that seem to determine the individuals understanding of the game dynamics. High relevance is portrayed by 20 and low relevance by 1.

Transparency of Model

The transparency determines both the improvement of the mental model as well as the understanding of it. In the model, this constant takes a high value of 20 and a low value of 1. One might argue for a dynamic formulation of this variable where the Quality of the Mental Model improves the Transparency of the game. We have however conceived the parameter as an indication of the physical structure and lay-out of the game so that the difficult version of the game it takes the value of 10, and the easy the value of 5.

Perceived Scoring Potential

The perceived scoring potential is measured in terms of points. It is assumed that two factors contribute to the Perceived scoring potential.
a) The previous scores, modified by a factor for aggressiveness or ambition
b) The difference between current scores and previous scores, so that previous downward jumps in scores modifies the perceived potential more that previous small steps in the same direction. (This formulation corresponds to a simplified version of the trend macro (Richardson and Pugh 1981)).

MODEL RUNS

Group 1 is portrayed in figure 5. The run is characterized by a parametric Transparency of Model that decreases from 10 to 5 after game number 5. The empirical reference mode is reproduced, in particular we see the upward jump in scores and a near stall in Quality of Mental Model improvement after the less transparent demand pattern is introduced. Figure 6 shows a more detailed look into the same situation. These figures are indicative of average group scores. By interviewing the participants, one would be able to determine individual scoring (as well as learning) patterns by making parametric their preparedness and aggressiveness levels. It should also be noted that DT = .25 games which corresponds to a stepwise increment in understanding during the games.

Figures 7 and 8 portray a situation where group 2 is parametrized with an initial Model Transparency of 5. Otherwise this run is identical to the previous ones. Note that final score is above 250, whereas it is below 150 in figure 5.

Note also that our built-in switch for ending the game is not turned on by our simulation. Our model formulation fails to capture the fact that people tend to be satisfied with their scores after 3-5 games with the same input.

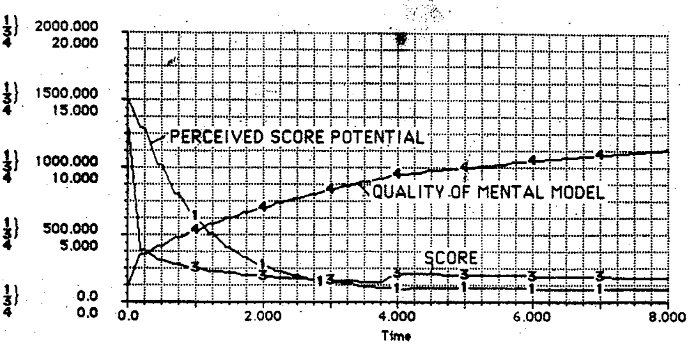


Figure 5: Group 1, entire game

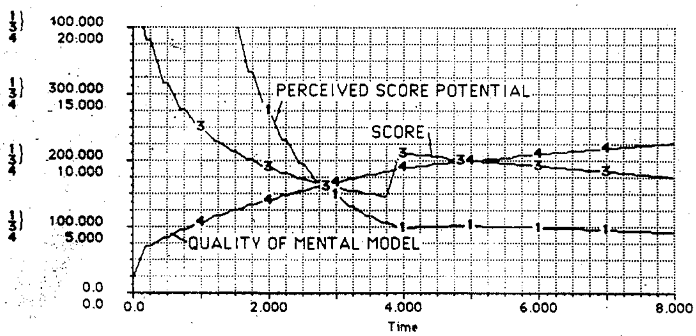


Figure 6: Group 1, a more detailed look.

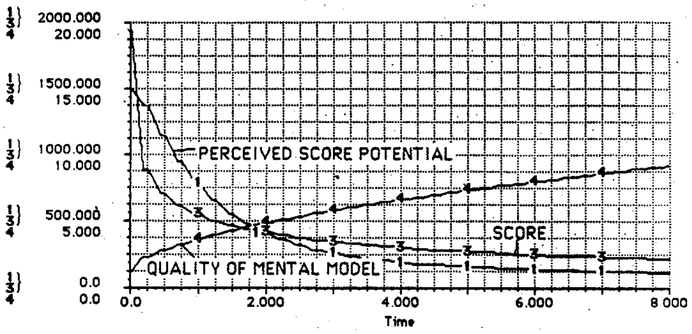


Figure 7: Group 2

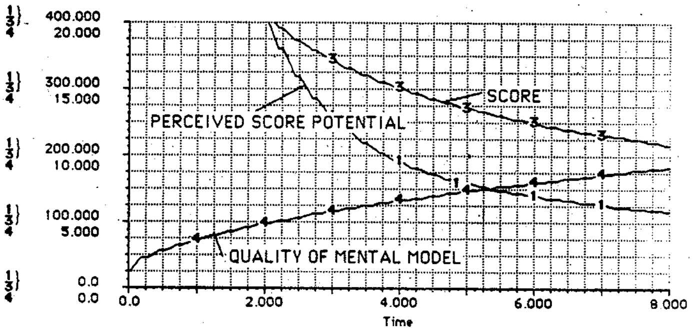


Figure 8: A closer look at Group 2 scores.

COMPARISON BETWEEN THE EXPERIMENT AND THE MODEL RESULTS.

The fit between figure 2 and figures 5 and 7 is very accurate. The closeness between experimental and synthetical results indicates that our theory of the mental process involved in this learning situation is right. However, our sample being small, the statistical conclusion validity is extremely low. Even rude statistical indicators of confidence interval and deviation cannot be used in a reasonable way.

FUTURE RESEARCH

Future research such first be investigatory, i.e. it should consist of similar experimental situations with different Model Transparencies and Initial Relevance. One could conceive of a 3 by 2 matrix where one would have three levels of the 2 mentioned parameters. To enable such a study, and also in order to improve reliability, one would have to increase the experimental subject pool in both quantitative and qualitative terms. Such experiments should be carried out with the same simulation games as provided for in this setting. The larger scale experiment would in addition to fine-tuning the formulation hopefully also enable formulating some of the constants as dynamic variables.

However, this research has a more concrete goal in improving our understanding of how we should direct our efforts in communicating complex structures. A second study should therefore look into what would be learned if the students, were given a verbal description of the malaise and tools to dynamically describe the situation. If our assumption is right, we would see substantial improvements in the Mental Model's quality at the individual level. - Hopefully inadequate conceptions and distorted mental models will not lead to flawed formulations. But the author's experience suggest that the expected linearity and uni-directional causality might lead participants to describe the system without taking into account the circular nature of causality.

REFERENCES

Bakken, B E — "A System Dynamics Approach to Group Formation and productivity", D-memo, System Dynamics Group, MIT 1987

Forrester, J W — "Industrial Dynamics", MIT press, Cambridge 1961

Kontradieff, N D — "The long wave in economic life", Review of Economic Statistics, November 1935

Newell, A and Simon, H — "Computer Science as empirical inquiry; Symbols and Search", Communications of the Association for Computing Machines 19, March 1976

Piaget, J — "Logique et connaissance scientifique"; Gallimard, Paris 1936

Papert, S — "Mindstorms", Basic Books, New York, 1980

Richardson, G P and Pugh, A L — "Introduction to System Dynamics Modelling with DYNAMO", MIT-press, Cambridge, 1981

Richmond, B — "Generalization and individual uniqueness. Modeling the Milgram experiments.", D-2508 System Dynamics Group, MIT 1976

Simon, H — "Models of bounded rationality", MIT-press, Cambridge 1983

Skinner, B F — "About behaviorism", Random House, New York 1974

Sprigge, T S — "Theories of Existence", Penguin Books, London, 1986

Sterman, J D — "STRATEGEM-2", Simulation and games (16)2, 1985

"Testing behavioral simulation models with direct experiment", WP 1752-86, Sloan School, MIT 1986