EFFECTS OF DESIGN CHARACTERISTICS OF INFORMATION PRESENTATION ON PERFORMANCE AND LEARNING IN A DYNAMIC DECISION ENVIRONMENT

Master thesis submission for the EMSD program

By Victor Alexiev

ABSTRACT

This paper examines the effects of using good information presentation design practices on the performance and learning of people interacting with a digital and dynamic decision environment. The research presented here has important implications for the design and development of effective interfaces for Management Flight Simulators and Interactive Learning Environments. It also connects to the Misperception of Feedback hypothesis, and suggests that better design and configuration of flight simulator data displays could reduce cognitive efforts, and increase performance and learning effects in complex, dynamic decision context. Prior research suggests that information presentation has strong influence on the choice of decision strategies and hence – on performance. However, these effects have not been sufficiently researched from a System Dynamics Modelling perspective. An inquiry into the subject is critical as Interactive Learning Environments are an important communication channel and a key delivery for many System Dynamics interventions. To compare the effects of information presentation on performance and learning, this author has performed an experiment, attempting to test for the effects of using an improved interface against the one from the original Boom and Bust simulator from Paich and Sterman (1992). The control group was exposed to a simulation interface similar to the one from the original paper, while the treatment group was exposed to an alternative interface, designed using Tufte's (2011) principles of good design for the visual display of quantitative information. Results from the experiment showed some support for the hypothesis that participant performance will be influenced by variations in the design of the information display. Furthermore, evidence was found that improvements in the interface could also lead to stronger learning effects from repetitive interaction with the simulator.

1. INTRODUCTION

Together with the work on biases and fallacies done within the domain of behavioural decision science (e.g. Kahneman, Slovic, & Tversky, 1982), research on the misperception of feedback has exposed numerous shortcomings of human judgment in dynamic decisions setting. Those findings have motivated scholars to look for remedies and aids that can support and improve the decision process and augment our cognitive capacity. The universe of decision support solutions created for this purpose spans from direct expert interventions in the decision process to the use of computerized decision support systems and interactive learning environments. The purpose of the latter is to provide a correction mechanism for our faulty mental models, by granting us a fail-safe, feedback-rich learning environment where decision-makers can test their theories about the underlying structure of the problem.

In experimental setting, the use of Interactive Learning Environments (or ILE) or Management Flight Simulators (or MFS) has enabled us to study in depth the decision-making process and its reaction to outcome feedback from a simulation environment. As a result, the *Misperception of Feedback* hypothesis has emerged as an important theme in *System Dynamics (*Sterman, 1989; Sterman, 1992; Paich and Sterman, 1993; Moxnes, 2004; Moxnes & Saysel, 2009; and others). Most frequently, literature on the topic was focused on exposing the inability of people to perform, learn, and adapt within complex and dynamic feedback environments. The major reasons to which this flaw is attributed are the "Misperception of time lags" in the system and the "Open Loop Explanations of Dynamics" (Sterman, 1989, pp. 334-335). There is also much research on developing methodologies to remedy the Misperception of Feedback phenomenon by information sharing (Moyaux, Chaib-draa, & D'Amours, 2007), educating the subjects about the system's structural components (Moxnes, 1998), or by using metaphors and analogies (Moxnes & Saysel, 2009). Yet, the majority of studies consistently show that decision makers operating in dynamic environments fail to benefit from learning and continue to perform "poor relative to normative standards or even relative to simple heuristics, especially when decisions have indirect, delayed, nonlinear, and multiple feedback effects" (Paich & Sterman, 1993, p. 1440). A common factor in most research is that experiments rely on some kind of interface to the simulation environment, the aim of which is to deliver outcome feedback to decision makers. While the importance of how information is fed back in an ILE context is clear, little academic effort has been devoted to examining the effects of data display designs on decision-making.

Interactive Learning Environments enable us to provide instant and interactive outcome feedback to the decision-maker or learner. This feedback is produced through simulating the actions (decisions) of the decision-maker(s) through the interconnected logical assumptions of the dynamic model. This makes it easier for the decision-maker(s) to experience the dynamics caused by their own decision policies and enables the phenomenon of "Learning by playing around" (Andersen, Chung, Richardson, & Stewart, 1990). With their interactive nature, ILEs help learners understand how "decisions made today alter the environment, giving rise to information upon which tomorrow's decisions are based" (Paich & Sterman, 1993, p. 1440). The use of digital displays has augmented the possibilities of formatting this information, which has some effect on the information acquisition and interpretation process. In fact, past studies on the relationship between information presentation and decision making have indicated that the format and organization of information displays influences the choice of decision strategy and hence – performance (Jarvenpaa, 1989; Einhorn & Hogarth, 1981; Payne, 1982; Kleinmuntz &

Schkade, 1993; Speier, Vessey, & Valacich, 2003; and others). However, those studies have focused predominantly on static decision task, where all information is available from the beginning and where no outcome feedback from individual decisions is presented.

In contrast, this paper focuses on the implications of the design of information displays for dynamic decision tasks for which "performance is determined by the cognitive processes related to problem solving" (Atkins, Wood, & Rutgers, 2002). The research presented here is based on an experiment, which aims to test if good design practices would result in improved decision-making leading to performance gains. Such practices are those that focus on reducing the cognitive effort and driving the decision-makers' attention towards the most-important outcome feedback cues. Findings from this research could help us design better interfaces for our ILEs, which facilitate the performance and learning processes in an interactive simulation context. Furthermore, discovering meaningful differences between the traditional ILE designs and the improved ones could mean that part of the performance losses from misperception of feedback could be remedied by delivering outcome feedback in a better format. Last but not least, if different decision strategies are induced by variation in the data display, then this calls for more attention and the development of best practices for the design of ILE interfaces, which prevent information overload and biases.

This paper proceeds as follows: First, relevant literature on Misperception of Feedback (or MOF) is reviewed in order to provide background understanding of existing research in the field. This overview includes several methods proposed by scholars for assisting decision-makers in building better mental models of the problem structure and reducing the effects of MOF. Next, considering the common use of information displays in MOF experiments as part of the ILE interface design, a brief presentation of literature examining the effects of information presentation on the choice of decision strategies and performance is provided. Since no works from the aforementioned research domains make firm suggestions on what good information presentation is, some literature on good practices from the domain of visual display of quantitative information is also summarized at the end of the next section. Second, building on the reviewed literature, this author provides a simple theoretical framework on how applying good information visualisation principles could affect decision-making within ILEs. Four hypotheses are formulated, which regard the *choice of decision strategies*, *performance* in the problem context, and the learning effects of information presentation. Third, the experiment that was designed for testing the hypotheses developed in Section 3 is described in detail. This includes the model to be used, the description of the two treatment alternatives, the participant recruitment process, and the experiment task. Fourth, the results from the experiment are presented and interpreted and hypothesis testing is performed to check which hypotheses, developed in Section 3, are supported by the results. Lastly, findings are summarized in a brief conclusion and the limitations of this work are discussed, together with suggestions for further research.

2. Relevant literature

This section summarizes the relevant literature on the misperception of feedback, with a focus on research involving ILEs. Furthermore, it reviews studies examining the effects of information presentation on decision making, especially in terms of decision performance and choice of decision strategy. Lastly, it presents an overview of good practices for design of information displays, that were relevant to the design of the alternative interface employed in the experiment that this paper describes.

2.1. MISPERCEPTION OF FEEDBACK

Ever since Herbert Simon coined the term "*bounded rationality*" in *Models of Man* (Simon, 1957), testing and exploring the limited capabilities of human cognition has been a common theme in social science research. Over the years, the efforts of studies in psychology and economics resulted in a long list of deviations from the predictions of rational models of behaviour stemming from a number of fallacies and biases produced by cognitive limitations (see Tversky & Kahneman, 1974; Kahneman, Slovic, & Tversky, 1982; and others). To a large extent, research in the area was focused on "static and discrete tasks" (Sterman, 1989, p. 321) and hence, failed to show what the implications of bounded rationality are on a systemic level and in a dynamic setting.

In attempt to shed more light on the shortcomings of dynamic decision context, Sterman (1989) studied the implications of individual decision making on a system's behaviour using an experimental setting that employed the popular *Beer Distribution Game*[1]. Results from the Beer Game generally show that given a relatively straightforward task, subjects would produce output dynamics that differs "significantly and systematically from optimal behaviour" (Sterman, 1989, p. 322). Numerous trials of the experiment produced oscillations, mostly with increasing amplification throughout the supply line. This occurred despite the fact that customer orders increase only once in the game and remain constant until its end (Sterman, 1989, p. 328). In his analysis of a sample of 11 beer-game trials drawn out of 48, Sterman (1989) hypothesized and found strong support that the aforementioned oscillation is produced by the use of a simple anchoring and adjustment heuristic. This decision rule was coupled with a failure to account for outcome feedback, which he termed "misperception of feedback" (Sterman, 1989, p. 334). It is important to note that Sterman's analysis was biased towards the best performing group as the final sample of 11 trials consisted of "those who understood and performed best in the game" (Sterman, 1989, p. 328). This made Sterman's findings even more alarming. To explain the poor results and the oscillations, Sterman (1989) pointed out two critical misperceptions that present impediments to experiential learning from dynamic decision environments:

1. "Misperception of Time Lags", characterized by (a) underestimating the time between placing an order and its delivery, which generally leads to the build-up of backlog and (b) ignoring the supply line continuing to place orders until they start arriving, which leads to a build-up of inventory (Sterman, 1989, p. 334);
2. "'Open loop' Explanations of Dynamics", the majority of subjects in the experiment maintained that fluctuations in the supply line were exogenously caused by oscillating demand, while demand in fact remained constant for most of the time (Sterman, 1989, p. 336);

In summary, Sterman 1989 showed how the simple mental models employed in a complex and dynamic setting are flawed and tend to "cause systematically dysfunctional behaviour" (Paich & Sterman, 1993, p. 1440). Basing their arguments on those findings, Paich and Sterman (1993) set out to explore the effects of feedback complexity on subject performance within a simulated market environment. In addition, Paich and Sterman (1993) attempted to determine if learning effects can be induced by repetitive interaction with the experimental market simulator. For the purpose of their research, student participants were

[1] The Beer Distribution Game was developed by MIT's SD group in the early 1960s as part of Forrester's work on industrial dynamics (Sterman, 1992)

asked to manage the introduction of a new product in a dynamic simulated market. To detect learning effects, subjects were asked to perform the task repeatedly and their results were recorded. Paich and Sterman (1993) used the performance results from each attempt as a proportion to the potential to measure performance, while learning was measured as the change of this measure from attempt to attempt.

Paich and Sterman (1993) hypothesized that performance, measured in cumulative profit relative to benchmark potential, should decrease with the increase of feedback complexity, but nevertheless – increase from attempt to attempt under the effects of learning. In their research, feedback complexity was defined as the number and strength of implicit (indirect) feedback loops. They found that "[t]he negative effects of feedback complexity on performance were not moderated by experience, even though average performance improved" (Paich and Sterman, 1993, 1439). Hence, while repetitive attempts generally led to better performance, the majority of participants still performed poorly, regardless of the number of opportunities they had for learning (Paich and Sterman, 1993, p. 1453). Paich and Sterman (1993) concluded that stronger feedback processes lead to poorer performance in dynamic environments. Moreover, they found no evidence "that subjects improved their ability to manage the environments with high feedback complexity as they gained experience, despite improvement on average" (Paich and Sterman, 1993, p. 1460). To Paich and Sterman (1993), this was a hint that the majority of the learning might come from the subjects getting accustomed to the general pattern of market dynamics, rather than gaining insight into system structure.

Unlike previous experiments, exposing inability of subjects to manage dynamic systems, Paich and Sterman (1993) regarded their setting to be "realistic and well matched to the interests and training of the subjects" (p. 1440). It is fair to say that realism has implications for the feedback complexity of the simulation and the Flight Simulator interface used to intake decisions and report information back to the participants. Moxnes (2004) had similar considerations about experiments that study Misperception of Feedback, especially in the context of renewable resource management. According to him, "laboratory experiments used thus far have been characterised by considerable complexity and ambiguity about model structure and parameters" (Moxnes, 2004, p. 139). To examine the implications of the Misperception of Feedback in a simpler and more straightforward setting, Moxnes (2004) designed an experiment in which subjects could fully reconstruct the underlying mental model using the instructions provided. As a result, "observed subject behaviour can be compared to optimal normative behaviour" (Moxnes, 2004, p. 140). This study differentiates from past research on misperception of basic dynamics (e.g. Sweeney, and Sterman, 2002) as its subjects interacted with a management flight simulator "with information feedback and repeated decisions" (Moxnes, 2004, p. 140). Moxnes' experiment involved two treatment groups – one managing a one-stock model, and the second managing a two stock model of the same renewable resource problem (Moxnes, 2004). The findings from the experiment fully supported the misperception of feedback hypothesis with the two-stock treatment group having a larger deviation from the normative performance. Furthermore, using repeated trials Moxnes concluded that outcome feedback from the repetition "is not sufficient to achieve rapid learning over time and over repeated trials" (Moxnes, 2004, p. 158).

The grim implications of Moxnes' (2004) results are that people seem to be unable to effectively manage even the simplest one-stock and two flows dynamics as they are "not able to formulate an appropriate [mental] model for the decision problem" (Moxnes, 2004, p.

150). An important question that emerges from this is: *How can we facilitate the formulation of an appropriate mental model in such context?* One suggestion from Moxnes and Saysel (2009) is the use of analogies or metaphors. In their experiment studying the effects of such analogies, Moxnes and Saysel (2009) employed three information treatments to condition the subjects and help them form a better understanding of a CO2 stock management problem. The information treatments used were an air mattress analogy (T1), a balloon analogy (T2), a phase diagram (T3), and outcome feedback (T4) (Moxnes & Saysel, 2009, pp. 21-25). Results from the experiment provided support for a "highly significant and largely positive effect on performance" (Moxnes & Saysel, 2009, p. 28) for information treatments T1 and T4. The implications of this are that the use of appropriate analogies and the delivery of outcome feedback have strong effect on the ability of subjects to formulate an appropriate mental model for the experimental context.

In summary, the impression from Misperception of Feedback literature is that the majority of tasks employed are generally complex, involve multiple interdependent components, and relatively raw data presentation format. Findings indicate that performance and learning in complex dynamic environments seems to be poor. Moreover, most of the performance gains tend to be a result of getting used to the particular simulation environment and reusing information about previous cues instead of making inferences about the causal structure of the problem. Attempts have shown that there are ways to remedy this limitation, but little effort has been made to research how information presentation aspects within the flight simulator environment affect learning and performance.

2.2. Effects of information presentation on decision making

Digital, interactive learning environments or flight simulators are a way for decision support experts to hand-over the responsibility over the learning experience back to the learner (Lawless & Brown, 1997). However, due to the technocratic nature of their creators, ILEs tend to focus more on transferring the problem representation to the computer environment, and less on ensuring its quantitative and logical correctness. Consequently, an important aspect that is often neglected or down-prioritized is the necessity for "making the [decision] environment more conductive to effective decision making" (Kleinmuntz & Schkade, 1993, p. 221). In an ILE setting, the decision environment is represented by the information display interface, which presents the dynamic outcome feedback.

With that in mind, Kleinmuntz and Schakade (1993) argue that the design of information displays has an important influence over the choice of decision strategy and hence – on decision performance. While early research in the area was focused predominantly on comparing tabular and graphical displays of data (e.g. Dickson, Gerardine & DeSanctis, 1986), the evolution of digital displays and computerized decision support systems has made the possible variations in the "visual representation of decision problems virtually infinite" (Kleinmuntz & Schkade, 1993, p. 221). The same applies to the presentation of data generated by simulation, which is necessary for the analysis of the decision problems at hand. To handle this complexity, Kleinmuntz and Schkade (1993) focus on what they believe to be the three fundamental characteristics of visual representation – form, organization, and sequence. For them *form* encompasses numerical, verbal, and pictorial information presentation, where pictorial consists of charts, maps, or other visual symbols (Kleinmuntz & Schkade, 1993, p. 221). Organization refers to the structuring of information, which could be hierarchical, matrix, groups, or other patterns (Kleinmuntz & Schkade, 1993,

p. 222). Lastly, sequence regards the order in which different pieces of information are presented to the decision maker (Kleinmuntz & Schkade, 1993, p. 222).

Separate studies on the form, organization, and sequencing of information indicate that they have measurable effects on decision performance. Research suggests that the reasons behind this boils down to a common factor – "decision makers respond adaptively to variations in information displays, using different decision processes depending on the different arrangement of form, organization, and sequence" (Kleinmuntz & Schkade, 1993, p. 222). With regards to form of information presentation, Dickson et al. (1986) employed a set of three experiments to study the effectiveness of graphs for decision support in comparison to tabular representation of data. They found that the graphical presentation of data was superior in cases where "*analysing time-dependent patterns* was important", and when "large amounts of data had to be presented to prompt the recollection of specific facts" (Dickson, DeSanctis, & McBride, 1986, p. 46). Nevertheless, the superiority of graphs as a data-communication device was deemed disputable in their research. This finding is in line with an earlier study on the subject performed by Lucas and Nielsen (1980). In it, the researchers found very little support at the 10% confidence interval for performance and learning benefits from samples receiving graphical versus numerical (table) feedback from a computer simulation of a logistics problem (Lucas & Nielsen, 1980, p. 989).

Another study by Jarvenpaa (1989) examined the implication of information organization on decision making. Results from her experiment showed strong evidence that the organization of information displays has an effect of the acquisition and evaluation of information coupled with weak evidence for effects on the decision time (Jarvenpaa, 1989). Jarvenpaa (1989) found no evidence that organization of information displays has an effect on decision accuracy (performance). Lastly, Hogarth and Einhorn (1992) studied the conditions under which different sequencing of information has an effect on the updating of belief. In their study, the authors built a model, which showed how "task variables and processes interact in producing order effects in belief updating" (Hogarth & Einhorn, 1992, p. 40). In summary, past research provides evidence that form, organization, and sequencing of information have an important effect on decision process but these factors play out differently depending on the context. For example, organization has a strong effect on information acquisition, while form influences the combination and evaluation of information (Kleinmuntz & Schkade, 1993, p. 224). Sequencing of information is generally found to have "fewer and smaller effects than organization" (Kleinmuntz & Schkade, 1993, p. 224).

Theorizing about the causes for effects of information presentation on the choice of decision strategy, Kleinmuntz and Schkade (1993) look at *procedural knowledge*[2] and its relationship to adaptive strategy. The authors suggest that the display characteristics influence the formation of anticipated effort and accuracy for each possible decision strategy. The total set of available decision strategies is based on the subject's procedural knowledge formed by past experience in analogous situations (Kleinmuntz & Schkade, 1993, p. 224). Consequently, decision makers chose a strategy based on a cost-benefit (effort-accuracy) heuristic and apply it to the problem. The experienced effort from applying the selected strategy, combined with the experienced accuracy through the outcome feedback, could then lead to a re-evaluation of the strategy next time a similar problem is faced. Thus, procedural

[2] "The knowledge that a decision maker possesses about strategies and their effectiveness in various tasks and settings" (Kleinmuntz & Schkade, 1993, p. 225)

knowledge is expanded by the experience within the decision task. Einhorn and Hogarth (1981) suggest that each decision strategy can viewed as a multidimensional object with its dimensions reflecting the costs and benefits balancing heuristic a decision maker employs to evaluate strategies. The main dimensions that Einhorn and Hogarth (1981) point out are *(a) probability of error, (b) size of error, (c) speed of decision, (d) justifiability, (e) computational effort, (f) search costs, and (g) awareness of conflict*. Either or all of those dimensions can be influenced by the form, organization, or sequencing of information.

In the realm of dynamic decision making, Atkins et al. (2002) employed an experiment to study the influence of different feedback formats on performance in the context of a simple inventory management task. The two alternative treatments employed were graphical (line over time) and tabular data presentation format. Subjects were asked to perform the experiment repeatedly so that learning effects can be detected (Atkins, Wood, & Rutgers, 2002, p. 596). The study found no significant difference between the samples for the main trial. However, contrary to the researchers' expectations, some evidence of performance improvement from repeated attempts was found for the group assigned to the tabular outcome feedback, while no such evidence was found for graphical feedback (Atkins, Wood, & Rutgers, 2002, p. 596). It is important to note that findings from the Atkins et al (2002) experiment should be considered with care as their total sample amounted to 18 people.

The general impression from existing research in the area is that findings are contradictory and highly contextual. Furthermore, the implications of dynamic information presentation, conditional on past decisions are not thoroughly examined. In fact dynamic decision-making literature "has largely neglected the influence of feedback formats on task performance" (Atkins, Wood, & Rutgers, 2002; Vicente, 1996). Moreover, only a handful of studies look into the joint effect of information format, organization, and sequencing (Kleinmuntz & Schkade, 1993). In addition, existing research has relied mostly on comparing categories of displays formats and their effects of performance, but no inquiry has been done in the combined effect of different formats. Lastly, no study reviewed for this paper has examined best practice literature on effective data presentation.

2.3. Information Presentation – Good Practices

Images have been used to represent data for a relatively long time. However, "the use of abstract, non-representational pictures to show numbers is a surprisingly recent invention" (Tufte, 2011, p. 11). The pioneer works on systematizing and improving knowledge in the area were conducted mostly by William Playfair (1759-1823), who sought to replace numerical representation of data with visual displays. According to Playfair, "[i]nformation that is imperfectly acquired is generally imperfectly retained; a man who has carefully investigated a printed table, finds, when done, that he only has a very faint and partial idea of what he has read" (Playfair, 1786, p. 3). To improve the acquisition of large quantities of information, Playfair developed a novel, for his time, charting method, which he termed *linear arithmetic* (Tufte, 2011). The major benefit from this new representation for Playfair was that "...on inspecting any one of these Charts attentively, a sufficiently distinct impression will be made, to remain unimpaired for a considerable time, and the idea, which does remain will be simple and complete, at once including the duration and the amount" (Playfair, 1786, p. 4).

In the centuries following Playfair's work, graphical display of data has increasingly gained importance and popularity and has become an inseparable part from the analysis of data. However, we sometimes tend to forget that when reasoning about quantitative evidence, “certain methods for displaying and analyzing data are better than others” (Tufte, 1997, p. 27). Hence, there is need for base principles that can help us design better visual explanations. In his work, Tufte (2011) summarized a set of principles for graphical excellence that should guide the creation, presentation, and interpretation of data graphics. To him, graphical excellence was “a matter of substance, of statistics, and of design” (Tufte, 2011, p. 51) and its essence was to provide the viewer with “the greatest number of [correct] ideas in the shortest time with the least ink in the smallest place” (Tufte, 2011, p. 51). The experimental setting of this research is aimed at measuring the effects on performance and decision strategy, stemming from the abidance by those base principles.

3. THEORY DEVELOPMENT

In a dynamic setting the decision environment evolves over time. The ability to make correct assessments of those changes, and formulate an appropriate mental model for the causal relationships behind them, is core to building valid understanding of the system (Sterman, 1994). Tufte (1997) claims that “[t]o understand is to know what cause provokes what effect, by what means, at what rate”. But how can we obtain and communicate such knowledge through an interactive learning environment in a way which affects decision making?

In a System Dynamics we build management flight simulator and employ their model interface to provide “structured experiences” (Lane, 1995, p. 607) to decision makers and hence facilitate learning through delivering outcome feedback (Davidsen, 2000; Lane, 1995). Generally, flight simulators and ILEs are used to “influence the formation of mental models” and sometimes for “research and validation” (Davidsen, 2000). The information communicated thorough the simulation interface should help decision makers formulate an appropriate mental model for dealing with the situation at hand (Davidsen, 2000). However, Moxnes and Saysel (2009) suggest that we cannot expect people to form a correct mental model without guidance (Moxnes & Saysel, 2009). This paper suggests that the careful design of the information display, and hence – the overall simulation interface, is a way to provide such guidance without human intervention. It is important to note that the ILE interface, although composed of separate elements, can also be viewed as a single communication medium, which affects the anticipated decision effort through its complexity. Hence, the presentation of each data point is as important as the overlay of the interface.

The SD literature makes only a few suggestions on best practices for simulation interface design, which aim at improving the user’s learning experience (Andersen, Chung, Richardson, & Stewart, 1990). However, interfaces employed in seminal studies still tend to be with rather technical and, in the eyes of the non-professional user, complex and complicated to deal with. Reviewing simulator interfaces employed in studies examining the misperception of feedback hypothesis and the inability of subjects to benefit from learning in dynamic simulation environments without external support confirms that. On the other hand, studies of what decision support should be provided tend to focus on the use of metaphors and abstractions aimed at facilitating the creation of a valid mental model (Moxnes & Saysel, 2009). Other methods rely on training the understanding of the decision maker on the underlying dynamics prior to the simulation (Moxnes, 1998). Having described how information presentation affects decision process and performance, this author believes that performance and learning gains could also be achieved by applying design principles to the interfaces used in Interactive Learning Environments.

Kleinmuntz and Schkade (1993), and others have shown that the way information is fed back to decision makers could have substantial effects on the choice of decision strategies. This is rooted in the theory that decision makers use adaptive strategies as response to problem complexity, response mode, similarity of alternatives, and characteristics of the information display (Payne, 1982). This has strong implications for learning in a dynamic environment as the formulated strategies change over time based on outcome feedback from the previously deployed strategies. For example, if two variations in the form and organization of information presentation lead to the formation of two different pairs of *Anticipated Effort* and *Anticipated Accuracy* (one for each display), then the chosen strategies within those two environments could also differ. If different strategies lead to different outcomes, then the two designs will also result in different *Experienced Effort* and *Experienced Accuracy*. Hence, it is possible that after the interaction is complete, the same decision-maker attempting to solve an identical problem through two different interfaces would have a different cost-benefit evaluation of the same set of possible strategies. As a result, (s)he would obtain different conclusions and thus – procedural knowledge about the applicability of their set of strategies to the same problem (See *Figure 1*).

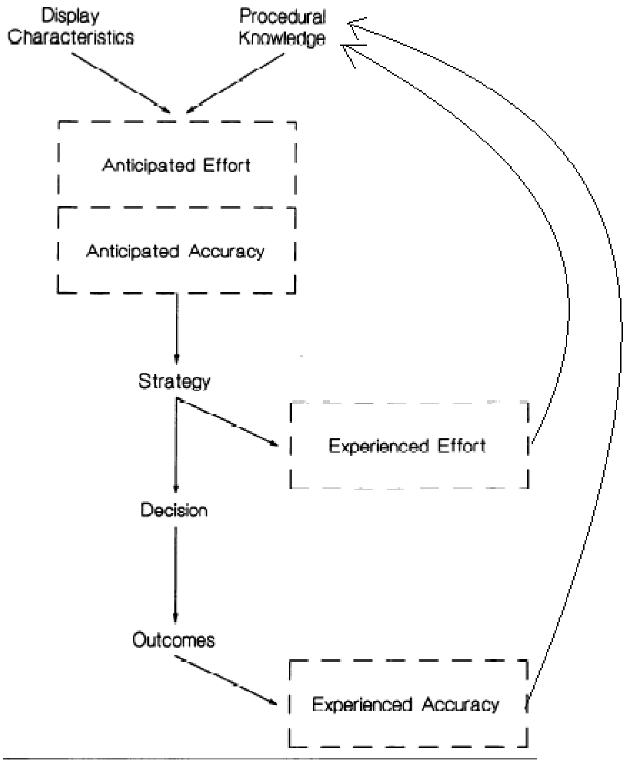


FIGURE 1: ADAPTED FROM KLEINMUNTZ & SCHKADE (1993). LINK FROM EXPERIENCED EFFORT AND EXPEERIENCED ACCURACY TO PROCEDURAL KNOWLEDGE ADDED BY CURRENT AUTHOR

Applied to repetitive interaction with a management flight simulator, the aforementioned theory serves as basis for the formulation of the following two hypotheses:

H1: Difference in the information presentation design and data displays will lead to the choice of different decision strategies

H2: Difference in the information presentation design and data display will lead to the choice of a different set of strategies in repetitive attempts of the task

If H1 is correct, it is fair to assume that the selection of alternative strategies within the two decision environments would produce different outcome and hence different learning contributing to a change in procedural knowledge. Since procedural knowledge will be employed next time the problem is faced, this difference would lead to a variation in the approach to the very same problem. Consequently, H2 suggests that the bundles of strategies employed in repetitive attempts of the problem situation will remain different, due to the difference in changes of decision-maker' procedural knowledge. This corresponds to the suggestion that in their second attempt, experiment participants will apply a diverse set of decision strategies in the two treatment groups as response to their *Experienced Effort* and

Accuracy from the first simulation. Such a change is important as “radical and sudden shifts in individual strategies may also indicate learning” (Moxnes, 2004, p. 147).

For the purpose of this research, **information presentation** refers to the form of outcome feedback on individual variables of interest while **data display** regards the overall organization of the ILE interface. **Improvement** of the aforementioned aspects refers to the application of Tufte’s (2011) design excellence principles and the general body of knowledge for good data presentation. **Decision strategy** is considered as the bundle of decisions employed to address the problem. In this case, it is measured by the number of changes applied for price and capacity (see section 4.4. for details), the mean value of those changes, and their standard deviation. **Cognitive effort** is the total decision time for each attempt.

Past research indicates that component characteristics of information displays (form and organization in this case) should influence decision process through the adaptive mechanism of balancing the desire to maximize accuracy while minimizing effort. Hence, carefully designing the information display, one can encourage the decision maker to use a good decision process (Kleinmuntz & Schkade, 1993). Variation in decision strategies, especially in a relatively simple feedback environment, is likely to produce measurable differences in performance outcomes (Kleinmuntz & Schkade, 1993, p. 225). If the selected decision strategies are based on a correct understanding of the problem situation, then better strategies should lead to better performance. Atkins, Wood, and Rutgers (2002) suggest that characteristics of outcome feedback, including the effort required for its interpretation, reflect the complexity of the overall task. As the literature review indicated, a negative relationship is found between task complexity and performance (Paich & Sterman, 1993). In addition, research performed by Speyer, Vessey, and Valacich (2003) suggests that while information acquisition is better addressed by tables in moderately complex tasks, as task complexity increases, a level will be reached at which graphs outperform tables. Hence, it is reasonable to assume, that applying design principles to the form and organization of information display to reduce complexity should lead to improvement of performance. Therefore:

> **H3**: Improvement on the information presentation design and data displays will lead to an improvement in performance

Here, **performance** is considered in the context of the specific objective, given to the experiment participants. It was measured as *Cumulative Profit* for the total simulation period.

Performance should improve with experience (Paich & Sterman, 1993). If assumptions in H2 are correct, then procedural knowledge would be developed more effectively from interactions with a better designed interface. Hence, it would be fair to deduce that performance differences between the two treatment groups should increase in the second interaction with the simulator due to learning effects. If the between-sample difference in the second attempt performance is in favour of the improved interface, then it could be attributed to learning effects stemming from the design of the information display.

> **H4**: Improvement on the information presentation design and data displays will have a positive effect on learning

It is important to note that “improvements over trials could also be the result of trial and error with no deeper learning involved” (Moxnes, 2004, p. 147). This claim is also supported by Paich and Sterman (1993), who suggest that performance improvements might also result from the fact that subjects "become increasingly familiar with the task and information display".

4. Experiment Design

To test the hypotheses outlined in the previous section, this researcher has formulated an experimental setting, which relies on the use of an Interactive Learning Environment. The ILE consists of an interface built on top of a System Dynamics simulation model. Both the interface and the model were developed using the Powersim8® modelling software. The experiment includes two treatments – a base treatment (T0), where the interface was modelled after Paich and Sterman's Boom and Bust experiment (1993) and an alternative treatment (T1), with an interface developed using Tufte's suggestions for good data presentation. It is important to mention that Forio's implementation[3] of the Boom and Bust model interface was used as a reference on how data presentation should be handled in the base treatment. Participants, recruited for the experiments were given a set of instructions (see *Appendix I*) and were allowed to ask questions about the interface or the instructions before starting the simulation. No information about the system structure was revealed outside of what was given as initial instructions and available through the simulator interface. Participants were randomly assigned to either of the treatments.

Once instructions were read, each participant was asked to confirm if they understand the decisions they need to make and their objective for the simulation game. Then they were introduced to the simulator and the researcher moved away. Since no special room was provided for the experiment the researcher mostly moved to the other end of the table or slightly away from the subject. After the first attempt was completed and data was copied, participants were asked if they could perform the same simulation again. The purpose was to collect data for changes in performance. Participants were not initially informed that they would be asked to perform the simulation more than once in order to avoid the conditioning that they should use their first attempt to educate their decisions for the second. Since no incentive was provided, some participants did not wish to perform the simulation again due to time constraints. The majority did. Participants performed a maximum of 3 trials, but the sample doing the experiment more than twice was too small to use (6 people). A graphical representation of the experiment set-up is displayed in *Figure 2* below.

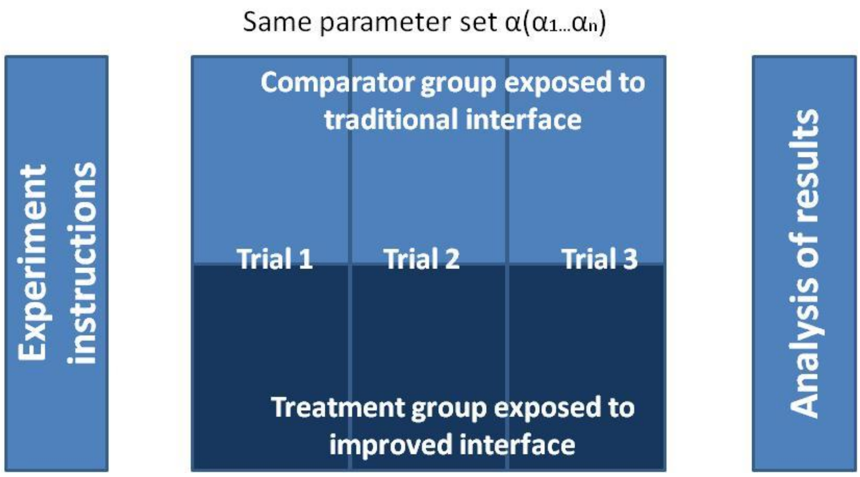


FIGURE 2: EXPERIMENTAL SET-UP

[3] See, http://forio.com/sim-store/demos/boom-and-bust.html

4.1. Model

The core driver behind the experiment was a System Dynamics simulation model based loosely on Paich and Sterman's Boom and Bust, and failure to learn in experimental markets (1993) paper. Similarly to the original model, the one employed in this experiment was characterized by two sectors – firm and market. Unlike in the original, this model assumed that there is no competition and hence – no competitive dynamics for the duration of the simulation. Participants were clearly informed about that by stating that their product was protected with a patent and hat no substitutes. The main reason behind this assumption was to simplify the dynamics and reduce feedback complexity stemming from the model in order to focus on task complexity aspects of the interface. Moreover, this researcher aimed to reduce the amount of information that the experiment participants needed to take care of in order to enable them to create effective strategies. Since the purpose of this experiment was not to test the misperception of feedback hypothesis, which has already gotten sufficient support, this researcher considered that removing the competition will make it easier to see if experiment participants are able to grasp the basic market dynamics better through the alternative interface.

The market segment in the model is characterized by a simple mechanism for generating orders based on word of mouth, expected delivery time, price, and potential customers. The major driver was a modified Bass diffusion model (reference) driven by the word of mouth phenomenon. "*Potential customer"* order at a rate based on:

- a "*normal sales*" parameter that indicates a base sales rate
- a "*word of mouth*" parameter, driven by the number of recent purchases
- an effect of "*perceived delivery*" delay on demand, which was a graphical convertor of *expected delivery delay* and acted as a discounting or scaling parameter to the orders
- an effect of "*product price*" on demand, driven by the difference between normal price and actual product price (pricing will be described separately)
- a simple multiplier consisting of "*potential customers*" divided by "*total market*" that represents that higher market saturation makes further penetration more difficult;

For simplicity, "*total market*" was assumed to be constant, which is a reasonable assumption considering the 5-year timeframe. Once orders were made they enter a *backlog,* which is cleared by shipments. Buyers who receive their shipments become "*customers"*. Customers discard the product within 12 months, on average, and are moved back to "*potential customers*".

Word of mouth is generated by customers who have purchased the product within an average period of 3 months, those are called recent purchases and are, supposedly, the ones who are still excited about their purchase. Recent purchases are increased by shipments and are decreased by rate of customers who are getting used to the product and are thus, less excited about it. Word of mouth is scaled by three constants:

- Fraction of customers who are willing to promote the product
- Number of unique monthly social contacts per customer who promotes the product
- Probability of adoption, representing the probability that each of those unique contacts will be willing to become a customer as a result of the word of mouth effect;

Customers, generated by the "*word of mouth"* effect are then added to the "*normal sales*" to produce the base orders, and their sum is scaled by the price and *expected delivery delay* effects.

The firm segment is characterized by capacity and pricing. The capacity structure relies on decision makers to set target capacity, which is the amount of productive capacity they would like to have. The model then compares this amount to the capacity that is currently on order or already deployed and forms a gap, which needs to be closed within the capacity order time, which averages 3 months. Once installed, capacity becomes productive and takes on average 36 months to become obsolete. Subjects were notified that the model will take care of ordering replacement capacity so they don't need to think about it. The model starts with 1 unit of capacity installed. Each unit of capacity produces 500 widgets per month and all monthly production enters the inventory, which is initially set at 0 widgets. Inventory is decreased by shipments, which are determined based on the average time to ship, which is determined based on the expected delivery time, given by the ratio between expected backlog and inventory. The average delivery time cannot be lower than 0.5 months.

The other parameter determined by the decision-maker is "*pricing*". Once set, prices are compared to normal price to produce an "*effect of product price on sales*". This effect represents the price elasticity of demand and is also a graphical converter. Based on prices and the number of orders, the model calculates revenues. Total costs are calculated based on the purchases of new capacity, the costs of maintaining capacity and producing widgets, and the costs of sales and shipment. Sales costs assume a fraction of each sale go to the sales personnel.

4.2. Treatments

The two treatments in the experiment differ only by the interface design for their ILE. The models running behind the interface are identical. Furthermore, both participant groups received exactly the same information treatment containing the same instructions on a single page (see *Appendix I*) and a run-through through all variables displayed in their interface.

The no-treatment group, T0, was exposed to a learning environment designed following Forio's implementation of Paich and Sterman's Boom and Bust (1993) mode. A short e-mail exchange with John Sterman himself (in private communication) indicated that this interface was close-enough to the original one, but the original was not made available. A screenshot of the T0 design is available in *Appendix IIa*. The simulation interface was modified slightly to reflect the removal of the competitive dynamics and the marketing investments from the model employed in the experiment.

The treatment group, T1, was exposed to an alternative interface design, displaying the same set of variables. The alternative interface was based on best practices for "*Good Design*" from Tufte (2011) and is available in *Appendix IIb*. The core principle employed in the design was to provide "the greatest number of ideas in the shortest time with the least ink in the smallest place" (Tufte, 2011, p. 51). This implied:

- Reorganization of the information display in order to enable the decision-maker to cover the whole set of information within one screen. Hence, avoiding scrolling and, hopefully, reducing decision effort.

- Focus – this researcher attempted to focus the decision-makers attention on the most key variables by was focused on the most significant variables that represent the dynamics of the system and its performance. This was done by giving them most prominent positioning on the information display. Hence, "*monthly orders*", "*monthly production*", "*cash*", "*net income*", "*inventory*", and "order *backlog*" took prime graphical position in the T1 interface.
- Correct representation – since most variables in SD have a dynamic character, time series plots are the most commonly used presentation format for outcome feedback. However, "time-series displays are at their best for big data sets with real variability – why waste the power of data graphics on simple linear changes, which can usually be better summarized in one or two numbers?" (Tufte, 2011, p. 30). In short – small and non-comparative data sets with (relatively) small variability should be represented in tables. Hence, "*unit price*", "*unit cost*", "*market saturation*", "*investment cost*", and "*delivery time*" had numerical representation in the T1 interface.
- Grouping – The objective of the outcome feedback in an ILE setting is to prompt decision makers to think about causality (Tufte, 1997). "The problem with time-series is that the simple passage of time is not a good explanatory variable: descriptive chronology is not causal explanation" (Tufte, 2011, p. 37). To convey insights through the visual display, we need to urge the decision maker into comparison between variables and between the before and after state of one and the same variable. Hence, related variables were plotted together. Therefore, "*monthly orders*" and "*monthly production*", "*inventory*", and "*order backlog*", and "*unit price*" and "*unit cost*" were grouped together to enforce comparison and causal reasoning.

In their paper, Speier, Vessey, and Valacich (2003) quote past research suggesting that decision makers narrow their attention to focus on relevant cues, and are more likely to focus on "conspicuous" information when cognitive processing demands are high. Hence, the described reorganization and simplification of information display can be expected to influence the decision-process of experiment participants by affecting the information acquisition effort and the causal reasoning.

4.3. Subject groups and Recruitment

Paich and Sterman (1993) claim that "[m]any prior experiments used abstract task or tasks not relevant to the subjects' training and experience". To match the task to the interest and background of the experiment sample, participants were recruited from the Norwegian Business School (NHH) in Bergen. Since the experiment timing coincided with the exam sessions in the Business School, the participants were drawn randomly from the university cafeteria, library, and dorms. The experiment was performed individually with each participant on the spot where they were recruited. A total of 35 people took part, and the majority of them played the simulation game twice. Unfortunately, a very small amount agreed to do a third trial since there were no incentives provided. Hence, the analysis was limited only to data from the two trials.

All except two of the participants had a business or economics related degree. The majority of subjects were at the last year of their bachelor or the second year of their master programs. There was one Phd student, 5 people in their second year of bachelor, and 7 people in their first year of masters. The sample consisted of 10 females and 25 males. Unfortunately a balance between male and female participants was not achieved. However, a test for difference between samples did not indicate performance discrepancy between males and females. Allocation to treatment groups was done randomly and as a result, 17 people were allocated to the no-treatment group (T0), while 18 were allocated to the treatment group (T1). The average time to complete the first attempt was 570 seconds (SD 257 sec), while for the second attempt it was 307 seconds (SD 249 sec). Since the experiment recruitment was relatively random, the group was quite diverse.

4.4. Task Description

Prior research, with a few exceptions (e.g. Moxnes, 2004 and 2009) employ tasks that seem to be fairly complex, involving multiple policy decisions, multiple feedback interdependencies and etc. This author chose to simplify the task in order to detect the effects of the treatment in a more basic context. Since "feedback relates to the properties of the task system and, therefore, is an aspect of task complexity in dynamic tasks" (Atkins, Wood, & Rutgers, 2002), the task simplification was achieved by reducing the number of feedback effects that play out. Hence, the competition segment and marketing budget decisions were fully removed.

After being given the 1 page description of the context and objectives, subjects were introduced to the ILE and asked to formulate and execute a policy, having in mind that once they simulate, they will receive outcome feedback for the following 3 months after accepting the decision. Participants had a total of 20 decisions to make per attempt, amounting to a 60-month simulation time. They could not ask any questions to the researcher and were not aware that they would be given a second attempt to improve their performance. The role of the researcher was only to record the total simulation time and to copy the decision and performance variables once the simulation is complete.

The decisions participants could make included changing their product price and/or their desired capacity on every 3-month period. Desired capacity was described as the amount of machines they would like to have in operation. The model would then take care to buy or sell machines in order to bring the total number to the desired level. Price was the amount of money they wanted to charge for their product and changed immediately. After accepting the decisions and simulating the next 3 month period, the participants would see how all variables in the information display develop in response to their decisions. Then they could adjust their policy in response for the next 3-month period.

5. Results and Analysis

Out of the 35 participants, 22 performed the experiment more than once. Hence, tests which regard any learning effects (H2, and H4) can be performed on the 22 entries. Nevertheless, tests comparing the strategies employed (H1) under different information presentation treatments and their performance implications (H3) can consider the full sample of 35. In *Table 1* on the next page, we can see some summary statistics for key experiment variables. From it, we can notice some signs of difference between the treatment groups, but a more profound inquiry is necessary in order to validate this. Furthermore, we see that the sample size difference has increased in the second trial, which is due to shortcomings of the recruitment process. Hence, we must be aware that potential biases might be introduced by this difference.

	Variable	Treatment group	N	Mean	StDev	Min	Q1	Median	Q3	Maximum
Trial 1	Total Time	both	35	570,3	263,9	230,0	350,0	510,0	796,0	1270,0
	Cumulative Profit	both	35	(334,004)	1,994,335	(8,685,415)	(157,792)	258,945	693,141	1,108,504
	Total Time	T0	17	599,9	300,4	230,0	338,5	510,0	835,0	1270,0
		T1	18	542,3	229,4	230,0	347,5	535,0	682,5	1050,0
	Cumulative Profit	T0	17	(277,794)	1,326,826	(3,810,832)	(683,276)	(44,746)	667,401	1,060,767
		T1	18	(387,091)	2,508,324	(8,685,415)	187,666	422,419	743,217	1,108,504
Trial 2	Total Time	both	22	305,4	146,0	130,0	203,8	240,0	385,0	590,0
	Cumulative Profit	both	22	511,728	412,735	(507,648)	299,888	578,442	764,512	1,142,536
	Total Time	T0	9	315,4	163,9	150,0	204,5	230,0	470,0	590,0
		T1	13	298,5	138,8	130,0	202,5	250,0	400,0	590,0
	Cumulative Profit	T0	9	287,433	307,979	(222,793)	18,949	323,875	572,907	690,293
		T1	13	667,010	413,553	(507,648)	541,971	728,532	935,101	1,142,536

Table 1: Summary statistics for the experiment

To test H1, we need to look for difference between the decision strategies employed by member of T0 and T1. In attempt to simplify the problem and enable for more conventional statistical testing, the individual decisions for each subject were aggregated into 6 variables (3 for pricing and 3 for capacity decisions) that should represent the overall decision strategies. For decisions available in the experiment, the variables are:

- Mean price change (AVGp1 for 1st attempt and AVGp2 for 2nd attempt) describes the average value of all incremental price changes. Incremental implies that every next change is considered as a deviation from the previous price. If price did not change from one period to the next, then its value is not counted towards the average.
- Price change SD (SDp1 for 1st attempt and SDp2 for 2nd attempt) or the standard deviation of price changes measures the variability of changes within the same strategy (decision bundle). Higher SD shows us that the subject has varied their price more, while a lower one indicates that price changes were of a similar size.
- Number of price changes (Countp1 for 1st attempt and Countp2 for 2nd attempt) measures the total number of price changes for the whole simulation period.
- Mean Capacity Change (AVGc1 for 1st attempt and AVGc2 for 2nd attempt) describes the average value of all incremental capacity changes
- Capacity Change SD (SDc1 for 1st attempt and SDc2 for 2nd attempt) measures the variability of capacity changes within the bundle of decisions
- Number of capacity changes (Countc1 for 1st attempt and Countc2 for 2nd attempt) measures the total number of capacity changes for the whole simulation period

A categorical box-plot for the Price (*Figure 3*) and Capacity (*Figure 4*) variables for the first attempts shows us that there are actually some small differences between treatment groups with regard to strategy bundles. Moreover, there seem to be multiple outliers that are quite separate from the rest of the sample population. It seems like T1 has induced slightly more variability of price changes than T0, while T0 has induced a slightly higher number of price changes than T1. To explore this further, a Kruskal-Wallis test was used to examine the difference between the sample medians. The reason for employing a non-parametrical test was that none of the variables examined was following a normal distribution (Kruskal & Wallis, 1952). The test showed no evidence supporting the existence of difference between the medians for Mean Price Change and Number of Price changes for T0 and T1. However, some evidence was found that suggests that there is difference in the Standard Deviation of price changes where the 0 hypothesis of no difference between the treatment groups was rejected with P = 0,017. No evidence was found for differences between T0 and T1 for any of the Capacity variables. Hence, there is not enough evidence to support H1.

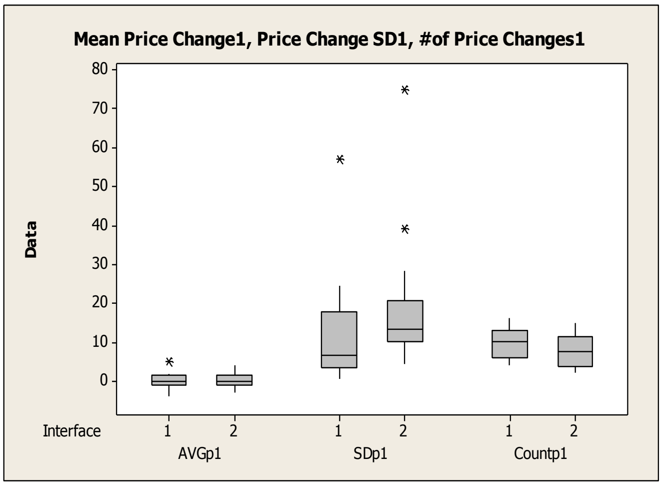


FIGURE 3

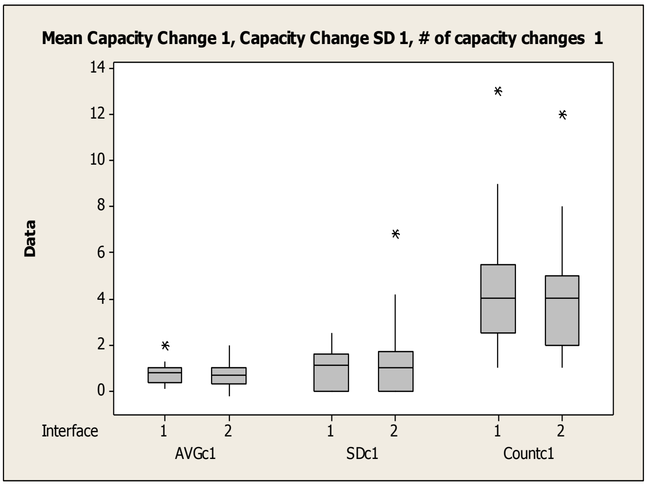


FIGURE 4

Performing the same test for H2 showed no difference between T0 and T1 for any of the price and capacity variables in the second attempts (see *Figure 5* and *6* on the next page for the box plots of price and capacity strategy variables). Hence H2 can also be rejected.

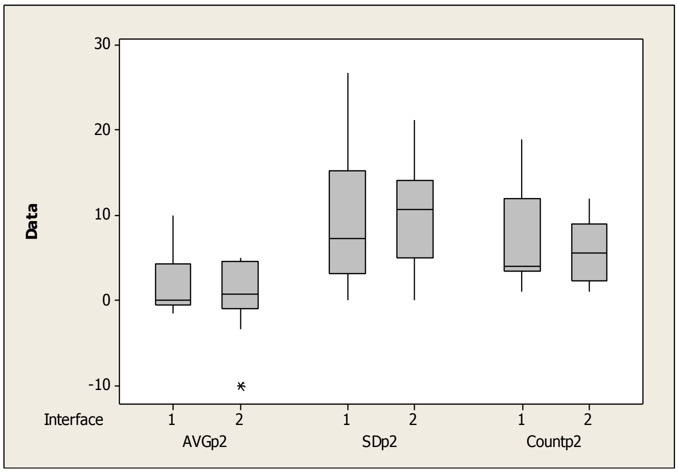


FIGURE 5

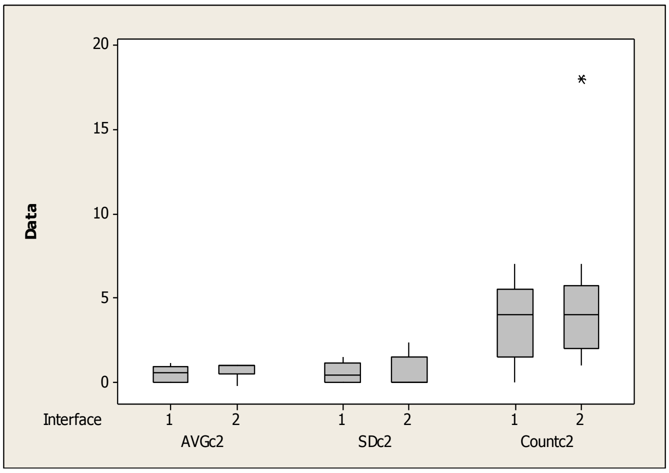


FIGURE 6

Examining H3, we need to compare decision performance in terms of cumulative profit between T0 and T1. Before doing statistical testing, a categorical box plot for performance is produced to compare the two samples (Figure 7). What immediately captures the eye is the number of outliers, which are significantly apart from the rest of the observations. Further exploration of the data shows that the three most significant outliers were in the lower quartile when it comes to time spent in the simulation. In addition, looking at the individual decisions we can see that in the most extreme case the decision-maker (Subject 13) tolerated a negative price margin throughout the whole simulation and kept increasing capacity and lowering price, although his/her inventory was growing and he was losing money. With the outliers included in the sample, the test for performance difference between T0 and T1 was not sufficient to reject the 0 hypothesis of no difference at 95%

confidence level, but it was close to the 90%. However, after removing all 5 outliers, support was found with P=0,049, suggesting a 95% confidence interval for the difference of M(T0) - M(T1) of (-848 439; -28 479), N1 (T0) =15 and N2 (T1) = 15.

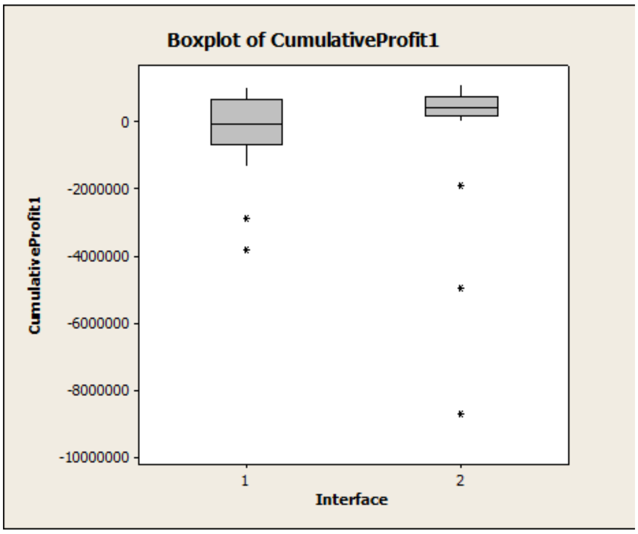


FIGURE 7

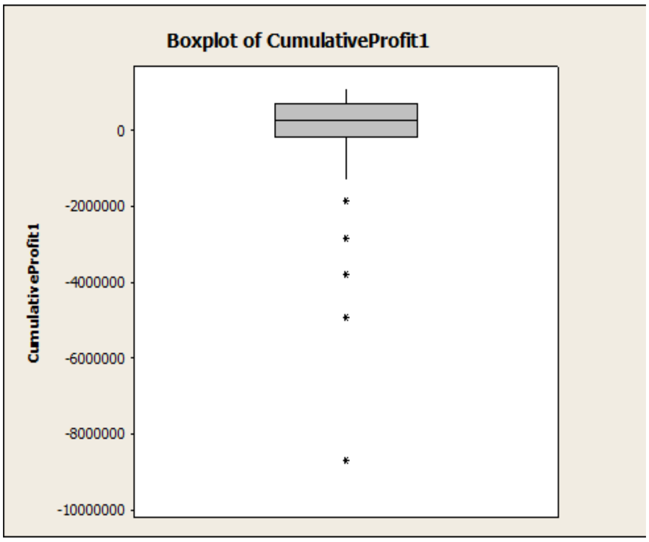


FIGURE 8

Concerning H4, we need to look at the difference between performance in T1 and T0 for the second attempt with the simulator. If this difference is larger than the one for the first attempt, then we will have support for H4. Looking at the categorical box-plot of observations of Cumulative Profit from the second attempt (*Figure 9*) we can see that there is only one outlier, which is an observation for the Interface 2 group (T1). This is a tolerable amount of outliers so the observation is kept for the statistical test. As the sample distribution was also not normal for this variable, a Kruskal-Wallis non-parametric test was applied again to test for differences between the medians. The test was able to reject the 0-Hipothesis with P = 0.01, giving confidence intervals for the difference of (-701 759; -57 395).While the lower boundary is higher than for Cumulative Profit 1, we see that the median and the upper boundary are lower. Hence there is evidence in support of H4, suggesting that participants exposed to T1 have experienced higher learning effects than those of T0. Nevertheless, it is important to mention that learning effects might be due to participants from T1 memorizing the effective strategies from the first trial better than in T0.

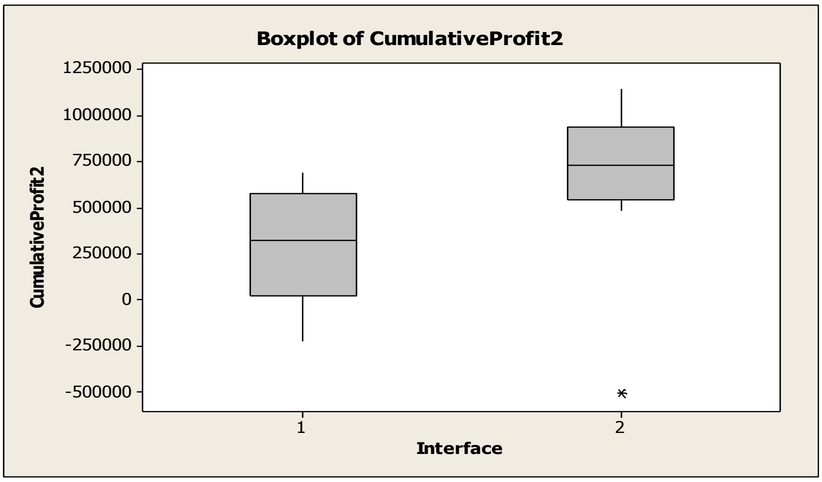


FIGURE 9

From a qualitative perspective, the experiment induced some interesting behaviour and comments regarding the employed simulation and its interfaces. Most participants were excited and impressed with the notion of being able to receive realistic and immediate outcome feedback from their decisions. However, they all expressed feeling of being overwhelmed with the quantity of data and the complexity of the data display, stating that "*There are simply too many things to keep an eye on*". This was most prominent in the no-treatment group, where participants complained about having too much information bunched together, and having to scroll down. A major part of the participants had problems understanding the way inventory or backlog worked. About 50% clearly stated that they wanted to get their backlog down to 0, not understanding that every incoming order passes through the backlog and that at any time in a continuous process, there are some orders that have not been fulfilled yet. This pushed some of the participants into building large inventories in attempt to clear their backlog. Another common misunderstanding was the fact that lowering price to boost sales would work only as a strategy if there is available inventory to meet the sales increase. Since customers were sensitive to time delay, lowering the price without having backup inventory lead to a rapid spike in expected time to deliver, which pushed away some of the customers. Only a few of the experiment participants grasped this notion and managed their pricing strategy together with their production strategy. In fact, only about a quarter of the participants expressed awareness that they had to manage both pricing and production in conjunction with each other. About half of the participants embarked first on a price searching strategy, trying to find the optimum market price and then tried to produce large quantities to sell at this price.

6. Conclusions and Limitations

Information presentation characteristics are shown to have important implications for decision-making in static environments. However, the review of relevant literature identified very few studies looking at the choice of strategy, performance, and learning implication of outcome feedback format in a dynamic environment characterized by feedback loops. This paper presents an experiment that aimed to test if applying information presentation design principles to an Interactive Learning Environment interface could affect (a) The strategies chosen to deal with the problem at hand, (b) The performance within the problem context, and (c) The learning effects from multiple attempts on the problem. To examine the aforementioned aspects, this researcher exposed 35 experiment participants to two alternative treatments. The no treatment group was exposed to an interface similar to the original from Paich and Sterman's (1993) Boom and Bust model, and the treatment group was exposed to an interface on which data display design best practices were applied to improve the quality of information presentation. To test for learning effects, part of the sample (22 people) was asked to perform the experiment again.

Results from the experiment did not provide support for the existence of a difference between the bundles of strategies chosen within the two treatment groups (H1). No support was found for the formulation of different procedural knowledge as a result from the interaction with the different interfaces either (H2). Hence, participant's strategy bundles, employed for dealing with the problem at hand, were not proven to be influenced by the way data was fed-back to them. It is possible for this finding to stem from a limitation of the method chosen to test for differences. Since the 20 individual decisions for each participant were bundled together in 3 single measures, the timing component of the individual decisions was left out. As timing of decisions has an important implication within a dynamic setting, in retrospection this might not have been the most appropriate method for examining the difference between strategies. In fact, an interval plot of decisions within individual timeslots shows that there might actually be differences between the groups (see *Appendix IIIA and B*). In the graphs provided within Appendix IIIA and IIIB we can clearly see that the overall increases and decreases in price and capacity happen at different times for the two different treatment groups.

Considering performance, the experiment results provided some evidence for performance differences between T0 and T1 for the first simulation attempt (H3). However, those performance differences were detected only after extreme outliers were removed due to their abnormally low values. The reasoning behind this was that participants who displayed this abnormal performance took too little time to perform the experiment and hence might not have taken it seriously. This might have been due to a misinterpretation of the objective, or simply due to lack of motivation stemming from limitations of the recruitment process. In summary, H3 is not supported counting the outliers, and is supported without them.

With regards to learning, there was sufficient evidence that the group allocated to T1 performed better in their second trial than participants from T0. While this result might be simply due to the fact that subjects in T1 were able to better remember the general dynamics of the underlying system, it does in fact indicate that the improved interface has facilitated the learning process. Although the purpose of most SD interventions, with or without an ILE, is to enable decision makers to learn about the underlying system structure, it would be beneficial if the simulator interface facilitates that process through reducing the decision effort. Having evidence in support of H3 and H4 also indicates that the strategies employed by T0 and T1 must have been different, since they led to different performance. Support for this is given in Appendix III A and B, where we can see a box-plot of the quarterly individual pricing decisions from the first and second attempt of T0 (Interface 1) and T1 (Interface 2).

This research has shown that information presentation design characteristics can have an impact on performance, learning, and choice of decision strategy. Nevertheless, its results should be considered with caution. Due to limitations of the research scheduling, the centralized organization of the experiment was impossible. Hence, the sample of participants was quite diverse, in terms of their study majors and years in university. Moreover, since there was no dedicated room for the experiment and participants were pulled out of their daily tasks, their motivation to perform well can be put to doubt. At the same time, the gender distribution of the samples was not biased, and hence – more male than female participants were recruited. This however, did not show to have an effect on the performance results. In addition, since there was no budget allocated to the experiment, no incentives were provided to the participants. Hence, motivation was further decreased and only 63% of the participants chose to do the experiment more than once. This reduced the size of the sample for testing H2 and H4 and introduced a bigger difference between sample sizes for T1 and T0. Hence, this could have lowered the strength of the results. Furthermore, the modifications applied to simplify the original model limit the possibility of comparison across different studies. Namely, a direct comparison with results from Paich and Sterman's Boom and Bust experiment (1993) is impossible. Overall, both T0 and T1 samples performed well with only a few dramatic exceptions.

Nevertheless, the study shows that the careful design of information presentation could potentially influence the decision-making process and lead to performance gains within Interactive Learning Environments. To turn this around, part of the misperception of feedback within ILE contexts could possibly be attributed to the poor display of outcome feedback. In addition, regardless of the consideration if learning differences occur because of the ease of getting used to the interface or because of some deeper understanding of causality evidence of such differences was found. One could argue that positive learning effects matter as long as a logical connection is established between the actions (decisions) performed within the environment and the observed behaviour. So long as this connection is facilitated through the interface design, then researchers and practitioners should take more care when preparing our ILE's for an intervention or a learning session.

Further research into the field could try to identify what is the effect of each of the specific interface design characteristics on the decision-making process. Moreover, it could try to dwell deeper into the differences between individual decision-points to see if interface design facilitates the early spotting of some behavioural cues. Timing of decisions is another aspect that requires further investigation as it might provide more substantial proof of the rise of alternative decision strategies as a result of different interface design. In any case, further research should rely on a better experiment set-up and a bigger and more balanced sample in order to provide more valid and reliable results.

WORKS CITED

Andersen, D., Chung, I., Richardson, G., & Stewart, T. (1990). Issues in Designing Interactive Games Based on System Dynamics Models. *System Dynamics Review* , 31-45.

Atkins, P., Wood, R., & Rutgers, P. (2002). The effects of feedback format on dynamic decision making. *Organizational Behavior and Human Decision Processes Vol. 88* , 587–604.

Davidsen, P. (2000). Issues in the Design and Use of System-Dynamics-Based Interactive Learning Environments. *Simulation Gaming, v31* , 170-177.

Dickson, G. W., DeSanctis, G., & McBride, D. J. (1986). Understanding The Effects of Computer Graphics for Decision Support: A Cumulative Experimental Approach. *Management of Computing, V26, N1* , 40-47.

Einhorn, H., & Hogarth, R. (1981). Behavioural Decision Theory: Processes of Judgment and Choice. *Annual Review of Psychology, V32* , 53-88.

Hogarth, R., & Einhorn, H. (1992). Order Effects in Belief Updating: The Belief Adjustment Model. *Cognitive Psychology, v24* , 1-55.

Jarvenpaa, S. L. (1989). The Effect of Task Demands and Graphical Format on Information Processing Strategies. *Management Science, Vol. 35, No. 3* , 285-303.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

Kleinmuntz, D. N., & Schkade, D. A. (1993). Information Displays and Decision Processes. *Psychological Science, Vol. 4, No. 4* , 221-227.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association, V47, N260* , 583-621.

Lane, D. C. (1995). On a Resurgence of Management Simulations and Games. *The Journal of the Operational Research Society, V46, N5* , 604-625.

Lawless, K. A., & Brown, S. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science, v25* , 117-131.

Lucas, H., & Nielsen, N. (1980). The impact of the mode of information presentation on learninig and performance. *Management Science, v26, N10* , 982-993.

Moxnes, E. (2004). Misperception of basic dynamics: the case of renewable resource management. *System Dynamics Review v20, 2* , 139-150.

Moxnes, E. (1998). Overexploitation of renewable resources: The role of misperceptions. *Journal of Economic Behavior & Organization v37* , 107-127.

Moxnes, E., & Saysel, A. K. (2009). Misperceptions of global climate change: information policies. *Climatic Change, v93* , 15-37.

Moyaux, T., Chaib-draa, B., & D'Amours, S. (2007). Information Sharing as a Coordination Mechanism for Reducing the Bullwhip E®ect in a Supply Chain. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 37, NO. 3,* , 396-407.

Paich, M., & Sterman, J. (1993). Boom, bust, and failures to learn in experimental markets. *Management Science Vol. 39, No. 12* , 1439-1458.

Payne, J. (1982). Contingent Decision Behavior. *Psychological Bulletin* , 382-402.

Playfair, W. (1786). *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century.* London: Unknown.

Simon, H. A. (1957). *Models of Man: Social and Rational.* New York: John Wiley and Sons, Inc.

Slovic, P. (1995). The construction of preference. *American Psychologist* , 364-371.

Speier, C., Vessey, I., & Valacich, J. (2003). The Effects of Interruptions, Task Complexity, and Informati on Presentationon Computer-Supported Decision-Making Performance. *Decision Sciences, V34, N4* , 771-797.

Sterman, J. (1994). Learning in and about complex systems. *System Dynamics Review V10, N2* , 291-330.

Sterman, J. (1989). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decisionmaking Experiment. *Management science V:35/3* , 321-339.

Sterman, J. (1992). Teaching Takes Off: Flight Simulators for Management Education". *OR/MS Today* , 40 -44.

Sterman, J., & Sweeney, L. (2002). Cloudy skies: assessing public understanding of global warming. *System Dynamics Review v18* , 207–240.

Tufte, E. (2011). *The visual Display of Quantitative Information. Second Edition.* Cheshire, Connecticut: Graphics Press.

Tufte, E. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, Connecticut: 1997.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185(4157)* , 1124-31.

Vicente, K. (1996). Improving dynamic decision making in complex systems through ecological interface design: a research overview. *System Dynamics Review, v12* , 251–279.

APPENDIX I – INSTRUCTION TO PARTICIPANTS

Production company management simulation

You've been hired as a CEO of a new electronics manufacturing company that develops widgets for a technology market. According to your business expertise, your firm's market is subject to S-shaped development (e.g. markets experiencing similar behavior are microwaves, fridges, laptops, and etc). Such markets are characterized by powerful positive feedbacks such as learning curves, scale economies, and network effects, implying that the initial market leader can gain cumulative advantage allowing them to dominate the market.

You must make choices regarding price and capacity expansion as to maximize cumulative profit over a time period of 5 years (60 months). You can make your choices once every 3 months and will then be provided with a report on the development of your company within three months after your decisions have been made.

Consumers in your market buy a single unit (per consumer) of your product and, after its useful lifetime expires, discard it and become, once again, potential customers. You know that there is a limited ultimate market, but you don't know exactly how large that market is. As this is a new product, you are also ignorant of other crucial facts, such as how often consumers will repurchase the product (the average product lifetime), the response of the customers to the availability of the product (e.g. expected delivery delays), the effect of different pricing on customer's purchase behavior, and the shape of your learning curve. Your task in the game is to devise a robust strategy for managing the firm that allows you to maximize your total cash at the end of the simulated period in a dynamic environment based on the information you receive as feedback from the simulator's interface.

Since your product is quite innovative and has been patented, you have no direct competitors in the market for the entire timeframe of the simulation.

In addition to the opportunity to learn about substantive issues in corporate strategy related to product life cycle management and positive feedback economics, the assignment gives you firsthand experience with the use of simulation games or "management flight simulators" as tools for learning about complex dynamics.

Production data:

- You start with 1 machine (unit of capacity) at hand.
- 1 unit of capacity is capable of producing 500 widgets per month
- The average time from ordering a new machine (unit of capacity) to getting it delivered, installed and fully operational is 3 months
- The useful life of a machine is 36 months. Machines will renew automatically to sustain target capacity
- All produced widgets within a month will be stored in your warehouse and represent your inventory

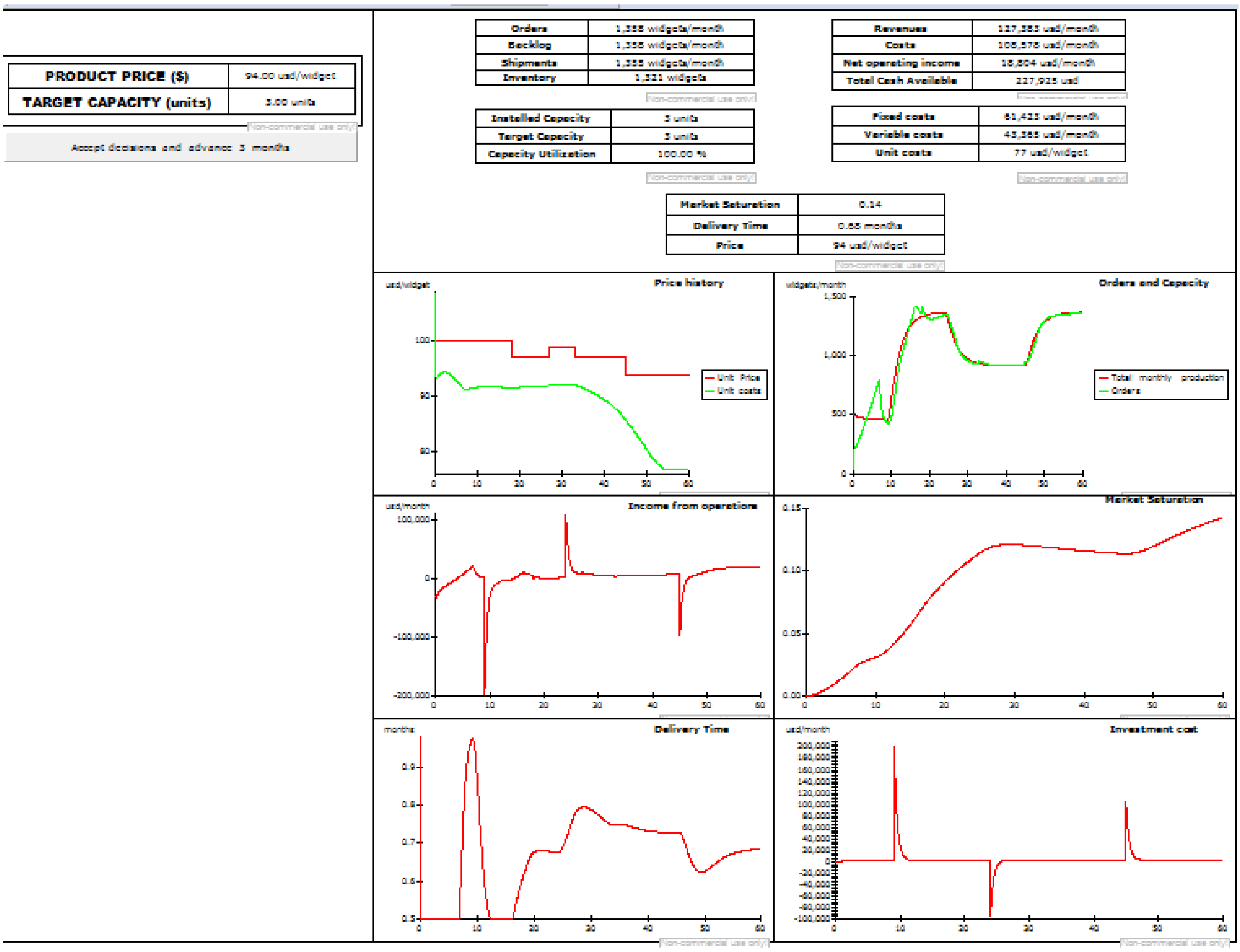
Customer segment:

- Customers who order the product expect it to be delivered within 1 month
- An increase in the expected time to deliver (a delivery delay) will reduce customer's willingness to buy your product. A decrease will have the opposite effect up to a point
- Orders that have not yet been fulfilled (product has not been delivered) are stored in a backlog
- The backlog is cleared by shipments (deliveries)
- Customers talk to each-other and customers who have recently purchased your product are likely to motivate other potential customers to buy it

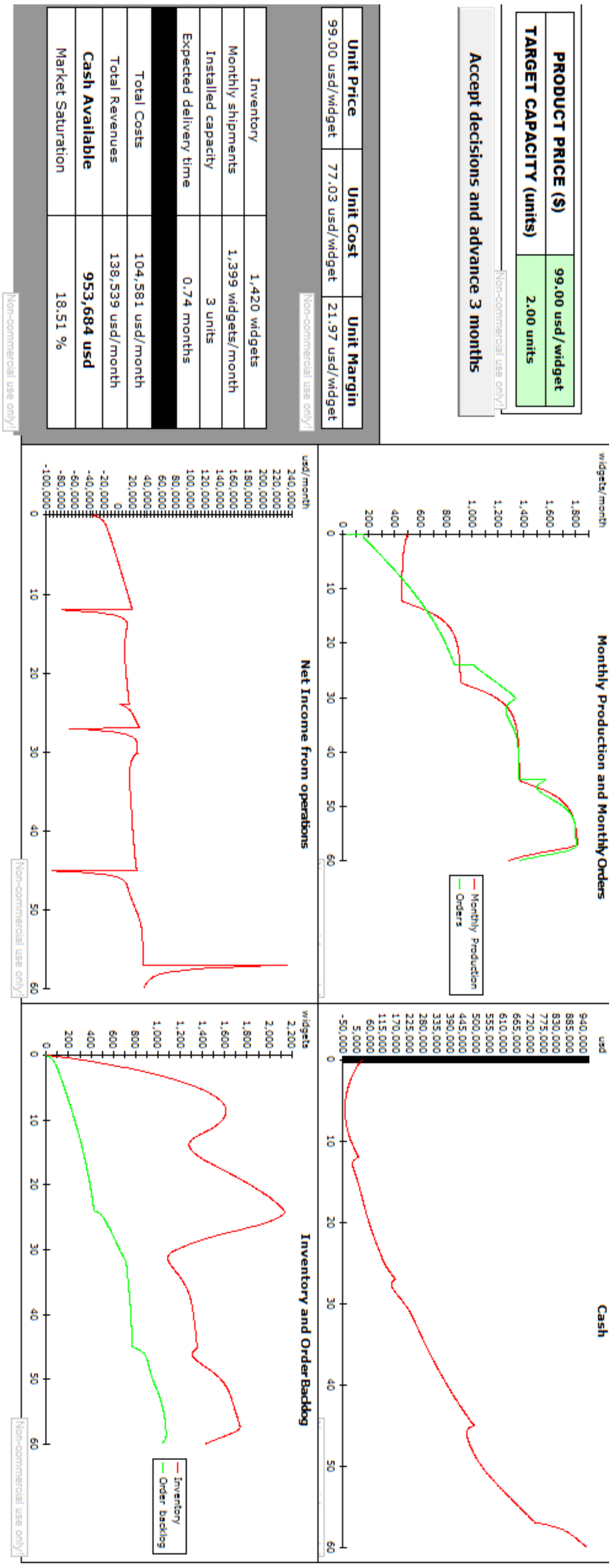
Costs:

- A new machine costs 50,000 usd per unit
- Machines have a rough operating cost of 30,000 usd per month if operating at full capacity. Lower capacity lowers the operating costs proportionally
- The cost per shipment (transportation cost and overhead) is a constant of 5 usd per widget
- Your sales cost is a constant of 25 usd per widget
- The inventory holding cost is a constant of 2 usd per widget per month
- You start with a total of 30,000 usd of cash available to run the business
- Your cash can go negative (taking credits from the bank) without incurring interest

APPENDIX IIA – BASE INTERFACE (T0)

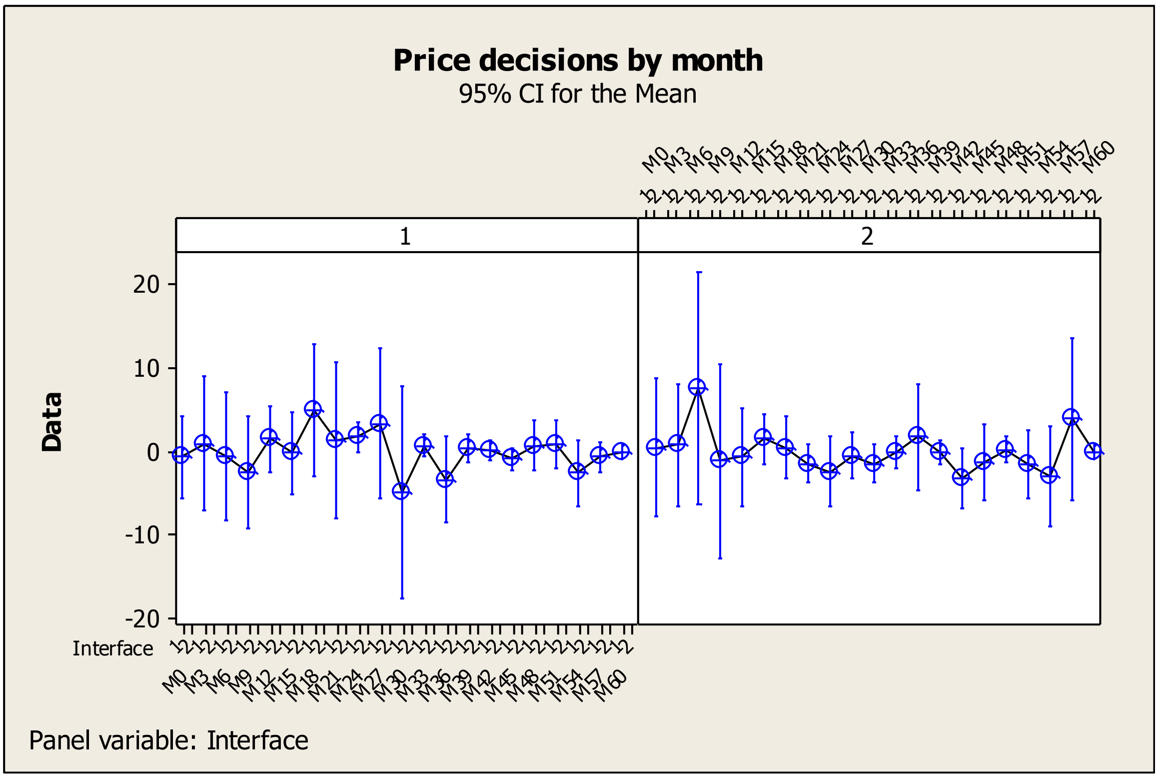


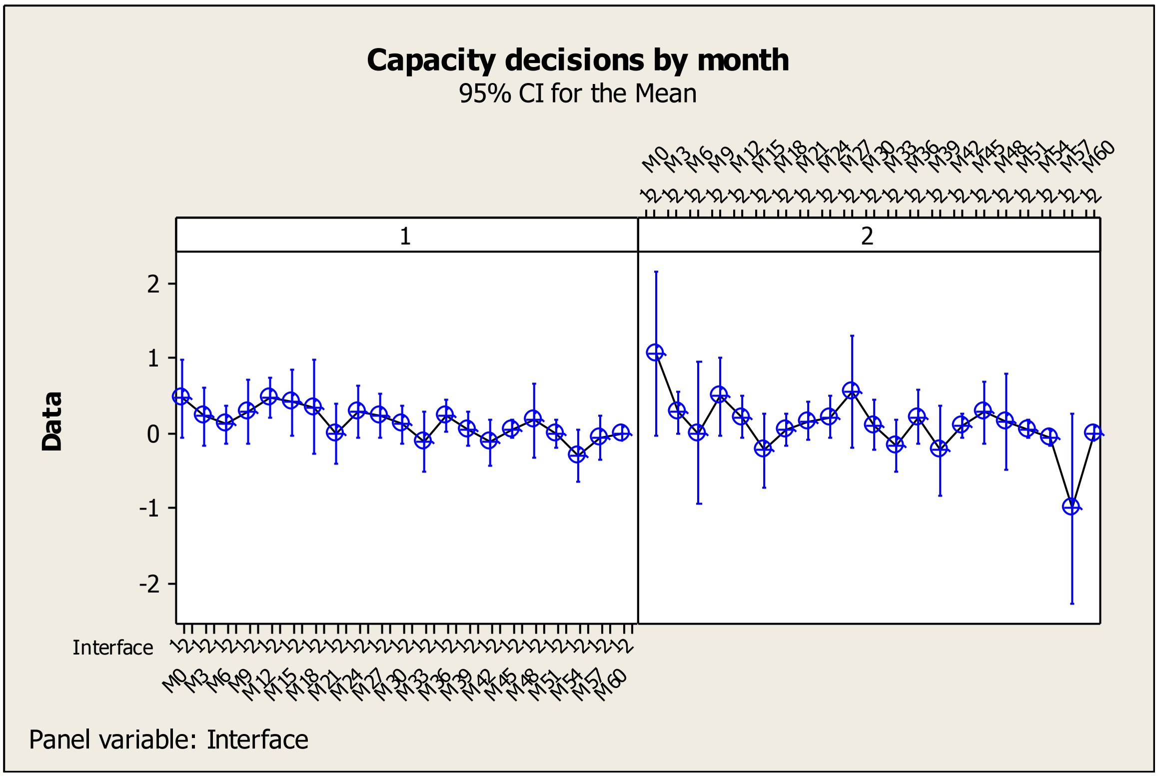
APPENDIX IIB – ALTERNATIVE INTERFACE (T1)



APPENDIX IIIA

TIMING OF DECISIONS FOR THE FIRST ATTEMPT FOR T0 AND T1





APPENDIX IIIB

TIMING OF DECISIONS FOR SECOND ATTEMPT FOR T0 AND T1