How best to present model structure to decision makers? An experimental study of the effectiveness of the Forio model explorer

William Schoenberg

Forio Online Simulations
333 Bryant Street
Suite 370
San Francisco CA 94107
415-440-7500 ext 87
bschoenberg@forio.com

Abstract

Making decisions in a complex dynamic environment is very difficult, and often decision makers do not make optimal choices (Moxnes, 2004). Typically system dynamicists attempt to reveal model structure to decision makers so that they can better understand the system and make better choices. For this study I test the effectiveness of a tool called the model explorer vs. a stock and flow diagram to see which is more effective at helping decision makers make better choices in complex dynamic environments. The results of this study were inconclusive. The model explorer participants succeeded at the task 33% of the time vs. 12.5% of the time for the stock and flow participants. There were not enough overall participants for this result to be statistically significant. Also in this experiment I devise a method for testing for type 1 error due to trial and error and discover that it is quite effective at catching participants whose results are based on guessing rather then understanding of the dynamics of the task.

1. INTRODUCTION

Many decision makers have a difficult time making decisions in complex dynamic systems (Sterman 1989; Moxnes, 2004; Jensen 2005). As a result many decision makers make sub-optimal decisions when faced with complex dynamic systems. Since system dynamicists generally work on complex dynamic problems we have come up with a series of tools and techniques to help decision makers perform better at these

tasks (Spector, 2000). The most popular of the techniques in use make use of either causal loop diagrams (CLDs), or stock and flow diagrams (SFDs) (Alessi, 2000). Both of these diagrams help to simplify and explain the causal structure underlying the complex dynamic systems that decision makers work in everyday through the use of arrows to represent causal structure (both CLDs and SFDs), and symbols to represent the stocks and flows within a system (SFDs only). In this paper I test the effectiveness of a new technique to help decision makers perform better in complex dynamic systems, using a tool called the model explorer created by Forio Online Simulations, versus that of a SFD.

The model explorer is a relatively new tool. Construction started in 2008, and it is still being improved today. The model explorer takes a system dynamics model and computer generates diagrams consisting of bubbles and arrows to represent the causal structure of the model (Fig 1). The model explorer allows the user to explore simultaneously the causal structure of a system dynamics model as well as the output of that model. The transparency that the model explorer provides into the model is intended to help its users better understand the simulation it is displaying (Größler et al. 2000). The user can view model structure in one of two ways. The first way for users to explore model structure is to enter a single variable and see its nearest inputs and outputs. The second form of interaction is by having the user input a starting and an ending variable to see all the interconnections between the two.
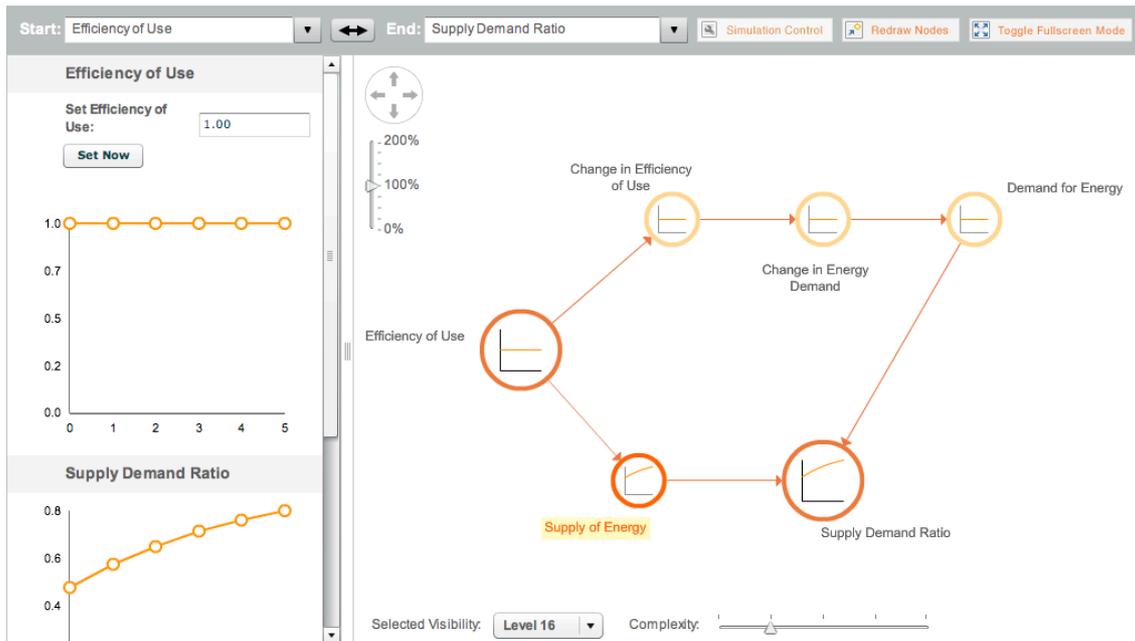
Figure 1: The Forio model explorer. In this figure it is displaying a diagram from the flubnium task. It is showing the second level of complexity for the diagram of Efficiency of Use to Supply Demand Ratio

When the user only enters a single variable the model explorer displays all of the inputs to, and outputs of that variable. If all of the inputs and outputs of that variable is less then 20 it will then display the next layer of inputs (for those variables which influence the original variable) and the next layer of outputs (for those variables which are influenced by the original variable). What this means is that only the most closely related portion of model structure is displayed when viewing any single variable, which keeps diagram size small so that information is digestible. When in single variable mode users can click on any other variable on the screen to generate a new single variable diagram of the inputs and outputs of the newly selected variable so that it is relatively easy to navigate through a causal chain in any direction.

The second mode of interaction in the model explorer is using A to B mode, where users are asked to enter in two variables. The first variable is the starting location and the second is the ending location. The model explorer then generates a diagram of all the causal links between the start and end variables. Users are also allowed to enter in the same variable for A as for B, which auto generates a diagram of all loops involving that variable. As you may very well imagine for large models, that may have dozens of

variables separating the staring location from the ending location, the diagrams produced can be quite large and overwhelming. Therefore there is a 'complexity' slider, which as you move it from left to right adds further and further distant (indirect) links to the diagram, or if you slide it from right to left, removes the indirect links in preference to the most direct links. In this way only the most direct links are shown in order to keep diagram complexity lower, making its message much more clearer to users (Sawicka and Molkenthin, 2004). Also, it is possible to scrub through the full range of complexity in any diagram to see the whole picture. The final feature of note in this mode is that a user may select any one variable on the path between A and B and the Model explorer will highlight all of the causal chains that have this new variable (C) in their path. This feature allows users to easily visualize how many causal chains flow from A to B through C, or for when A and B are the same, how many loops pass through C.

The approach of the model explorer to helping decision makers understand complex dynamic systems is a continuation of the current approach of diagramming and drawing pictures of the complexity in order to increase its understand-ability to decision makers (Larkin and Simon, 1987). The model explorer adds the ability to auto-generate the diagrams, view model behaviour and structure at the same time, as well as the ability to view just the desired slices of model structure. For these reasons I believe that it is an enhancement of the current techniques for explaining and conceptualizing complex dynamic systems.

For this study I have performed a experiment where users were divided into two treatment groups. Both groups were given the same decision making task, the only difference being that first group was given a model explorer as their visual aid for making decisions and the second group was given a stock and flow diagram. My hypothesis is that model explorer group will be better able to accomplish the task relative to the SFD group.

2. EXPERIMENTAL DESIGN

2.1 The Task

In order to investigate the effectiveness of the model explorer I created a task that the participants of the experiment had to perform. This task needed to fulfil the following requirements:

1. Be representative of a complex dynamic system where there are decisions that need to be made by people.
2. Be simple enough to solve without knowing anything about system dynamics or having any specialized training or education.
3. Be difficult enough to solve so that the answer could not be guessed by anyone negating the need for model structure visualization.
4. Be easily re-parameterized so that it is easy to confirm false positives by having participant who gets it right once, run it again with a new set of assumptions.

The task I created for use in this study places the user in the role of a government regulator for a new energy source called flubnium. The job of the participant is to regulate the efficiency of the conversion from flubnium (the raw material) into energy. The goal of the participant is to balance the demand for energy with the supply of flubnium.

The task is based on the theory behind Jevons paradox which states that as the efficiency of an energy source is increased the demand for the energy produced from that source rises, and does not fall as would be expected (Jevons 1906). This satisfies my first requirement that the task be based on a complex dynamic problem. I used a made up material called flubnium in order to help people concentrate on the facts and figures of the simulation as opposed to relying on outside experience or feelings to make their decisions.

In order to keep the task simple enough to be solved without knowing anything about system dynamics, or having training in energy markets; I kept the model small. The model was just 14 equations (including constants), 2 stocks and 1 feedback loop that was connected through the user (Appendix A). This had the benefit of making the answer to the task calculate-able using pencil and paper, using only addition,

subtraction, multiplication and division. Parameterizations of the task were chosen that kept all key values and answers as whole numbers so that they could be easily calculated by hand. This ease of calculation means this task meets my 2nd requirement.

In order to make sure that the task was not too easy, and that the answer could not be easily guessed by just looking at the instructions again Jevons paradox was used. In this manner changing the efficiency of use (the only participant controlled parameter) affected both the supply of energy and the demand for energy (see section 3.1 for model equations, and Appendix A for full model). This non-intuitiveness ensured that the answer was not too easily achievable, which means it satisfies my 3rd requirement.

Finally the model behind this task is very easily re-parameterized so that it can be re-used to test if the participant guessed the right answer or if they actually understood the relationships and can reproduce the correct answer under a different set of assumptions. This satisfies my 4th and final requirement for the task.

2.2 The Experimental Procedure

The experiment was run online over the course of three days, from March 10th 2011 at 5pm PST to Sunday March 13th 2011 at 8am PST. Participants were hired from Craigslist, using an ad placed in the gigs / computer gigs section, and paid $20 for their participation. Participants were told that they were to be playing an online experiment that would take no longer then 1 hour to complete and that all results would be anonymous. Participants were sent a URL to the simulation along with unique user names and passwords and were able to login at their leisure anytime during the experimental period. The experimental user interface was designed according to human computer interface design principals described by Howie et. al. in order to increase the chance that participants would succeed in the task (2000).

Participants were randomly assigned to two treatment groups based off of the order that they responded to the ad. The first user was placed in the treatment group that received a model explorer as a visual aid, the second user in the treatment group that received a SFD. This pattern was kept up for all 17 participants. There were 9 in the model explorer group and, 8 in the SFD group.

Once logged in, participants were presented with the full case information (Fig 2, Appendix B). The case information page revealed to the user the back-story that they were to confine their decision making to. This page told them what their role in the simulation was, what their goal in the simulation was, and the parameterization of the case which they could use to mathematically solve to achieve the goal. The information given to the user was the following:

1. The current supply of flubnium.
2. The current demand for energy.
3. The time to close the gap between supply and demand.
4. The additional demand created by a doubling of efficiency.
5. The initial efficiency of use.
6. The number of rounds to complete the task.

Figure 2: The full case information for the flubnium task. Notice how it warns participants to focus only on model results, not on past experience. This page lists out all info needed to solve the case.

After viewing the case information, participants were then brought to a page called 'Technical Instructions' which told the participant what was expected of them (Fig 3,

Full text Appendix C). This page told the participant that the experiment would take 1 hour to complete, that they would have 2 chances to complete it, that there would be a pre-simulation as well as post-simulation survey, and most importantly that they were going to be presented with a model structure visualization which was going to reveal to them important information that was necessary to get the correct answer.

Figure 3: Technical Instructions for the simulation. This page tells participants what will be expected of them over the next hour.

After viewing the technical instructions users were then presented with a pre-simulation survey that asked the following questions:

1. Describe what your goal in the simulation is?
2. Have you ever heard of a concept called Jevons Paradox? If so please describe your understanding of it.
3. Do you have any interest in, or experience with energy markets?
4. How many years of post-secondary (College or University) education do you have?
5. What is your age?
6. What is your sex?
7. Describe your strategy for reaching the goal, and why.

8. Have you ever played a simulation game before?

These questions were used to test age, sex and education level as a reason for experimental differences, as well as to gauge the participants level of understanding of the task, its goal, and its setting.

After answering the pre-simulation survey the participants were then brought to the decision dashboard based on which treatment group they were in (Fig 4, Fig 5). On the page from top to bottom, left to right, was the following:

1. A label saying "Decision Dashboard"
2. Text showing the current value of Energy Supply and Energy Demand
3. Text showing the game number, round number, and participant identifier
4. Model visualization (which took up the majority of the screen real estate)
5. Instructions on how to use the visualization
6. A button leading to 4 line charts of results labelled "To Results"
7. A button leading to the instructions labelled "To Instructions"
8. A text input for entering the Efficiency of Use decision
9. A button which advanced the simulation labelled "Advance"
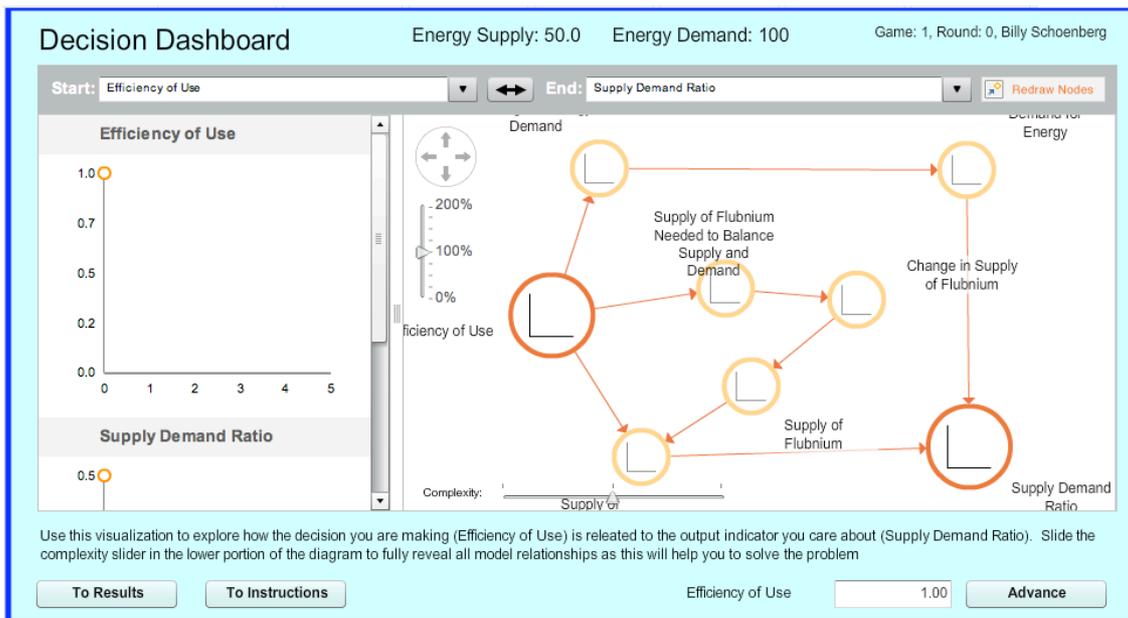
Figure 4: This figure shows the main decision screen for participants in the model explorer group. Notice how the model explorer is the most prominent item on the page.

Decision Dashboard   Energy Supply: 50.0   Energy Demand: 100   Game: 1, Round: 0, Billy Schoenberg
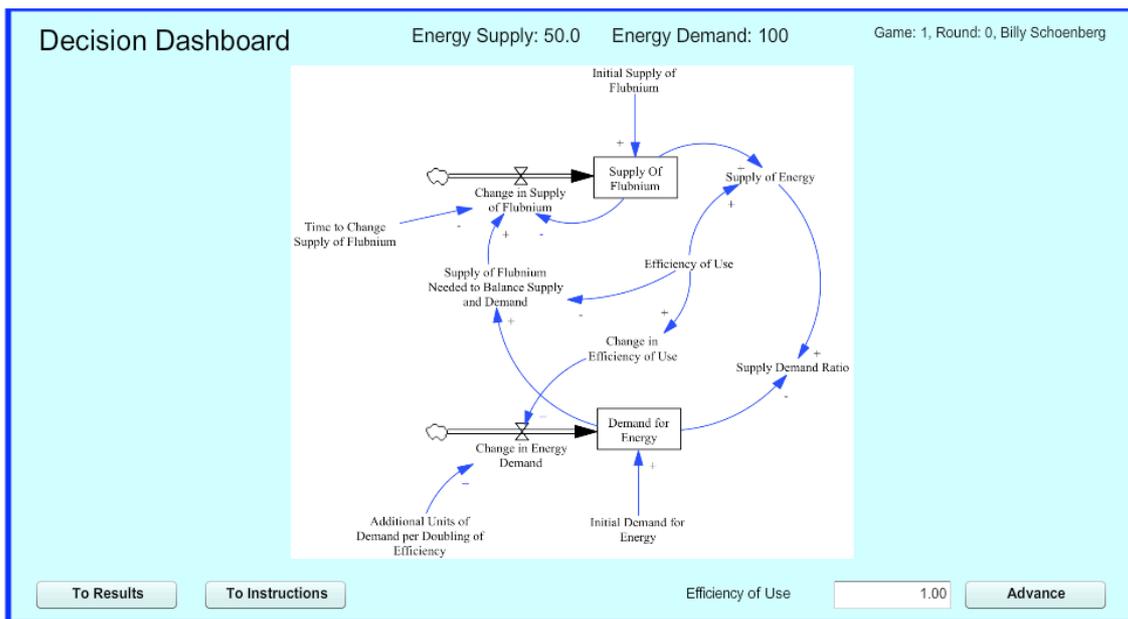
Figure 5: This figure shows the main decision screen for participants in the SFD group. Notice how the SFD is the most prominent item on the page.

After completing a game participants were brought to a post-simulation survey that was designed to gauge their understanding of what they had just done, any strategy changes they were going to make, as well as gauging the usefulness of the visualization in their solution of the task. The questions are as follows:

1. Do you think you reached your goal this game? Explain.
2. Did the strategy you described for meeting your goal work for you? Why or why not?
3. Did the model visualization help you when you were playing the game? Why or why not?
4. Would you change your strategy next game? If so describe the new strategy and why

After playing the game a second time participants were then brought to the same post-simulation survey. If they were unable to solve the task the second time they played they were thanked for their participation and logged out of the simulation.

If the participant solved the task correctly they were brought to a page congratulating them for getting the correct answer, and presenting them with the same case with a new parameterization (Fig. 6). Participants then played the game as they did in the first and second round and were again asked to fill out the post simulation survey, and then they were thanked for their efforts and logged out. The reason for the third round was to verify that participants could replicate the correct decision making skills in a re-parameterized environment to cut down on the rate of type 1 error from trial and error/ luck.
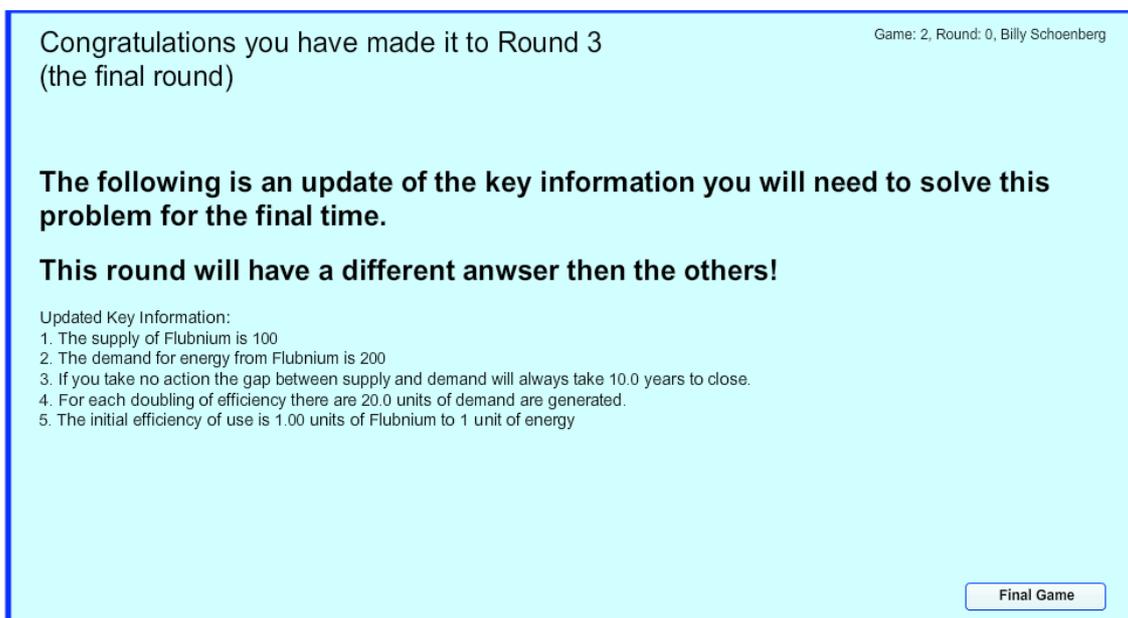

Figure 6: The screen shown on entering the third and final round. Notice how the user is told that this round will have a different parameterization.

2.2 The parameterizations and correct responses

Both treatment groups were subjected to the same series of parameterizations. The following table describes the simulation setup.

Table 1: Parameterizations of the flubnium task used during for this study by round.

Variable	Rounds 1 and 2	Round 3
Initial Supply of Flubnium	50	100
Initial Demand for Energy	100	200
Time to close the gap	5	10
Additional Demand per doubling of efficiency	50	20
Initial Efficiency of Use	1	1
Number of Rounds	5	5

The participants using the following set of formulas could calculate the correct value for efficiency of use:

1. Supply Gap = Demand for Energy – Supply of Energy
2. Supply of Energy (t + 1) = ((Supply Gap / Time to close gap) + Supply of Energy) * Efficiency of Use
3. Demand for Energy (t + 1) = Demand for Energy + (Additional Demand * (Efficiency of Use – 1))

For rounds 1 and 2 the correct answer was an efficiency of 5 if entered in the first time step, and for round 3 it was 2. If an incorrect value were entered the equations above would have to be calculated on the values generated in order to come up with the correct value. A correct answer could always be achieved by entering in the value of efficiency of use produced using the above series of equations.

For the purposes of this study participants will be judged as having a correct answer if they are able to at the end of the simulation in rounds 2 and 3 have supply and demand be no further then 5 units apart. If participants only achieve this in round 1 or 2, their correct answer is regarded as a false positive and ignored.

2.3. Hypotheses

Since the model explorer is a continuation of the ideas behind the stock and flow diagram I believe that the model explorer group will perform better on the flubnium task as compared to those with the stock and flow diagram

Therefore my main hypothesis can be formally stated as:

1. *Let PSSFD be the proportion of successful participants in the SFD group to the total number of SFD participants.*

2. *Let PSME be the proportion of successful participants in the model explorer to the total number of model explorer participants.*

H_0: *PSSFD will be greater then or equal to PSME using a significance level of .05.*

H_1: *PSSFD will be less then PSME using a significance level of .05.*

In addition, I believe that the model explorer will be more helpful to those that succeed in the flubnium task as compared to those who succeed using the stock and flow diagram.

This hypothesis can be formally stated as:

1. *Let PHRSSFD be the proportion of responses that the visualization was helpful by successful participants of the SFD group to the total number of responses to the visualization question by SFD group participants.*

2. *Let PHRSME be the proportion of responses that the visualization was helpful by successful participants of the model explorer group to the total number of responses to the visualization question by model explorer group participants.*

H_0: *PHRSSFD will be greater then or equal to PHRSME using a significance level of .05.*

H_1: *PHRSSFD will be less then PHRSME using a significance level of .05.*

Finally, I believe that those who fail in the flubnium task in either treatment group will find the model visualization to be less helpful then those who succeed in either treatment group.

My final hypothesis can be formally stated as:

1. *Let PNHRF be the proportion of responses that the visualization was not helpful all participants who were not successful to the total number of responses to the visualization question by all participants who were not successful.*

2. *Let PNHRS be the proportion of responses that the visualization was not helpful by all participants who were successful to the total number of responses to the visualization question by all participants who were successful.*

H_0: *PNHRF will be greater then or equal to PNHRS using a significance level of .05.*

H_1: *PNHRF will be less then PNHRS using a significance level of .05.*

3. The Model

The Flubnium task is based on a simple 14 equation, 2 stock model (Appendix A). The two stocks are the demand for energy, and the supply of flubnium. The supply of flubnium closes its gap with demand based on the variable time to change supply.

3.1 The equations for the supply of flubnium

1. Supply of Flubnium = STOCK(Change in Supply of Flubnium, Initial Supply of Flubnium)
2. Change in Supply of Flubnium = (Supply of Flubnium Needed to Balance Supply and Demand - Supply of Flubnium) / Time to Change Supply of Flubnium

3. Supply of Flubnium Needed to Balance Supply and Demand = Demand for Energy / Efficiency of Use

The supply of flubnium is a stock that is modified by the change in supply of flubnium. The change in supply of flubnium is calculated as the goal (supply of flubnium needed to balance supply and demand) minus the gap (supply of flubnium) divided by the time to change the supply of flubnium. This means that the system (by itself) will never close the entire gap between supply and demand. The goal for supply is calculated as the demand for energy divided by the efficiency of use (participant decision). This equation translates the current demand for energy back into units of energy producing material.

3.2 The equations for the demand for energy

1. Demand for Energy = ACCUM(Change in Energy Demand, Initial Demand for Energy)
2. Change in Energy Demand = Change in Efficiency of Use * Additional Units of Demand per Doubling of Efficiency
3. Change in Efficiency of Use = Efficiency of Use - PREVIOUS(Efficiency of Use, Initial Efficiency of Use)

The demand for energy is a special kind of stock called an ACCUM in the Forio modelling language. An ACCUM works just like a stock accumulating the values of its first parameter (change in energy demand), using the second parameter (initial demand for energy) as its initial value, but the key difference is that an ACCUM does not wait until after the time-step to add the value of the first parameter to the accumulation like a stock does. The reason an ACCUM was used was so that a change in efficiency of use by a participant would reflect a change in the value of demand for energy in the next year, rather then two years in the future. This helped to close the gap between action from the participant (in the form of entering a value) and the reaction of the system so that participants were more easily able to tell the consequences of their decisions.

The change in energy demand was calculated by multiplying the change in efficiency of use by the additional units of demand per doubling of efficiency. This is where the crux of Jevons paradox is produced. Each time the efficiency of use is changed, there is also a change in the demand for energy based on the additional units of demand parameter.

The change in the efficiency of use is simply calculated as the current user entered value of efficiency of use subtracted by the previously entered value of efficiency of use. The PREVIOUS function takes two arguments; the variable to return the previous value of, and the initial value of the variable so that the function may return a correct value in the first time-step. In this model the initial efficiency of use is always 1, meaning one unit of flubnium yields 1 unit of energy.

3.3 Equations for the key indicators

The key indicators in this model are the supply demand ratio and an output called correct which reports whether or not the participant has correctly balanced supply and demand.

1. Supply Demand Ratio = Supply of Energy / Demand for Energy
2. Correct = IF(ABS(Supply of Energy – Demand for Energy) > 5, 0 ,1)

The supply demand ratio is simply calculated as the ratio of supply of energy to the demand for energy. The output correct is returns a 1 if the user has balanced supply and demand, and 0 if they have not.

4. RESULTS

Out of the 9 participants in the model explorer group 3 were able to balance supply and demand to within 5 units in both rounds 2 and 3. For the SFD group only 1 out of 8 was able to do the same. Figures 7, 8 and 9 show how the model explorer participants performed relative to the optimum, which is bolded and highlighted in red. As you can there was one false positive in the model explorer group (M. User 4) but the

other three participants who reached round 3 very closely matched the optimal solution (Fig. 9).
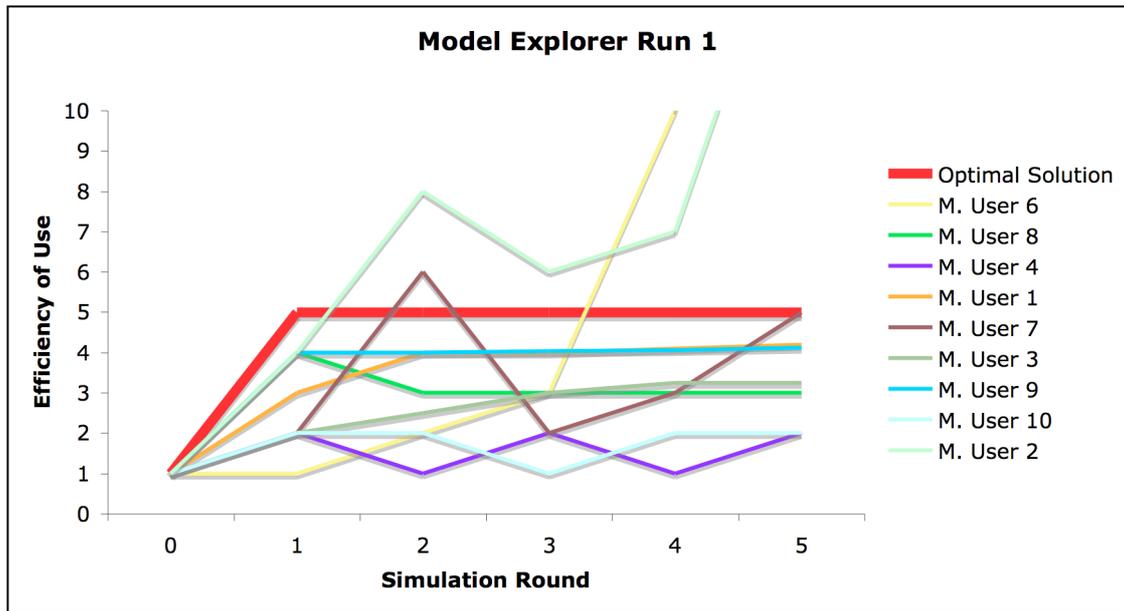
Figure 7: Results from the model explorer group round 1 relative to optimum.
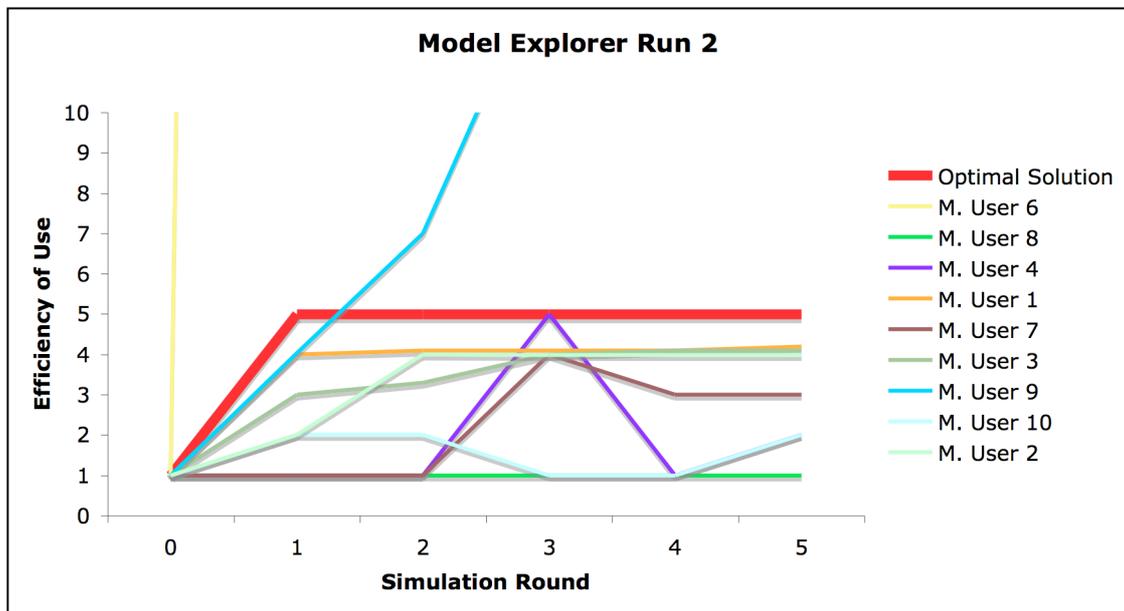
Figure 8: Results from the model explorer group round 2 relative to optimum
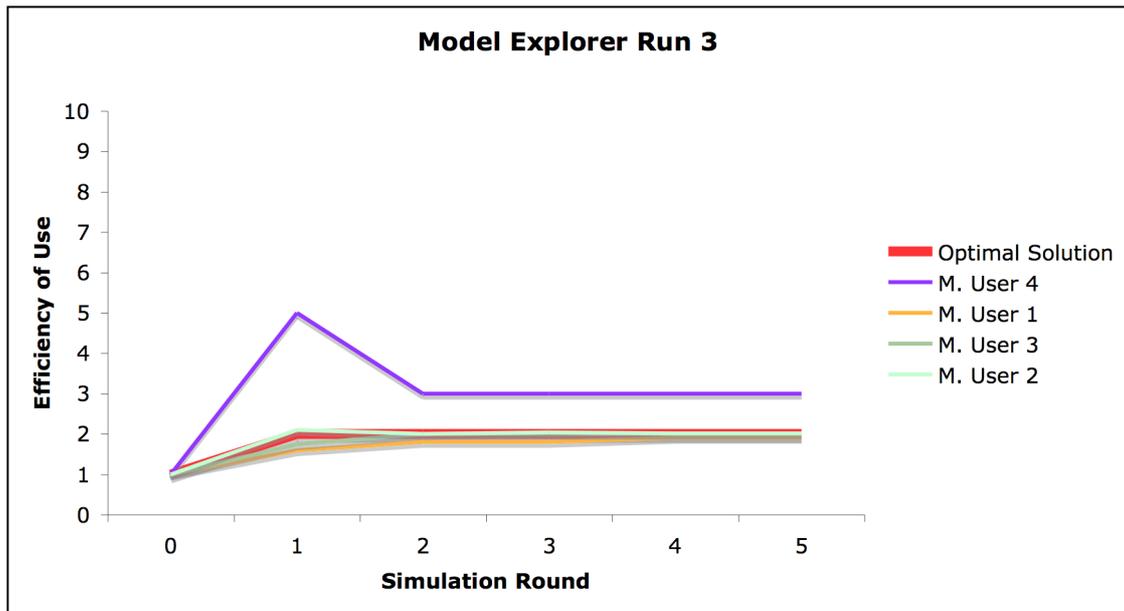
Figure 9: Results from the model explorer group round 3 relative to optimum

For the SFD group there were 3 users who reached round 3, but two of them were unable to replicate their results in the final round, and were therefore considered false positives, and not marked as having solved the task correctly. Figures 10, 11, and 12 show the decisions made by the SFD group relative to the optimum, which is bolded and highlighted in red.
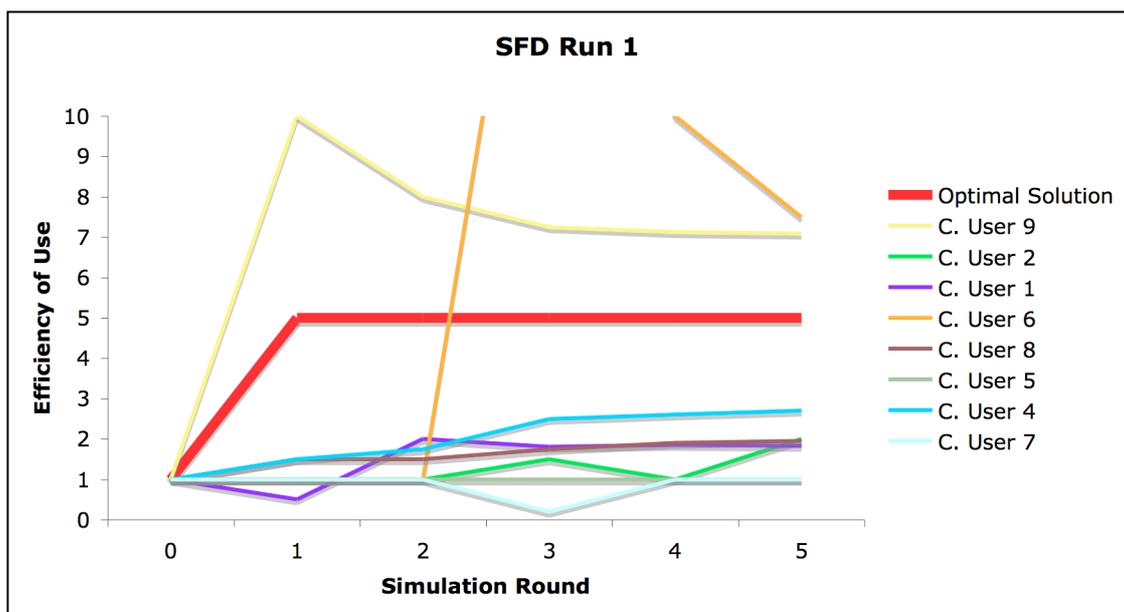
Figure 10: Results from SFD group round 1 relative to optimum
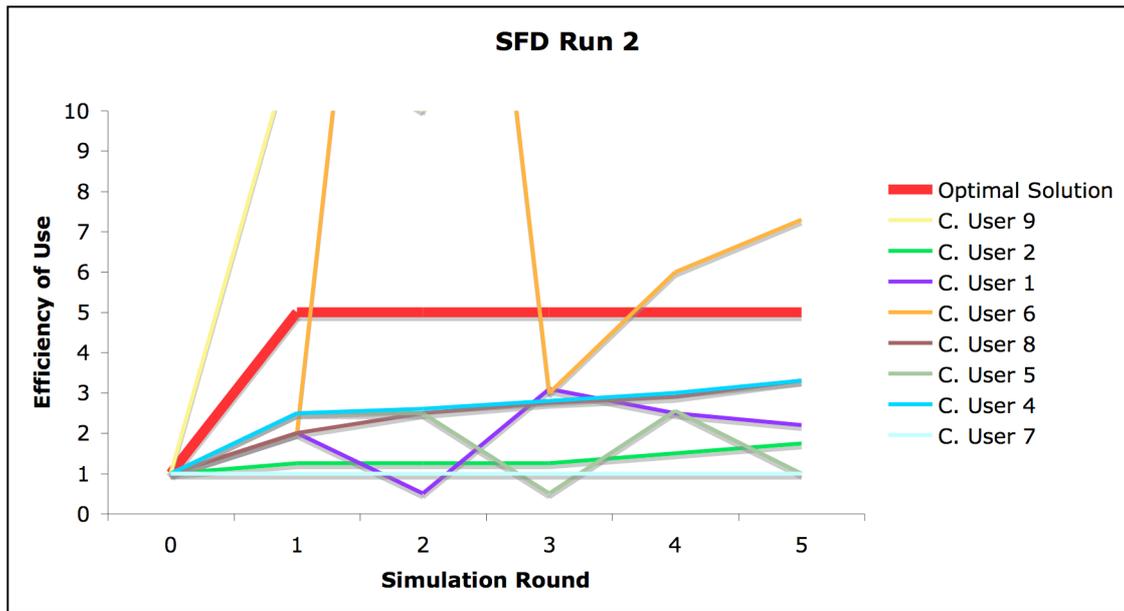
Figure 11: Results from SFD group round 2 relative to optimum
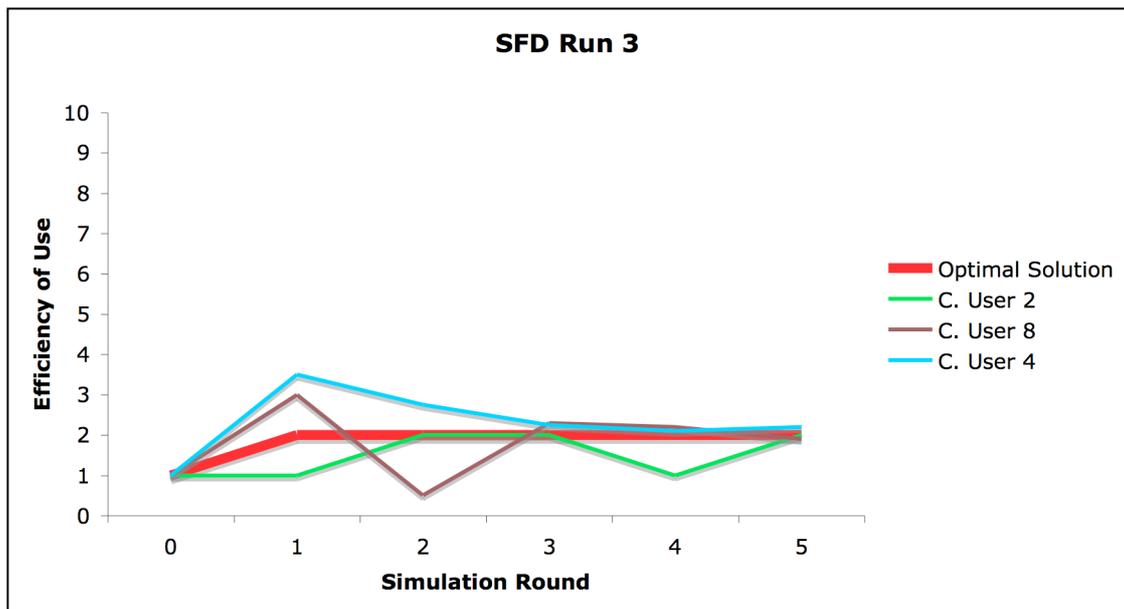
Figure 12: Results from SFD group round 3 relative to optimum

Out of the 17 total participants who performed the task, 4 got correct answers, and 13 were incorrect. The average chance of a model explorer participant succeeding was 1 out of 3, while for the SFD group the chance was 1 out of 8 (Table 2). Using Fisher's exact test for calculating the Chi squared on a 2x2 contingency table the first hypothesis

yielded a one-tailed p-value of .3353, which fails to pass the .05 significance level, required to reject the null hypothesis.

Table 2: Success by treatment group. Proportions of each treatment group are in parenthesis

Treatment Group	Correct Answer	Incorrect Answer
SFD	1 (12.5%)	7 (87.5%)
Model Explorer	3 (33.3%)	6 (66.6%)

Of those in the model explorer group who would go on to succeed in the task there were 2 responses that the model explorer was helpful, and 7 responses that it was not. This is in contrast to those who would go on to fail in the task where there were 5 responses that said the model explorer was helpful, and 8 that said it was not. (Table 3)

*Table 3: Visualization question responses by those in the model explorer group. In parenthesis is the percent of total.*

	Useful	Not Useful	Total
Fail	5 (38.46%)	8 (61.54%)	13
Success	2 (22.22%)	7 (77.78%)	9
Combined	7 (31.82%)	15 (68.18%)	22

The one participant in the SFD group who would go on to succeed in the task never said that the SFD was helpful, and gave 3 responses that it was not. This matches those who would go on to fail in the task where there was 1 response that said the SFD was helpful, and 15 that said it was not (Table 4).

Table 4: Visualization question responses by those in the SFD group. In parenthesis is the percent of total.

	Useful	Not Useful	Total
Fail	1 (6.25%)	15 (93.75%)	16

Success	0 (0.00%)	3 (100.00%)	3
Combined	1 (5.26%)	18 (94.74%)	19

Table 5: Visualization question responses by those in all groups. In parenthesis is the percent of total.

	Useful	**Not Useful**	**Total**
Fail	6 (20.69%)	23 (79.31%)	29
Success	2 (16.67%)	10 (83.33%)	12
Combined	8 (19.51%)	33 (80.49%)	41

Testing hypotheses 2 and 3 using Fisher's exact test for calculating the Chi squared on a 2x2 contingency table yielded a one-tailed p-value of .5455 for hypothesis 2 and .5691 for hypothesis 3 which both fail to pass the .05 significance level required to reject the null hypothesises.

5. DISCUSSION

The results of this experiment were statistically inconclusive. Statistically the successful proportion of the model explorer participants was not higher then the successful proportion of the SFD group (P value .3533). In addition, there was not a statistically significant result that there was a higher proportion of responses that the model visualization was helpful among those who succeeded in the model explorer group versus the portion of responses that the model visualization was helpful by those in the SFD group (P value .5455). Finally, there was not a higher proportion of responses that the model visualization was not useful among the population of both groups who failed versus the proportion of the population of both groups that was successful (P value .5691). Since the number of participants was relatively small these results are not unexpected. In order to get statistically significant results for my main hypothesis the sample size would have to be approximately quadrupled assuming the ratios of success in the model explorer group and SFD group were held constant.

The most interesting result that came out of this experiment was the observation that out of the 7 participants from both groups that made it to round 3 only 4 were

successful in performing the task again with a different parameterization. This means that if there was no test for type 1 error there would have been an additional 75% percent of participants who succeeded. This would have changed the outcome of this experiment, in this case making the ratio of successful SFD participants higher then the ratio of the successful model explorer participants. This is important because many experimental studies conducted by others in a similar manner fail to test for Type 1 error, which as we can see here has a large effect on experimental outcomes.

A strong example of a participant in this experiment who was clearly applying a strategy they did not understand to solving the re-parameterized flubnium task was M. User 4. It is clear from their first decision in round 3 to set the efficiency of use to 5 (the correct strategy from rounds 1 and 2) that they were replaying their trial and error success from round 2 to round 3. M. User 4 tried to make a correction to that strategy in time 2 of round 3, but it was not enough (Figure 9).

The implications of being able to test for type 1 error in an experimental study like this are very important. Testing for type 1 error successfully would allow experimental designs to be created which allow participants to learn how the experiment works by performing the tasks that the experiment asks without introducing error from those participants who will figure out the correct response by trial and error. This is important because some people learn best through experimentation, and experiments where participants cannot experiment themselves put these types of people at a disadvantage (Bakken et al., 1992). While this does not eliminate the effects of trial and error on experimental outcomes, it does allow experimental designers to try to control for it. Type 1 error from trial and error based strategies can still affect experimental results because some participants may guess the right answer twice in a row. Also a re-parameterization should not fundamentally change the correct strategy in an experiment, which means that during the portion of the experiment when participants can explore the model they may come up with strategies that they find work but do not understand why they work. As long as those participants are still be able to successfully apply those trial and error strategies to a new parameterization they will be counted as being successful. Since the goal of many experiments is to test if participants were able to learn a concept and understand why it occurs the successful

responses generated by those who have found the correct strategy but do not know why it works would still be considered type 1 error and they are not captured by the re-parameterization test for type 1 error.

A second interesting result was the relatively low proportion of participants who reported that the model visualization was helpful, especially in the stock and flow group where after only 1 of the 18 runs completed was the SFD referred to as helpful. In the model explorer group there were 7 responses that the model explorer was helpful out of a total of 22 runs. The combined results for both groups was that after only 19.51% of the runs was the model visualization reported to be useful. This is a very low percentage if you consider that of those who were successful at the task, only 16.67% of their responses were that the visualization was helpful. This means that the model visualization was said to be more helpful to those who ended up failing the task (20.69% of all responses) (Tables 3,4,5).

These results may have implications on how model structure is presented to those unfamiliar with system dynamics. It reinforces the findings that facilitation is still needed to help decision makers understand model structure. With more instruction on the model visualizations participants may have been better able to utilize them, which may have affected the outcome of this experiment, but it is unclear if that can make up for the lack of facilitation especially related to teaching how to use the model visualizations.

6. CONCLUSIONS

The purpose of this experiment was to determine if the Forio model explorer was a more effective tool then a traditional stock and flow diagram to help decision makers perform better at complex dynamic experiments. To test this idea I built an online simulation, which was run by 17 participants. I found that with my sample size the model explorer was not more effective then the SFD. I also found that approximately 43% of would be successes were actually false positives. Finally I found a surprisingly low proportion of participants found the model visualization tools available to them as useful in solving the task, even more so among those who succeeded.

Any future testing done using experimental methods like mine, should attempt to test for type 1 error in order to rule out false successes by those using trial and error. When re-testing the hypothesis of this experiment a larger sample size is needed, at least 68 participants so there is a chance for statistically significant results to occur. Also any further experimentation should focus on how best to present the model visualization in a non-facilitated environment so that there is the largest chance that it is understood and found useful by the participants.

7. REFERENCES

Alessi, S. (2000) "Designing Educational Support in System-Dynamics-Based Interactive Learning Environments" Simulation & Gaming **31**(2): 178-196.

Bakken, B., Gould, J., and Kim, D. (1992) "Experimentation in learning organizations: A management flight simulator approach" European Journal of Operational Research **59**(1):157-182

Größler, A., Maier F.H., and Milling P.M. (2000) "Enhancing learning capabilities by providing transparency in business simulators." Simulation & Gaming **31**(2):257-278

Howie, E., Sy, S., Ford, L., and Vicente, K.J. (2000) "Human-computer interface design can reduce misperceptions of feedback." System Dynamics Review **16**(3):151-171

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

Jevons, W.S. (1906) "The Coal Question An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of our Coal-Mines" Macmillian and co. London England.

Larkin, J.H., and Simon, H.A. (1987) "Why a diagram is (sometimes) worth ten thousand words." Cognitive Science **11**:65-99.

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

Sawicka, A., Molkenthin, R. (2004) "Cognitive load dynamics: How to increase effectiveness of SD-based learning environments." 23rd International Conference of the System Dynamics Society, Boston, MA.

Spector, J. M. (2000). "System Dynamics and Interactive Learning Environments: Lessons Learned and Implications for the Future" Simulation & Gaming **31**(2): 528-535

Sterman, J. D. (1989). "Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment." Management Science **35**(3): 321-339.

APPENDIX

A: The Full Model

M ExecuteDecisionImmediately = false
M StartTime = 0
M EndTime = 5
M TimeStep = 1
M NumberFormat = s3

D Your Age = 0
P Your Age.SaveResult = true
P Your Age.VisibilityLevel = 16

D Your Sex = 0
P Your Sex.SaveResult = true
P Your Sex.VisibilityLevel = 16

D Your Years in Uni = 0
P Your Years in Uni.SaveResult = true
P Your Years in Uni.VisibilityLevel = 16

D Initial Supply of Flubnium = 50
P Initial Supply of Flubnium.ExecuteDecisionImmediately = true
P Initial Supply of Flubnium.SaveResult = true
P Initial Supply of Flubnium.VisibilityLevel = 16

D Initial Demand for Energy = 100
P Initial Demand for Energy.ExecuteDecisionImmediately = true
P Initial Demand for Energy.SaveResult = true
P Initial Demand for Energy.VisibilityLevel = 16

D Time to Change Supply of Flubnium = 5
P Time to Change Supply of Flubnium.ExecuteDecisionImmediately = true
P Time to Change Supply of Flubnium.VisibilityLevel = 16

D Additional Units of Demand per Doubling of Efficiency = 50
P Additional Units of Demand per Doubling of Efficiency.ExecuteDecisionImmediately = true

P Additional Units of Demand per Doubling of Efficiency.VisibilityLevel = 16

D Efficiency of Use = 1
P Efficiency of Use.SaveResult = true

V Initial Efficiency of Use = Initial(Efficiency of Use)
P Initial Efficiency of Use.VisibilityLevel = 16

V Supply of Flubnium Needed to Balance Supply and Demand = Demand for Energy/Efficiency of Use

V Change in Supply of Flubnium = (Supply of Flubnium Needed to Balance Supply and Demand - Supply of Flubnium)/Time to Change Supply of Flubnium

V Supply of Flubnium = STOCK(Change in Supply of Flubnium,Initial Supply of Flubnium)
P Supply of Flubnium.SaveResult = true

V Supply of Energy = Efficiency of Use * Supply of Flubnium
P Supply of Energy.SaveResult = true

V Change in Efficiency of Use = Efficiency of Use - PREVIOUS(Efficiency of Use,Initial Efficiency of Use)
P Change in Efficiency of Use.VisibilityLevel = 16

V Change in Energy Demand = Change in Efficiency of Use * Additional Units of Demand per Doubling of Efficiency

V Demand for Energy = ACCUM(Change in Energy Demand, Initial Demand for Energy)
P Demand for Energy.SaveResult = true

V Supply Demand Ratio = Supply of Energy / Demand for Energy
P Supply Demand Ratio.SaveResult = true

V Correct = IF(ABS(Supply of Energy - Demand for Energy) > 5 , 0, 1)
P Correct.VisibilityLevel = 16

B: Full Case Information

In this simulation you are playing the role of a government regulator. In this simulation humans have discovered a new source of energy called 'Flubnium'. Flubnium is mined from the earth and converted to energy which is demanded in the market. In this game you are tasked with the role of dictating how efficient the conversion process from Flubnium to energy is. Your goal is to balance supply and demand.

Make your decisions based only on the information presented on this page, do not rely on any other sources

Key information:

1. The supply of Flubnium is 50.0

2. The demand for energy from Flubnium is 100

3. If you take no action the gap between supply and demand will always take 5.00 years to close

4. For each doubling of efficiency there are 50.0 units of demand are generated.

5. The initial efficiency of use is 1.00 units of Flubnium to 1 unit of energy

5. You have 5 rounds to succeed balancing supply and demand.

C: Technical Instructions

Over the next hour you will be playing a simulation based on the case presented to you on the previous page. You will be given two chances to balance the energy supply from Flubnium with the demand for that energy. On the next page you will be presented with a questioneer which will help to establish your baseline understanding of the problem and some basic demographic information.

After answering those questions you will be presented with a decision dashboard where you will be able to view key indicators from the Flubnium market, and make your decision. After making your decisions each year you will be presented with a model structure visualization aid to help you understand the structure of the simulation model which this game is based on. You will be allowed to review this diagram as you

make your decision. The diagram presented to you is of critical importance and will reveal to you key information that you can use to solve the problem. After each game (5.00 rounds apiece) you will be given a questionnaire and then asked to proceed to the next game.

Remember it is important that you keep your answers to yourself and if you have any questions to direct them to wasbridge@gmail.com.

Please do not close the browser until you are done with the final round of the experiment, and please do not have any other programs and webpages open in the background.

Thank you and good luck!