# Mental Models, Decision Making and Performance in Complex Tasks

Michael Shayne Gary
Australian Graduate School of Management
University of New South Wales
Sydney NSW 2052 Australia
email: sgary@agsm.edu.au
Telephone: 61 2 9931-9247

Robert E. Wood
University of Sydney
Sydney NSW 2042 Australia
email: r.wood@econ.usyd.edu.au
Telephone: 61 2 93561 0038

ABSTRACT

Previous studies have used the mental models construct as an ex-post explanation for poor performance on complex tasks, but this relationship has remained untested. This experimental study measured and tested the role of mental models in a complex decision environment. Participants worked on a product lifecycle management simulation under one of two levels of complexity across three phases of decision trial blocks spanning fifteen weeks. The results indicate that ability and task complexity are significant predictors of mental model accuracy, and that mental model accuracy and complexity are significant predictors of performance. Finding empirical support for the connection between mental model accuracy and performance, and measuring the magnitude of this effect is a key step forward in understanding why decision makers perform so poorly in complex decision environments. The results suggest there is potential to increase performance in such contexts , such as our increasingly large and complex organizations, by up to 100% through improving decision making. Validation of measures of mental model accuracy will enable researchers to incorporate this variable into their study designs in future research, and begin to identify levers for improving causal inferences, mental model accuracy, decision heuristics and performance.

Key words: decision making, mental models, feedback, system dynamics, simulation

Dynamic decision making tasks are part of our everyday lives. They range from managing a small team, managing a large project, managing a fish stock or other natural resource, managing a firm, to managing a national economy. All of the empirical evidence suggests that individual decision making in dynamic tasks is far from optimal and even quite poor relative to simple decision heuristics (Atkins, Wood, & Rutgers, 2002; Brehmer, Hagafors, & Johansson, 1980; Hogarth & Makridakis, 1981; Kleinmuntz, 1985; Paiche & Sterman, 1993; Sterman, 1989a). Previous research also indicates that learning in dynamic decision-making tasks plateaus rapidly and experience does not improve the effectiveness of decisions to any great degree (Atkins et al., 2002; Paiche & Sterman, 1993). There are many hypotheses but very few empirical tests of the causes of poor performance on complex decision-making tasks. One prominent explanation for poor performance on complex tasks is that humans develop incomplete and inaccurate mental models of dynamic, complex decision environments resulting in misperceptions of feedback between decisions and the environment (Diehl & Sterman, 1995; Goodman, Hendrickx, & Wood, 2004; Moxnes, 1998; Paiche & Sterman, 1993; Sengupta & Abdel-Hamid, 1993; Sterman, 1989a, 1989b).

The mental model construct used in ex post explanations of performance on dynamic decision tasks typically includes knowledge of the underlying causal relationships that make up the deep structure of the task, and an understanding of the impact of decisions within that structure (Diehl & Sterman, 1995; Kleinmuntz, 1985). Using this reasoning, decision makers have a difficult time managing complex tasks effectively because the underlying causal relationships, which may include time delays, feedback effects and nonlinearities, are difficult to detect and integrate into mental models. This explanation for poor performance in complex decision environments sounds plausible, but has not been subjected to empirical testing. Until we understand the mechanisms responsible for poor decision making and performance in such contexts, it will be very difficult to design interventions and strategies to enhance performance in complex dynamic decision environments. Our study is the first to explicitly measure mental model accuracy and test the impact of mental model accuracy on performance. An experimental approach enables us to study decision making and performance on a dynamic decision making task where we objectively and perfectly understand the underlying causal relationships and can compute the optimal performance. Measuring the mental model accuracy and establishing the impact on performance provides a start in identifying strategies for improving understanding, decision making and performance in complex decision environments.

In the following sections, we start with a description of dynamic decision environments, and subsequently review the relevant psychological research on explanations for poor performance on complex tasks. This is followed by general descriptions of mental models and decision heuristics, along with the arguments and hypotheses for the proposed relationships between the study variables. Figure 1 provides a schematic of the proposed relationships.

----------------------------------

Insert Figure 1 Here

----------------------------------

**THEORETICAL BACKGROUND AND HYPOTHESES**

Dynamic decision tasks are those that require a series of decisions rather than a single decision; these decisions are interdependent, and the environment changes as a consequence of both the decision-maker's actions as well as other external factors (Brehmer, 1992; Edwards, 1962). Complexity in these environments derives from the number of interdependent variables, the presence of time delays separating decisions from their resulting impacts, nonlinear relationships between variables, and multiple side effects of each decision. While previous research has attributed poor performance to the effect of task complexity on mental models, few studies have manipulated task complexity and those that have (e.g. Atkins et al., 2002; Diehl & Sterman, 1995; Paiche & Sterman, 1993) have not directly tested the effects of task complexity on mental model accuracy.

One dimension of complexity is the number of interdependent variables in the decision environment. Environments with a large number of variables increase the cognitive load of decision makers attempting to learn about potential interdependencies between variables (Sweller, 1988; Sweller, Chandler, Tierney, & Cooper, 1990). Another dimension of complexity is the number of decisions that must be made each time period. Attending to a higher number of decisions each time period increases the cognitive processing requirements for decision makers. Yet another dimension of complexity is feedback complexity, which is related to the strength, or open loop gain, of feedback loops in the decision environment. Decision environments with the same feedback loops but differences in the strength of those loops can differ dramatically in the experienced difficulty of the decision process (Atkins et al., 2002; Diehl & Sterman, 1995; Paiche & Sterman, 1993). For example, Paiche and Sterman (1993) found that the strength of feedback loops related to word of mouth and average lifetime of a product had a significant effect on performance in the management of a product lifecycle.

When the cognitive load of a task exceeds working memory capacity during the acquisition or learning phase, the development of mental models is impaired (Sweller, 1988, 1994). The cognitive load of a task refers to the total amount of mental activity imposed on working memory at an instance in time. Information may only be stored in long-term memory after first being attended to, and processed by, working memory. Working memory, however, is extremely limited in capacity and when the information to-be-learned in the construction of a mental model includes a large number of elements and interactions among those elements, such as multiple feedback loops, the cognitive load imposed on an individual quickly exceeds that capacity.

To sum up our arguments to this point, the complexity in a decision environment determines the cognitive load associated with the performance of a task and, when that load exceeds working memory capacity, the formation of accurate and complete mental models is impeded. The resulting inaccurate and deficient mental models result in poor longer-term performance.

## Mental Models

There is now considerable evidence that the core determinant of skilled performance is the knowledge base accumulated in long term memory stored in the form of a hierarchical information network (Sweller, 1988; Sweller et al., 1990). Previous research spanning psychology, administrative and organization theory, economics, political science, computer science and cognitive science have used a variety of terms for this knowledge base, including mental models, schemas, causal maps, cognitive maps, and belief systems (Cooper & Sweller, 1987; Diehl & Sterman, 1995; Hodgkinson, Maule, & Bown, 2004; Huff, 1990; Simon, 1982; Sweller et al., 1990). Although there are some differences in the knowledge content associated with some of these terms, these distinctions are beyond the scope of this paper. We have used the label mental model to encompass all of these terms referring to the knowledge base stored in long term memory.

Under this broad designation mental models include a range of knowledge or beliefs about concepts, known ‘facts’, images, perceived causal relationships, and decision heuristics about the world or of a particular task or system. These mental models become more detailed and complex as more extensive knowledge is acquired in a given content area. Mental models serve several functions in judgment and decision processes. For example, they provide a framework for filtering and interpreting new information and determining appropriate responses to that information. In this context, accurate mental models of the decision environment result in more appropriate and effective decisions and therefore better outcomes. The idea that people rely on mental models for deductive reasoning, inference and decision making can be traced at least as far back as Craik’s (1943) suggestion that the mind constructs “small-scale models” of reality.

In this study, we focus on two components of knowledge in decision makers’ mental models- knowledge about the causal relationships at work in a decision environment and the decision heuristics or rules of thumb adopted to automate and simplify decision making. As a result of differences in the type and amount of experience in a given decision environment, individuals will learn different chunks of knowledge and will develop different mental models of the causal relationships for the same task. We hypothesize that an individual’s performance on a task is partly a function of the accuracy of that individual’s perceived causal relationships between variables in the decision environment they have inferred through experience in the task domain. High performers will know more about the causal relationships between variables in a decision environment.

## Decision Heuristics

At any point in time, an individual's current mental model containing the perceived causal relationships at work in the decision environment provides a framework for decision-making and may influence the responses an individual makes to the different situations they encounter. However, over time responses to specific situations become codified into decision routines or heuristics that are executed automatically, without high levels of concentration or reference to the specific causal relationships between variables in the decision environment. Once decision heuristics are stored in the mental models associated with a task, they are automatically evoked in response to task situations, make little demand on working memory, and facilitate fast responses. These automatic responses may result in either good performance or poor performance, depending on whether or not the heuristics guiding the automated responses are effective or not. As an example of effective heuristics, math experts are fast and accurate in solving math problems because they have a much wider set of routines to apply and these routines are developed to the point that they are applied automatically, without conscious effort (Sweller, 1988).

Research on cognitive processes in judgment and choice has identified a wide range of general heuristics that are evident across a range of decision tasks, some of which are noted for their dysfunctional consequences (Brehmer, 1994; Brehmer et al., 1980; Kleinmuntz, 1985; Tversky & Kahneman, 1974), while others are noted for their facilitation of adaptive responses to situations (Gigerenzer, Todd, & Group, 1999). There is little published research on the formation and effects of task specific heuristics. Some research in educational psychology on learning of mathematics and other subjects have identified dysfunctional decision heuristics that can impede learning and the development of expertise in those specific domains (Sweller, 2003).

Task specific decision heuristics are of interest in the learning of dynamic decision environments because they are the component of mental models that may have the most direct relationship with the responses chosen. Therefore, the broad claim that people have difficulty learning how to respond effectively to dynamic decision environments may be more finely focused on the development and effects of specific decision heuristics. There are several reasons to expect that decision heuristics will be developed and applied on dynamic decision tasks. Primarily, the cognitive loads associated with decisions in complex and dynamic environments creates a strong incentive for the adoption of simplifying decision rules that avoid the processing of available information and thus bring the cognitive load to manageable levels.

Based on this line reasoning, we would expect people to develop task specific heuristics for the different decision responses required on a dynamic decision task and that these responses will be related to some small subset of the situational cues that are most strongly and/or most obviously related to those actions (Cyert & March, 1963; Forrester, 1961; Morecroft, 1985; Simon, 1982). Current research does not provide us the grounds for predicting how many cues. We would also expect that, once developed, the application of these heuristics would be related to performance. Also of interest is the question of how long it takes to develop a decision heuristic and whether the heuristic continues to be applied following poor performance outcomes.

Decision heuristics are only partly based on the perceived causal relationships of the decision environment that are encoded in mental models. We hypothesize that more accurate mental models will result in more effective decision rules and higher performance. However, not all such knowledge developed and stored in long-term memory leads to greater expertise in the performance of a task. Decision heuristics can develop from repeated exposures to task situations, independently of any knowledge of the causal relationships that underpin the decision environment. Once the decision heuristics that guide decision-making become automatic, they may no longer be accessible to conscious recognition and recall.

The formation and continuous updating of mental models in dynamic decision environments requires an iterative process of drawing causal inferences about the underlying relationships at work in the system. At the same time they are learning the causal relationships between variables, individuals will be developing decision heuristics or response routines for the situations that they encounter repeatedly. Higher levels of complexity in such environments increase the difficulty of drawing accurate causal inferences due to the cognitive load associated with environments involving a large number of interdependent variables, time delays and nonlinearities. When the cognitive load associated with a task exceeds working memory capacity, the development of decision heuristics can reduce the task demands to manageable levels.

Based on the preceding set of arguments for the relationships shown in Figure 1, we advanced the following hypotheses.

H1: Mental model accuracy will be negatively related to the level of complexity of the decision environment. More accurate mental models will be developed for less complex tasks than for more complex tasks.

H2: Accuracy of mental models will be positively related to subsequent performance on both immediate-transfer and delayed-transfer tasks, after controlling for task complexity, cognitive ability and motivation.

H3: The effects of task complexity on performance will be positively mediated through mental model accuracy.

H4: Participants will develop simple heuristics for each of the major decisions. These simple heuristics will explain the vast majority of variance in participants' decisions. The content of the decision heuristics (information weights) will provide insight into how much weight decision makers put on selected cues for each decision.

## METHOD

### Participants

Second year MBA students with no prior experience on the simulation were invited to participate. The 63 participants included 47 male and 16 female volunteers, with an average age of 30. Participants were randomly assigned to either the low complexity (n = 31) or the high complexity (n = 32) group. Participants were paid a fixed amount for their participation in the experiment. In addition, a small donation was paid to a nominated club or charity for the 43 students who also participated in the delayed-transfer stage 15 weeks later.

### Task and Procedures

The study task was an interactive computer-based management simulation. The task is based on managing a new product through the product lifecycle, and was modified from a task utilized in previous research (Paiche & Sterman, 1993). Participants manage decision variables, such as price and production capacity, with the goal of maximizing cumulative profit from the sales of their product through a forty-quarter simulation.

Participants completed three phases; a learning phase, an immediate-transfer testing phase, and a delayed-transfer testing phase. The learning phase and immediate-transfer phase were completed in an initial experimental session in a lab. Assessments of the self-efficacy and mental models of participants were completed after the learning phase in the initial session. Participants completed the initial experimental session in groups of 15 to 20. During that session, each participant was seated at a separate computer and the space between computers was great enough so that participants could not see other screens. The delayed-transfer task was completed fifteen weeks later.

The learning phase included three blocks of 40 decision trials for participants to learn about and become familiar with the decision environment. After each decision trial, participants received feedback on their results for that trial plus their cumulative performance to that point.  This feedback was presented in both table and graphical format in order to control for the effects of feedback format (Atkins et al., 2002). Following the learning phase, participants were asked to complete a series of questionnaires to assess their self-efficacy and mental models of the task. After completing the questionnaires, participants proceeded to the immediate-transfer phase, in which they completed three more blocks of 40 decision trials on the same decision environment they had encountered in the learning phase. Participants were under no strict time pressure and completed each phase at their own pace. On average, the initial experimental session took approximately three hours, which included 60 minutes on the learning phase, 75 minutes to complete the self efficacy questionnaire and mental model assessment, and 45 minutes to complete the immediate-transfer task. Upon completing the immediate-transfer phase, participants left the lab and were paid for their participation in the study.

The delayed-transfer phase was completed fifteen weeks later using a web-based version of the simulation. This phase involved logging into the web-based simulation from remote locations and completing three more blocks of 40 trials using the same decision environment the participants used in previous phases.

## Complexity Interventions

Decision complexity was manipulated to be one of two levels (low and high) by varying the number of decision variables to be managed, the level of competition in the market, and the number of relationships that link the decision variables to observable market outcomes. The lower level of decision complexity was fully nested within the higher level of complexity. In the low complexity version of the task, there were two decision variables- price and target capacity- and 19 interconnected variables. There was no competitor in the low complexity version of the task. In the high complexity version of the task, there were three decision variables- price, target capacity, and marketing spend- and over 30 interconnected variables, including a competitor sector that influenced the relationships between the decision variables and the different market responses. Participants were randomly assigned to one of the two complexity levels and they worked on that level of complexity during the learning phase and on both transfer tasks.

## Measures

Mental Models. The contents of mental models were assessed through a questionnaire that was developed to assess participants' recognition and recall of the causal relationships between variables in the decision environment. Each item in the mental model questionnaire tested participants' recall of a bivariate causal relationship between a pair of variables from the management simulation, including the sign or polarity if there was a relationship. The items in the questionnaire covered the exhaustive set of actual relationships in each of the complexity conditions along with several items for which no relationship existed in the decision environment. A schematic of the full set of causal relationships in the low decision complexity condition is shown in Figure 2. Figure 3 provides a segment of the questionnaire instructions along with the first three items used to collect participants knowledge of causal relationships.

---------------------------------------

Insert Figure 2 & Figure 3 Here

---------------------------------------

The items in the questionnaire tested participants' recall of relationships between variables within and across four distinct sectors in the decision environments, including Customers (Cust), Operations (Ops), and Pricing, Marketing and Financials (PMF) sectors, which were present in both complexity conditions and the Competitive (Comp) sector, which was present in the high complexity but not the low complexity decision environment. The questionnaires completed by participants in both complexity conditions included 30 items on the relationships between variables in the three sectors that were common to both decision environments, broken down as follows: Relationships within the Cust sector - 5 items; Relationships within the Ops sector - 7 items; Relationships within the PMF sector - 9 items; Relationships between the Ops, PMF and Cust sectors - 5 items; and Relationships between the Ops and PMF sectors - 4 items. The questionnaire completed by participants in the high complexity condition included a further 24 items, including assessments of the following: Relationships within the Comp sector - 8 items; Relationships between the Cust and Comp sectors - 9 items; additional relationships within the Cust sector - 4; additional relationships within the PMF sector - 1; and additional relationships between the Ops, PMF, and Cust sectors - 2.

All relationships among variables in the management simulation are known with absolute certainty, and each item on the questionnaire was scored as correct or incorrect for each participant. There are nine possible ways to answer each influence diagram item- four possibilities with one directed arrow in two feasible directions and two viable polarities, one possibility of no direct causal relationship between the two variables (indicated by writing NONE between the variables), four possibilities of two directed arrows (two-way dependency in a feedback loop) and two possible polarities- indicating a random answer strategy on the questionnaire would result in a score of 11% accuracy.

*Mental Model Accuracy* was the percentage of items on the questionnaire answered correctly in each condition. That is the total absolute knowledge score divided by the total number of items on the questionnaire, which differed between the low and high complexity conditions as described above. Thus mental model accuracy was a measure of the proportion of the complete causal structure of the decision environment understood by participants. The possible scores range from 0 to 1, where a score of 1 indicates perfect knowledge of the structure for the assigned level of complexity.

Mental model accuracy scores were also calculated for each of the subsets of items related to the different sectors of the simulation model, as described above, and for the 30 items that assessed the relationships that were common to both the low and high complexity decision environments.[1]

*Decision heuristics* were identified through estimation of the predictors for the levels of price and target production capacity set by participants' on each trial. Price and target production capacity

[1] Two other questionnaires were administered to participants after the learning phase: 1) a True/False questionnaire presented a series of items about the causal relationships between variables in the task and 2) a Graphical Scenario questionnaire presented the graph of one variable over time from the task and asked subjects to choose from a multiple choice of answers for the evolution of a second variable in the task. Items from the True/False questionnaire were used to cross-validate the reliability of responses obtained on the Influence Diagram questionnaire. The Graph Scenario measure was included as a task specific ability measure. Analyses employing these two variables do not alter the results obtained and are available from the first author.

were the two trial-by-trial decisions made by participants in both the low and high decision complexity environments. We adopted the decision rules from Paiche and Sterman (1993) and estimated the information weights for each cue of the decision rule using OLS regression. The specific form of the equation and the cues used for each heuristic are discussed in detail in the results section.

*Performance* was measured for each of the nine blocks of trials by the cumulative profit at the end of the 40th and last decision trial in each block. The potential achievable cumulative profit was different in the high and low complexity task conditions, and therefore we assessed subjects' raw performance relative to benchmarks for the high and low conditions. The cumulative profit benchmarks were found through single point optimization using a modified Powell search implemented in Vensim simulation software. To find the benchmark profit, Marketing Spending was fixed at 5% of revenue throughout the simulation; this value was already fixed at 5% in the low complexity condition. Capacity was determined by a perfect foresight rule in which capacity always matched demand in both the low and high complexity conditions. Finally, benchmark profit was found by finding the single Price level that optimized profits over the entire simulation. This optimal pricing mechanism is very simplistic since price does not change throughout the simulation in response to changing capacity, backlog, order demand, or any other variable in the decision environment. Therefore, the calculated cumulative profit benchmark is clearly not a global optimum for the task, but is instead simply a consistently calculated benchmark[2].

The nine blocks of performance included three blocks completed during the learning phases; three blocks on the immediate-transfer task, which was completed at the end of the initial experimental session following the assessments of mental models and the motivational control variable; and three blocks on the delayed-transfer task, which was completed fifteen weeks after the initial experimental session.

## Control Variables

*General cognitive ability* was indexed through participants' scores on the GMAT. Because of the potential importance of cognitive ability to the learning of complex tasks, the GMAT measure was included to ensure that random allocation of participants to the two complexity conditions effectively removed cognitive ability as an explanation for differences between the two groups. Also, the inclusion of the GMAT measure enabled statistical estimation of the effects of cognitive ability in the predictions of mental models and performance.

*Perceived self-efficacy* is an established motivational predictor of performance on complex tasks and the constituent processes, such as search, information processing and memory processes

[2] Several different cumulative profit benchmarks were calculated, including the naïve strategy benchmark reported in Paiche and Sterman (1993), and the results were not sensitive to the different benchmarks.

that can affect learning (Bandura, 1997). Also, levels of decision complexity have been shown to influence the motivational reactions to tasks (Wood, Bandura, & Bailey, 1990). Therefore, self-efficacy was incorporated as a control variable to ensure that differences in the mental models of participants in low and high complexity decision environments at the end of the learning phase were not solely attributable to motivational differences. Perceived self-efficacy was measured with a 10-item scale covering a broad range of activities participants needed to manage throughout the simulation. The format followed the approach presented by Bandura (1997), which has been validated in numerous empirical studies. For each item, participants first recorded whether or not they understood what was required to manage the activity - yes or no - and then recorded their confidence in their capabilities on a 10-point scale where 1 = “very low confidence” and 10 = “very high confidence.”

Analyses

Hypothesis 1 was tested using an Independent-Samples T-Test for differences between the low and high complexity groups. Generalized Linear Models were used to test hypothesis 2 regarding the effects of mental model accuracy on immediate and delayed-transfer performance. Due to unequal sample sizes for the immediate and delayed-transfer phases, the analysis was separated into two parts, with immediate-transfer performance and delayed-transfer performance as dependent variables for two separate models. Task complexity was included as a between subjects fixed factor in the models, while self-efficacy and GMAT were included as covariates in order to control for differences that may have been due to motivation and general cognitive ability.

Hypothesis 3 predicted that the effects of task complexity on immediate and delayed-transfer performance would be mediated through the accuracy of participants’ mental models of the decision environment. To test this hypothesis, we estimated a mediated path model using a causal steps regression approach supplemented with Sobel tests of the cross products for indirect effects (MacKinnon, Warsi, & Dwyer, 1995). The same causal steps and indirect effects were applied to test for the linkages from cognitive ability to mental model accuracy to performance, which are shown in Figure 1 but not the subject of specific hypotheses.

To test Hypothesis 4, information weights for target capacity and price decision rules, identified by Paiche and Sterman (1993), were estimated using OLS regression. The information weights estimated for each decision heuristic in this study are compared with those from the Paiche and Sterman (1993).

**RESULTS**

Descriptive Statistics

The correlations, means, and standard deviations for the study variables are shown in Table 1. The decision environment complexity variable was dummy coded so that 0 = low complexity and 1 = high complexity. There are strong autocorrelations in the performance for the 9 blocks of trials across the learning, immediate-transfer and delayed-transfer tasks, thus indicating that the individual determinants of performance were relatively consistent from one block to the next. The pattern of correlations between the study variables is consistent with the hypothesized relationships shown in Figure 1. Complexity is negatively related to self-efficacy and to performance on each block, as expected. Also, mental model accuracy is negatively related to the level of decision environment complexity.

-----------------------------------

Insert Table 1 Here

-----------------------------------

General cognitive ability was not significantly related to performance on any of the 9 trial blocks, but was positively related to mental model accuracy across the two complexity conditions ($r = .35$, $p<.01$). Thus, it appears that the cognitive abilities tapped by the GMAT relate to the development of task knowledge but not, directly, to the application of that knowledge.

By way of contrast, the self-efficacy control variable had a direct and sustained relationship with performance on the three blocks for the immediate-transfer task and the three blocks for the delayed-transfer phase, which were completed fifteen weeks after the self-efficacy measure (all r’s > .27, $p<.05$). Participants’ self-efficacy for their version of the task was influenced by their prior performance attainments over the three blocks of trails on the learning phase (r’s > .25, $p<.05$) and those who worked on the more complex version were less confident about their ability to manage the product lifecycle effectively than those who operated in the less complex decision environment ($r = -.33$, $p<.01$).

Most importantly, mental model accuracy was a strong and consistent predictor of performance on each block of the immediate-transfer task (r’s <= .35, $p<.01$ for blocks 4, 5 and 6, respectively) and on the three blocks of the delayed-transfer task ($r = .32$, $p<.05$, $r = .40$, $p<.01$ and $r = .46$, $p<.01$, for blocks FT1, FT2, and FT3 respectively).

Figure 4 illustrates the mean performance on each of the nine trial blocks completed for the high and low complexity groups. Dotted lines divide the Learning, Immediate-transfer, and Delayed-transfer phases of the experiment. In the learning phase, performance improves considerably from trial block 1 to block 3, but further improvements in performance are very incremental. Performance is relatively stable throughout the Immediate-transfer phase, indicating learning plateaus relatively quickly in the experiment for both groups. Relative to performance in the Immediate-transfer phase, performance falls slightly in the Delayed-transfer phase. T tests show that the differences in performance between the low complexity and high complexity groups shown in Figure 4 are significant on each of the trial blocks ($p$'s $< .001$).

----------------------------------

Insert Figure 4 Here

----------------------------------

Tests of Hypotheses

Figure 5 illustrates the difference between mental model accuracy for participants in the low and high complexity groups, which supports hypothesis 1. Participants in the low complexity condition developed more accurate mental models of their decision environment than participants in the high complexity condition ($t = 2.33$, $p<.05$). The mental model scores for participants in both the low and high complexity conditions were significantly different from the random-answer base rate of 11%. Analyses of the seven categories of items included in the mental model questionnaire described in the Method section showed that the greater accuracy of mental models for participants in the low complexity condition were due to the cumulative effect of small differences across the different components of the decision environment.

----------------------------------

Insert Figure 5 Here

----------------------------------

Hypothesis 2 was supported for both the near and delayed-transfer task. As shown in Table 2, mental model accuracy was a highly significant predictor of performance on the immediate-transfer task trial blocks ($F = 7.662$, $p<.01$), after controlling for task complexity, motivation and general cognitive ability. Task complexity was also a significant predictor in the full model ($F = 92.144$, $p<.001$), thus indicating that complexity had a direct effect on performance that was not fully mediated through the accuracy of the mental models developed. This is consistent with the relationships shown in Figure 1. Self-Efficacy and GMAT did not contribute significantly to explaining the variance in immediate-transfer performance. These results indicate that mental model accuracy is a more direct predictor of performance on the immediate-transfer task than general cognitive ability or motivation. This may be because the effects of motivation and cognitive ability on performance are fully mediated through their effects on the development of mental models and, contrary to the model shown in Figure 1, general cognitive ability has no direct effects on performance. These points are taken up later, in the discussion of the analyses for hypothesis 3.

-------------------------------------

Insert Table 2 & Table 3 Here

-------------------------------------

The results for the analyses of delayed-transfer performance are shown in Table 3. Mental model accuracy was again a highly significant ($F = 7.963$, $p < .01$) predictor of performance on all three blocks of the delayed-transfer task, after controlling for task complexity, motivation and general cognitive ability. Thus the measure of mental model accuracy remains a successful predictor of performance over a period of fifteen weeks after the initial exposure to the task. These results suggest that the measure of participants' mental model has successfully tapped the knowledge of relationships that were stored in their long-term memories following completion of the learning phase. Again, task complexity also had a strong direct effect on delayed-transfer performance ($F = 45.324$, $p<.001$). Self-Efficacy and GMAT again did not contribute significantly to explaining the variance in delayed-transfer performance, after controlling for task complexity and mental model accuracy.

Figure 6a shows the bivariate correlations (above the path arrows) and path coefficients (below the path arrows) for the hypothesized relationships in Figure 1 that are significant in the prediction of immediate-transfer performance. Self-efficacy was dropped from the model after extensive analyses demonstrated this variable had no significant impact in predicting either mental model accuracy or performance and no substantive effect on the estimates of other variables in the model. General cognitive ability did not have a significant direct effect on immediate-transfer performance ($r = .08$, ns), however as shown by the following analyses, including the Sobel test, it did have an indirect effect on immediate-transfer performance through mental model accuracy (MacKinnon et al., 1995). Task complexity ($b = -.31$, $p<.001$) and general cognitive ability ($b =.36$, $p<.001$) were highly significant predictors of mental model accuracy ($\overline{R}^2 = .194$, $F = 8.467$, $p<.001$), and mental model accuracy was a significant predictor of immediate-transfer performance ($b =.22$, $p<.005$) after controlling for the effects of task complexity ($b = -.60$, $p<.001$) and general cognitive ability ($b = .01$, ns); with an $\overline{R}^2 = .460$ ($F = 18.355$, $p<.001$).

-----------------------------------

Insert Figure 6 Here

-----------------------------------

The Sobel test of the cross products of the paths from cognitive ability to mental model accuracy, and from mental model accuracy to immediate-transfer performance was highly significant ($z = 2.528$, $p < .01$)[3]. The Sobel test of the pathways from task complexity to mental model accuracy, and from mental model accuracy to immediate-transfer performance provides only partial support for the hypothesised indirect effect ($z = -1.684$, $p<.10$).

The results for delayed-transfer performance replicated those discussed above for immediate-transfer performance and the path model is illustrated in Figure 6b. Mental model accuracy was a significant predictor of delayed-transfer performance ($b = .299$, $p<.05$) after controlling for the effects

[3] The Sobel test mediation results reported in this section were also tested using Goodman (I) and Goodman (II) tests (MacKinnon et al., 1995), and the results of all three tests were consistent.

of task complexity (b = -.518, p<.001) and general cognitive ability (b = .005, ns) with an $\overline{R}^2 = .410$ (F = 10.728, p<.001). The Sobel test of the pathways from general cognitive ability to mental model accuracy, and from mental model accuracy to delayed-transfer performance was again highly significant (z = 2.452, p < .01). The Sobel test of the cross products of the paths from task complexity to mental model accuracy, and from mental model accuracy to delayed-transfer performance again supports a conclusion of weak partial mediation (z = -1.740, p<.10). Therefore, while the indirect effects shown for general cognitive ability were more strongly supported, the indirect effects of task complexity on both near and delayed-transfer performance through mental model accuracy, as shown in Figure 1, only received partial support.

Modeling decision heuristics

In our tests for possible decision heuristics for target capacity and prices, we started with the decision rules identified by Paiche and Sterman (1993). The decision rules were based on: "participants written reports of their strategies, prior models of similar decisions in the literature, and the feedback structure of the task" (Paiche & Sterman, 1993: 1450). The task used in the high complexity condition in this study is a slightly modified version of the task used by Paiche and Sterman[4], and this presents an opportunity to subject the decision rules to further testing and comparison. The beta coefficients, or information weights, for the estimated rules show how much weight the participants put on the selected cues in their capacity and pricing decisions. We also tested for changes in the information weights across trial blocks to see if the weight assigned to each of the cues changes with experience.

The three cues in the target capacity decision rule were: actual demand ($D_{t-1}$), demand growth rate ($g_{t-1}$), and the ratio of order backlog to actual production capacity ($B_{t-1}/C_{t-1}$)[5]. The two cues for the price heuristic were: unit variable cost ($UVC_t$) and a markup based on the ratio of order backlog to current production capacity ($B_t/C_t$). The information weights for the capacity and pricing decision rules were estimated separately for each trial block for each participant using OLS regression. The Durbin Watson test statistic revealed positive autocorrelation in the residuals for both heuristics. To correct for first-order autocorrelation, the one time period lagged variable for each decision variable –

[4] The high complexity task included a decision variable for Marketing Spend that was not used as a decision variable in Paiche and Sterman (1993), but was included as a constant parameter in their task. The parameters for word of mouth and average replacement time were equivalent to the most difficult market response scenario from Paiche and Sterman. The competitor pricing policy was a fixed mark-up over unit production cost, and resulted in continuously falling competitor prices as unit costs fell due to the learning curve effect. Finally, the instantaneous values of rates were reported to subjects instead of averaged rates from a reporting sector.

[5] Paiche and Sterman's decision rule for target capacity was: $C_t^* = s^*[D_0^{e(1-\alpha_0)} D_{t-1}^{\alpha_0}](1+g_{t-1})^{\alpha_1}(B_t/C_t)^{\alpha_2}$ . Where $s^*$ is a constant target market share of 50%, $D^e$ is the prior estimate of market demand, $D_{t-1}$ is the actual demand lagged by one time period), $g_{t-1}$ is lagged demand growth, and the ratio of backlog/capacity. This heuristic was estimated in their study and in our study as: $\log(C_t^*) = c + a_0 \log(D_{t-1}) + a_1 \log(1+g_{t-1}) + a_2 \log(B_t/C_t) + \varepsilon_1$ .

Lag Target Capacity or Lag Price – was included in the models. The results of the analyses for the low complexity and high complexity conditions are shown in Table 4 along with the results reported by Paiche and Sterman (1993) for comparison.

----------------------------------

Insert Table 4 Here

----------------------------------

The results, which support Hypothesis 4, indicate that each decision heuristic captures the majority of the variance in participants' decisions in both the high and low complexity conditions. The mean $R^2$'s for the high and low complexity conditions are 0.83 and 0.90 respectively for the Target Capacity Heuristic and 0.91 and 0.95 for the Price Heuristic. The signs of the coefficients for the different predictors of decisions in both complexity conditions were the same as those found in the Paiche and Sterman (1993) study, except for the negative sign of the intercept in the high complexity price heuristic. The relative magnitudes of the information weights for each decision heuristic were also similar across the two studies and different complexity conditions.

Paiche and Sterman (1993) found that subjects' target capacity decisions were primarily based on their prior expectation of market demand and only secondarily on actual demand. The prior expectation of market demand was a function of the intercept term ($a_0$ = 8.414) and the information weight for actual market demand ($a_1$ = .383)[6]. Subjects in their study were largely insensitive to growth in demand and the demand/supply balance measured by the ratio of backlog/capacity. Table 4 illustrates that participants' target capacity decisions for both the high and low complexity conditions in the current study were also primarily based on their prior expectations of market demand. The intercept term was a significant predictor of target capacity decisions in more than 86% of the instances ($c_{LC}$ = 3.792, $p<.000$; $c_{HC}$ = 3.870, $p<.000$), and actual industry demand had a weaker effect on participants' capacity decisions ($a_{0LC}$ = .109, $p<.10$; $a_{0HC}$ = .062, $p<.10$) and was not significant in over 56% of the cases. Information about the ratio of backlog/capacity had a significant impact on target capacity decisions in almost 65% of the cases and was given moderate weight in the decision heuristic ($a_{2LC}$ = .207, $p<.05$; $a_{2HC}$ = .221, $p<.05$). Participants were insensitive to the demand growth rate in setting target capacity decisions ($a_{1LC}$ = .118, ns; $a_{1HC}$ = .129, ns).

For the price heuristic, unit cost was a significant predictor of participants' pricing decisions in more than 70% of the instances ($b_{1LC}$ = .159, $p<.01$; $b_{1HC}$ = .369, $p<.01$). In contrast, the backlog/capacity ratio had little effect on participants' pricing decisions in either complexity condition ($b_{2LC}$ = .027, $p<.05$; $b_{2HC}$ = .005, $p<.10$). These results for both the target capacity and price heuristics support those found in Paiche and Sterman (1993). Further analysis of the information weights for the target capacity and price heuristics in both the low and high complexity conditions

[6] From the two equations in footnote 5 for the target capacity heuristic, $D^e$ can be calculated using: $c = (1-\alpha_0)\ln(s^* D^e)$ .

examined the heuristic formation and learning process. In particular, we investigated the degree to which participants' adjusted their decision heuristics with experience on the simulation. After participants' initial learning of the simulation on trial blocks one and two, the information weights for the different decision cues were very stable thereafter in the immediate and delayed-transfer phases.

**DISCUSSION**

Previous studies have hypothesized that poor performance on complex tasks is due to the formation of incomplete and inaccurate mental models resulting in misperceptions of feedback between decisions and the environment (Diehl & Sterman, 1995; Moxnes, 1998; Paiche & Sterman, 1993; Sterman, 1989a, 1989b). Our study is the first to explicitly test the impact of mental model accuracy on performance. The results indicate that more accurate mental models do indeed result in higher performance and that the formation of mental models is influenced by cognitive ability and complexity of the task.

The high complexity task in this study has a greater number of interdependent variables in the decision environment and a greater number of decisions that must be managed by participants each trial. Participants formed less accurate mental models on the more complex version of the task, which is consistent with the arguments for the misperceptions of feedback (Paiche & Sterman, 1993). Additional complexity also increases the cognitive load that must be processed as individuals learn how their decisions impact on the environment and strive to understand the complex world they are managing. Increases in cognitive load can also interfere with the development of accurate mental models of the decision environment (Sweller, 1988). Whatever the exact cognitive mechanism, the results of this study have shown how the resulting knowledge deficits have significant implications for an individual's subsequent performance on immediate and delayed-transfer tasks. Participants in the low complexity condition formed more accurate mental models and achieved significantly higher performance relative to subjects in the high complexity condition. These results are consistent with and extend the findings from previous research that high levels of feedback complexity negatively impact performance (Diehl & Sterman, 1995; Paiche & Sterman, 1993). It seems multiple dimensions of complexity impair causal inference and the formation of accurate mental models.

We also found a positive relationship between participants' GMAT scores and the accuracy of their mental models. General ability measures such as GMAT tap a wide range of analytical, logical, quantitative, verbal language, and standardized test taking skills of individuals. Our results indicate that at least some portion of the general skills assessed in the GMAT are also directly related to the formation of more accurate mental models. Interestingly, there was no significant direct effect of GMAT on performance. Instead, the effects of general cognitive ability are mediated through the formation of mental models.

There is now considerable evidence that deficiencies in human judgment and decision making result in poor performance in complex tasks relative to optimal or even relative to naïve decision rules (Atkins et al., 2002; Brehmer, 1992; Diehl & Sterman, 1995; Goodman et al., 2004; Hogarth & Makridakis, 1981; Kleinmuntz, 1985; Moxnes, 1998; Sterman, 1989a; Tversky & Kahneman, 1974). Our findings are very consistent with this stream of research. On average, participants across both conditions earned cumulative profits that were roughly 50% of the benchmark. Participants in the high complexity condition, on average, earned cumulative profits that were less than 30% of the benchmark. These results along with those of previous research indicate there is enormous potential to improve decision making and performance in complex decision environments. The largest potential for improvements may lie within our increasingly large and extremely complex organizations. The link between experimental studies of individual choice and decision making in organizations is obvious; individuals within organizations are responsible for the myriad of organizational decisions that are made each day. Cognitive limits impair our ability to mentally compute optimal solutions for problems involving high-order, nonlinear differential equations. Instead, we adopt decision heuristics or policies that guide decision making using simple rules of thumb (Cyert & March, 1963; Forrester, 1961; Morecroft, 1985; Simon, 1982). Given the complexity of organizations, it should not be a surprise that individuals or teams of individuals typically adopt suboptimal policies based on deficient mental models. Perhaps we could improve the performance of our organizations by up to 100% if we can identify factors that enhance mental model accuracy and decision making effectiveness.

The measurement instrument of mental models introduced in this paper will enable researchers to incorporate this variable into their study designs in future research. Further work is certainly necessary to validate the measurement instrument, but the initial results are very encouraging and provide an important first step. Such close examination of decision-makers mental models has the potential to open up a whole new area of research to unpack these mental models. Future studies can also explore the relationship between mental models and the decision heuristics adopted in complex tasks. Understanding this link is crucial to improving performance. Finally, further studies can also employ our measure of mental models to investigate the evolution of mental models with increased experience.

With an established measure of mental models that predicts performance on complex decision tasks, researchers are much better placed to study the mechanisms that lead to under-developed mental models. Understanding these mechanisms has important implications for interventions that are targeted at improving human performance on dynamic decision tasks. For example, two hypothesised mechanisms for poor performance on complex tasks that lead to quite different sets of implications for performance improvement are the misperceptions of feedback (Paiche & Sterman, 1993) and cognitive load theory (Cooper & Sweller, 1987; Sweller, 1988). If decision makers do not understand or misperceive the underlying causal relationships in a complex decision environment, then interventions clarifying the causal structure should improve performance. Causal relationships of a complex task can be presented in a variety of formats in an effort to enhance the formation of accurate mental models, and there is some support that such interventions improve performance (Sengupta & Abdel-Hamid, 1993). On the other hand, interventions targeted to reduce cognitive load might emphasize robust search and exploration strategies for learning about a complex task. Serial learning about the component parts of a complex task, in a staged learning strategy, may minimize cognitive load and help avoid overwhelming our cognitive processing capabilities (Sweller et al., 1990). Previous research employed an alternative intervention to prevent cognitive overload and found that providing information about successful decision heuristics as support for decision making improved performance (Sengupta & Abdel-Hamid, 1993). More recent research finds that providing feedback about causal relationships impacts both learning and performance; information about the causal relationships improve performance in the short run but not in the long run (Goodman et al., 2004; Langley & Morecroft, 2004). Decision makers with information about causal relationships in a complex environment may not explore enough of the problem space to develop complete mental models. More research is needed to disentangle the factors impacting learning and the formation of accurate mental models.

**References**


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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*(3), 211-241.

Brehmer, B. (1994). The Psychology of Linear Judgment Models. *Acta Psychologica, 87*(2-3), 137-154.

Brehmer, B., Hagafors, R., & Johansson, R. (1980). Cognitive Skills in Judgment: Subjects' Ability to Use Information About Weights, Function Forms, and Organizing Principles. *Organizational Behavior and Human Performance, 26*(3), 373.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology, 79*, 347-362.

Craik, K. (1943). *The Nature of Explanation*. Cambridge: Cambridge University Press.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm* (2nd ed.). Cambridge, MA. 1992.: Blackwell Publishers Inc.

Diehl, E., & Sterman, J. D. (1995). Effects of Feedback Complexity on Dynamic Decision Making. *Organizational Behavior and Human Decision Processes, 62*(2), 198-215.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human factors, 4*, 59-73.

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA.: M.I.T Press.

Gigerenzer, G., Todd, P., & Group, t. A. R. (1999). *Simple Heuristics that Make us Smart*. Oxford: OxfordUniversity Press.

Goodman, J. S., Hendrickx, M., & Wood, R. E. (2004). Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology, 89*(2), 248.

Hodgkinson, G. P., Maule, A. J., & Bown, N. J. (2004). Causal Cognitive Mapping in the Organizational Strategy Field: A Comparison of Alternative Elicitation Procedures. *Organizational Research Methods, 7*(1), 3.

Hogarth, R. M., & Makridakis, S. (1981). The Value of Decision Making in a Complex Environment: An Experimental Approach. *Management Science, 27*(1), 93-107.

Huff, A. S. (1990). *Mapping strategic thought*. New York and Chichester: Wiley.

Kleinmuntz, D. N. (1985). Cognitive Heuristics and Feedback in a Dynamic Decision Environment. *Management Science, 31*(6), 680-702.

Langley, P. A., & Morecroft, J. D. W. (2004). Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research, 155*, 715-732.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*(1), 41-62.

Morecroft, J. D. (1985). Rationality in the Analysis of Behavioral Simulation Models. *Management Science, 31*(7), 900-916.

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

Paiche, M., & Sterman, J. D. (1993). Boom, Bust, and failures to learn in experimental markets. *Management Science, 39*(12), 1439-1458.

Sengupta, K., & Abdel-Hamid, T. K. (1993). Alternative conceptions of feedback in dynamic decision environments: An experimental investigation. *Management Science, 39*(4), 411-428.

Simon, H. A. (1982). *Models of Bounded Rationality Volume 2: Behavioral Economics and Business Organization*. Cambridge, MA: MIT Press.

Sterman, J. D. (1989a). Misperceptions of Feedback in Dynamic Decision making. *Organizational Behavior and Human Decision Processes, 43*(3), 301-335.

Sterman, J. D. (1989b). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Experiment. *Management Science, 35*(3), 321-339.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295-312.

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 43, pp. 215-266). San Diego: Academic Press.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive Load as a Factor in the Structuring of Technical Material,. *Journal of Experimental Psychology: General, 119*(2), 176-192.

Tversky, A., & Kahneman, D. (1974). Judgement Under Uncertainty: Heuristics and Biases. *Science, 185*(27 September), 1124-1131.

Wood, R. E., Bandura, A., & Bailey, T. (1990). Mechanisms governing organizational performance in complex decision-making environments. *Organizational Behavior and Human Decision Processes, 46*, 181-201.

**Table 1 Correlations, Means and Standard Deviations for study variables.**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. GMAT | 1 | | | | | | | | | | | | |
| 2. Complexity | .019 | 1 | | | | | | | | | | | |
| 3. Perf1 | .173 | -.501** | 1 | | | | | | | | | | |
| 4. Perf2 | .101 | -.578** | .428** | 1 | | | | | | | | | |
| 5. Perf3 | .025 | -.707** | .453** | .741** | 1 | | | | | | | | |
| 6. Perf4 | .116 | -.646** | .471** | .724** | .869** | 1 | | | | | | | |
| 7. Perf5 | .132 | -.626** | .498** | .736** | .846** | .922** | 1 | | | | | | |
| 8. Perf6 | .078 | -.664** | .509** | .750** | .828** | .869** | .903** | 1 | | | | | |
| 9. Perf_FT1 | -.020 | -.682** | .420** | .550** | .582** | .625** | .613** | .687** | 1 | | | | |
| 10. Perf_FT2 | .127 | -.617** | .390* | .581** | .655** | .703** | .658** | .737** | .781** | 1 | | | |
| 11. Perf_FT3 | .138 | -.608** | .459** | .537** | .592** | .618** | .626** | .709** | .847** | .790** | 1 | | |
| 12. Self-Efficacy | .137 | -.328** | .266* | .283* | .248* | .291* | .266* | .281* | .347* | .400** | .326* | 1 | |
| 13. Mental Model Acc | .354** | -.301* | .243 | .379** | .356** | .353** | .364** | .397** | .319* | .403** | .456** | .230 | 1 |
| **Total** | | | | | | | | | | | | | |
| Mean | 642.22 | 0.51 | 0.04 | 0.32 | 0.43 | 0.46 | 0.51 | 0.51 | 0.43 | 0.49 | 0.47 | 5.66 | 0.52 |
| Std. Deviation | 54.30 | 0.50 | 0.78 | 0.43 | 0.38 | 0.37 | 0.36 | 0.37 | 0.37 | 0.43 | 0.46 | 1.28 | 0.12 |
| N | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 62 | 43 | 42 | 43 | 63 | 63 |
| **Low Complexity** | | | | | | | | | | | | | |
| Mean | 641.19 | | 0.43 | 0.57 | 0.70 | 0.71 | 0.73 | 0.75 | 0.65 | 0.71 | 0.71 | 6.08 | 0.55 |
| Std. Deviation | 56.72 | | 0.34 | 0.38 | 0.32 | 0.32 | 0.34 | 0.33 | 0.32 | 0.30 | 0.33 | 1.23 | 0.12 |
| N | 31 | | 31 | 31 | 31 | 31 | 31 | 31 | 24 | 24 | 24 | 31 | 31 |
| **High Complexity** | | | | | | | | | | | | | |
| Mean | 643.22 | | -0.34 | 0.08 | 0.17 | 0.23 | 0.29 | 0.26 | 0.15 | 0.19 | 0.16 | 5.25 | 0.48 |
| Std. Deviation | 52.73 | | 0.89 | 0.33 | 0.21 | 0.24 | 0.21 | 0.21 | 0.20 | 0.38 | 0.41 | 1.20 | 0.11 |
| N | 32 | | 32 | 32 | 32 | 32 | 31 | 31 | 19 | 18 | 19 | 32 | 32 |

** $p < 0.01$, 2-tailed.
* $p < 0.05$, 2-tailed.

**Figure 1 Hypothesized Study Main Effects and Mediated Relationships**

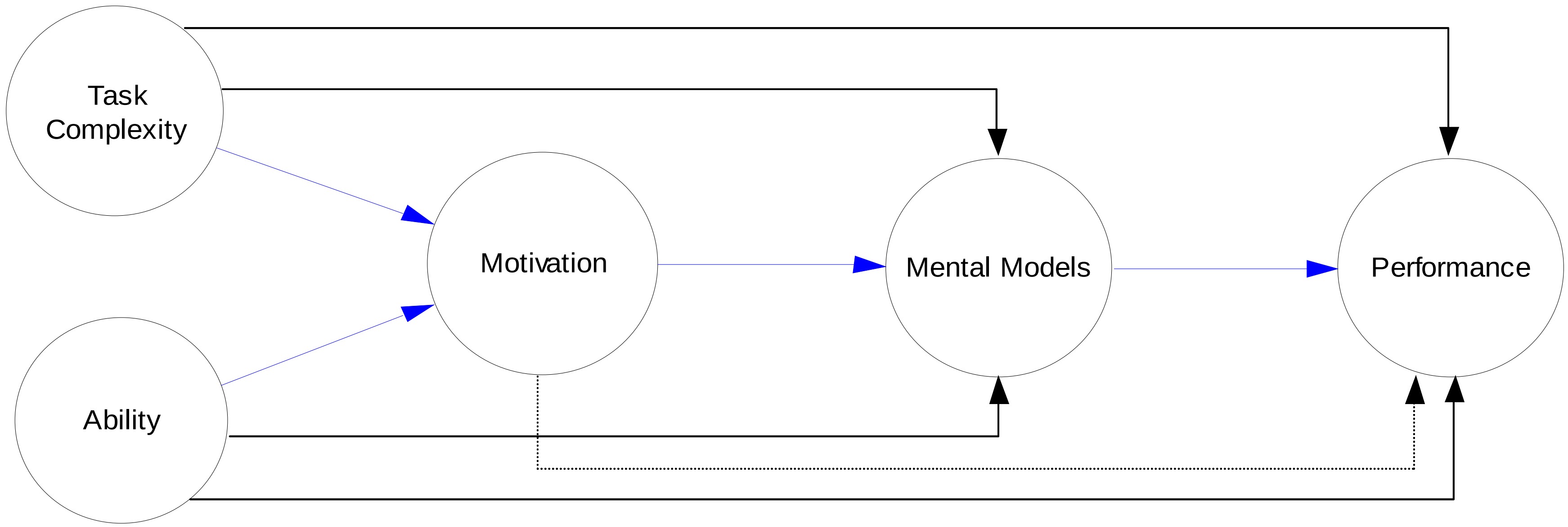


**Figure 2 Influence (Causal Loop) Diagram for Low Complexity Task Condition**

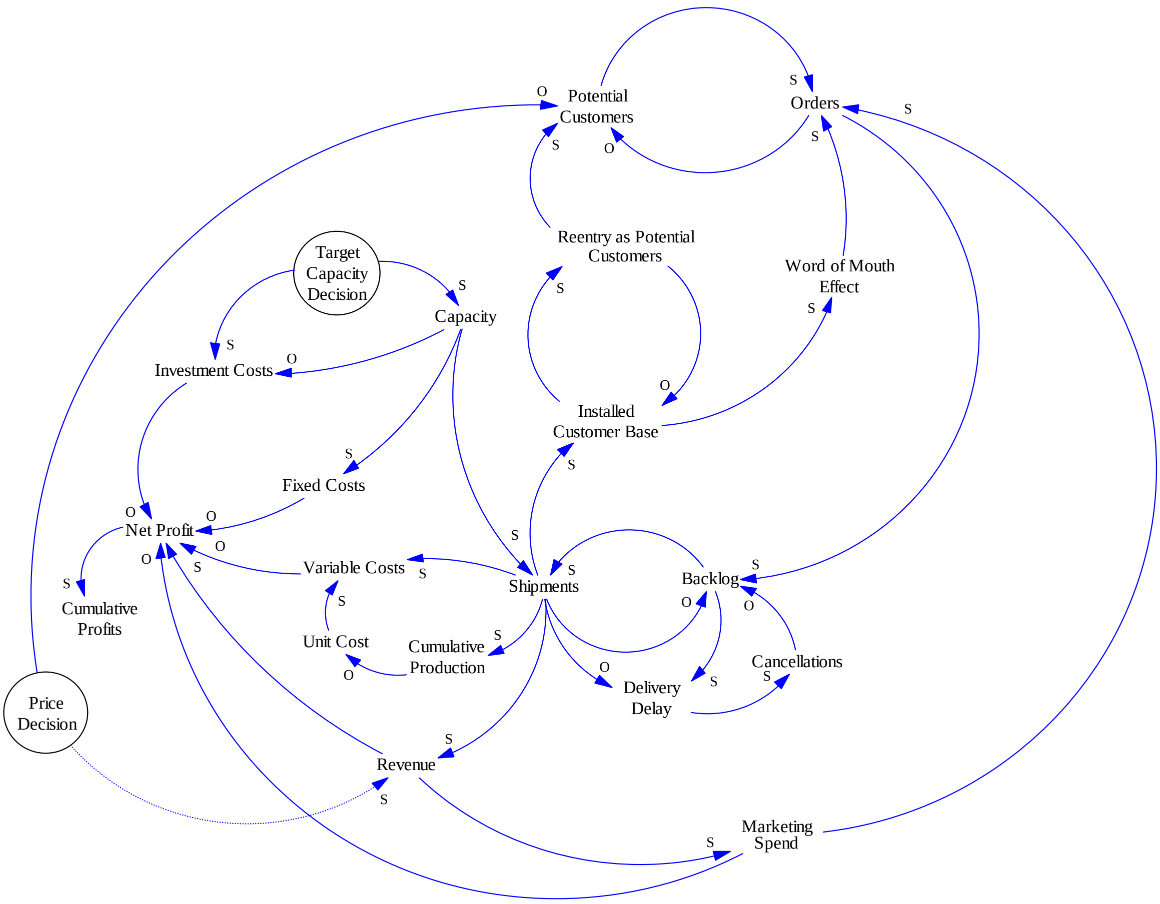


**Figure 3 Segment from the Influence Diagram Questionnaire**


This arrow indicates that an increase in X results in an increase in Y above what it would have been (all else equal). On the other hand, a decrease in X results in a decrease in Y below what it would have been (all else equal). X and Y move in the SAME direction.


In contrast, this arrow indicates X and Y move in the OPPOSITE direction. For example, an increase in X results in a decrease in Y below what it would have been (all else equal). On the other hand, a decrease in X results in an increase in Y above what it would have been (all else equal).

Think about the relationships between these variables that you believe are embedded in the simulator. Relying only on your experience with the simulated firm, draw the appropriate influence arrow(s) for each variable pair and indicate whether the causal influence is in the same or opposite direction using an ‘S’ or ‘O’ at the end of the arrow. Identify any cases in which there is two-way dependency between the variables by drawing the appropriate arrows representing the two-way loop of influence. Focus only on direct relationships and ignore any intervening variables that may result in indirect influence arrows. If there is no direct relationship between the variable pair, write ‘NONE’ between the two variables. If you do not have any idea about the correct answer, then write ‘Do Not Know’ instead of guessing randomly.

| | | |
|---|---|---|
| 1. | Orders | Backlog |
| 2. | Shipments | Backlog |
| 3. | Backlog | Delivery Delay |

**Figure 4 Performance relative to benchmark for low and high complexity groups across learning, immediate-transfer and delayed-transfer trial blocks**

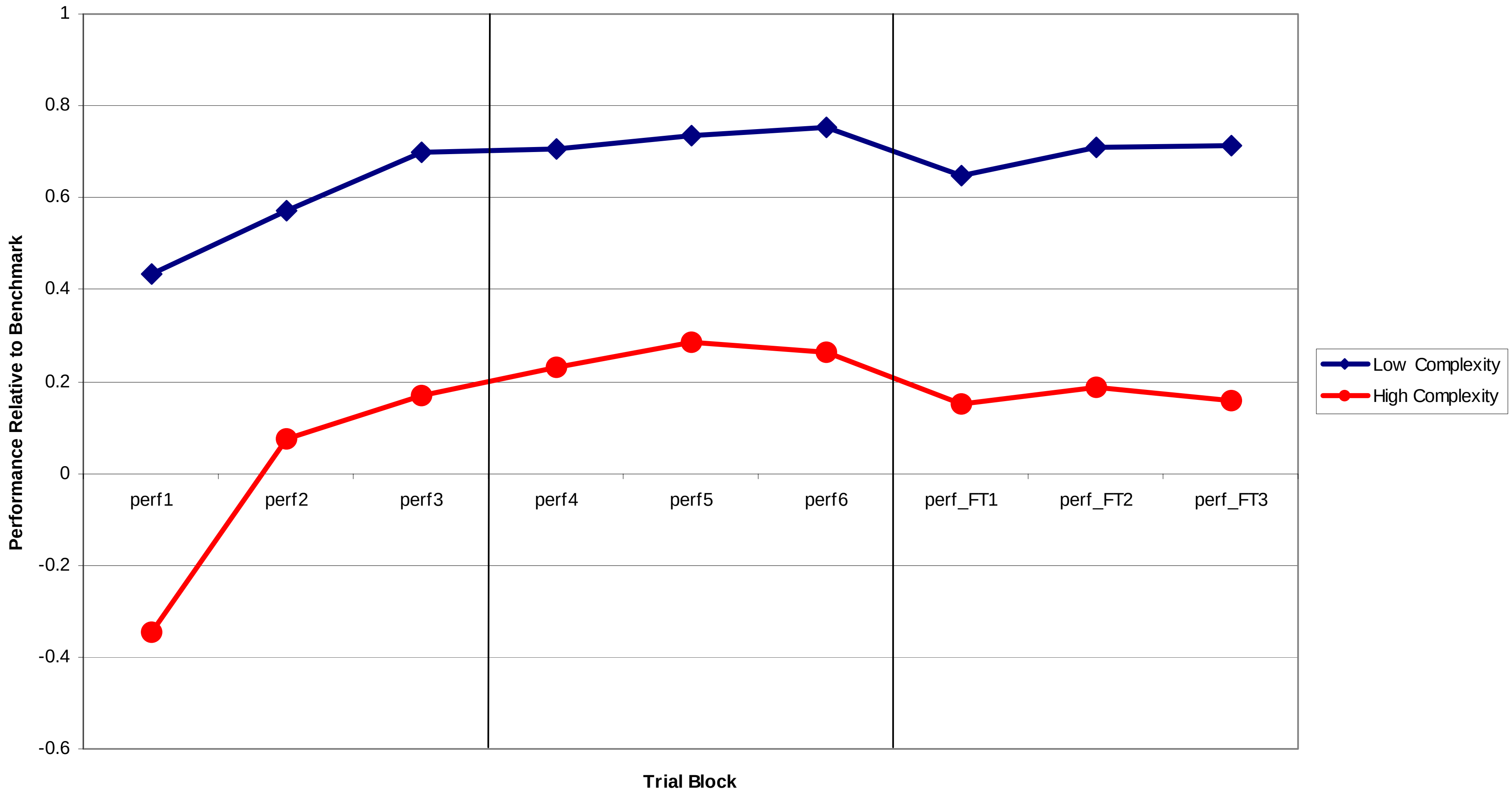


**Figure 5 Mental model accuracy as a percentage of items correct in the low and high complexity conditions**

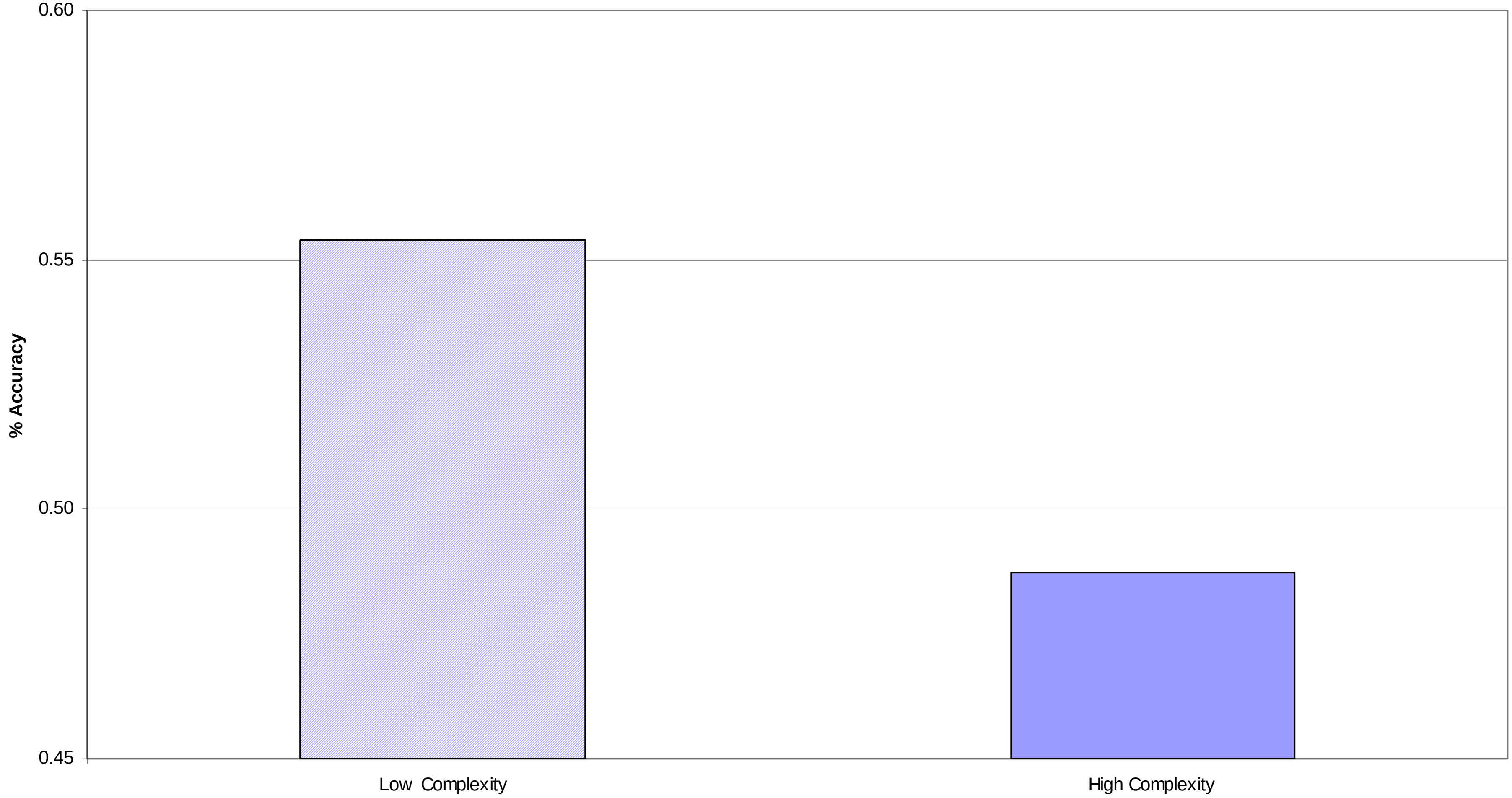


**Table 2 Generalized Linear Model Results for Immediate-transfer Performance as Dependent Variable**

Dependent Variable: Perf456

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 11.282(a) | 4 | 2.821 | 37.675 | .000 |
| Intercept | .019 | 1 | .019 | .260 | .611 |
| GMAT | .063 | 1 | .063 | .838 | .361 |
| Mental model accuracy | .574 | 1 | .574 | 7.662 | .006 |
| Complexity | 6.899 | 1 | 6.899 | 92.144 | .000 |
| Self-efficacy | .031 | 1 | .031 | .407 | .524 |
| Error | 13.551 | 181 | .075 | | |
| Total | 70.519 | 186 | | | |
| Corrected Total | 24.833 | 185 | | | |

a R Squared = .454 (Adjusted R Squared = .442)

**Table 3 Generalized Linear Model Results for Delayed-transfer Performance as Dependent Variable**

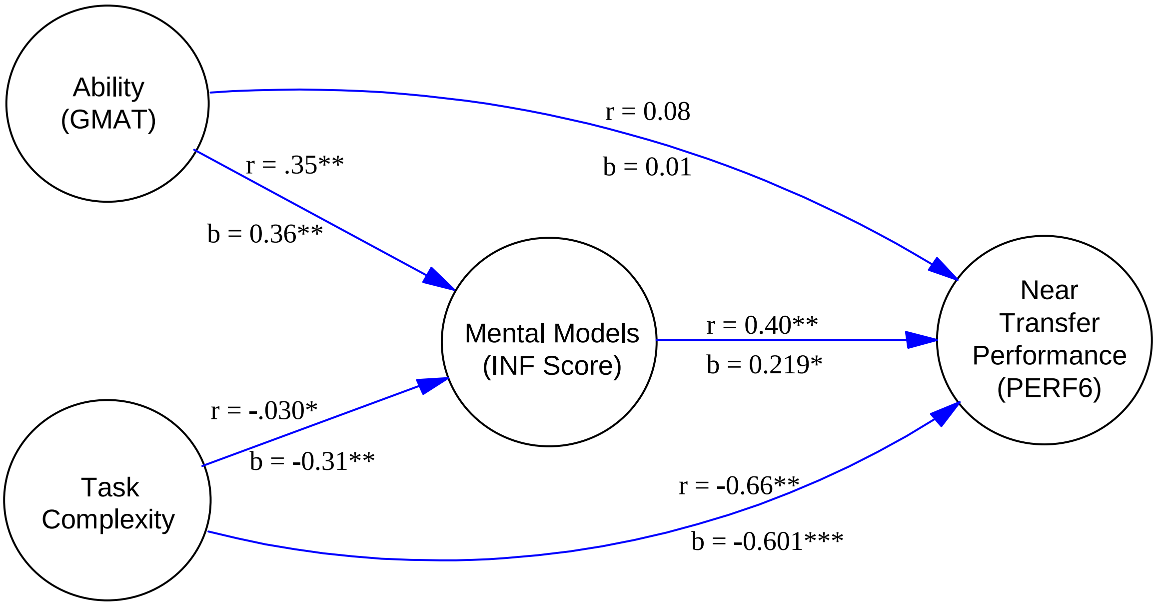
Dependent Variable: PerfFT123

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 9.743(a) | 4 | 2.436 | 24.605 | .000 |
| Intercept | .048 | 1 | .048 | .487 | .487 |
| GMAT | .008 | 1 | .008 | .079 | .778 |
| Mental model accuracy | .788 | 1 | .788 | 7.963 | .006 |
| Complexity | 4.487 | 1 | 4.487 | 45.324 | .000 |
| Self-efficacy | .054 | 1 | .054 | .547 | .461 |
| Error | 12.177 | 123 | .099 | | |
| Total | 49.026 | 128 | | | |
| Corrected Total | 21.921 | 127 | | | |

a R Squared = .444 (Adjusted R Squared = .426)

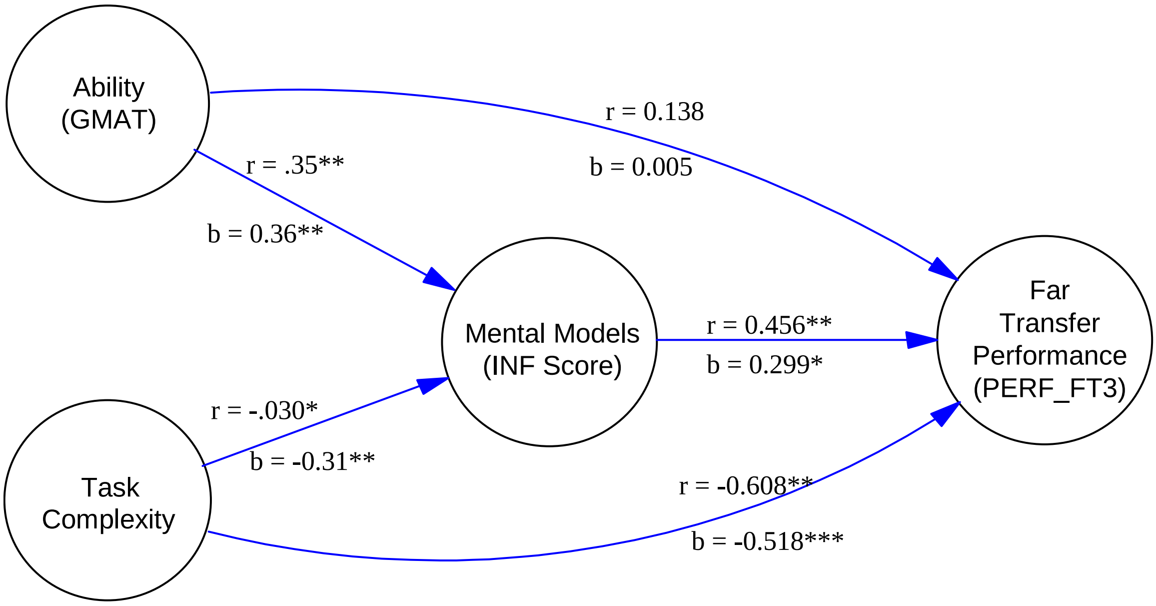
**Figure 6 Path model results for (a) Immediate-transfer and (b) Delayed-transfer**

(6a) Immediate-transfer Path Model




*** p < .001   ** p < .01   * p < .05

(6b) Delayed-transfer Path Model




*** p < .001 ** p < .01 * p < .05

**Table 4 Estimated Information Weights for Price and Target Capacity Decision Heuristics**

| Parameter | Mean reported by Paiche & Sterman (1993) | Mean | Std Dev | Median p-value | % NS |
|---|---|---|---|---|---|
| **LOW COMPLEXITY** | | | | | |
| Target Capacity Heuristic[1]: | | | | | |
| Intercept (c) | | 3.7926 | 2.7268 | 0.0000 | 0.1318 |
| Industry Demand ($a_0$) | | 0.1090 | 0.1750 | 0.0896 | 0.5698 |
| Demand Growth Rate ($a_1$) | | 0.1182 | 0.2551 | 0.1388 | 0.5891 |
| Backlog/Capacity ($a_2$) | | 0.2066 | 0.3855 | 0.0265 | 0.4574 |
| Lag Target Capacity ($\rho$) | | 0.6196 | 0.2547 | 0.0000 | 0.0891 |
| Adj. $R^2$ | | 0.8959 | | | |
| Price Heuristic[2]: | | | | | |
| Intercept ($b_0$) | | 0.4496 | 1.4506 | 0.0324 | 0.4612 |
| Unit Variable Cost ($b_1$) | | 0.1586 | 0.2487 | 0.0049 | 0.2984 |
| Backlog/Capacity ($b_2$) | | 0.0269 | 0.0311 | 0.0310 | 0.4535 |
| Lag Price ($\rho$) | | 0.7813 | 0.1949 | 0.0000 | 0.0116 |
| Adj. $R^2$ | | 0.9059 | | | |
| **HIGH COMPLEXITY** | | | | | |
| Target Capacity Heuristic: | | | | | |
| Intercept (c) | 8.414 | 3.8701 | 3.4409 | 0.0000 | 0.1318 |
| Industry Demand ($a_0$) | 0.383 | 0.0617 | 0.2994 | 0.0896 | 0.5698 |
| Demand Growth Rate ($a_1$) | 0.036 | 0.1286 | 0.2859 | 0.1388 | 0.5891 |
| Backlog/Capacity ($a_2$) | 0.318 | 0.2207 | 0.3828 | 0.0265 | 0.4574 |
| Lag Target Capacity ($\rho_{TC}$) | 0.560 | 0.6532 | 0.2480 | 0.0000 | 0.0891 |
| Adj. $R^2$ | 0.872 | 0.8340 | | | |
| Price Heuristic: | | | | | |
| Intercept ($b_0$) | 3.125 | -0.0790 | 0.7252 | 0.0498 | 0.4979 |
| Unit Variable Cost ($b_1$) | 0.259 | 0.3692 | 0.2919 | 0.0057 | 0.2675 |
| Backlog/Capacity ($b_2$) | 0.016 | 0.0053 | 0.0299 | 0.0809 | 0.5597 |
| Lag Price ($\rho_{Pr}$) | 0.781 | 0.6750 | 0.1802 | 0.0000 | 0.0247 |
| Adj. $R^2$ | 0.947 | 0.9511 | | | |

[1] The model estimated for the target capacity heuristic in both complexity conditions was:

$$\log(C_t^*) = c + a_0 \log(D_{t-1}) + a_1 \log(1+g_{t-1}) + a_2 \log(B_{t-1}/C_{t-1}) + \rho_{TC} C_{t-1}^* + \varepsilon_1$$

[2] The model estimated for the price heuristic in both complexity conditions was:

$$\log(P_t) = b_0 + b_1 \log(UVC_{t-1}) + b_2 \log(B_{t-1}/C_{t-1}) + \rho_{\Pr} P_{t-1} + \varepsilon_2$$