Learning to Control Dynamic Systems: Information Utilization and Future Planning

Wai-Tat Fu (wfu@uiuc.edu)
Human Factors Division and Beckman Institute
University of Illinois at Urbana-Champaign
1 Airport Road, Savoy, IL 61874

Cleotilde Gonzalez (conzalez@andrew.cmu.edu)
Department of Social and Decision Sciences
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Abstract

Dynamic systems involve states that change both autonomously and as a consequence of the learner's actions. Research has shown surprisingly poor performance when people learn to control dynamic systems. Many researchers have proposed that learners often misperceive the feedback provided by the dynamic system, although it is still unclear how the feedback is misperceived and what can be done to improve learning. In two experiments, we studied learning behavior in a dynamic system called the beer game. We found that performance did improve through repeated practice, but subjects had a strong tendency to ignore the temporal dynamics in the system. Concurrent verbal reports indicated that performance improved mostly through better utilization of information related to the temporal dynamics of the system. As a consequence, subjects learned to anticipate changes in the system. In the second experiment, we provided only information that was critical for their decisions and found that initial performance was significantly better, indicating faster learning. It is concluded that poor utilization of information that is critical to the temporal dynamics of the system and insufficient anticipation of system changes are the major problems for learning in dynamic systems.

Introduction

Recent research on decision making has shed significant light on human behavior in a variety of microeconomic contexts. Despite its success of explaining behavior in static and discrete decisions or judgments, relatively little work has been done to study decision behavior in dynamic systems. One of the characteristics of dynamic systems is that actions of the decision maker often cause, either directly or indirectly, changes in the system itself, which then affect the effects of future actions. For example, a firm's decision to increase production feeds back through the market to influence the demand and price of goods; greater output may also tighten the markets for labor and raw materials to which competitors may react – all influencing future production decisions. Such multiple feedbacks, some of them with time delays, are arguably the norm rather than the exception in real problems of decision making.

Behavioral research on dynamic systems shows surprisingly poor performance, sometimes even with practice (e.g., Berry & Broadbent, 1984; Brehmer, 1992; Jensen & Brehmer, 2003; Kerstholt & Raaijmakers, 1997; Sterman, 1989). One possible reason for the poor performance is the individuals' inability to incorporate delayed feedback into their decisions (Brehmer & Allard, 1991). In particular, people often fail to account for actions which had been initiated but not yet had their effects. In many cases, people have to respond to a new stimulus before they receive feedback from the previous trial. In addition, people often attribute the dynamics they experience to external events, when in fact these dynamics are internally generated by their own actions. This "open-loop" mental model (Sterman, 1989) is believed to hinder learning of the temporal dynamics of complex systems. Indeed, we believe that the temporal dynamics inherent in the system create a very difficult credit-assignment problem for the learners: positive or negative outcomes of actions have to be associated with the corresponding earlier actions (e.g., see Fu & Anderson, 2006 a, b) so that better choice of actions can be made in the future. The credit-assignment problem not only requires a reinforcement-learning mechanism (Fu & Anderson, 2006b), but also an appropriate mental representation of how the states of the dynamic system may change wit different actions.

We chose to study subjects' behavior in a simulated supply chain management system. Supply chain management is a common and simple concept: your customer orders products from you; you keep track of what you're selling, you order enough raw materials from your suppliers to meet your customers' demand and keep your inventory and backorders as low as possible. Although the concept is simple, the dynamics of the input-output relationship in the whole supply chain have known to cause significant difficulties for people to perform optimally (e.g., Croson & Donohue, 2002). Coordination and communication between suppliers, manufactures, and wholesalers is often considered the main difficulty in supply chain management (Croson and Donohue, 2000). However,

our focus in this paper is on the psychology of decision making that emerge from individual behavior.

Misperception of feedback

In its strongest form, the misperception of feedback hypothesis implies that people simply cannot learn to control dynamically complex systems. Indeed, researchers often demonstrate that individuals cannot understand the 'basic building blocks' of systems thinking such as the concept of stocks and flows (e.g., Sweeney & Sterman, 2000). On the other hand, significant learning is observed in complex dynamic systems, suggesting that although people may not understand the building blocks of dynamic systems, extended practice may give them the opportunity to accumulate experiences with the relationships between control inputs and system outputs, utilize relevant information that will affect their performance and dynamics of the system, and how to engage in future planning to anticipate common situations (Kerstholt and Raaijmakers, 1997). The goal of this paper is to collect detailed protocol data to understand the changes underlying the improvement in performance as subjects learn to control a dynamic system.

One possible explanation for the negative effects of information delays is that people do not detect there are feedback delays despite the fact that they had all the information that they need to infer them. For example, in the experiments by Brehmer and Allard (1991), although subjects consistently reported having detected that outcomes of their actions were delayed, most subjects could not infer the *nature* of the delays and failed to adopt the appropriate strategy to compensate for the delays. In fact, Brehmer and Allard found that subjects simply adopted the same strategy in situations when there were significant delays *and* in situations when there was no delay.

We focus on two questions related to the learning of temporal dynamics: (1) what information do people utilize to make decisions in a dynamic situation, and how the utilization of information change with experience, and (2) what are the major differences in terms of strategies or processes when we compare learning behavior between a static and a dynamic situation. To preview our results, we found that people tended to ignore the temporal dynamics initially, and as a consequence failed to utilize information that indirectly influenced the outcome of their decisions. In addition, we found that future planning was essential to anticipate changes as well as outcomes of actions in dynamic systems, and it often took a significant amount of experience for people to learn to engage in future planning.

Supply Chain Management: The Beer Game

We collected empirical data from individuals as they performed a supply chain management task called the "Beer Game" (Sterman, 1989). The beer game represents a simplified supply chain consisting of a single retailer who supplies beer to a consumer (simulated as an external demand function), a single wholesaler who supplies beer to a retailer, a distributor who supplies the wholesaler, and a factory that brews the beer (it obtains it from an inexhaustible external supply) and supplies the distributor. We developed a computerized version of the beer game that was used in all the experiments reported in this paper. A screenshot of this simulation is presented in Figure 1.

In the original version of the game, individuals play the game in groups of four, with each participant playing the role of one of the four facilities. Their goal is to minimize the cost for the entire supply chain. Each player contributes to this goal by ordering beer from their respective supplier in a manner that maintains enough beer in their respective inventory to meet the demand from their respective customer (i.e., the facility they supply, or the consumer in the case of the retailer).

The customer's order is filled with available inventory, and then the player orders more beer from their supplier to replenish the loss from their inventory. Difficulties arise when players must anticipate demand, as there is a one-week delay between when an order is placed and when the supplier receives the order. Assuming that the supplier has enough inventories, there is an additional two-week transportation delay before the player receives the ordered beer. If the supplier's inventory is too small to fill the order, additional delays will occur.

Costs accrue as follows. Each week, each player is charged a 50¢ holding fee for each case of beer in their inventory. If inventory is too small to meet demand, the shortage is backlogged to be filled as soon as possible. Players are charged a weekly $1 shortage fee for each case of backordered beer. The basic strategy, therefore, is to minimize inventory while avoiding backorders. The dynamics of the beer game make successful performance difficult.

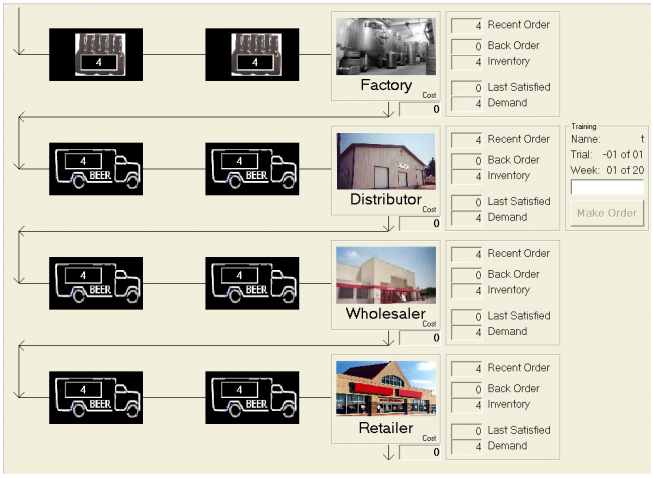

Figure 1.A screenshot of the beer game simulation.

The Temporal Dynamics in the Beer Game

Figure 2 shows the simplified temporal dynamics involved in the beer game when the player is deciding on how much to order from the factory (i.e., the player works

as the distributor). Once the order is placed, it will go to the factory, but there is a one-week delay (i.e., the "recent order" box) before it reaches the factory. When the factory sends out the beer, it will be in the supply line (i.e., the arrows and the trucks between the Factory and the Distributor in Figure 1) and it takes two weeks before the beer can be used to satisfy the demand from the wholesaler. The current inventory (if any) or backorder will be updated after beer is sent to the wholesaler.

In the experiments reported here, subjects played the role of the distributor, and decided how much to order from the factory. The order took one week to reach the factory. After the order was received, the factory sent the beer to the distributor, which took 2 weeks. The dotted arrow indicates a common misperception, that the order placed would directly influence the inventory/backorder without delay, ignoring the temporal dynamics in the system.

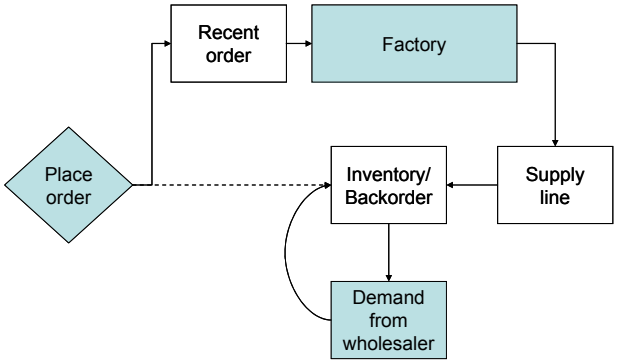

Figure 2. The temporal dynamics in the beer game and the common misperception by subjects.

The retailer, wholesaler and factory were played by the same computer algorithm, which ordered the demand associated to the position. Thus, variability was not added to the external customer demand as it propagated upstream through the supply chain.

Previous studies show that people tended to under-weigh the supply line, which eventually led to large fluctuations in inventories (Sterman, 1989; Croson & Donohue, 2002). In other words, instead of developing the mental model as shown in Figure 2, people tend to assume a direct influence from the order to the inventory/backorder (i.e., the dotted arrow in Figure 2), instead of the delayed influence through the factory and the supply line. The underweighting of the supply line seems to be persistent and robust across participants. Surprisingly, although the underweighting of the supply line was identified as the major reason of poor performance in supply chain management, there is, to our knowledge, still no good understanding of why this underweighting occurs.

The Bullwhip Effect

The bullwhip effect is a real-world phenomenon that involves oscillations of net inventory (i.e., inventory – backorders) at each level of the supply chain and amplification of those oscillations as one moves farther up the chain (Croson and Donohue, 2002). The amplification and oscillations are very costly, but unfortunately the bullwhip effect is very common and it has been treated as an inevitable effect of every supply chain.

In multiple experiments, Sterman (1989, 2004) has demonstrated the bullwhip effect in the laboratory, using the beer game. Analyzing individual behavior he has concluded that individuals do not *learn* to control the system because they often fail to account adequately for the supply line (they misperceive the feedback provided by dynamic systems). Thus, in responding to high demand, players increase their orders too much leading to excess inventory.

Researchers have identified several causes for the bullwhip effect. Rational decision makers must use current demand to forecast future demand in an effort to control the impact of order delays, transport delays, production delays, etc. on inventory. Forecasts based on simple ordering formulae (e.g., moving averages) lead to the bullwhip effect. Ordering in batches (e.g., monthly instead of daily) can also create the bullwhip effect. Other causes include fluctuating prices which lead to forward buying, and rationing where suppliers divide limited inventory among customers who then inflate their orders to get a bigger share (Croson & Donohue, 2002).

The Beer Game is much simpler than real world supply chains. Since prices are fixed players have no incentive for forward buying. The frequency with which orders are placed is fixed at one per week. This prevents order batching. Each facility in the supply chain has only one customer. Thus rationing is not possible. In addition, in the scenario commonly used in the beer game experiments, external consumer demand starts at a constant of 4 cases of beer per week and then jumps to a constant of 8 cases per week at the fifth week and remains there for the remainder of what is typically a 52 week scenario.

Experiment 1

Our experiment required playing the beer game for 20 trials, where each trial used the standard 52-week scenario. The experiment, therefore, required a total of 1,040 ordering decisions in contrast to the typical single-trial experiment that requires a one-time run of 52 weeks and thus 52 ordering decisions.

Method

20 subjects were recruited from the Carnegie Mellon University community. Subjects were paid $15 for their participation.

To familiarize subjects with the system they played a short 20-week training scenario with a constant demand increase from 4 to 8 at week 5. The purpose of the training scenario was to illustrate how to order from the factory and how the inventory and backorder were calculated as they progressed. All other questions are also answered during this time. To stimulate active learning and to encourage subjects to aim at reducing the total cost, subjects were required to keep the inventory and backorder below 20, and the total cost below 100. If subjects failed to meet any of

these criteria, they were asked to repeat the training scenario. Subjects repeated the training scenario 3.4 times on average.

After finishing the training scenario, subjects played the standard 52-week scenario for 20 trials. In a standard scenario, the demand from the wholesaler started at 4, increased to 8, back to 4, decreased to 2, and then went back to 4 and stayed there until the end of the trial. The weeks at which the demand change occurred were noisy, so that across trials, subjects could not simply recall when the changes would occur. Specifically, an integer was randomly selected from the range from -2 to +2 and the selected integer was added to the weeks when changes occurred.

To understand changes in performance with practice, concurrent verbal protocols were collected from half of the subjects during trial 1, 11, and 20. Subjects were asked to "think aloud" when they were playing the games in these trials. They were specifically told to mention all information they were utilizing on the screen as well as in memory, all mental calculations, and all reasoning that they used during the task. All verbal utterances, the screen, and their actions were recorded as "movie" files by a computer program.

Results

Figure 3 shows the mean net inventory of the subjects in trial 1, 11, and 20. Subjects in trial 1 had large fluctuations in their net inventory. The largest fluctuation was near week 20, where the demand was decreased from 8 to 4 for the first time, and then further decreased to 2 in week 30. The inventory peaked at approximately 30 in trial 1, but the peak was reduced to approximately 10 in trial 11. In trial 20, the peak was further reduced to below 5 throughout the weeks. The mean total cost after 52 weeks was 372.35, 152.54, and 91.01 for trial 1, 11, and 20 respectively. A two-tailed paired t-test shows that the differences were significant ($t(10)=6.24$, $p<0.001$; $t(10)=2.64$, $p<0.05$), indicating learning across trials. Since the cost for inventory was lower than that for backorder, there was a bias towards keeping an inventory, as shown in Figure 3.

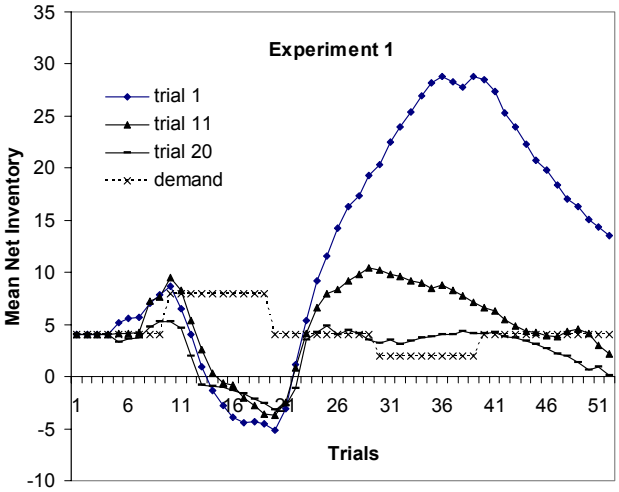

Figure 3. The demand and the mean net inventory in trial 1, 11, and 20 in Experiment 1.

The verbal protocols were transcribed and coded according to what information was used to determine the order to the factory. We found that they fell into 8 categories: Inventory/Backorder, recent order, total cost (of distributor), supply line, demand for the distributor (demand_distributor), demand for the retailer (demand_retailer), demand for others (demand_others), and inventory and backorder for others (inventory_others). We also noted instances where subjects explicitly expressed future planning or prediction. For example, as when subjects expressed things of the kind "I know that the demand is going to increase in a few weeks". The coding results from the verbal protocols are shown in Table 1.

Table 1 shows that in trial 1, subjects were mostly utilizing the demand from the wholesaler (i.e., demand_distributor), and their own inventories and/or backorders, and relatively under-utilized the supply line, and the demand to the retailer. This is consistent with previous findings (e.g., Sterman, 1989) that subjects tended to ignore the temporal dynamics of the system and attempted to find direct, immediate feedback between actions and their effects. In this case, subjects almost ignored the fact that to reduce the inventories/backorders, the supply line was as important as the current demand and the inventory/backorder. The fact that they under-utilized the demand to the retailer suggested that they were not looking ahead to how demand might have changed in the near future. In fact, we believe the high fluctuations shown in Figure 3 could be well explained by this lack of future planning in the first trial.

Table 1. The mean frequencies of use of information for each order to the factory in trial 1, 11, and 20.

Categories	Trial 1	Trial 11	Trial 20
Future planning	0.5	4	7
Demand_distributor	17.9	13.5	8
Supply Line	2.1	5.2	6.6
Demand_retailer	0.6	4.3	5.2
Inventory/Backorder	16.9	9.8	5.3
Total cost	1.1	0.2	0.1
Recent orders	2	0.4	0
Demand_others	0.8	0	0
Inventory_others	2	0.2	0

With practice, we found it striking that subject started to increasingly utilize the supply line and the demand to the retailer, and they had also learned to have more future planning, suggesting that they were adapting to the temporal dynamics of the system. They had also reduced the utilization of the demand from the wholesaler and the inventories/backorders. The utilization of other "irrelevant" information, such as the demand and inventories to others (e.g., the wholesaler or factory), were also reduced. The differences in future planning and inventory/backorder were significant ($t(10)=3.56$, $p<0.01$; $t(10)=2.12$, $p<0.05$ and $t(10)=2.30$, $p<0.05$; $t(10)=0.05$ respectively). Although the other differences were not significant due to the small

number of subjects, the general pattern was consistent with this interpretation.

The results from Experiment 1 showed that with extended practice, subjects learned to improved performance significantly, even when the demand changes were variable. This finding supported the notion that subjects had learned to generalize from experience and were able to anticipate demand changes. Results from verbal protocols showed that initially, subjects utilized information that was not critical for their decisions and rarely engaged in future planning. With practice, subjects learned to utilize most of the critical information and ignored most of the "irrelevant" information, and were engaged in future planning to anticipate changes in external demand.

Experiment 2

Results from Experiment 1 suggest that subjects might have misperceived the system dynamics by ignoring the temporal dynamics and assuming immediate effects of actions. As a result, they did not fully utilize information that was important in controlling their own inventories/backorders. With practice, subjects learned to utilize the important information and ignore "irrelevant" information (information that did not directly affect the decision on how many to order from the factory). In Experiment 2, we further tested the idea that poor performance in early trials was caused by poor utilization of relevant information. In Experiment 2, we removed most of the information that was not critical for the calculation for the decision (i.e., the demand, recent orders, costs, and inventories/backorders of the retailers, the wholesaler, and the factory.) and just provided subjects with the important information. Our prediction was that providing only relevant information would help subjects to focus on figuring out the temporal dynamics of the system, resulting in better performance (low oscillations).

Method

20 subjects were recruited from the Carnegie Mellon University community. Subjects were paid $15 for their participation. The procedures were the same as those in Experiment 1, except that only the demand from the wholesaler, their own inventories/backorders, the supply line, and the demand to the retailer were provided to the subjects. All other information was not available on the screen.

Results

Figure 4 shows the mean net inventories in trial 1, 11, and 20 in Experiment 2. Comparing it to the oscillations in Figure 3, one can clearly see that the fluctuations in trial 1 were much lower. Indeed, after 52 weeks, the total cost for trial 1 was 218.16. The difference in the first trial between the two experiments was significant ($t(19)=5.37$, $p < 0.001$), indicating that performance when only important information was shown was better than when information for all players was also provided in Experiment 1. The total costs for trial 11 and 20 were 156.87 and 85.67 respectively. The differences of total costs between the three trials were significant ($t(19)=2.46$, $p<0.05$; $t(19)=4.79$, $p<0.001$), indicating learning across trials.

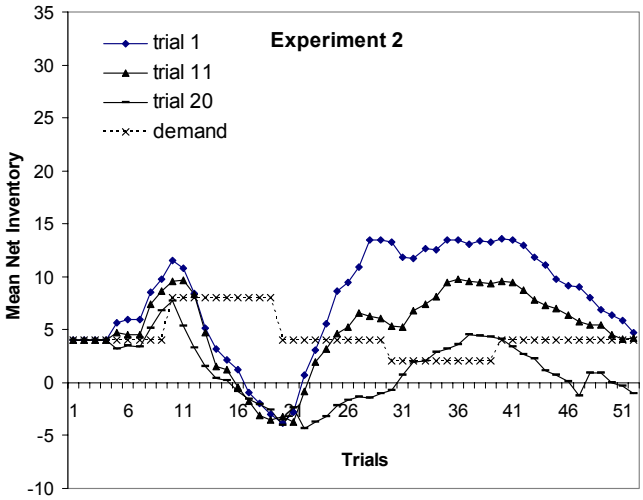

Figure 4. The demand and the mean net inventory in trial 1, 11, and 20 in Experiment 2.

Figure 5 shows the learning trends for both experiments. We can clearly see the difference between the two experiments in the first 10 trials. In Experiment 1, when irrelevant information was available, subjects started out with a much higher total cost, and slowly started to reduce the total cost across trials. On the other hand, in Experiment 2, when only relevant information was available, subjects started out with a much smaller total cost and learned across trials to reduce it. At around trial 10, subjects in both experiments reached roughly the same level, although they kept improving at a similar rate from that point onwards.

As we showed in the analyses of the verbal protocols, subjects in Experiment 1 started out utilizing their own inventory/backorder and the demand from the wholesaler. This was a clear indication that subjects were ignoring the temporal dynamics and were assuming a "closed-loop" system with static relationship between their actions and their effects. The lack of understanding of the temporal dynamics was also supported by their lack of anticipation of the changes in customer demand, which eventually propagated through the supply chain and affect the demand from the wholesaler and their inventories.

The difference in the first 10 trials between the two experiments suggests that the absence of irrelevant information helps subjects to learn the temporal dynamics of the system. In fact, it probably took subjects roughly 10 trials to figure out what information was important, and perhaps after that they started to understand how the supply line and customer demand may affect the temporal dynamics of the system.

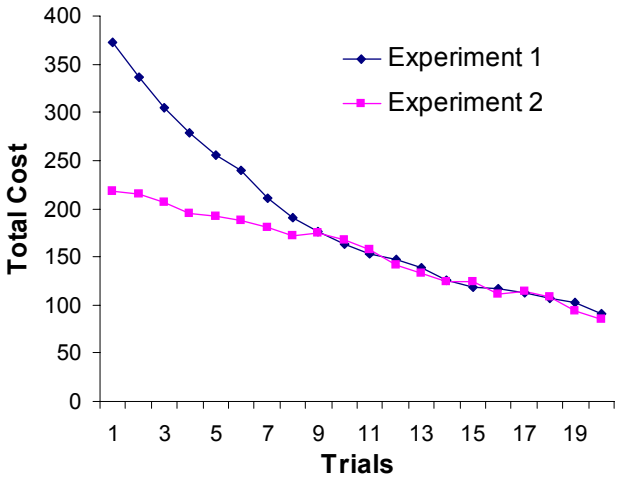

Figure 5. The total cost at the end of each of the 20 trials in both experiments.

Discussions

Consistent with previous results (Steman, 1989), we found that subjects had trouble dealing with the long time delays between placing and receiving orders – the supply line. The results show that initially, most subjects failed to account for the supply line adequately, which has been considered a major cause for poor performance. However, we found that, with practice, subjects learned to utilize the supply line and to anticipate the customer demand, and learned to ignore other information, especially those that were not critical for their decisions on how many to order from the factory.

The results reinforce and extend prior work in dynamic decision-making (Brehmer, 1987; Hogarth, 1981, Sterman, 1989). A heuristic may produce stable behavior in one setting and oscillation in another solely as a function of the feedback structure in which it is embedded. In general, we found that subjects had a strong tendency to assume a static environment in which directional input-output (i.e., action and its effect) exists. This is consistent with previous results in which subjects were found to *implicitly* learn to perform better (by simple input-output association) without *explicit* awareness of the dynamics of the systems (e.g., Berry & Broadbent, 1984). In our studies, we found that it took subjects a long period of training to (1) utilize the right set of information that are relevant to the temporal dynamics of the system, and (2) anticipate future demand changes by having sufficient future planning to incorporate their understanding of the dynamics into the appropriate actions.

Our results show that the current approach is useful in understanding complex dynamic systems. We found that the change in utilization of information as a major factor for poor learning, which had been neglected by previous research using simple aggregate methods such as regression and statistical techniques in operations research (e.g., Croson & Donohue, 2002). Our results show that misperception of feedback is an incomplete explanation of learning difficulties: people need to develop the strategies needed to *compensate* for the delays. The development of these compensation strategies require clear understanding of the states of the systems which requires prolonged experiences. Indeed, our results suggest that people often have difficulty recognizing what information is relevant for the temporal dynamics and how those information may help them anticipate future changes in the system. We show that by directing the focus on relevant information, learning can be much more effective.

Acknowledgments

This research was partially supported by the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory (DAAD19-01-2-0009) and the Multidisciplinary University Research Initiative grant from the Army Research Office (W911NF-05-1-0153). We thank Jack Lim for programming the beer game simulation, Polina Vanyukov for collecting and transcribing the verbal protocols, and Aanand Radia for collecting part of the data in Experiment 1.

References

Berry, D.C. & Broadbent, D.E. (1984). On the relationship between task performance and associated verbalized knowledge. Quarterly Journal of Experimental Psychology, 36, 209-231.

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

Brehmer, B., & Allard, R. (1991). Real-time, dynamic decision making: The effects of complexity and feedback delays. In J. Rasmussen, B Brehmer, & J. Leplat (Eds.), *Distributed Decision Making: Cognitive models of cooperative work*. New York: Wiley.

Croson, R. & Donohue, K. (2002). Experimental economics and supply chain management. Interfaces, 32, 74-82.

Fu, W.-T. & Anderson, J. R. (2006a). From Recurrent Choice to Skilled Learning: A Reinforcement Learning Model . *Journal of Experimental Psychology: General*, 2.

Fu, W.-T., & Anderson, J. R. (2006b). *Solving the credit-assignment problem: The interaction of explicit and implicit learning with internal and external state information.* In Proceedings of the 28th Annual Conference of the Cognitive Science Society.

Jensen, E., & Brehmer, B. (2003). Understanding and control of a simple dynamic system. System Dynamics Review, 19, 119-137.

Kerstholt, J.H. & Raaijmakers J.G.W. (1997). Decision making in dynamic task environments. In R. Ranyard, W.R. Crozier, & O. Svenson (Eds.), Decision making: Cognitive models and explanations. Ablex: Norwood, NJ.

Sterman, J. (1989). Misperceptions of feedback in dynamic decision making. Organizational Behavior and Human Decision Processes, 43(3), 301-335.

Sweeney, L.B., & Sterman, J.D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. System Dynamics Review, 16, 249-286.