2009

Supporting decision making in a complex world

Jonathan Tall
Louisiana State University and Agricultural and Mechanical College, jtall1@tigers.lsu.edu

SUPPORTING DECISION MAKING IN A COMPLEX WORLD

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

In

The Department of Psychology

By
Jonathan Tall
B.S., Northwestern State University, 2003
M.A., Louisiana State University, 2006
May, 2009

TABLE OF CONTENTS

LIST OF TABLES…..……... iv

LIST OF FIGURES……... v

ABSTRACT……. vi

INTRODUCTION……. 1
Features of DDM……..……. 3
Previous Research…….……. 4
Increasing Complexity..…….…… 6
Diagnostic Medical Task..……..…….…..…… 8
Current Study…...…….. 9

EXPERIMENT 1……. 13
Methods..…….…… 13
Participants and Design……...…… 13
Task and Procedure…….…….. 13
Results and Discussion…… 14
Prescription Test……..…… 15
Mental Model Test……...…… 17

EXPERIMENT 2……. 21
Methods..…….…… 21
Participants and Design…..…….…… 21
Task and Procedure…….…….……...……..……. 22
Results and Discussion…… 22
Prescription Test……..…… 22
Mental Model Test…….…….…… 24

EXPERIMENT 3……. 27
Methods..…….…. 29
Participants and Design…….……...…… 29
Task and Procedure…….……..…….…….... 29
Results and Discussion…… 30
Prescription Test……..…… 31
Mental Model Test…….……..… 35

GENERAL DISCUSSION…….. 38

REFERENCES…… 42

APPENDIX A: ON-SCREEN TEST INSTRUCTIONS……..…… 45

APPENDIX B: SAMPLE SIZE CHANGES……..…….…… 46

VITA…… 47

LIST OF TABLES

1. Main effect template for patients in Exp. 1 & 2……..…….…….…… 17
2. Exp. 1 overall prescription test results…….……. 18
3. Exp. 1 correct prescription test results…….……. 18
4. Exp. 1 mental model test results……..……. 20
5. Exp. 2 overall prescription test results……. 23
6. Exp. 2 correct prescription test results……. 23
7. Exp. 2 mental model test results…….……...…..…….…….…… 26
8. Main effect template for 10 out of 15 patients in Exp. 3.……. 30
9. Main effect template for last 5 patients in Exp. 3……. 30
10. Exp. 3 accuracy prescription test results…….….. 31
11. Exp. 3 prescription rate prescription test results…….….. 33
12. Exp. 3 precision prescription test results…….…. 34
13. Exp. 3 mental model test results……..……. 36

LIST OF FIGURES

1. Main game screen…..……... 15

2. Prescription test…….…… 16

3. Mental model test……...…… 17

ABSTRACT

Recent research has questioned whether explicit thinking is necessary or even useful for complex decision making (Gladwell, 2005; Dijksterhuis & Nordgren, 2006; Newell, Wong, Cheung, & Rakow, in press). The present research approaches this issue by examining how different types of decision support facilitate/hinder performance in a diagnostic medical task. The results from 3 experiments indicate that providing an external memory aid improves performance in complex tasks. Additional support in the form of a coding procedure improved awareness of the magnitude of drug effects, but did not improve detection of negative side effects. The results suggest that while performance is improved, participants prescribed the correct treatments for the wrong reasons. Important differences in task representation (and resulting knowledge) due to the different forms of decision support are also discussed.

INTRODUCTION

Recent research has questioned whether explicit thinking is necessary or even useful for complex decision making (Gladwell, 2005; Dijksterhuis & Nordgren, 2006; Newell, Wong, Cheung, & Rakow, in press). It has been argued that better decisions can be made if one is given time to consolidate information implicitly, either by preoccupying oneself with another task or by sleeping, or by making an immediate judgment without reflection (Dijksterhuis & Nordgren, 2006; Newell, Wong, Cheung, & Rakow, in press). This is contrary to the common sense notion that explicit, well-thought decisions are superior to 'snap' judgments. To help illustrate, explicit processing is typically characterized as conscious, controlled, and deliberate processing. Implicit processing, in contrast, is said to be intuitive, automatic, and involve experiential processing. Thus, the contention revolves around whether or not one should 'think' about things before making a decision.

Curiously, this debate appears focused on an either/or perspective. That is, one should focus predominantly on one mode or the other. Consequently, by focusing on the 'wrong' mode, we are likely to arrive at a poor decision. However, it is difficult to conceive of a situation in which we use only one mode of processing. If we engage both modes when experiencing and evaluating the world, then it may be more productive to consider how the interaction of the two systems affects learning. As such, it is not that we engage in one mode or the other, so much as the degree to which we engage in each (Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989; Stanley, Mathews, Buss, & Kotler-Cope, 1989). Thus, any failures are not due to one mode of processing alone, but rather are failures of the two modes to work together effectively (Sallas, Mathews, Lane, & Sun, 2007; Sun, Zhang, Slusarz, & Mathews, 2007). Such failures would be examples of negative interactions, while examples of the systems productively benefiting from each other would be positive interactions.

One way in which one might alter these interactions is by supplementing the underlying processes with cognitive aids. Cognitive aids are external tools (e.g., calculators or check lists) that can be used to perform a task. They generally move the thinker toward relying more on the explicit mode because they stimulate explicit reflection. However since both modes depend on the experience of training, it is possible for cognitive aids to either facilitate or impair both types of learning. For example, cognitive aids which alter ones' representation of events could either impair or facilitate both modes of learning. If performance depends on effectively using both modes of thinking, then it should be possible to supplement processing with appropriate cognitive aids to create positive interactions (Mathews et al., 1989; Stanley et al., 1989; Sallas et al., 2007; Sun et al., 2007).

In order to make valid generalizations, it is important that any task used simulates some of the richness of complexity found in the world. In the real-world, the decisions that we make are often based on multiple factors. Proper management of resources is a good example. We can think of this not only in terms of a corporation with employees, product deadlines, etc., but also in more personal terms, like our own finances with bills, grocery expenses, travel expenses, whether income is bi-weekly or monthly, etc. These decisions are made complex not only by the number of features one must evaluate, but also through the influence of time on the decision sequence. For example, the amount owed on next month's bill is due in part to whether or not the bill was paid this month. There are also 'side effects' to each decision, such as the loss of power, because of a delinquent bill.

One way to capture these rich complexities in the laboratory is through the use of dynamic decision making (DDM) experiments. DDM experiments involve a series of interdependent decisions that are made in an environment that changes as a function of the

decision sequence, independent of the decision sequence, or in both ways (Edwards, 1962). DDM is commonly studied through the use of 'microworlds.' Microworlds (Turkle, 1984) are simulations that share many of the features that are believed to be present in real-world scenarios. Thus, they allow us to study decision making in a variety of complex scenarios.

Features of DDM

To better understand how they can aid in the study of decision making in a complex world, it is important to highlight the major features of DDM experiments. These features are dynamics, complexity, opaqueness, and dynamic complexity (Brehmer, 1992; Diehl & Sterman, 1995; Gonzalez et al., 2005). Dynamics refers to the time element of the system. That is, the state of the system at time t is dependent upon the state at time $t - 1$ (Rouse, 1981). One example of an area of interest with respect to the dynamics of a system is the amount and quality of feedback over time (feedback loops). Complexity can be defined by multiple features. The number of variables, the number of variable interactions, and the types of variable interactions all contribute to system complexity (Gonzales et al., 2005). Opaqueness refers to the 'invisibility' of one or more aspects of the system (Brehmer, 1992). Sterman (1989) defines the dynamic complexity as combining aspects of dynamics and complexity in order to highlight their effects on information available from the feedback loop.

One example of a DDM task is the process control task (Berry & Broadbent, 1984; Stanley et al., 1989; Lane, Mathews, Sallas, Prattini, & Sun, 2008). The process control task has had many cover stories, such as controlling: a sugar factory, a nuclear reactor, or a computer personality named 'Clegg'. During the task, participants are asked to maintain a particular level of output (i.e. sugar pellets) by manipulating a single input (i.e. number of workers). The program operates according to a simple formula: output = |2*(workers) – (previous output) +

(noise)|. Noise in the task, when included, was either +1000, -1000, or 0 with equal probability. Participants were given immediate feedback on each trial.

Previous Research

Previous research using the process control task has yielded evidence for positive facilitation through cognitive aids. Roussel (1999) used the sugar factory version of the process control task. He attempted to aid learning in the task by providing cognitive support. One of his support groups was given a hint (a list of 3 valid solutions to particular states). It was assumed that this small set of answers would assist participants in learning how to control the reactor. Indeed, this is what Roussel (1999) found. The hint group outperformed an experiential control group that received no support at all. This data suggests that cognitive support can lead to increased performance in DDM.

Lane et al. (2008) used the nuclear reactor version of the process control task to perform a follow-up study to Roussel (1999). Though Roussel (1999) had demonstrated that the hint could lead to better performance, he did not demonstrate whether the improved performance was the result of having 3 correct answers (and thus getting 25% of the states correct all of the time), or the result of improved learning from the decision support. Lane et al.'s (2008) design included 3 separate groups: an experiential group (which just performed the task without aid), a quiz-only group, and a quiz + hint group. The latter two groups were periodically interrupted and asked a series of questions about how to solve the task. The results from the experiment showed that while the quiz caused the latter two groups to reflect on the task (and hence perform better than the experiential group), the hint + quiz group performed the best overall. Moreover, analysis of the data showed that this improvement occurred in the non-hint states as well. That is, improvement was not restricted to the small subset of correct answers provided by the hint, but

rather the hint was used as a guide to aid participants in learning to control the reactor. This data further supports the hypothesis that cognitive support can lead to increased performance in DDM.

Mathews (unpublished) took a slightly different approach. This study used the cover story of controlling the computer personality 'Clegg.' The key manipulation involved the degree of support that the participant received. Participants received either: no support, reflection, reflection + hint, or a partial formula. In the two reflection groups, participants were asked to write a set of instructions for someone else describing how to control Clegg. The reflection + hint group received the same type of hint as the Roussel (1999) and Lane et al. (2008). Finally, the partial formula group was given the formula: (N) * (input) [+, -, x, or /] (previous state) = (new state). They were told that 'N' stood for some number. Thus, they only needed to figure out 'N' and one operator. The results showed that when there was no noise in the system, each group performed similarly. However, when noise was used in the program, some types of support hindered performance. The reflection group and partial formula group both performed worse than the experiential group. The reflection + hint group however still outperformed all of the other groups.

The differences in performance in the Mathews (unpublished) study highlight an intriguing interaction between decision support and task complexity. As complexity changed in the task (qua change in variability due to noise), the quality of the benefit each support provided changed. One possible explanation deals with task representation. The representation that one has affects the salience (or obscurity) of particular features within the task. These changes can also facilitate a change in the strategy employed as what becomes deemed as the 'relevant' features change.

To relate back to the DDM structure, these are changes along the opaqueness and dynamic complexity dimensions. For example, the partial formula group clearly sees the relationship between the variables. That is, there is a single underlying formula that ties each of the input and output states. This removes some of the opaqueness. However, the increased salience of the relationship blinds the participant to influence of noise (since noise is not listed as a feature of the formula given to them). This blindness can be attributed to a change in dynamic complexity. That is, participants are likely to interpret feedback differently. Rather than seeing that their hypothesis is getting them 'roughly there', the increased salience of the formula might incline them to reject anything that does not produce a clear 1-to-1 mapping.

Based on this interpretation, what is guiding the effectiveness of the decision support is not merely the providing of aid, but rather how that aid supports the relevant underlying processing. Since the type of aid can influence the type of processing used, it is important to understand the processing requirements of the task. In other words, by understanding the requisite requirements, one can better tailor support for that task and improve the quality of decisions made within it. Similarly, one can come to understand the task requirements by the aid that benefits performance in it. Before the issues are addressed further, it is important to first examine how complexity and processing interact more broadly.

Increasing Complexity

Each of the previous studies have demonstrated that some form of decision support can aid in improving performance in complex tasks like DDM. However, there remains a question of whether or not decision support will continue to improve performance as the task continues to increase in complexity. Dienes and Berry (1997) and Reber (1989) argue that implicit learning is more robust than explicit learning. This notion is amplified by Dijksterhuis and Nordgren (2006)

who propose that explicit processing is good for simple decisions, but complex decision making is better left to implicit processes. They argue that as complexity increases in a system one should shift to more implicit processing of information, and thus performance should not improve with decision support that encourages explicit reflective thought.

One way to understand this argument is by thinking in terms of representation and processing limitations. Explicit processing relies on limited resources. As task complexity increases, there is much more information to monitor. By shifting one's representation of the system, you can limit the amount of information that you need to monitor. However, as discussed previously, incomplete task representation can hinder performance. Implicit processing is robust with respect to variability and is a much higher capacity system. If uninfluenced by explicit processes, it should lead to more accurate decisions (Dijksterhuis & Nordgren, 2006).

There is, of course, evidence for the opposite. Raab (2003) conducted a study of implicit and explicit learning in sports (handball, volleyball, and basketball). Participants watched the sports played on video. They later were asked about the tactics employed by each team. The results showed that implicit learning excelled during situations of low tactical complexity. During highly complex situations, participants benefitted from additional instructions that helped explicit learning by calling attention to key elements in the situation.

Mathews (unpublished), also, found decision support to be helpful in complex situations. All of the groups in this study performed relatively equal when there was no noise in the system. However, when noise was introduced, there was a sharp change. The experiential group performed better than most of the other groups, but not better than the reflection + hint group. Therefore, there at least seems to be some evidence to suggest that some forms of decision support can still provide aid (positive facilitation) as complexity increases. More critically,

improvement is restricted to forms of decision support that best augment the underlying processing requirements of the task.

These processing requirements may not necessarily be implicit or explicit in the strict sense. Instead, support can facilitate or hinder both types of processing based on the degree to which in encourages reliance on one type over the other. For example, support that heavily engages strategic processing encourages explicit processing. This could change the representation of the task and call more attention to critical features within the task. The increased attention to the relevant features could then allow quicker implicit pattern abstraction from the feedback loop. Thus, a more explicit type of support could improve both systems and increase overall performance through positive facilitation. The goal of the present research is explore this relationship by examining the effects of cognitive support on learning in complex tasks.

Diagnostic Medical Task

To help address the issue of decision support in complex tasks, we have designed a new DDM task. The task was designed to resemble some of the features encountered in real medical practice. A doctor typically meets with many patients, most of whom do not see the doctor on regular intervals. Doctors usually have a fair (but limited) amount of drugs to select from when trying to mitigate a health problem. One important feature in determining this choice is the effect that the drug has on the patient's other health measures. That is, one does not want to replace a problem with a new one.

To capture these features, the diagnostic medical task (DMT) requires participants to assume the role of doctor. In this role, the participants take on the responsibility of trying to improve the health of their patients. More specifically, they are trying to reach a designated

‘optimal’ level in one health measure (Blood sugar), while trying to maintain a ‘safe’ level in each of the other health measure of the patient. This is accomplished through simulating meetings with patients during which the participant is allowed to make prescriptions from a list of 5 drugs (inputs). They are allowed to prescribe 1 or 2 of the drugs on each visit, and must monitor the effect of the prescription on 5 health measures (outputs). Participants judge each output level according to a list of zones on the bottom of the monitor. Each patient’s reaction to prescribed medications is governed by a template that is unknown to the participant. Patients can either share the same template, or have unique templates. Furthermore, the combinations of medications do not always operate by summing the effects of the individual drugs. In some cases, they may have effects that are completely unrelated. Finally, participants must meet with all of the patients on a given round, before seeing the result of their prescriptions. Since the patient order is randomized on each round, the feedback delay for a particular patient can be between 1 and 28 trials.

Two types of tests were conducted to assess performance and learning in the task. A prescription test was given to assess the participant’s ability to prescribe the correct medication to the proper individual. This was followed by a second test, the mental model test. It determined the participant’s explicit knowledge about the effects of each drug.

Current Study

The current study uses the increased complexity offered by the DMT to study decision support. Participants were assigned to training groups that differed in the type of decision support received during (and prior to) training. A no support group that received no support was compared to the various support conditions in each experiment.

In Experiment 1, two levels of decision support were provided. The first type of support, called ‘list’, is primarily a memory aid. Participants were given a list of all the drug combinations that they used to record their knowledge of drug effects during training. The second type of support group (list + strategy) was given the same list as the previous group, but was also given additional instructions on how to use the list effectively to keep track of their knowledge and eliminate non-effective drug combinations over time. The participants were also taught a means of coding the magnitude of the main effect of the drug in accordance with the five possible output zones. The five zones were: Excellent (90-100), Good (70-89), Acceptable (50-69), Poor (20-49), and Critical (0-19). This procedure should help participants develop a representation of each treatment’s effect at the appropriate level of abstraction. Such a representation would remove some of the opaqueness due to noise in the system. That is, patients receiving the same treatment do not stay at the same level (number within the zone) as each other, nor do they stay at the same level visit by visit. Participants using the above mentioned coding procedure should be relatively resilient to these fluctuations due to the appropriate level of abstraction (the zones). Additionally, they were given strategic information. That is, they were taught to narrow their pool of drug choices by eliminating the drugs in a systematic fashion. On the first reflective period, they would eliminate all drugs with ‘Critical’ effects. On the next reflective period, they would eliminate any drugs with ‘Poor’ effects, and so on. This strategy should help them to effectively reduce the pool of drugs that they are hypothesis testing during the practice phase.

To help frame the manipulation, one can conceive of the no support condition as relying predominantly on implicit learning. Without support, the potentially long feedback delays make keeping track of the visit by visit changes for each patient (in addition to updating one’s knowledge about the 15 combinations) difficult for explicit processes. In addition to enhancing

memory, each of the support conditions push participants further towards the explicit end of the spectrum because they are encouraged to reflect about the drug effects. For example, the list condition provides very general/abstract information (in terms of drug effects). It affords a slight advantage explicitly over the no support group because one can record and reflect upon this general knowledge. The list + strategy condition provides the most explicit support. The task is structured around very specific information (i.e. the output zones). The strategy implementation should promote a better representation of the task by limiting focus to the relevant output while the range of input choices is effectively reduced. It was hypothesized that performance would improve with each additional level of cognitive support. However, the results did not support this hypothesis. Instead, the moderate level of support (list only) yielded superior performance in the task as compared to the list + strategy.

Experiment 2 examines the effect of teaching a rich coding scheme without using the elimination strategy that failed to enhance performance in Experiment 1. Experiment 2 replicated the format of Experiment 1 with the exception of the list + strategy condition. Rather than employing a strategy, a new condition (list + code) was used in which participants received the same list as in Experiment 1, but were given additional instructions for a specific coding procedure. Participants were taught to record information about the specific zone each drug moved the patient into as well as information about whether or not there were any side effects of the drug. This condition would provide more salient information about each drug's effects, while preserving the participant's ability to employ their own 'natural' strategies. It was again hypothesized that performance will improve with each additional level of support.

Experiment 3 attempted to extend the findings of the previous experiments to an even more complex task that is more similar to real medical practice and shed more light on the

relationship between type of task complexity and useful task support. Unlike Experiments 1 & 2, Exp. 3 partitioned patients into two template groups (i.e., typical and atypical patients) that are unknown to the participant. Therefore, participants will have to determine the proper prescription on a patient by patient basis. This is more akin to the real-world where all patients do not always respond in the same way to the same treatment. The new variable of patient type increases the level of opaqueness in the task by adding variability to the effects of key drug combinations. It is hypothesized that adding a form of decision support that helps make this variability more salient will enhance performance. In addition to the no support and list conditions, we tested a different form of support, designed to make these patient differences less opaque, a matrix. This new matrix condition received a matrix of patients and drugs, while the matrix + code condition received both the matrix and the same coding procedure as the list + code condition. The purpose of these manipulations will be discussed in more detail in the introduction to Exp. 3.

EXPERIMENT 1

Methods

Participants and Design. All 88 participants were undergraduates enrolled in psychology courses at Louisiana State University. They participated in the experiment to partially meet course requirements. There were three training conditions: No support (control), List, and List + strategy. Participants were randomly assigned to one of the three conditions. There were 29 in the no support group, 30 in the list group, and 29 in the list + strategy group.

Task and Procedure. In this task, participants play the role of a doctor. Their goal is to improve the health of their patients. This was accomplished by learning the effects of various medications, and then prescribing the best medication for each person. The participants were not told ahead of time about any of the drug effects. Their goal was to find the medication that improved the main output (Blood Sugar) to 'excellent', while maintaining the other 4 health measures at 'acceptable' or better. Separate instructions were given to each condition. The no support condition received information about the task. The list condition received information about the task and about the external aid (such as when they would be allowed to use and what information that were allowed to record in it). The list + strategy condition received all of the previous instructions, but were additionally given information on how best to code information on their external aid, as well as a strategy to help eliminate non-working drug combinations. After receiving task instructions and watching a brief tutorial, participants started the learning phase. This phase was divided into rounds which involved "meeting" with each of the patients (N = 15) under their care (See **Figure 1**). The order of patients as well as the starting values on the 5 health measures (output variables) was randomized. Each round, the order of patients was again randomized. On each trial, participants indicated the drug or pair of drugs to be used with

the individual patient. For the first two rounds, all participants were required to "continue the previous doctor's prescriptions" which were provided in a list with the instructions. This was done to ensure that each participant was exposed to the full effects of each drug combination. All drug combinations took 2 rounds before reaching full potency. In the support conditions, participants were interrupted every 3 rounds and given an opportunity to record information. During this time, the simulation screen was obscured and they were allowed to make notes on the list provided (in accordance with the instructions), and could review the information from previous recordings. The list was then put away and training resumed. After completing the learning phase (40 minutes), participants were given two tests. Additional instructions preceded each test and appeared on-screen (See **Appendix A**). The first test determined their ability to prescribe the correct medications for each patient (See **Figure 2**). Then on a second test, they were evaluated on their knowledge of the impact of the drugs on all measures of behavior (See **Figure 3**). For each measure, they indicated the magnitude and direction of a drug's impact on a 5-point scale that ranged from highly negative to highly positive (-2 <-> +2).

Performance on the task was primarily assessed by the number of correct prescriptions made on the prescription test. A second measure of participants' knowledge was gained by analyzing the accuracy of their specific knowledge of the drug effects on all of the output variables using what we call the mental model test.

Results and Discussion

Analysis of the results is focused on the 3 drug combinations which brought 'Blood Sugar' into the desired range. The 3 drug combinations of interest are: DE (this drug combination was excellent for 'Blood Sugar' and also had an excellent side effect); AD (this drug combination was excellent for 'Blood Sugar' and had no side effect); and BC (this drug

combination acted as a lure. It was excellent for 'Blood Sugar' but had a very strong negative side effect). DE (positive side effect) and AD (No side effect) were considered correct answers on the prescription test (for a complete list of drug effects, see **Table 1**). The analysis is also divided up by test.

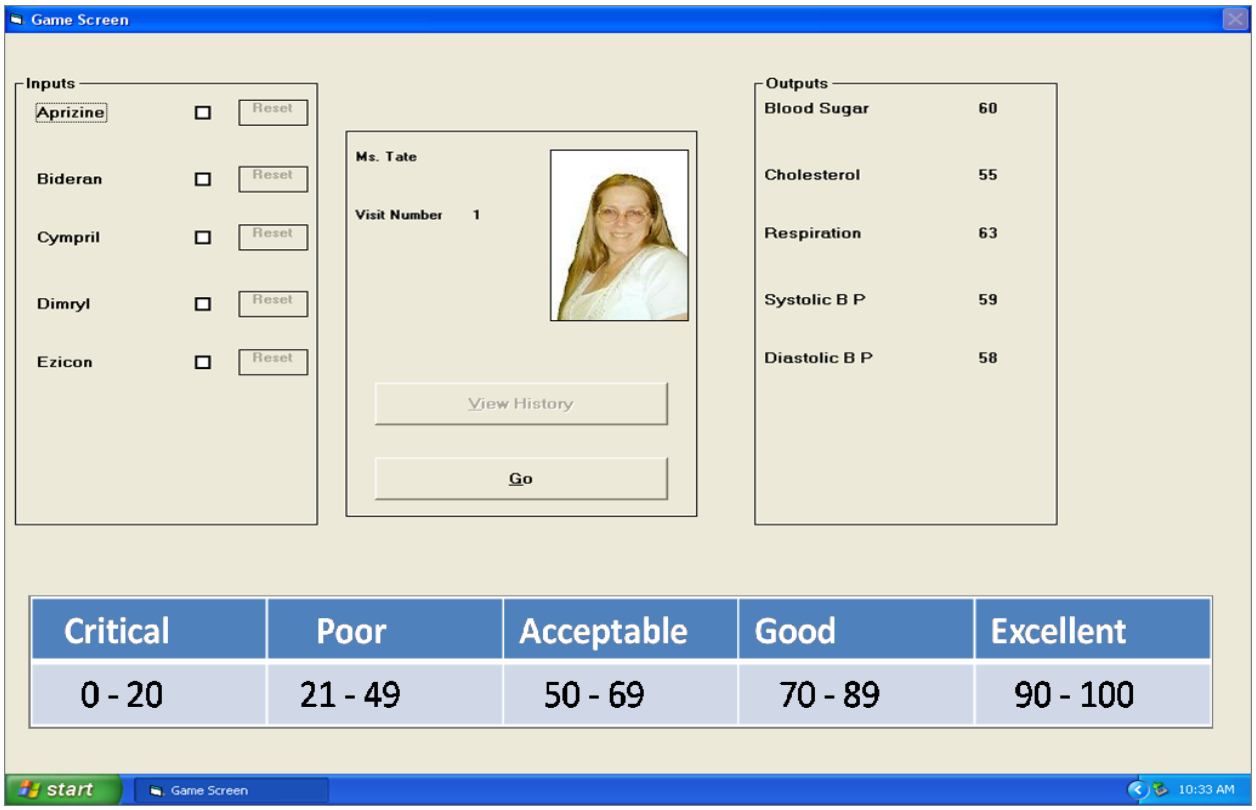

Figure 1. Main game screen. Zones are added to picture, but were actually on the bottom of the monitor.

Prescription Test. On the prescription test, we compared the percentage of time that each group prescribed each of the drugs that had a strong positive effect on the main output variable (see **Table 2**). A one-way ANOVA revealed a significant effect of group on prescribing the correct drugs, $F(2, 87) = 5.78$, $p < .01$. There were no significant differences found with respect to the lure drug, $F(2, 87) = 0.78$, $p > .05$. Post-hoc tests were conducted using Fisher's LSD. The post-hoc tests revealed the no support group prescribed one of the correct drugs significantly less than the list group and the list + strategy group ($p=.001$ and $p<.05$, respectively). A separate one-

way ANOVA was conducted on each of the correct drugs individually (see **Table 3**). There were no significant differences found with respect to the DE (positive side effect) combination, $F(2, 87) = 0.21$, $p > .8$. The one-way ANOVA on the AD (No side effect) combination, however, was significant, $F(2, 87) = 3.53$, $p < .05$. Post-hoc tests using Fisher's LSD only showed a difference between the no support and list conditions ($p<.01$). The list condition prescribed the AD (No side effect) combination significantly more often than the no support condition did.

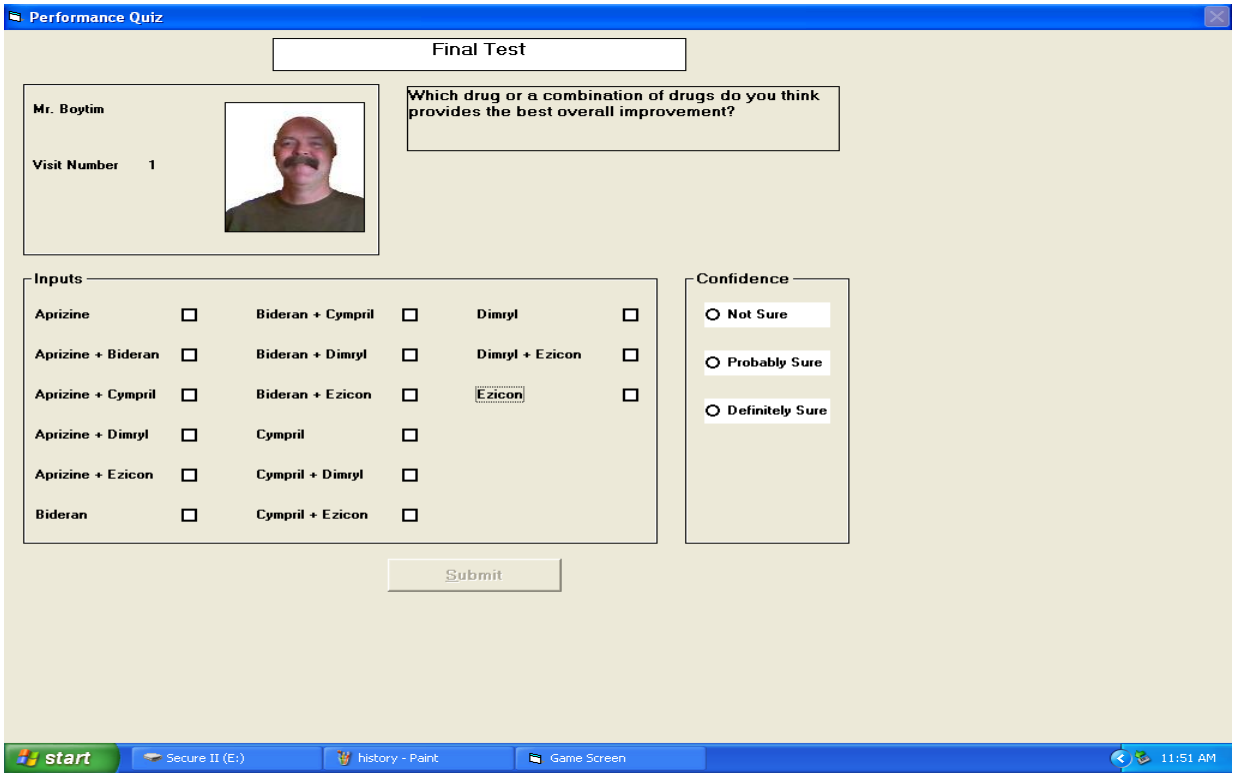

Figure 2. Prescription test

These results suggest that the cognitive support provided did increase performance in the task. That is, participants in each of the support conditions prescribed a correct combination more often than the no support condition. Individual analysis of the correct drug combinations showed that while the list + strategy condition required both correct drug prescriptions to be totaled in order to show a difference, the list condition was far superior to the no support group at prescribing AD (No side effect) alone.

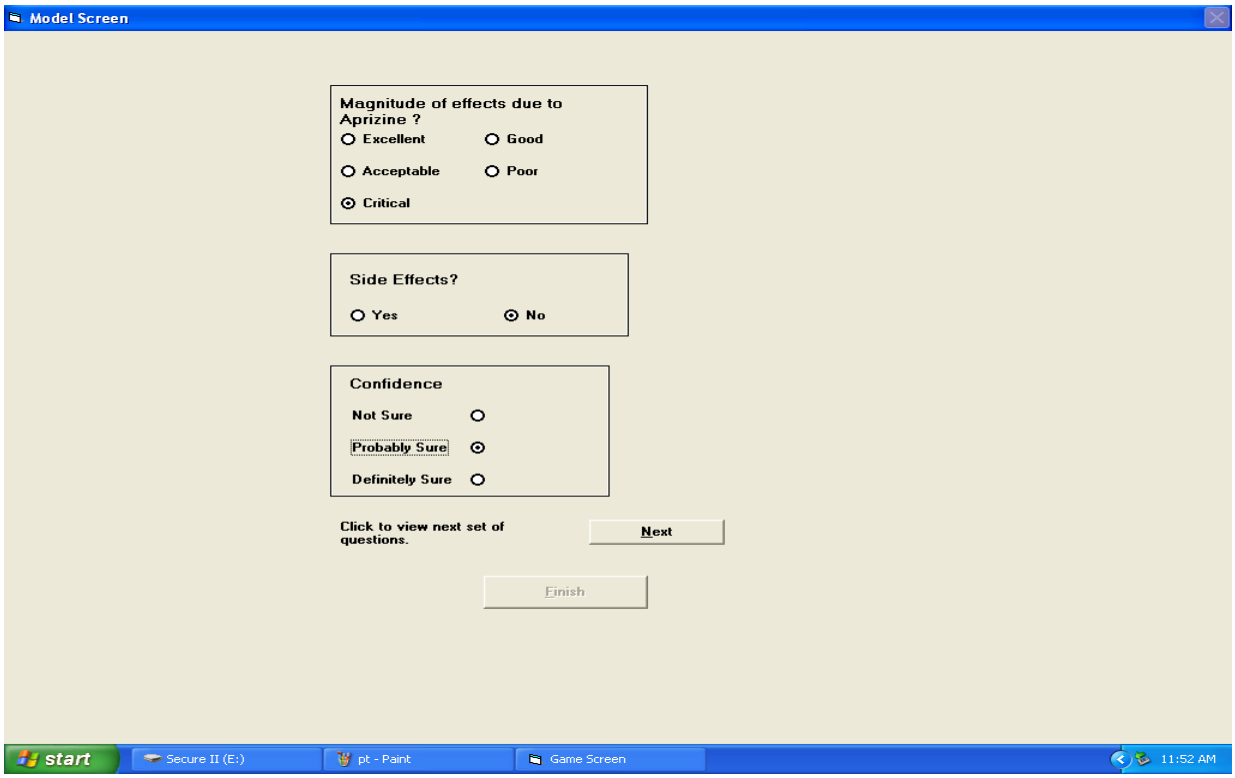

Figure 3. Mental model test

Table 1. Main effect template for patients in Exp. 1 & 2. Side effect is listed in parentheses. Scale is + 2 to – 2. +2 being a strong positive effect (excellent zone) and -2 being a strong negative effect (critical zone).

	Aprizine	**Bideran**	**Cympril**	**Dimryl**	**Ezicon**
Aprizine	0 (0)	0 (0)	0 (0)	+2 (0)	0 (-2)
Bideran	xxxxxxxxxxx	0 (0)	+2 (-2)	0 (0)	0 (0)
Cympril	xxxxxxxxxxx	xxxxxxxxxxx	-2 (+2)	0 (0)	0 (+2)
Dimryl	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	0 (0)	+2 (+2)
Ezicon	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	-2 (0)

Mental Model Test. Next, we turn to the mental model test (see **Table 4**). The mental model test assessed participants' perceived effect of the drugs. A one-way ANOVA revealed a significant effect of group on perceived main effect the AD (No side effect) drug combination,

$F(2, 87) = 3.40$, $p < .05$. However, there were again no significant differences found with respect to the lure drug, $F(2, 87) = 1.74$, $p > .15$. Nor, were there any significant differences between the groups with respect to the DE (positive side effect) combination, $F(2, 87) = 0.09$, $p > .9$. Post-hoc tests using Fisher's LSD on the AD (No side effect) combination again revealed only significant differences between the list group and the no support group ($p = .011$). Separate analyses on the perceived side effect of each drug combination did not reveal any significant differences.

Table 2. Exp. 1 prescription test results. This is a breakdown of the correct drugs column from the previous figure. Number listed is the percentage of time the correct drugs or lure drugs were prescribed. Standard error is in parentheses.

	Prescription test	
	Correct Drugs	**Lure Drug**
No support	49.20 (7.17)	7.13 (2.33)
List	80.44 (5.19)	13.78 (4.79)
List + Strategy	69.66 (7.33)	8.74 (4.23)

Table 3. Exp. 1 prescription test results. Number listed is the percentage of time the correct drugs were prescribed. Standard error is in parentheses.

	Prescription test	
	DE (positive side effect)	**AD (No side effect)**
No support	18.62 (5.06)	30.57 (6.83)
List	22.22 (6.62)	58.22 (7.02)
List + Strategy	24.60 (7.55)	45.06 (8.22)

The results of the mental model test show that the advantage of the cognitive support provided was primarily limited to the list condition. While this still supports the idea that cognitive support can improve performance on complex tasks, it is unclear why performance was

hindered so much in the list + strategy condition (as compared to the list alone). While the list + strategy condition performed nearly as well as the list condition on the prescription test, the perceived main effect of the AD (No side effect) combination was marginally closer to that reported by the no support condition. This is even more troubling if you consider that the list + strategy condition kept track of the exact zones the main effect was moved to by each drug. If anything, this should have resulted in more precise knowledge about the AD (No side effect) combination's effect. It is also not clear why the perceived main effect of the DE (positive side effect) combination was so low in each group. This drug combination had clearly superior overall effects (once the side effect is considered), or at the very least, comparable effects to the AD (No side effect) combination (in the sense that it achieved the same main effect without a negative side effect). It is possible that the perceived effect is so low due to the fact that it was prescribed such a small percentage of the time. But, as previously stated, it remains unclear why this combination was prescribed so little. It is also interesting (although a bit disturbing in its practical implications) that although the lure drug was not often prescribed, there was no evidence that anyone was aware of the negative side effect on the mental model test. Thus while decision support enhanced performance, it did not increase awareness of a negative side effect.

Regardless, the data provide evidence that even as we increase complexity (DMT versus process control), decision support (providing a list) can still lead to improved performance. This is contrary to arguments presented earlier by Dijksterhuis and Nordgren (2006), and instead, adds to the growing body of data showing improvement from decision support in DDM.

Table 4. Exp. 1 mental model test results. Number listed is on a scale from -2 to +2. -2 being a strong negative effect (critical zone) and +2 being a strong positive effect (excellent zone). Standard error is in parentheses.

		Mental model test		
		DE (positive side effect)	**AD (No side)**	**Lure (negative side)**
Main Effect	**No support**	0.48 (0.21)	0.90 (0.23)	0.03 (0.21)
	List	0.40 (0.22)	1.60 (0.16)	0.10 (0.22)
	List + Strategy	0.34 (0.26)	1.24 (0.18)	0.55 (0.21)
Side Effect	**No support**	0.03 (0.12)	0.03 (0.08)	-0.14 (0.13)
	List	0.10 (0.11)	0.13 (0.08)	0.00 (0.10)
	List + Strategy	-0.17 (0.09)	0.00 (0.05)	0.03 (0.10)

EXPERIMENT 2

The previous experiment demonstrated that decision support can lead to improved performance under increasingly complex conditions. It is not clear from the previous results why the list + strategy group performed as they did. Perhaps the strategy interrupted the natural, more implicit, processes normally used by our participants. Experiment 2 was designed to address this issue. It replicates the procedure of Experiment 1 with one significant change. The list + strategy condition was changed to a list + code condition. In this new condition, participants received the same list as the list condition from Experiment 1. Additionally, they received instructions on how to code the main effect of the drug by zone (as the list + strategy condition did in Experiment 1) and they were given additional information on how to code side effects.

The goal of this manipulation is two-fold. First, the coding procedure succeeds in changing the representation of the task in the desired manner. That is, it makes some of the features of the task more salient than in the other conditions (e.g. Participants keep track of specific knowledge about the main effect and also specifically track side effects). This should alter their interpretation of feedback by calling their attention to the relevant details. If true, then the hypothesis from Experiment 1 should hold: performance should increase with each level of support. Second, the coding procedure leaves the task open to participants' use of their own 'native' strategies. If the strategy issue from Experiment 1 was interfering with natural strategies, this group will perform well on the task, because they could better remember specific side effects on the mental model test.

Methods

Participants and Design. All 113 participants were undergraduates enrolled in psychology courses at Louisiana State University. They participated in the experiment to partially meet

course requirements. As with Experiment 1, there were three training conditions: No support (control), List, and List + code. Participants were randomly assigned to one of the three conditions. There were 38 in the no support group, 37 in the list group, and 38 in the list + code group. A discussion of the sample size change can be found in **Appendix B**.

Task and Procedure. The task was identical to Experiment 1 with the exception of the new list + code condition. This condition was provided a list. On the list, they were instructed to record the zone that the main effect reached as a consequence of each drug combination. Additionally, they were told that if they believed the drug had a side effect, they should record this as well. Side effects were indicated by a '+' if the drug had a positive side effect, and a '-' if the drug had a negative side effect. Finally, they were instructed that the side effect might not match the drug's main effect. In other words, a drug with a large negative main effect could have a positive side effect or negative side effect, and likewise for the other 4 magnitude main effects as well.

Results and Discussion

Prescription Test. Analyses were conducted in the same manner as Experiment 1. The mean percentages for the prescription test are listed in **Table 5**. A one-way ANOVA revealed a significant effect of group on prescribing the correct drugs, $F(2, 112) = 6.34$, $p < .01$. Although the difference was in the expected direction, less prescribing of the lure drug in the list code condition, the difference was not significant, $F(2, 112) = 2.43$, $p = .09$. The post-hoc tests using Fisher's LSD revealed the no support group prescribed one of the correct drugs significantly less than the list + code group only ($p=.001$). A separate one-way ANOVA was conducted on each of the correct drugs individually (see **Table 6**). There were no significant differences found with

respect to the DE (positive side effect) combination, $F(2, 112) = 0.93$, $p > .35$, or the AD (No side effect) combination, $F(2, 112) = 2.61$, $p = .08$.

Since the coding procedure was only given to the list + code condition, it was predicted that the coding condition would be more aware of the side effect than the other two conditions. A planned contrast was conducted comparing the list + code condition to the other two conditions and revealed the list + code condition to have prescribed the lure condition significantly less $t(110) = -2.15$, $p < .05$.

Table 5. Exp. 2 prescription test results. Number listed is the percentage of time the correct drugs or lure drugs were prescribed. Standard error is in parentheses.

	Prescription test	
	Correct Drugs	**Lure Drug**
No support	60.00 (6.34)	11.23 (3.37)
List	70.63 (5.50)	9.19 (3.75)
List + Code	86.67 (3.96)	2.28 (1.49)

Table 6. Exp. 2 prescription test results. This is a breakdown of the correct drugs column from the previous figure. Number listed is the percentage of time the correct drugs or lure drugs were prescribed. Standard error is in parentheses.

	Prescription test	
	DE (positive side effect)	**AD (No side effect)**
No support	21.75 (5.48)	38.25 (6.29)
List	17.84 (5.24)	52.79 (6.31)
List + Code	28.77 (6.41)	57.89 (6.40)

This pattern of results demonstrates superiority of the list + code over the no support group in terms of correct prescriptions. Though not significant, the list + code condition did outperform the list condition as well. This performance boost was primarily derived from both an increase in the prescription rates of the superior DE (positive side effect) combination, and the clear reduction in the amount of time they prescribed the lure.

Mental Model Test. The mental model test data are reported in **Table 7**. A one-way ANOVA revealed a significant effect of group on perceived main effect the AD (No side effect) drug combination, $F(2, 112) = 4.59$, $p = .012$. A one-way ANOVA on the DE (positive side effect) combination was nearly significant between groups, $F(2, 112) = 3.02$, $p = .053$. However, there were again no significant differences found with respect to the lure drug, $F(2, 112) = 1.56$, $p > .21$. Post-hoc tests using Fisher's LSD on the AD (No side effect) combination again revealed the list + code group reported significantly higher main effects for the AD (No side effect) combination than the no support group ($p = .003$). Separate analyses on the perceived side effect of each drug combination did not reveal any significant differences for the AD (No side effect) combination, $F(2, 112) = 0.12$, $p > .88$, or the lure combination, $F(2, 112) = 0.65$, $p > .50$. There were however significant differences found between groups in the one-way ANOVA on the DE (positive side effect) combination, $F(2, 112) = 4.55$, $p =. 013$. Additional post hoc tests revealed the list + code group reported a higher magnitude side effect for the DE (positive side effect) combination than the list group.

Again, since the coding procedure was only given to the list + code condition, another planned contrast was conducted comparing the list + code condition to the other two conditions and revealed the list + code condition to have reported a significantly higher main effect of the DE (positive side effect) drug combination, $t(110) = 2.46$, $p < .05$.

The results of the mental model test demonstrate a clear advantage of the list + code condition over the no support condition in terms of awareness of the positive side effect of DE (positive side effect), however teaching them to explicitly encode side effects did not improve their performance in terms of noting negative side effects of the lure on the mental model test. The noticing of this side effect probably led to the increased prescribing of the DE (positive side effect) combination on the prescription test, as well as the higher main effect reported for the DE (positive side effect) combination by the list + code condition over both the no support and list conditions.

These trends in the results support the earlier prediction that performance would improve with each additional level of support. The additional coding procedure aided in recognition of the positive side effect, but curiously failed to improve detection of the negative side effect. These results also support the hypothesis that the problems with the list + strategy condition in Experiment 1 was likely due to the adoption of the recommended strategy which may have interfered with natural implicit learning processes. Further research will be needed to clarify why this interference occurred.

Table 7. Exp. 2 mental model test results. Number listed is on a scale from -2 to +2. -2 being a strong negative effect (critical zone) and +2 being a strong positive effect (excellent zone). Standard error is in parentheses.

		Mental model test		
		DE (Positive side effect)	**AD (No side effect)**	**Lure (Negative Side)**
Main Effect	**No support**	0.16 (0.21)	1.03 (0.19)	0.26 (0.21)
	List	0.16 (0.20)	1.43 (0.16)	0.30 (0.19)
	List + Code	0.76 (0.19)	1.68 (0.11)	-0.11 (0.12)
Side Effect	**No support**	0.08 (0.09)	0.08 (0.10)	-0.11 (0.07)
	List	-0.14 (0.10)	0.08 (0.11)	0.00 (0.12)
	List + Code	0.29 (0.11)	0.03 (0.03)	-0.16 (0.10)

EXPERIMENT 3

Experiment 3 was meant to contribute more towards an understanding of the relationship between support and processing in complex tasks. It also extends the findings of the previous experiments by adding an element of real-world complexity, individual differences. In the previous experiments, all patients responded to the prescribed medications in the same way. Thus, participants could test many hypotheses at the same time. Moreover, once they found a combination that worked, that combination worked for all of their patients. By adding a second template, the current experiment will allow for more variation in how the patients respond to the prescribed medications. Participants will then have to explore their options on a patient by patient basis. This more closely resembles how people respond to treatments in the real-world.

It should be noted that adding individual differences in the form of patient types to the experiment introduces an additional level of variability into the system that is very opaque. Recall that Mathews (unpublished) found that when noise was introduced into the system, some types of decision support were disrupted. Although individual differences are not the same as the random noise from that study, they do add an additional level of variability that may appear to the participant, at first, like random noise. We hypothesize that decision support will work best when it is designed to make opaque variables more salient.

We test this notion by introducing a new type of decision support designed to make patient differences more salient. In addition to the two support conditions from Experiment 2, two new support conditions (matrix and matrix + code) were added. Unlike the list conditions which only have a list of the possible drug combinations, the table conditions are given a matrix with a list of patients across one axis and the list of drug combinations across the other axis. The goal of this manipulation is to determine whether having an external device is sufficient to

improve performance, or if the quality of the external device is important. The new aids mirror previous support in the type of information that can be recorded, but differ in the level of specificity included (i.e. patient level knowledge). Importantly, the two new conditions still preserve the participant's ability to employ their own native strategies.

One way to frame this manipulation is to view it in terms of the DDM feature opaqueness. By adding individual differences, a level of complexity is added that is opaque to the user. For example, the no support, list, and list + code conditions have no clear way of keeping track of patient differences. It is believed that without matrix support, participants will 'average' drug effects across patients and, thus, fail to learn about patient types. Motivation for this hypothesis comes from previous research in which participants unintentionally integrated knowledge of effects across different output variables (Tall, Mathews, Lane, & Sallas, 2007). Similarly, it is posited that participants might average drug effects across patients without recognizing patient types. In contrast, the matrix conditions will make individual differences more visible (less opaque).

To clarify our hypotheses, it is predicted that: 1) the matrix conditions should exhibit more accurate performance on the prescription test than the other conditions. That is, these two conditions should have better patient level knowledge. 2) The code conditions should demonstrate clearer knowledge of side effects on the mental model test. 3) The matrix + code condition should have the best overall performance on the task. 4) The no support condition should be good at estimating average effects of drugs, but be the worst at detecting and correctly prescribing for patient types.

Methods

Participants and Design. All 328 participants were undergraduates enrolled in psychology courses at Louisiana State University. They participated in the experiment to partially meet course requirements. There were 6 training conditions: No support (control), List, List + code, Matrix, and Matrix + Code. Participants were randomly assigned to one of the six conditions. There were 56 in the no support condition, 57 in the list condition, 55 in the list + code condition, 55 in the matrix condition, and 50 in the matrix + code condition

Task and Procedure. The task and procedure are identical to those from Experiment 2 with three exceptions. First, as mentioned before, there are now two different patient templates. 10 out of 15 patients will operate according to one template, while the remaining 5 patients operate in accordance with the second template. The two patient templates differ on 2 critical drug combinations, AD (Majority drug) and CE (Minority drug) (See **Tables 8 & 9**). Also, it should be noted that each template has only one correct prescription. This was meant to alleviate some of the potential confusions from the previous two experiments in which there were technically two correct answers. Second, two additional explicit support conditions were added. These two conditions mirror the list and list + code conditions in terms of instructions, but differ in the type of aid provided to them. Performance in this experiment will primarily come from analysis of the prescription test. Participants will be scored on the correct ratio of prescriptions (10 and 5). Participants will also receive a second score that determines how many prescriptions were made to the right people. These scores will be used to separate their knowledge of the patient types (ratio) from their knowledge about the individual patients (total correct). An additional measure of performance will gained from the knowledge of overall drug effects and side effects on the mental model test.

Table 8. Main effect template for 10 out of 15 patients in Exp. 3. Side effect is listed in parentheses. Scale is + 2 to – 2. +2 again being a strong positive effect (excellent zone) and -2 being a strong negative effect (critical zone).

	Aprizine	**Bideran**	**Cympril**	**Dimryl**	**Ezicon**
Aprizine	0 (0)	0 (0)	0 (0)	+2 (0)	0 (-2)
Bideran	xxxxxxxxxxx	0 (0)	+2 (-2)	0 (0)	0 (0)
Cympril	xxxxxxxxxxx	xxxxxxxxxxx	-2 (+2)	0 (0)	0 (+2)
Dimryl	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	0 (0)	0 (0)
Ezicon	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	-2 (0)

Table 9. Main effect template for last 5 patients in Exp. 3. Side effect is listed in parentheses. Scale is + 2 to – 2. +2 again being a strong positive effect (excellent zone) and -2 being a strong negative effect (critical zone).

	Aprizine	**Bideran**	**Cympril**	**Dimryl**	**Ezicon**
Aprizine	0 (0)	0 (0)	0 (0)	0 (+2)	0 (-2)
Bideran	xxxxxxxxxxx	0 (0)	+2 (-2)	0 (0)	0 (0)
Cympril	xxxxxxxxxxx	xxxxxxxxxxx	-2 (+2)	0 (0)	+2 (0)
Dimryl	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	0 (0)	0 (0)
Ezicon	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx	-2 (0)

Results and Discussion

The results focus on three drugs combinations. The first working combination (major) succeeded in getting 'Blood Sugar' to the desired level for the majority template patients only. The second working combination (minor) was similarly effective for only the minority template patients. Finally, the lure combination from Experiments 1 & 2 was again analyzed, since it had

equal effect on both majority and minority template patients. The results will again be separated by test.

Also, because the number of rounds completed during the 40 minute practice phase was so dramatically different between groups in this experiment, ANCOVAs were used on all analyses. Number of rounds was used as a covariate, and the adjusted means are displayed in each of the figures. All post hoc analyses were conducted using Fisher's LSD.

Prescription Test. The first analyses concerned the participants overall performance (see **Table 10**). The ANCOVA revealed a significant effect of group. $F(6, 328) = 3.25$, $p < .005$. Post hoc analyses showed the no support condition prescribed the correct drugs to the right people significantly less often than the List condition ($p < .01$) and the Matrix + code condition ($p < .05$).

Table 10. Exp. 3 prescription test results. Number listed is the percentage of time the correct drug was prescribed to the correct person. Overall is across both templates, while major and minor represent the breakdown of those individual templates. Standard error is in parentheses.

	Prescription test		
	Correct (Overall)	**Correct (Major)**	**Correct (Minor)**
No support	35.3 (4.1)	42.0 (5.1)	21.8 (4.9)
List	50.7 (3.9)	56.5 (4.8)	39.1 (4.7)
List + Code	45.6 (4.0)	50.3 (4.9)	36.2 (4.8)
Matrix	45.8 (4.0)	51.2 (4.9)	35.1 (4.8)
Matrix + Code	47.5 (4.2)	53.5 (5.2)	35.6 (5.0)

Separate ANCOVAs were conducted for each patient template. While the groups did not significantly differ in the percentage of time they correctly prescribed the major drug to the majority patients, $F(6, 328) = 1.77$, $p > .10$; they did differ significantly in the percentage of time they were able to correctly match the minor drug to the minority patient, $F(6, 328) = 3.44$, $p < .005$. Curiously, post hoc tests showed the no support condition to have prescribed the minor drug correctly significantly less often than the list ($p < .01$) and the list + code ($p < .05$) only, but was trending towards being significantly less often than matrix ($p = .058$) and matrix + code ($p = .056$).

This pattern of results is quite complex. First, each of the support conditions prescribed the minor drug correctly more often than the no support condition. This much was expected. However, the list condition was the most accurate of the support conditions at prescribing the minor drug correctly. This is contrary to what was expected, since the list condition does not allow for any means of keeping track of individual patients. Second, the list condition was slightly superior to the matrix + code condition in terms of the total number of correct prescriptions. Again, this is unexpected because the most general form of support (list) is performing comparably to the most specific (matrix + code).

Further analysis of the data was conducted on the sheer number of times particular combinations were prescribed in general (see **Table 11**). There were no significant differences found between the groups relative to the amount of time they prescribed either the major drug, $F(6, 328) = 1.53$, $p > .15$, or the lure drug, $F(6, 328) = 1.70$, $p > .10$. However, significant differences in prescription rates were observed between groups for the minor drug, $F(6, 328) = 2.91$, $p < .01$. Interestingly, post hoc tests revealed the list condition to prescribe the minor drug more often than the no support ($p = .02$) and matrix + code conditions ($p = .009$). Though not

significant, the pattern of results is trending in a similar fashion for the list + code condition. The results suggest the list and the list + code conditions might be getting more minor prescriptions to the minority patients due to their increased rate of prescribing the drug. Thus, the list condition's correct prescriptions might be a function of broadly prescribing the two working drugs, while the matrix + code condition might be better at selectively identifying the proper drug for the proper patient.

Table 11. Exp. 3 prescription test results. Number listed is the percentage of time the drug was prescribed in general. Standard error is in parentheses.

	Prescription test		
	Major	**Minor**	**Lure**
No support	34.6 (3.9)	14.2 (2.5)	12.2 (3.1)
List	45.1 (3.7)	22.0 (2.3)	10.0 (3.0)
List + Code	43.8 (3.8)	19.2 (2.4)	17.5 (3.0)
Matrix	38.6 (3.8)	15.7 (2.4)	14.6 (3.0)
Matrix + Code	43.5 (4.0)	12.9 (2.5)	12.8 (3.2)

A follow-up analysis was conducted to determine precision of prescriptions (see **Table 12**). Here, precision is defined as the amount of time you correctly prescribed a drug to the proper patient type *as a function of the number of times you prescribed a working drug*. Note how this is different from the early overall correct assessment which evaluated the amount of time you prescribed a drug to the proper patient type *as a function of the number of patients*. Thus, while the later measure was always out of 15 (the total number of patients), the former

measure varied based upon how often you prescribed the drugs. For example, a participant could have prescribed a working drug only 2 times, but prescribed them to the proper two patients. This would result in 100% precision (2/2), but only 13% accuracy (2/15). Because the precision measure focuses on the number of times prescribed, it becomes a more stringent measure of patient level knowledge. The ANCOVA uncovered significant group differences in prescription precision, $F(6, 328) = 2.49$, $p < .02$. Post hoc tests confirmed that the matrix + code condition was more precise than the no support ($p = .013$) and list + code conditions ($p = .046$). Similarly, the matrix condition was shown to be more precise than the no support ($p = .014$) and, nearly, the list + code conditions ($p = .051$). Though the matrix + code and matrix condition were more precise than the list condition, it was not significant ($p = .085$ and $p = .093$, respectively). Still, the trending pattern of results is clear.

Table 12. Exp. 3 prescription test. Number listed is the percentage of time a working drug was correctly prescribed based on the number of times that it was used. Standard error is in parentheses.

	Prescription test (Precision)
No support	67.4 (3.4)
List	71.6 (3.2)
List + Code	70.4 (3.3)
Matrix	79.5 (3.3)
Matrix + Code	80.0 (3.5)

This pattern of results suggests that despite prescribing working drugs, like the minor combination, less often than the two list conditions, the two matrix conditions were highly

precise at getting those drugs to the right people *when* prescribed. This provides support for the idea that the two list conditions knew which drugs worked, but not necessarily on whom, and hence prescribed the combinations liberally. The two matrix conditions suffered a different fault. They were able to make clear differentiations between patients, but it appears as if they were unable to generalize any findings to other patients. In other words, their low prescription rate of the minor drug (yet high precision) could be interpreted as demonstrating that while knowing this drug works for *some* people, they were not likely to apply it to a patient they were unsure of. Combine this with the fact that they also did not prescribe the major drug any more often than the other groups, and the full picture emerges. Either they had reasonably precise knowledge of what worked for that patient, or they prescribed some other (non-working) combination.

Mental Model Test. The perceived effects of the 3 drug combinations of interest are displayed in **Table 13**. ANCOVAs were conducted first on the main effects of each drug of interest, then on their side effect. An ANCOVA on the perceived main effect of the major drug uncovered significant differences between the groups, $F(6, 328) = 2.36$, $p < .05$. Post hoc tests highlighted the matrix + code condition as reported significantly higher main effect for the major drug than the no support ($p = .002$). It was also noted that the list condition reported significantly higher main effects for the major drug than the no support group ($p = .009$). Another ANCOVA was performed on the main effect of the minor drug combination, and it also revealed significant differences between groups, $F(6, 328) = 2.56$, $p < .05$. Post hoc tests showed the list condition as rating the main effect of the minor drug higher than the no support ($p = .043$) and matrix + code conditions ($p = .009$). Likewise, the list + code condition also rated the main effect higher than the no support ($p = .012$) and matrix + code conditions ($p = .002$). The ANCOVA on the main effect of the lure combination was not significant, $F(3, 328) = 1.04$, $p > . 35$.

When evaluating side effects, neither the ANCOVA on the major drug, nor the ANCOVA on the lure drug was significant. The ANCOVA on the minor drug did however reveal significant differences, $F(6, 328) = 2.15$, $p < .05$. Post hoc tests showed the no support condition to rate the side effect of the minor drug as significantly more positive than list + code ($p = .002$), matrix ($p = .011$), and matrix + code conditions ($p = .019$).

Table 13. Exp. 3 mental model test. Number listed is on a scale from -2 to +2. -2 being a strong negative effect (critical zone) and +2 being a strong positive effect (excellent zone). Standard error is in parentheses.

		Mental model test		
		Major	**Minor**	**Lure**
Main Effect	**No support**	0.97 (.143)	0.53 (.168)	0.32 (.166)
	List	1.47 (.136)	1.02 (.160)	0.18 (.158)
	List + Code	1.33 (.139)	1.13 (.164)	0.69 (.161)
	Matrix	1.33 (.139)	0.79 (.164)	0.55 (.161)
	Matrix + Code	1.63 (.146)	0.39 (.173)	0.48 (.170)
Side Effect	**No support**	-0.01 (.057)	0.18 (.066)	-0.19 (.074)
	List	0.13 (.055)	0.05 (.063)	-0.12 (.071)
	List + Code	0.49 (.056)	-0.11 (.064)	-0.02 (.072)
	Matrix	-0.06 (.056)	-0.06 (.064)	0.00 (.072)
	Matrix + Code	0.01 (.059)	-0.04 (.068)	-0.10 (.076)

These results unfortunately are difficult to interpret. The instructions provided prior to the test informed participants to report the effect each drug had on the majority of patients. The intention was for them to list the drug effects according to the majority template. It is possible however that some people misinterpreted these instructions, and averaged across all 15 patients.

Perhaps still, some may have disregarded the instructions and listed the drug by its strongest effect (e.g. reporting the major drug by the majority patients and the minor drug by the minority patients). The interpretation of the side effect of the minor drug is even more puzzling. While the ratings were slightly more positive in the no support condition, it is not clear why this distinction is meaningful.

In summary, the first hypothesis was not clearly supported nor denied. Recall that it was predicted that the matrix conditions would display better patient level knowledge. In terms of precision, this is true, while it is not true that they were more accurate. The matrix conditions showed more precise knowledge for the limited number of patients that they prescribed a working combination to, while the list conditions displayed a more general (drug focused) knowledge. The more liberal prescription rates in the list condition yielded higher accuracy than expected, but it is questionable whether or not this is derived from patient level knowledge. The second hypothesis regarded the code conditions as demonstrating clearer knowledge of side effects. While interpretation of the mental model test data is somewhat murky for the major and minor drug, the data from the lure combination was succinct. The coding did not produce the expected knowledge benefit. The third prediction was that the matrix + code condition would have the best overall performance on the task. Though the matrix + code condition did not differ significantly from the list condition, it still yielded marginally less overall accuracy than the list condition on the prescription test. Combined with the lack of clear mental model test data, it is not clear that this prediction was supported. Finally, the last prediction involved the no support condition demonstrating a good knowledge of the 'average' drug effects, while having poor patient knowledge on the prescription test. This prediction was demonstrated rather well by the sheer inability of this condition to match the accuracy and precision of the other 4 conditions.

GENERAL DISCUSSION

The current study examined several varieties of decision support. Clearly, an external memory aid was useful in each of the experiments. Providing participants with a list of all 15 possible drug combinations was very helpful across experiments. In addition to reducing cognitive load, the list appears to facilitate abstraction across patients. This is likely due to a reduction in opaqueness. Rather than looking for patient differences, focus is shifted to what a drug *does*. If a drug has a particular effect, it is likely to have that effect on most/all people. In the case of the first two experiments, this could remove patient variability as a factor altogether.

The matrix assisted performance differently. Rather than encouraging abstraction, it allowed for very precise encoding of information. Participants in these conditions were able to record and process very specific patient level information. Although this also removes opaqueness relative to patient differences, it does so by increasing the salience of the patient differences. This extra awareness of differences focuses attention at the individual level, and thus facilitates a specific (patient) level of encoding.

The inclusion of a coding procedure was designed to increase awareness of the magnitude of drug effects. This form of support managed to provide some help in each of the experiments. Participants receiving this support were less likely to prescribe the lure drug combination. However, this benefit was primarily implicit. That is, participants prescribed the lure combination less often (thereby recognizing it as defective), but did not reveal any explicit knowledge of the negative side effect. Instead, they simply reported the combination as having a lower main effect. This means that they were often prescribing the right treatment to their patients for the wrong reason. Further research is needed to determine if it is possible to facilitate an explicit awareness of these drug differences.

The final form of support used was an elimination strategy. While this manipulation improved performance over the no support group, participants in this condition did not reach the performance level of those who received the list only. It is unclear why the strategy interfered with the task. More research is needed to address this question.

Experiment 3 tried to bring the task closer to what occurs in the real-world. Since patients do not always respond the same way to the same treatment (as they did in our previous two experiments), Exp. 3 added different patient types. This naturally contributed an additional level of variability and complexity to the task. However, it allowed us to explore the relationship between decision support and processing in complex tasks with more fine-grained differences. First, we were able to include additional levels of decision support (i.e. matrix conditions). These conditions received a different type of aid that was aimed at reducing the opaqueness of patient types. Second, the prescription test allowed for more precise measures of knowledge. We were able to look at different levels of accuracy and precision with respect to prescription patterns.

The results showed that the decision support groups were all roughly equivalent in their overall accuracy of drug prescriptions. That is, they tended to prescribe the right drug to the right person about 50% of the time. This was a clear improvement over the unsupported condition. Furthermore, when broken down by patient type, it was found that this increase in accuracy came primarily from the support conditions' ability to recognize the minority template patients. It is curious that each of the support conditions seemed to show this increased awareness of the minority template considering that half of these conditions had no means of recording patient level information. In other words, providing one with the means to keep track of individuals does not seem to improve the accuracy rate when prescribing to different patient types.

Further analysis of the data showed that in addition to having comparable accuracy rates, the two list conditions seemed to prescribe the minority template drug more often than the two matrix conditions. This is an important finding because it shows that part of the accuracy of we find in the list conditions are due to more liberal prescription habits with the minority drug. It appears as though these conditions are learning information about drug trends and not patient trends. That is, participants in these conditions are gaining an awareness that the majority drug works for 'many people', and the minority drug works for 'some people.' Such knowledge would lead to increased amounts of prescribing the drug, but would only lead to a moderate level of accuracy. This latter point was backed up by an analysis of the precision with which each group made prescriptions. In other words, it was found that the matrix conditions may have prescribed drugs less often than the list conditions, but were highly accurate during the few times they did prescribe.

This is an important finding because it shows that despite having comparable accuracy rates, the two types of decisions support yielded quite different types of knowledge. In terms of task representation, the list provided the expected general (drug level) knowledge. Participants were aware of the drug effects, broadly. They knew these drugs worked for some people and were therefore likely to prescribe it to a patient that they were unsure of. The matrix, however, encouraged a very specific, but narrow, view of the task. By focusing on the patients, it appears as though participants in this group were limited in their ability to generalize across patient types. They were as likely to prescribe a non-working combination to a patient as often as a working combination, if they did not know what worked for the patient.

These important differences in representation did not appear to affect overall accuracy in this experiment. However, it is possible that over time differences in accuracy can emerge. While

the drug level representation more quickly reveals the pattern of responses, it is difficult to see how this form of representation will ever yield more than the hit-and-miss prescription rates observed here. Thus, while they may be more accurate at knowing the proportion of people to which each drug should be prescribed, it is not definite that proportional prescriptions will be accurate prescriptions. In contrast, it is possible that with more experience, patient level focus can lead to superior performance. It can be argued that the specific knowledge gained from this representation takes longer to develop. With more time, it is possible that the highly precise prescription rates will carry over to more and more patients. That is, one might pick up on either the general emerging pattern or simply learn about one additional patient at a time. Thus, more research is needed to address how fast knowledge is gained from the different representations, as well as, any limits on how much can be known.

REFERENCES

Berry, D. C., & Broadbent, D. E. (1998). On the relationship between task performance and verbalizable knowledge. *Quarterly Journal of Experimental Psychology, 36*, 209-231.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues.* Hove: Lawrence Erlbaum Associates Ltd.

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

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In: Frensch P. A., & Funke J. (Eds.), *Complex problem solving: the European perspective* Hillsdale, NJ: Lawrence Erlbaum Associates.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 198–215.

Dienes, Z., & Berry, D. C. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulleting and Review, 4*, 3-23.

Dienes, Z. & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 848-862.

Dienes, Z. & Fahey, R. (1995). The role of implicit memory in controlling a dynamic system. *Quarterly Journal of Experimental Psychology, 51A*, 593-614.

Dijksterhuis, A. & Nordgren, L. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1, 95-109.

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

Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 259-278.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Funke, J. (1995). Experimental research on complex problem solving. In: Frensch P. A., & Funke J. (Eds.), *Complex problem solving: the European perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gladwell, M. (2005). *Blink: The power of thinking without thinking.* London: Penguin.

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes, 96,* 142-154.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior,* 21, 273-286.

Lane, S. M., Mathews, R. C., Sallas, B., Prattini, R., & Sun, R. (2008). Facilitative interactions of model- and experience-based processes: Implications for type and flexibility of representation. *Memory and Cognition, 36(1),* 157-169.

Mathews, R. C. (unpublished). Effects of random noise on learning in a process control task.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(6),* 1083-1100.

Newell, B. R., Wong, K. Y., Cheung, J. C. H., & Rakow, T. (in press). Think, blink, or sleep on it? The impact of modes of thought on complex decision making. *The Quarterly Journal of Experimental Psychology*, 1-26.

Raab, M. (2003). Decision making in sports: Implicit and explicit learning is affected by complexity of situation. *International Journal of Sports and Exercise Psychology, 1,* 310-337.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219-235.

Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* New York: Oxford University Press.

Rouse, W. B. (1981). Human–computer interaction in the control of dynamic systems. *Computing Surveys*, 13, 71–99.

Roussel, L. G. (1999). *Facilitating knowledge integration and flexibility: The effects of Reflection and exposure to alternative models.* Unpublished doctoral dissertation. Louisiana State University.

Sallas**,** B., Mathews, R. C., Lane, S., Sun, R. (2007). Developing rich and quickly accessed knowledge of an artificial grammar. *Memory & Cognition, Vol. 35(8),* 2118-2133

Smith, E. R., & Collins, E. C. (in press). Dual process models in social psychology. In: Evans, J., & Frankish, K. (Eds.), *In two minds: Dual processes and beyond.* Oxford University Press.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989) Insight without awareness: On the interaction of verbalization, instruction, and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology, 41,* 553-577.

Stanovich, K. E. (in press) Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In: Evans, J., & Frankish, K. (Eds.), *In two minds: Dual processes and beyond.* Oxford University Press.

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

Sun, R., Zhang, X., Slusarz, P., & Mathews, R. (2007). The interaction of implicit learning, explicit hypothesis testing learning, and implicit-to-explicit knowledge extraction. *Neural Networks, 20,* 34-47.

Staudel, T. (1987). *Problem solving, emotions, and competence*. Regensburg, Germany: Roderer.

Tall, J., Mathews, R. C., Lane, S., & Sallas, B. (2007, Nov). *Learning the effects between multiple inputs and outputs: Integration effects in a diagnostic task.* Poster session presented at the 48th annual meeting of the Psychonomic Society, Long Beach, Ca.

Turkle, S. (1984). *The second self: Computers and the human spirit*. London: Granada.

APPENDIX A

ON-SCREEN TEST INSTRUCTIONS

Prescription Test Instructions

On this next screen, you will be asked to indicate the best drug for each patient. You will make your selection from a list of all possible combinations. You may only select one of these possible combinations. You will also rate your confidence that this is the best drug for that patient. When you have made your selection, click 'submit', and you will be taken to the next patient.

Mental Model Test Instructions

On this screen, you will need to summarize your knowledge of all of the drugs. You will be asked about the effects of each drug (or combination of drugs). Please respond with main effect that occurs –most commonly- among the patients.

APPENDIX B

SAMPLE SIZE CHANGES

Experiment 2

Data collection for this experiment was very slow throughout the semester. In an attempt to counter this, additional sessions were posted each week. A sudden surge in the participant pool occurred near the end of the semester. Since 8 participants could enroll per session, the extra sessions provided a small boost to the overall data pool for this experiment.

Experiment 3

Analysis of the previous experiments using 'G*Power 3' estimated power to be quite low [0.53] (Faul, Erdfelder, Lang, & Buchner, 2007). Therefore, sample size for Experiment 3 was dramatically increased to achieve an adequate estimated power [.95].

VITA

Jon Tall received his Bachelor of Science in psychology from Northwestern State University. He went on to receive a Master of Arts in philosophy of science from Louisiana State University. He is currently studies decision making as a graduate student in cognitive psychology at Louisiana State University.