The Effectiveness of Force Directed Graphs vs. Causal Loop Diagrams: An experimental study

William Schoenberg
Masters Student, University of Bergen
Software Engineer, Forio Business Simulations

Abstract

When it comes to making tough decisions in dynamic environments, decision makers usually do not make the optimal choices (Moxnes, 2004). In order to help decision makers understand the consequences of their decisions modelers usually reveal the structure of their models through Causal Loop Diagrams (CLD). Here I have run a small pilot experiment comparing an alternative method of model structure, model behavior visualization called Force Directed Graphs (FDG) in an attempt to determine which is the more effective aid to decision makers. Participants in this study were asked to make decisions in a dynamic system, and were given either a CLD of the underlying model, or a FDG as an aid. The results of this study were inconclusive as to which was more effective, but it appeared that FDG users had better strategy, but were on the whole unable to translate that into optimal decision making. This paper also discusses changes to be applied to its experimental design before this study can be run in full.

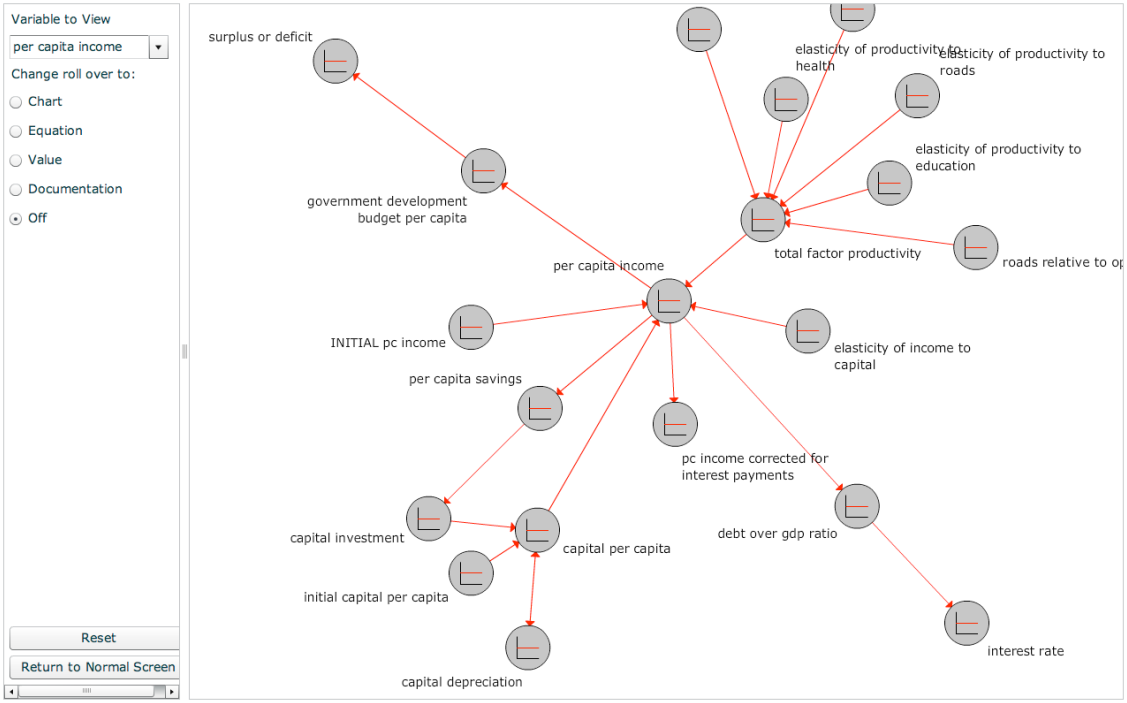
1. Introduction

Much of the general public has a very difficult time making good decisions when faced with a task that involves feedback dynamics (Brehmer 1992, Funke 1991, Jensen 2005, Moxnes 1998; Moxnes 2004, Moxnes and Saysel 2009, Rouwette, Größler et al. 2004, Sterman 1989a, Sterman 1989b, Sterman and Booth Sweeny 2007). This means that many managers of dynamic systems, whether they be supply chains, or ecosystems etc, fail to make good decisions, which would otherwise result in a better state of the system. In order to convey the dynamics of a complex system to the general public,

many system dynamicists use simplified causal loop diagrams (CLDs). Generally these CLDs link model structure to model behavior using pluses and minuses to represent the direction of the relationship. In this paper I will explore the effectiveness of an alternative to the CLD, called a force directed graph (FDG) developed by Forio Business Simulations.

The 'force directed' part of the name force directed graphs refers to the algorithm used to generate the diagram. This means that these diagrams are auto generated for any model, and are not laid out by hand. This algorithm borrows from physics the concepts of static electric charges and springs in order to create a 2D (potentially 3D) representation of nodes and springs which simulates Coulombs law and Hooke's law. A picture of the current version of an FDG is pictured below in figure 1, look to figure 3 for the version of the FDGs as used in this study. A picture of a FDG is shown below (Fig. 1).

Figure 1: *Picture of a Current (3/20/09, after pilot) FDG*



This picture shows a typical FDG. The grey circles represent variables, the red arrows represent the relationships between variables. Users can focus the diagram on a different variable by clicking once, or by selecting a different variable from the dropdown box on the left. The radio button group on the left shows all of the options available for behavior when rolling over a variable. (For a picture of the FDGs used in the study see figure 3)

In Figure 1, the grey circles represent the model variables. The red graph in them show the current behavior of that variable. The red arrows show the relationships with the directionality represented by the arrowhead, and by using colors, green for outflow and blue for inflow, when a user rolls their mouse over any particular variable. In addition, users are also given the ability to change the information displayed to them when they roll their mouse over a variable. Their options are: a chart of the current behavior of the variable (over time), the equation used to derive the variable, the specific value of the variable in the current time, any documentation provided by the model developer, or nothing. The force directed graph then displays either 1 or 2 degrees of causality dependant upon the number of variables that would be put on the screen. To explore the model, users click once on a node to reveal its connections, and hide the connections not directly relevant to it, but it continues to display (for as long as possible) all of the previous variables visited. The other way users may navigate through the model is to pick the variable they wish to see information for from the dropdown list on the left.

This approach to viewing model structure, and model behavior relation is quite different from the typical static CLD, Vensim tracing tree, or iThink story telling mode user interface. FDGs allow the user to explore model structure and current model behavior simultaneously for the entire model while the others generally do not.

For this paper I have performed a preliminary pilot experiment users were divided into 2 treatment groups. Both groups were given the same task to solve involving a dynamic system. The only difference between the two groups was that the first was given a CLD to explain the model structure, and the second was given an FDG to do the same. My hypothesis is that the group given the FDGs will on average make better decisions.

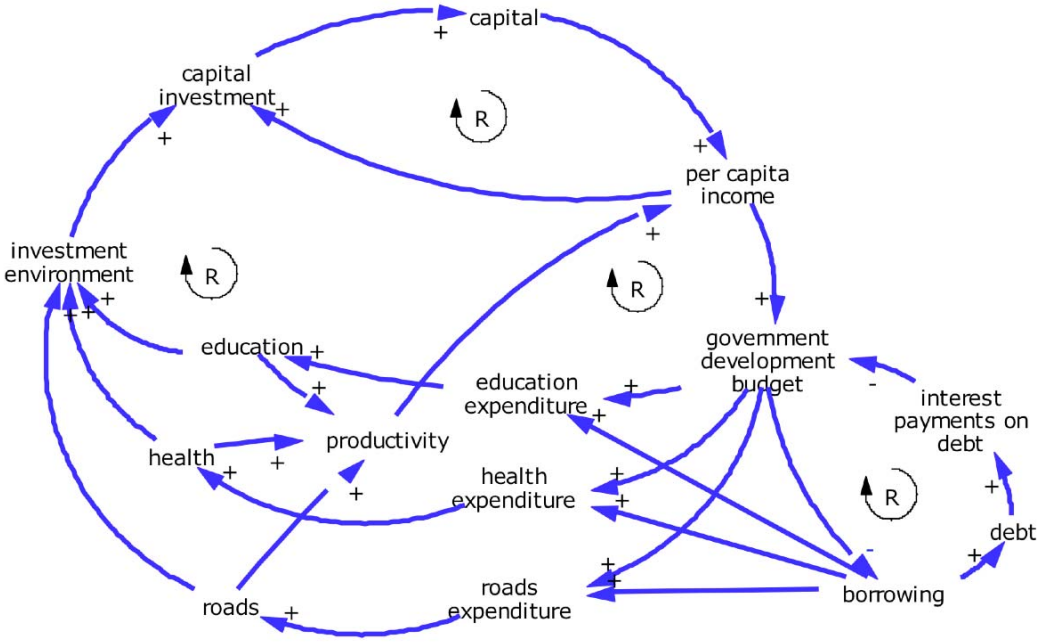
2. Model

Participants in this pilot experiment played a game based on the BLEND model (Bergen Learning Environment for National Development) (Kopainsky et al., forthcoming; Alessi et al. 2008). The original BLEND model has its roots in the T21

model developed by the Millennium Institute (Barney, 2002; Barney & Pedercini, 2003; Pedercini & Barney, accepted; Pedercini, Sanogo, & Camara, 2007;). This game was based on a model that was much smaller and simpler then either the full BLEND model or the T21 model. Birgit Kopainsky of UiB and Matteo Pedercini of the Millennium Institute developed this simplified BLEND model which is based on a on an model used for a cross country analysis that identified the role of a country's resources for its long term economic development (Pedercini 2009). The model depicts the development of per capita income over time as a consequence of a reinforcing process between government development budget and resource accumulation. All variables are calculated on a per capita basis (e.g. capital per capita, debt per capita, and government development budget per capita) so that demographic development does not have to be taken into account. This particular model was chosen because of its level of complexity (5 stocks, and 64 variables) was enough to require the use of a CLD or FDG in order to follow the model structure, but was not so large as to be impossible to understand by someone who was not intimately involved in the development process.

The premise behind the game was that the participant was the newly elected prime minister of a small hypothetical impoverished nation in sub-Saharan Africa. The explicit goal was to raise the level of per capita income in a sustainable manor, without accumulating a crushing amount of debt. Participants were judged on their ability to raise the per capita income minus the per capita interest payments (corrected per capita income). The game unfolded over a 50-year time period from 2010 to 2060 in 5 year increments. Each decision-making period the participant chose how much money per person to allocate for spending in three key areas: transportation, health care, or education. Based off of these three decisions the model was able to generate the corrected per capita income. A CLD of the model is shown below in Figure 2. For a more detailed description of how the model works see (Pedercini 2009, Kopainsky et al. forthcoming)

Figure 2: *Causal Loop Diagram of the BLEND model*



This CLD was the CLD used by participants in the CLD group of this study. It shows the major feedback loops in the model. Participants had control over the education expenditure, the heath expenditure and the roads expenditure.

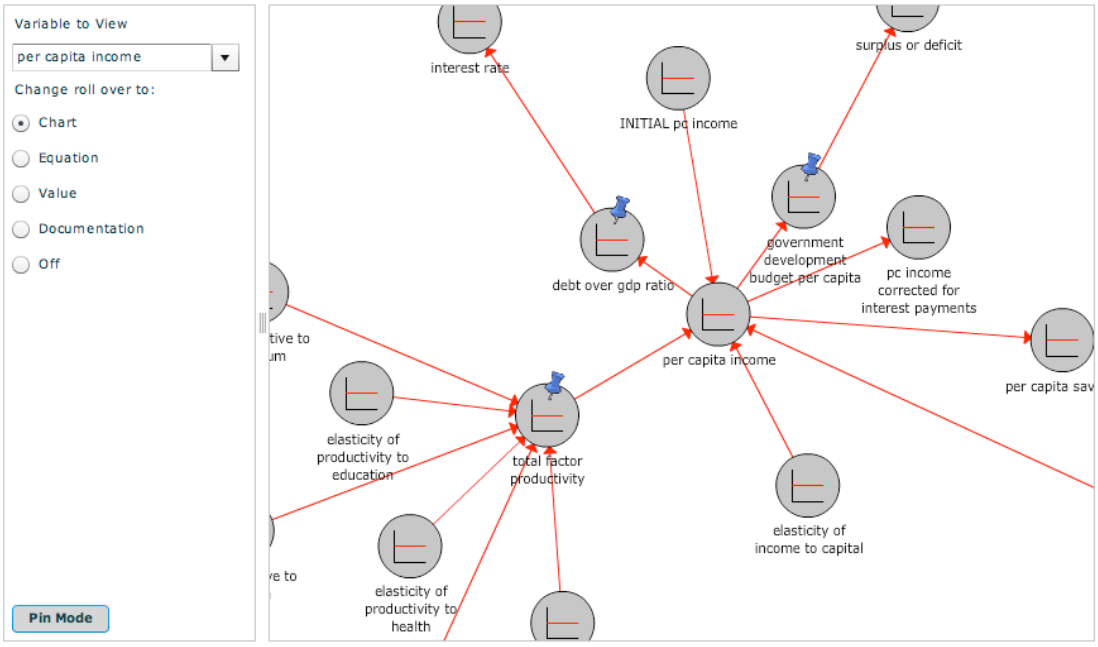
3. Experimental Design

3.1 The UI and the Task

In this pilot experiment participants were broken up into two groups. Both groups would be attempting to solve the same dynamic problem, using the same UI except for one key difference, whether or not they were able to view a CLD or an FDG, when prompted to explore model structure. In order to determine which group a participant belonged to, the participants randomly chose answer sheets to fill in which had written on them an A or a B, which signified which group they belonged to (blank questioner can be seen in appendix A). After selecting an answer sheet participants were directed to the correct URL for their version of the game UI. Once participants logged into the game they had 1 hour to complete the experiment.

After logging in, both treatment groups were brought to the same screen, which gave background information on the task in the form of questions and answers. On this page participants were immediate introduced to their task and given basic instructions on how to accomplish it. Next both groups were then guided into a more in-depth discussion of model structure, and terminology definitions describing concepts such as Revenue, Borrowing, Government Budget, and Debt and Interest. These descriptions were never more then 2 sentences, and were directly related to the model structure, and their task. After this both groups then were given information about the decisions that they had to make, the delay times for those decisions to affect their target indicator (corrected per capita income), and a text description of the model structure which linked each of the decisions to the target indicator. After reading this information participants in the CLD group were shown Figure 2, and those in the FDG group were shown an FDG centered at the uncorrected version of per capita income, (but also showing the corrected version). A screenshot of this FDG can be seen in figure 3.

Figure 3: *Picture of FDG groups starting FDG (From experiment UI).*



The blue push pins are placed by the user to keep variables in the spot specified. They are used here only to make sure the screenshot was able to show a larger diagram in this small picture. (They have been removed from subsequent versions of the FDG, as of 3/20/09 any variable stays where it is dragged, and labels are also force directed away from relationship arrows to avoid overlapping).

Before advancing into the start of the game participants were asked to answer 3 questions in order to determine their understanding of the task, and of delays. After answering these questions both group were then brought to the decision control panel pictured in Figure 4. Once in the decision control panel both groups had a multitude of options. Displayed on the page were: the three sliders for each of their decisions (the development expenditure for roads, health, and education), a table showing their balance sheet for the current year, which contained their development expenditure, their borrowing per capita, their debt per capita, and the Surplus or Deficit per capita, a reminder of their task, the ability to view the results page, the ability to review the previous instructions screens, and the ability to advance the model. After making their decisions participants were then brought to the results page, which showed a tab navigator with 4 tabs, dividing the results into categories. The first category was per capita income, which was a time series graph of the per capita income and the per capita income corrected on the same graph. The second category was investment indicators, which were the three investment indicators relative to their optimum all, displayed on the same graph over time. The third category was budget indicators where, a graph of the debt over GDP ratio, the per capita debt and the interest payments per capita were shown. The fourth category was previous decisions, which showed on one graph the previous spending decisions made by the participant. In addition, from the page participants were given a link to view the model, or make decisions. The make decisions link would bring them back to the decision control panel, and the view model link would bring those from the CLD group to the CLD shown in Figure 2, and for those in the FDG group to a fully explore-able FDG showing the in-progress run data.

After finishing the game for the first time users were asked to answer a second set of questions, much like the first in order to judge their grasp of the task, and delays, in order to determine if they had begun to learn. After answering these questions users were then allowed to run the game as many times as they desired of the rest of the hour in order to get the best score possible. After the 1 hour was up, participants then answer the same series of questions, and were asked to mark down how many times they had run the game. All user data was stored, but just the first run, and best run were collected and analyzed.

3.2 Benchmark

For this game the best strategy to take was to invest early in education, because it had the longest delay time, and to invest in roads as well in order to create short-term economic growth to help offset the costs of the investment into education. It was essential to not allow the debt to GDP ratio to rise above a maximum of 2. After 15 years of investment into education and roads only, health care would be in need of investment, so gradually participants should have increased their investment into health care by substituting out education investments. After 10 years of this, a budget surplus would begin to occur and spending could be increased in all three categories allowing participants to heavily increase the per capita income, without having the negative effects of debt and interest payments. A good score has the corrected per capita income over $700, and the best ever done using the aforementioned strategy was approximately $1,500 in corrected per capita income.

This strategy could be determined by studying the information given about the length of the delays, how each of the investment categories affected the rest of the model, and a basic understanding of budget portion of the model. All of this information was available to both treatment groups.

3.3 Hypothesis

Standard theory would suggest that CLDs are a very effective tool used to help people understand dynamic systems, in order to make better decisions. But it is my hypothesis that FDGs can do a better job (as compared to a CLD) with a model of limited complexity (less then 20 stocks and 100 variables), as the simplified BLEND model is.

Therefore my main hypothesis can be formally stated as:

H_0: *Participants will be able to achieve a higher corrected per capita income with a CLD available for consultation.*

H_1: Participants will be able to achieve a higher corrected per capita income with a FDG available for consultation.

I believe that the FDGs will allow participants to perform better because participants will be much better able to explore the model and learn at their own pace. Participants will also be better able to understand where behavior originates because participants will be able to track changes in behavior overtime, and throughout the causal structure simultaneously.

Secondly, I believe that having an FDG available (rather then a CLD) will allow participants to have a better chance of understanding model behavior, and a better decision making strategy on their first run, as compared to CLDs. This hypothesis can be formally stated as:

H_0: Participants will be able to achieve a higher corrected per capita income on their first run with a CLD available for consultation.

H_1: Participants will be able to achieve a higher corrected per capita income on their first run with a FDG available for consultation.

This ought to be the case because participants using the FDGs will have more information available to them about model structure then those who are using the CLD. Therefore they should be better able to trace the information given to them in the instructions, and put it to better use quicker.

3.4 Other design issues

The experiment took place from 6pm-7pm on a Thursday evening. All participants had no previous knowledge of the game or any training in system dynamics, but they all knew each other. Participants were both male and female, all well educated ranging in age from mid twenties, to mid thirties. In order to fulfill Vernon Smith's precepts, participants were motivated to perform well by the promise of a free dinner upon completion of the task, and the participant with the highest corrected per capita income

was announced after the experiment and applauded by all. Since the participants all knew each other, this was a reward in and of itself.

4. Results

In this pilot study there were 6 participants in total, 4 who were in the FDG group and 2 who were in the CLD group. The mean corrected per capita income for the CLD group on their first run was $286.71, while FDG group was -$1292.41. For the best run the CLD group mean was $968.39 versus $965.56 for the FDG group (see Table 1). The standard deviation from all four means was too high for these results to be statistically significant with an alpha of .05. Both cases failed to show any difference between the mean of the CLD and FDG groups. The two tailed P – values for these tests can be seen in Table 2.

Table 1: *Basic statistics for all four groups, comparing final corrected per capita income.*

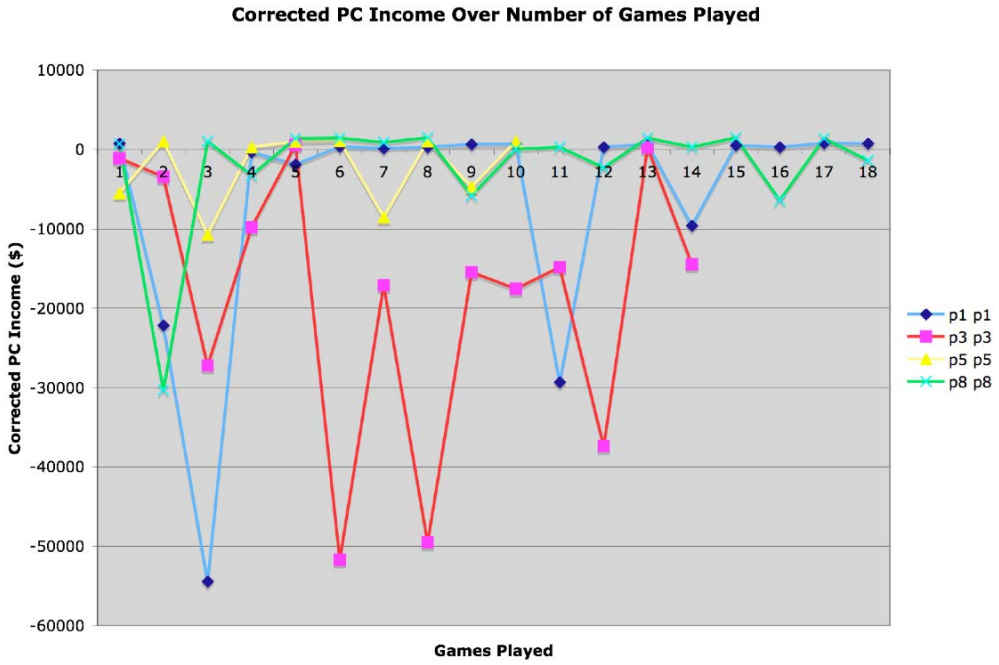
Group	Sample Size	Mean	Standard Deviation
CLD Run 1	2	$286.71	$280.47
FDG Run 1	4	-$1,292.41	$2,912.30
CLD Best Run	2	$968.39	$95.60
FDG Best Run	4	$956.56	$392.79

Table 2: *Results of two tailed equal variance P-test*

Test	P-Value	Pass-Fail
First run different	.5103	Fail
Best run different	.9929	Fail

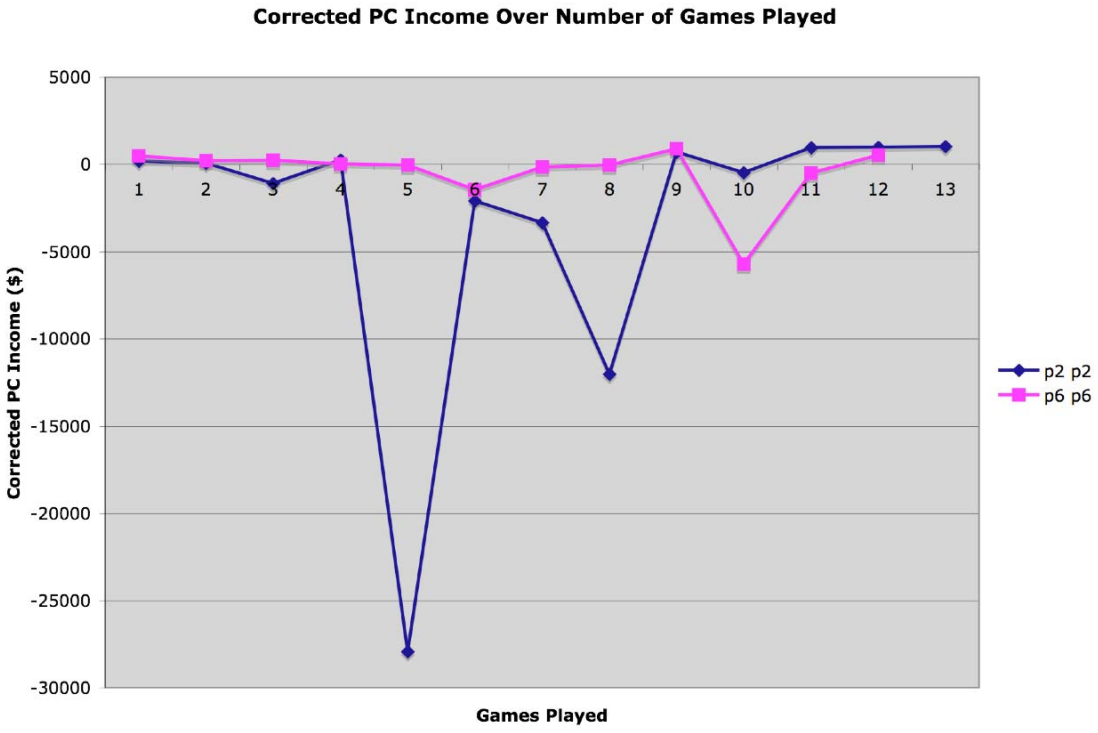
The development of each teams scores can be seen in figures 5 and 6. Three of the four FDG participants did worse on their second run then their first. It took approximately 4-6 runs for participants in this group to start getting results where the corrected per capita income was above $500 (figure 5). On the other hand the CLD participants fared about the same on their first 4-6 runs, and had a much smaller range of results. On average they also ran fewer runs then the FDG group.

Figure 5 : *Participant Behavior over time FDG group*



The graph above shows the final value of the corrected per capita income achieved by each participant in the FDG group. On the X-Axis is the number of the run, and on the Y-Axis is the indicator.

Figure 6: *Participant Behavior over time CLD group*



The graph above shows the final value of the corrected per capita income achieved by each participant in the CLD group. On the X-Axis is the number of the run, and on the Y-Axis is the indicator.

5. Discussion

The results of this pilot experiment were inconclusive; statistically there were no differences between the mean corrected per capita income for the CLD group vs. the FDG group for either their first or best runs. Since this is a pilot, these results were expected.

After reviewing the answers to the questions posed in questioner (appendix A), some differences between the two groups emerged. Before playing the game, but after viewing the diagram, users were asked to record their strategy. Neither participant in the CLD group was able to describe any part of the strategy revealed in section 3.2. The first and only time either participant in the CLD group described a workable strategy was after the first run where Participant 2 stated, "I was thinking about prioritizing long term investments like education, I also tried to not go into debt at all costs". But, when asked the same question after their final run the participant stated, "My priorities were mostly determined through trial and error."

In contrast, 2 of the 4 participants in the FDG group had a strategy close to that in section 3.4. Participant 1 stated,

> "Go into debt on all fronts, but focus on roads first, education less so, and health third. Then lessen roads and up education in the latter (part of) the first third. Then in the last third allow steady healthcare to continue up, but let education fall and road continue to fall." * anything in () inserted for clarity

This shows a good understanding of the investment priorities, and potentially of the delays, but where this strategy is fatally flawed is in its treatment of debt. When asked about what delays were, right after the question about strategy Participant 1 had this to say,

> "Kind of – I wanted immediate payoff from roads, and then want to surge education later to leverage the roads payoff. Education will take a while but become critical later"

This clearly shows that Participant 1 had some knowledge of delays, but this cannot be compared to a pre-FDG response since none was taken. This is an area that must be addressed in the experimental design before expanding the size of this study. In addition to Participant 1, Participant 8 had a very similar strategy, except focused more on early short-term gains through investment in transportation, in order to increase the amount of funding available for later in the game.

One of the most interesting results to notice in the mix of qualitative and quantitative data is that it took 2 runs for Participant 8 to turn their strategy knowledge into good results, while Participants 1, and 2 were never able to so. Participants 1 and 2 seem to either have an inability to correctly follow their own strategy, or they are too easily discouraged from following their strategy when their first implementation fails because of overspending early on (an integral part of Participant 1's stated strategy). Participant 8, who was able to turn their strategy into results was able to do so with the best results overall with a max corrected per capita income of $1,469.12, versus Participant 1 who resorted to trail and error and ended up with a corrected per capita income of $778.46 and Participant 2 who did the same and ended up with a max value of $1,037.54.

The first major flaw revealed through the survey questions and informal comments after the experiment was that participants tended to under utilize the View Model link on the results page. Users preferred to look at the model structure only a few times in the hour. Most participants mentioned using only one or two of the graphs shown in the results page along with a trial and error strategy as evidenced by Participant 2, from the CLD group who was quoted above espousing their use of trial and error, whereas in earlier questions this participant cited specific investment strategies. This behavior was not endemic to the CLD group either. Participant 1 of the FDG group (the same participant whose good strategy was quoted above) stated in their response to the final

question about strategy with the remark, “After randomly banging variables this way and that, I saw that the…” These responses then explain the extreme variation in corrected per capita income over the number of runs performed (figures 5 and 6).

In order to address these issues, the structure of the experiment must change. First, participants must only be allowed to run the model in 3 stages. The first stage will be a guided exploration, the second stage will be the first un-guided run, and the third stage will be a second guided run. The reasons for doing this is to avoid the trial and error behavior witnessed above, when users are given a large number of chances to perform (Diehl & Sterman 1995, Jensen & Brehmer 2003, Moxnes 1998; 2004, Paich & Sterman 1993, Sawicka et al 2005, Sawicka & Rydzak 2007). This will allow for more confidence in the results of any future trials. The reasoning for including 3 stages is so that participants are given some chance to learn over the course of the experiment, which in preliminary studies by Kopainsky shows that participants who have a chance to explore the model perform better then those who do not.

In order to have a better benchmark for pre-experiment knowledge so that learning can be better tracked, I will ask 4 sets of questions. The first set of questions will come before seeing the diagram, but after reading the instructions, the second set after seeing the diagram, but before exploring, the third set after the first run, and the fourth set after the second run.

In order to address the second issue, where participants under utilize the model structure diagram, the diagrams need to be positioned more prominently. For the second iteration of this experiment, I will have the screen where the diagram is shown be the first screen participants see after advancing the simulation, and I will force users to look at the diagram by locking it on the screen for a period of time. My hope is that this will force participants to take it into account when making decisions, by forcing them to actively look away from the computer if they do not want to use the diagram. I believe the major reason the diagrams were underutilized is because they were an extra click away, hidden by easier to digest, and smaller chunks of results.

Another technique I plan on introducing into the next round of testing is the use of video and sound to guide participants through their respective diagrams, and to explain the instructions. I believe that this will make the diagrams more engaging so that people will pay more attention to them, and it will prevent people from skipping or skimming through the instructions. I also believe that video is a better medium for instructional material then typed text, which I hope will help to increase participants understanding of the problem, and hopefully that will influence their performance.

6. References

Alessi, S. M., B. Kopainsky, P. I. Davidsen and M. Pedercini (2008). A system dynamics-based multi-user domain for improving national development planning. Annual Meeting of the American Educational Research Association. New York City, NY.

Barney, G. O. (2002). The Global 2000 report to the President and the Threshold 21 model: Influences of Dana Meadows and system dynamics. *System Dynamics Review, 18*(2), 123-136.

Barney, G. O., & Pedercini, M. (2003). *Models for national planning*. Paper presented at the 22nd International Conference of the System Dynamics Society.

Brehmer, B. (1992). "Dynamic decision making: Human control of complex systems." Acta Psychologica **81**: 211-241.

Diehl E, Sterman JD. 1995. Effects of Feedback Complexity on Dynamic Decision Making. *Organizational Behavior and Human Decision Processes* 62:198-215

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. Complex problem solving: principles and mechanisms. R. Sternberg and P. Frensch. Hillsdale, NJ, Lawrence Erlbaum**:** 185-222.

Jensen E, Brehmer B. 2003. Understanding and control of a simple dynamic system. *System Dynamics Review* 19:119-37

Jensen, E. (2005). "Learning and transfer from a simple dynamic system." Scandinavian Journal of Psychology **46**(2): 119-131.

Kopainsky, B., M. Pedercini, S. M. Alessi and P. I. Davidsen (forthcoming). "A blend of planning and learning: Simplifying a simulation model of national development." simulation & Gaming.

Moxnes E. 1998. Overexploitation of renewable resources: The role of misperceptions. *Journal of Economic Behavior & Organization* 37:107-27

Moxnes, E. (1998). "Not only the tragedy of the commons: Misperceptions of bioeconomics." Management Science **44**(9): 1234-1248.

Moxnes, E. (2004). "Misperceptions of basic dynamics: the case of renewable resource management." System Dynamics Review **20**(2): 139-162.

Moxnes, E. and A. K. Saysel (2009). "Misperceptions of global climate change: information policies." Climatic Change **93**(1): 15-37.

Paich M, Sterman JD. 1993. Boom, Bust, and Failures to Learn in Experimental Markets. In *Management Science*, p. 1439: INFORMS: Institute for Operations Research

Pedercini, M., & Barney, G. O. (accepted). Dynamic analysis of Millennium Development Goals (MDG) interventions: The Ghana case study. Accepted for publication in *Socio-Economic Planning Sciences*.

Pedercini, M., Sanogo, S., & Camara, K. (2007). *Threshold21 Mali: System dynamics based national development planning*. Paper presented at the 25th International Conference of the System Dynamics Society.

Pedercini, M. (2009). Modeling resource-based growth for development policy analysis. Department of Geography. Bergen, University of Bergen. **PhD**.

Rouwette, E. A. J. A., A. Größler and J. A. M. Vennix (2004). "Exploring influencing factors on rationality: a literature review of dynamic decision-making studies in system dynamics." Systems Research and Behavioral Science **21**(4): 351-370.

Sawicka A, Gonzalez JJ, Qian Y. 2005. Managing CSIRT Capacity as a Renewable Resource Management Challenge: An Experimental Study. *23rd International Conference of the System Dynamics Society*. Boston, GA, USA

Sawicka A, Rydzak F. 2007. Incorporating delays in the decision-making interface: An experimental study. *25th International Conference of the System Dynamics Society*. Boston, Ma

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. and L. Booth Sweeny (2007). "Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter." Climatic Change: 213-238.

Appendix A
National Development Game Survey Questions

Please fill in while playing the National Development Game.

Directions: You have 1 hour to play an game where you are given the job as the Prime Minister of Blendia. You will read all about your task when you play the game. You are allowed to play as many times as you want during this one hour. As with any study of this nature your results are strictly confidential and anonymized.

Question set 1: (To be answered before making any decisions, you will be referred to these questions from the interface)

1. How will you approach this problem? Please elaborate on what your investment priorities will be, and how will you determine those priories.

2.
(a) Do you believe that you understand the concept of delays?
(b) Describe your understanding in the simplest terms?

Question set 2: (To be answered after playing the first time)

1. What was your strategy? Please elaborate on what your investment priorities were, and how those priories were determined.

2. (Leave blank if same as above)
(a) Do you believe that you understand the concept of delays?
(b) Describe your understanding in the simplest terms?

3. Do you think you could do better if you play again?

Question set 3: (To be answered after playing for the last time)

Total number of times played: _____

1. What was your strategy? Please elaborate on what your investment priorities were, and how those priories were determined. Did your strategy change over time? How? Did it help? How many times did you have to play before you got a result you were happy with?

2. (Leave blank if same as above)
(a) Do you believe that you understand the concept of delays?
(b) Describe your understanding in the simplest terms?