Learning to Make Decisions in Dynamic Environments: ACT-R Plays the Beer Game

Michael K. Martin (mkmartin@andrew.cmu.edu)
Dynamic Decision Making Laboratory
Department of Social and Decision Sciences, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Cleotilde Gonzalez (conzalez@andrew.cmu.edu)
Dynamic Decision Making Laboratory
Department of Social and Decision Sciences, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Christian Lebiere (clebiere@maad.com)
Micro Analysis and Design
Boulder, CO, USA

Abstract

Sterman (1989) proposed that decision makers misperceive the feedback provided by dynamically complex environments, and questioned whether people can learn to make effective decisions in such environments. We provide empirical evidence of learning in a well-known dynamic environment called the beer game. We then describe a preliminary version of an instance-based, dynamic decision making model built using the ACT-R cognitive architecture. The model mimics the general patterns of human behavior observed for aggregate performance across trials and local performance within trials. Implications for research on dynamic decision making are summarized.

Introduction

Dynamic Decision Making (DDM) requires a series of interdependent decisions in an environment whose state evolves over time (see Brehmer, 1992, for a review of DDM). Dynamic decisions often involve choosing control inputs for a dynamic system in a manner that achieves or maintains a desired system state (e.g., a state of equilibrium).

The beer game is a dynamic system used extensively to study the way decision makers perform when confronted by dynamic complexity. Thousands of people from all over the world, ranging from high school students to chief executive officers and government officials, have played the beer game to learn the basic concepts of operations management (Sterman, 2004).

The beer game is not really about beer, and it is not really a game. It is a learning environment of the type called management flight simulators (Sterman, 2004). It provides players an interactive experience that demonstrates the impact of time delays and feedback loops on supply-chain management, and more generally, on coordination among levels in an organization.

In particular this game has been used to demonstrate the *bullwhip* effect, a costly real world phenomenon in which orders oscillate, in increasing amplitude, as one moves farther up the supply chain (Croson and Donohue, 2002). Sterman (1989, 2004) has demonstrated the bullwhip effect in multiple beer game experiments, and has concluded that individuals do not learn to control the system because they misperceive the feedback provided by dynamic systems. Similar results and misperception-of-feedback explanations can be found in other studies (see Croson & Donohue, 2002, for a review of beer game experiments).

We contend that participants in previous experiments performed poorly simply because they did not have enough practice with the system, giving them little opportunity to learn. Proficient DDM typically requires *extended practice* with a system, presumably because it gives decision makers a chance to learn the system dynamics important for control (Kerstholt and Raaijmakers, 1997).

This paper contributes to the current state of affairs in two ways. First, it provides evidence that people learn to adequately control the supply chain when given extended practice. Second, it offers an explanation as to *how* people learn to control the system by providing an ACT-R cognitive model of the learning process.

In the next section we describe the beer game and bullwhip effect in more detail. We then present our study on the effect of extended practice. Next we present the ACT-R cognitive model and comparisons between the model and human. Finally we conclude and present future directions for research.

The Beer Game

The beer game represents a simplified supply chain consisting of a single retailer who supplies beer to consumers (simulated as an external demand function), a single wholesaler who supplies beer to the 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.

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).

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. Each week, each player receives an order from their customer, starting with the retailer and working upstream in the supply chain toward the factory. 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 because 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 inventory, 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.

The Bullwhip Effect and Experimental Economics

Researchers have identified several causes for the bullwhip effect (Croson & Donohue, 2002). 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.

The beer game is much simpler than real world supply chains. Players have no incentive for forward buying because prices are fixed. Order batching is less likely because the frequency with which orders are placed is fixed at one per week. Rationing is not possible because each facility in the supply chain has only one customer. Finally, in the standard scenario, 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.

Sterman (1989) demonstrated that the bullwhip effect emerges even though the beer game presents participants with a nearly ideal supply chain; participants' orders oscillated, and grew in amplitude as orders propagated upstream. This produced oscillations in each participant's net inventory (i.e., inventory – backorders), which also grew in amplitude the farther the facility was from the external consumer. The end result was a supply chain whose operations costs exceeded "optimal" costs by almost 10-fold.

Based on this finding, along with similar findings from experiments with simulations of other supply chains, Sterman (1989) concluded that people misperceive the feedback provided by dynamic systems. According to the misperception of feedback hypothesis, people lack the cognitive machinery to comprehend the dynamic complexity produced by the causal and temporal relationships among system variables. Dynamic complexity is created by delays in a system's response (e.g., transport and order delays), feedback loops, stocks and flows, and nonlinear relationships among system variables. All are commonly found in dynamic systems, and all are present in the beer game.

Extended Practice Experiment

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., Jensen & Brehmer, 2003; Sweeney & Sterman, 2000). This position however, cannot explain how experts in the real world can perform effectively in highly complex dynamic systems such as air traffic control.

A possibility we address here is that although people may not understand the building blocks of dynamic systems, extended practice may help individuals learn to control a dynamic system because it gives them the opportunity to learn the relationships between control inputs and system outputs, and how to anticipate common situations (Kerstholt and Raaijmakers, 1997).

Our experiment required playing the beer game for 20 trials, where each trial used the standard 52-week scenario (described above). 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.

This experiment simplified game play in two ways. First, participants played alone rather than in teams. Participants played the role of the distributor and the computer played the remaining roles. Second, the computerized players simply ordered the demand. Thus, variability was not added to the external customer demand as it propagated upstream through the supply chain.

Method

Participants. Thirteen Carnegie Mellon University students participated for payment. Participants were paid a base rate of $10, plus performance bonuses of up to $16 (see below).

Procedure. We developed a computerized version of the beer game that presents information in the same way as the in the Systems Dynamics Group www site (http://beergame.mit.edu/). A screenshot of this simulation is presented in Figure 1.

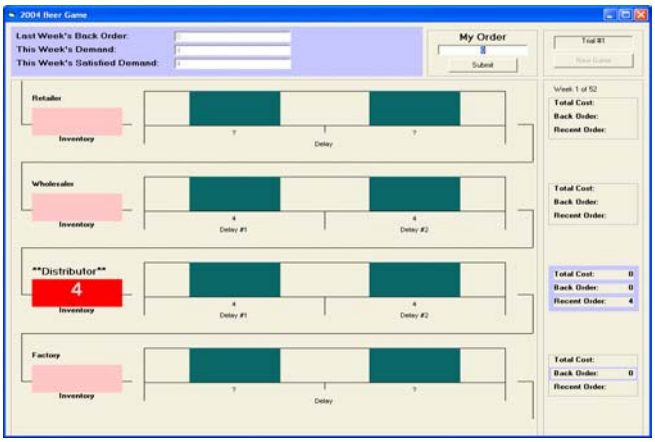

Figure 1: Screenshot of the Beer Game Simulation

The simulation provided information only about the inventory and supply line of the role played by the participant (distributor). Also, only the participant's cumulative cost was displayed. As in the www simulation, the last week's back order, and this week's demand and satisfied demands were displayed.

Participants played the 52-week scenario 20 times. They were instructed to minimize their total cost by ordering beer each week in a manner that allowed them to meet their customer's demand (i.e., the wholesaler's weekly orders). They were told about the cumulative weekly charges, the one week ordering delay, the two week transportation delay, and the possibility that if their supplier (i.e., the factory) could not fill their order, the transportation delay would be longer because of the time it takes the factory to transport raw materials..

The bonus pay schedule was then described. Trials were divided into four blocks of five. A $4 bonus was given for each block of trials in which the designated performance target was achieved at least once. Performance targets (total costs), based on 11 pilot study participants, grew more stringent over the time course of the experiment. The performance targets for blocks 1-4 were total costs of 750, 650, 550, and 450, respectively. (The minimum total cost possible was 396; there were no practical limitations on maximum total cost possible.)

To familiarize participants with the system they played a short 10-week scenario with random external demand. Questions were addressed during this time. Afterward, they played the standard scenario 20 times.

Results

One participant did not complete the 20 trials, so their data set was not considered subsequently. The data set of a second participant was removed after an outlier analysis.

Figure 2 shows the mean cost per trial. A one-way repeated-measures ANOVA using total cost as a dependent variable indicates that performance improved with practice, $F(19,190) = 3.4$, $p < .05$. Helmert contrasts (e.g., Judd & McClelland, 1989) indicate that performance gradually improved until about the ninth trial.

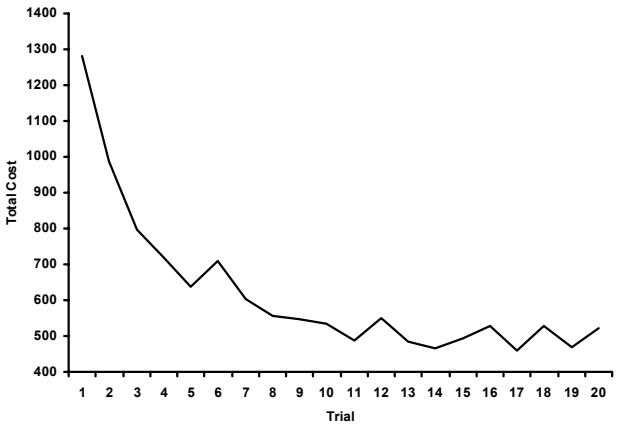

Figure 2: Cumulative Cost as a Function of Practice

Figures 3, 4 and 5 depict performance within trials 1, 9, and 20 respectively. Each shows net inventory (inventory – backorders) across the time course of the 52-week scenario. A net inventory of 0 is ideal.

As Figure 3 shows, our participants exhibited the same behavior as that reported in previous studies. The net inventory oscillates around the ideal of 0. The large deviations from 0, in turn, produce high total costs.

The 3-week delay between placing and receiving orders inevitably leads to back-orders when external consumer demand jumps from 4 to 8 cases per week. (The distributor sees the jump at week 7.) This sudden increase in demand creates a shortage which must be corrected by ordering more beer than indicated by current demand. Too much beer is ordered, creating a slight overshoot in ideal inventory as indicated by the second cycle of positive net inventory. To correct for the overshoot, orders are cut back below current demand, creating yet another cycle of inventory shortages.

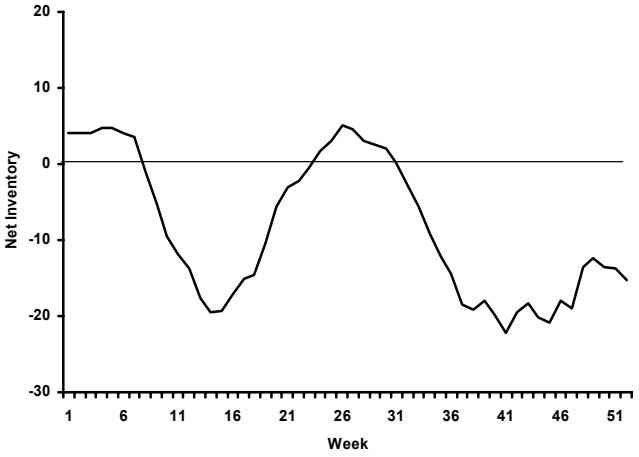

Figure 3: Net Inventory per Week in Trial 1

As with the control of any system with response delays, the only way to avoid oscillations in net inventory is to anticipate demand. Figure 4 shows that by Trial 9 the oscillations in net inventory are still present but participants have learned to dampen them. As can be seen, they anticipate the step increase in external consumer demand and build inventory prior to the increase in demand. The

build-up, however, is not yet sufficient, which leads to back-orders and negative net inventory. They continue to overcorrect for back-orders, as indicated by the second cycle of positive net inventory.

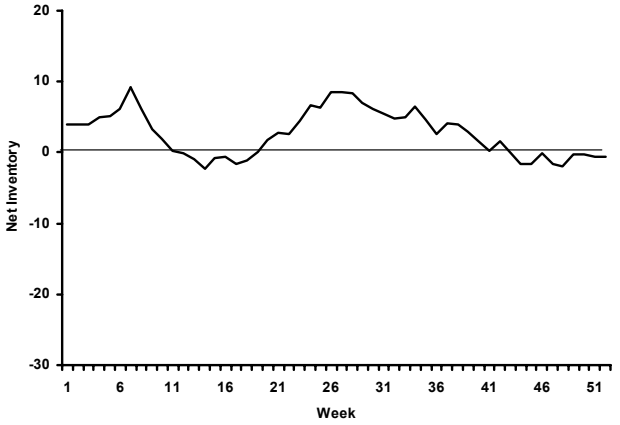

Figure 4: Net Inventory per Week in Trial 9

Figure 5 shows that participants have learned to mostly avoid oscillations in net inventory by Trial 20. The dampening of oscillations between Trials 9 and 20 seems to appear because participants have learned how to correct for back-orders without overshooting the desired net inventory of 0.

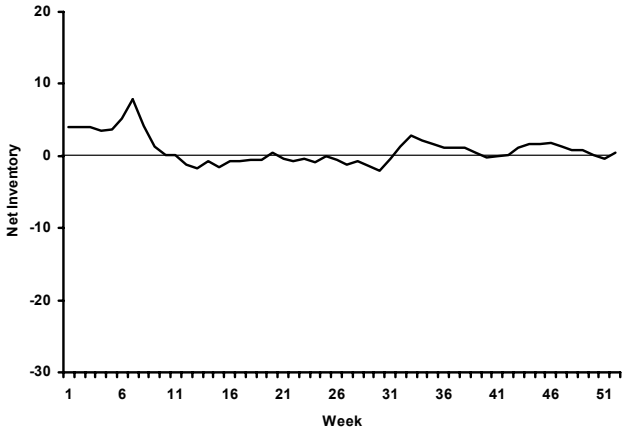

Figure 5: Net Inventory per Week in Trial 20

ACT-R Plays the Beer Game

Our participants learned to play the beer game. But what did they learn, and how did they do it? Gonzalez, Lerch, and Lebiere (2003) proposed Instance-Based Learning Theory (IBLT) to account for DDM performance and concurrent learning processes. IBLT has been successfully applied to multiple dynamic tasks including the Sugar Production Factory and the Tansportation task among others (see Gonzalez and Lebiere, in press).

The gist of IBLT is that dynamic decisions are made by comparing current situations with previously experienced situations. If a similar situation is recalled, the decision associated with that situation is used as an anchor that is adjusted to fit the current situation. Learning occurs as decision makers gradually shift from using simple decision making heuristics to the instance-based anchoring and adjustment process.

IBLT, as implemented in ACT-R, provides a simple explanation of the observed dissociation between verbalizable knowledge and DDM performance (e.g., Berry & Broadbent, 1984). According to IBLT each judgment of an alternative creates an instance, which is represented as a chunk in declarative memory in ACT-R. The slots in the chunks represent the situation, the decision made, and the expected utility of that decision. As declarative knowledge, each instance can be verbalized. However, the subsymbolic parameters that control the retrieval and application of instances (e.g., base-level activation, similarity among chunks, and strengths of association) are not consciously accessible. These subsymbolic parameters represent implicit knowledge of the system, and underlie DDM performance. The implication is that DDM tasks can be learned without explicitly encoding structural and temporal relationships among system variables.

In accordance with IBLT, we enforced the following constraints for modeling beer game performance in ACT-R. First, we represented information only if it was directly available to participants. Second, we represented information only if participants paid attention to it – as indicated by think-aloud protocols from two additional beer game participants. Third, we avoided clever engineering by using only those cognitive mechanisms inherent in ACT-R. This includes using recommended default values for all parameters.

We have also imposed two additional constraints on our modeling efforts to date. The declarative chunks described by Gonzalez et al. (2003) contained slots that represented expected utility. In that model, feedback mechanisms were used to adjust expected utilities. Subsequent application of those instances then depended on their expected utility. We do not include slots for expected utility in the beer game model because of the complications arising from delayed feedback, and the difficulties associated with determining utility. The second additional constraint is that the model reported here uses partial matching only. Base-level learning and blending mechanisms, as used in Gonzalez et al. (2003), have not been used so far.

Because the model operates in a task where contextual attributes vary continuously (e.g., the number of cases of beer in inventory, back-order, etc.), exact matches between context and relevant instances are rare. Partial matching provides a mechanism for retrieving chunks with attribute values that are similar to the current context. Thus, relevant chunks can be retrieved even though they do not exactly match the retrieval cues provided by the current context (i.e., the values of the slots in the goal buffer).

Specifically, the chunk with the highest match score will be retrieved if its activation is higher than the retrieval threshold (-1.0 in our case), where match score M_{ip} is a function of the activation of chunk i in production p (including transient activation noise, .25 in our case) and its degree of mismatch to the desired values:

$$M_{ip} = A_i - MP\sum_{v,d}(1 - Sim(v,d))$$

In the partial matching equation above, *MP* is a mismatch penalty constant (1.5 in our case), while *Sim(v,d)* represents the similarity between the desired value *v* in the goal and the actual value *d* in the retrieved chunk. We used a negatively accelerated similarity function.

The Model

Based on performance, it appears that participants learned: (1) to anticipate the increase in demand and (2) to adjust the size of their orders so that the amplitude of oscillations in net inventory progressively decrease. For our model, we started with the simple heuristic of ordering the demand to replace inventory losses. Verbal protocols indicated that participants frequently examined back-orders and/or inventory immediately after placing an order – even though the change due to that order would not occur until at least 3 weeks later. This observation prompted the addition of slots that represented the changes in back-order and inventory. We then added several more simple heuristics that increase or decrease the base order (i.e., order the demand) according to changes in back-order and/or inventory. These heuristics form the core of the model, and are engaged in the creation of all instances.

At the beginning of each ordering cycle, the model assesses changes in inventory and back-orders, and then attempts to retrieve a relevant instance from declarative memory. The retrieval cue is constructed by projecting the current state of the system onto the next state. That is, current inventory is multiplied by the inventory change that occurred upon entering the current state to produce an expected inventory. An expected back-order is constructed similarly. Expected inventory and expected back-order are then used as retrieval cues.

If the retrieval fails, the heuristics described above are applied to the current demand. If the retrieval is successful, three pieces of information from the projected state are used to construct the current order. First, the demand slot from the projected state indicates the expected demand. The expected demand becomes the current base order. (Notice that this is similar to the first heuristic we created, if it is recognized that expected demand equals current demand in unfamiliar situations.) Retrieval of expected demand thus provides a mechanism by which the model can learn to anticipate the increase in demand.

The next two pieces of information correspond to the changes in inventory and back-orders that produced the projected state. These may be thought of as the size of the adjustments that lead into the projected state, and thus the size of the adjustment that should be made to the current base order.

Results

The results reported herein use the mean of 11 simulated subjects based on the model described above, each playing the beer game 20 times in the standard scenario as human participants did.

The model's mean learning curve approximates the humans' mean learning curve in terms of Total Cost, r^2 = .875 (see Figure 6). The model does not perform quite as well as humans but it appears to learn more quickly than humans do. The addition of blending might be expected to help with both of these defects.

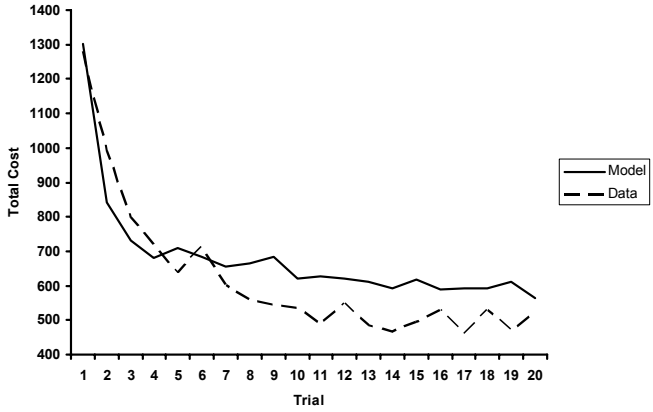

Figure 6: Practice Effect for Model and Humans

Building an ACT-R model that exhibits a learning curve for an aggregate performance measure (i.e., total cost) is fairly straightforward. It is more important for our current efforts that the model learns to control inventory in a manner consistent with that demonstrated by our participants. We can assess this by examining how the patterns of net inventory over weeks in the scenario match those produced by humans.

Figures 7, 8 and 9 depict the model's mean performance in terms of net inventory for trials 1, 9, and 20 respectively. The pattern of the model's performance in trial 1 (see Figure 7) closely mimics that produced by humans. It exhibits the large oscillations in net inventory, along with the overcorrections demonstrated by humans. One difference in the pattern is that the model's cycles of net inventory oscillations have greater amplitude than those of humans. The model also appears to be already learning to dampen the oscillations in net inventory, whereas humans demonstrated a second cycle that was roughly of the same amplitude as their first.

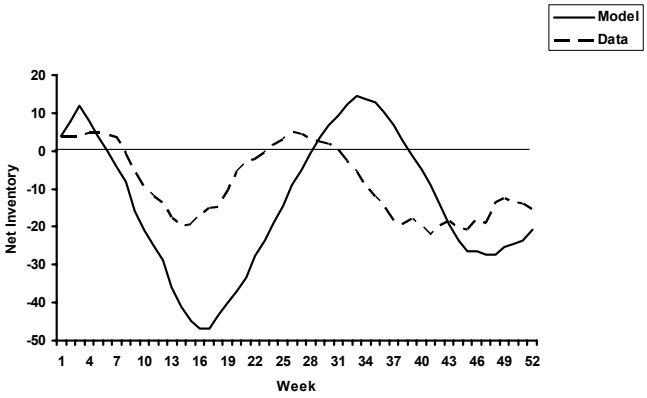

Figure 7: Model's Net Inventory per Week for Trial 1.

By trial 9 the model, like the humans, has learned to partially anticipate the increase in demand, and has learned how to decrease the amplitude of the oscillations in net inventory (see Figure 8). Overall, the pattern of the model's

performance is similar to that of humans. One difference is that humans tended to be biased toward a positive inventory, whereas the model appears to be biased toward a negative inventory. This is probably due to the fact that the model, at this point, does not take into account the difference in costs associated with inventory versus back-orders.

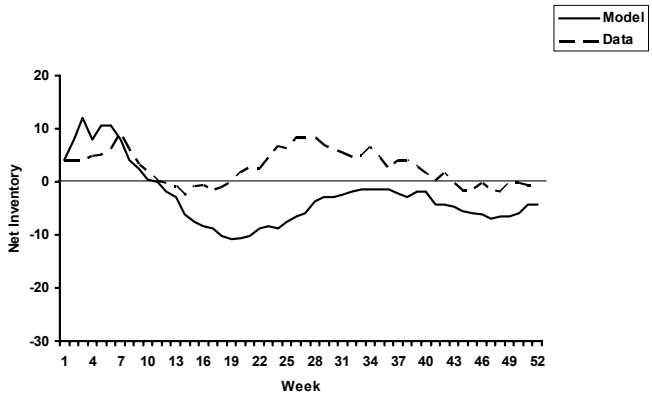

Figure 8: Model's Net Inventory per Week for Trial 9.

By Trial 20 the model's performance indicates further dampening of net inventory oscillations (see Figure 9).

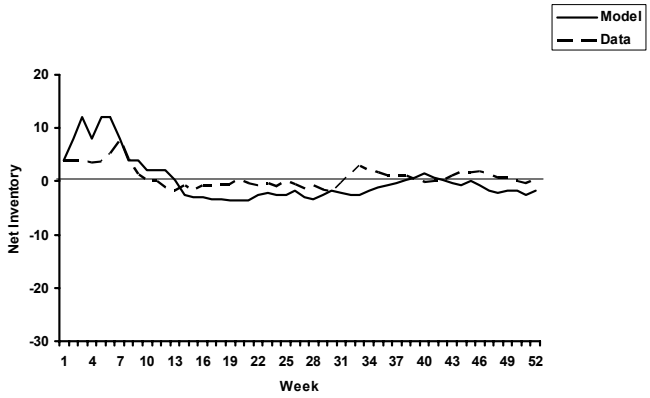

Figure 9: Model's Net Inventory per Week for Trial 20

Conclusions

Learning in dynamic environments is particularly challenging due to the complexity of dynamic problems and cognitive limitations, but our behavioral data showed considerable performance improvements with extended practice in a dynamic task. Our simplifications to the beer game removed the uncertainty in demand created by other players, raising a question of whether it is dynamic complexity or uncertainty that hinder learning.

The cognitive model and the closeness to human data have demonstrated that IBLT implemented on top of a cognitive architecture provides a constrained and reasonably accurate model of the learning process dynamic tasks. The results from the cognitive model support the prediction from IBLT that decision making in dynamic environments is a learning rather than an optimizing process. Humans learn to make better decisions by noticing the changes in an environment, storing examples of each situation experienced, and predicting future situations based on past experience.

Although encouraging, the results presented in this paper are however, far from conclusive. An interesting avenue for future research concerns the robustness of instance-based learning. If people primarily learn the input-output relationships in a dynamic environment rather than more abstract characteristics of dynamic systems, questions arise as to whether and how this type of learning transfers to varying environmental conditions. Our current experimental research is examining this, and is providing preliminary evidence of transfer of knowledge.

Acknowledgments

This research was supported by training grant 5-T32-MH19983 from the National Institute of Mental Health, and the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory (DAAD19-01-2-0009).

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.

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

Gonzalez, C. & Lebiere, C. (in press). Instance-based cognitive models of decision making. To appear in Zizzo, D. and Courakis, A. (Eds.). *Transfer of knowledge in economic decision making.* McMillan.

Gonzalez, C., Lerch, J.F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27,* 591-635.

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

Judd, C.M. & McClelland, G.H. (1989). *Data analysis: A model comparison approach.* Orlando, FL: Harcourt Brace Jovanovich.

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.

Sterman, J.D. (2004). Teaching takes off: Flight simulators for management education. Retrieved April 7, 2004, Massachusetts Institute of Technology, Sloan School of Management website: http://web.mit.edu/jsterman/www/SDG/beergame.html.

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