An Examination of Decision Aid Reliance

in a Dynamic Environment

John W. Briggs

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business

with a Major in

Accounting and Information Systems

Dr. Steven Sheetz, Chairman
Dr. Reza Barkhi
Dr. John Brozovsky
Dr. Larry Killough
Dr. Mary Beth Rosson

December 11, 2003

Blacksburg, Virginia

Keywords: Decision Aid, Reliance, Decision Making, Decision Support Systems

Copyright 2003, John W. Briggs

An Examination of Decision Aid Reliance in a Dynamic Environment

John W. Briggs

(ABSTRACT)

Computerized decision aids are powerful tools to assist with decision-making. Decision models are designed to incorporate and analyze available data in order to present a recommended solution to a problem. Business decision makers, including accountants, have much to gain from integrating decision support technology with their own skills and experience. Several studies have determined that there are many instances in which these decision aids perform favorably to human decision-makers. Despite this fact, studies have shown that reliance upon these aids is incomplete, even when they process data in a highly efficient manner.

On the other hand, decision aids have limitations. If such a decision support system is not updated to match changing conditions, relying on the aid can lead to suboptimal decision-making.

This study uses a laboratory experiment involving a managerial accounting task: prediction of manufacturing overhead costs. In the experimental scenario, a decision support system's recommended solutions become inaccurate due to a shift in environmental conditions. The first research objective is to determine whether subjects rely on the aid's advice before this change and, to their detriment, after the change. The second research objective is to examine whether the feedback environment, the timing of the decision aid's inclusion into the task, or the inherent confidence level of the task participant affect the tendency to rely on the aid in both of these environmental conditions.

The results of the study provide evidence that decision-makers rely on decision aids, and are susceptible to over-reliance on them. These findings add to the results of prior studies that only examine a single trial task. Additionally, it is determined that the timing of a decision aid's recommendation can affect the degree to which it is relied upon. Next, there is evidence that feedback environment can help reliance and mitigate over-reliance. There is no evidence that task confidence affects reliance. Lastly, decision aids result in longer amounts of time used to complete the task.

DEDICATION

I dedicate this work to my family. I thank my beloved wife, Angie, who endured as much as I did during this period. Her love and support are tremendous. My parents, sister, grandparents, and other family members were also constant sources of encouragement.

ACKNOWLEDGEMENTS

Words can barely describe how grateful I am to complete this task. And there are so many individuals to thank.

My dissertation chairperson, Dr. Steve Sheetz, provided much wisdom and knowledge. His hard work was infinitely valuable. He truly put me in a position to succeed.

Dr. John Brozovsky was also a great companion through this process. He was always available to help me through my entire doctoral program. Also, Dr. Larry Killough, Dr. Reza Barkhi, and Dr. Mary Beth Rosson brought expertise to my committee for which I am extremely thankful.

Dr. Robert Brown and Dr. Wayne Leininger served as guides for me throughout my studies. They were there to lead and direct me through the best of times and through other times. Dr. Cintia Easterwood generously allowed access to her students for my experiment. The entire accounting faculty at Virginia Tech provided tremendous knowledge, motivation, and help to me. I thank each one of you. The support staff of the Department of Accounting was also invaluable.

Next, I consider myself fortunate to have worked with all the doctoral students I met at Virginia Tech. I made many friends and learned much from you.

And above all I thank God for the guidance He provided to me during this challenging process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

The usefulness of a computerized decision aid can vary when the aid is used in a changing environment. Assuming that the aid is created from an analysis of past data, the aid's value to decision-makers falls when its logic is not updated to match changing conditions. Ideally, models are continuously updated to appropriately reflect changes in environments. But in reality, this may be too costly, or infeasible. This dissertation describes an experiment that places users in a situation in which they have access to a decision aid, and due to changes in environmental conditions, the aid's ability declines. The purpose of the study is first to examine whether decision-makers appropriately rely on the aid. Then, the study looks to see if reliance continues following the change. Lastly, the study examines factors that influence these reliance behaviors. Prior literature has concentrated on finding factors that motivate individuals to use an aid's recommendations. This study continues this search, and then takes the additional step of examining the causes of over-reliance on these recommendations when they are not valid. Two task-related variables and one behavioral trait are proposed to be related to the level of reliance, and then over-reliance following an environmental change. The task-related factors are the availability of a summarized type of feedback, and the timing of the decision aid's entrance into the task. The individual trait of interest is the level of task confidence the decision-maker possesses.

In business and in other decision settings, decision aids are increasingly becoming utilized. For example, there is a growing availability of computerized decision aids in accounting (Brown and Eining, 1997). Rapid technological advances have made information, often in the form of decision support systems, increasingly available. The purpose of these aids is to extend the cognitive limits of decision-makers, or to provide expertise to them. Decision aids are deemed valuable for their ability to increase decision-making accuracy and consistency. These aids are costly to organizations, as they are either developed internally or purchased. If employees fail to utilize effective decision aids, the organization may potentially suffer. On the other hand, over-reliance on these aids can inhibit learning (and therefore expertise development) and lead to poor decision-making.

Both of these issues (reliance and over-reliance) have been examined in the literature in various ways. Although several studies have examined the extent to which users rely on decision aids in different settings, little research has assessed the behavior of users when the aid's usefulness varies significantly. Glover, Prawitt, and Spilker (1997) study the case where novice decision-makers fail to recognize that a tax decision aid does not fit a given task and inappropriately use the aid. Kowalczyk and Wolfe (1998) investigate the incidence of anchoring on a decision aid's recommendation, finding that decision aid advice is anchored upon (in a single trial), as subjects fail to respond to subsequent refuting evidence. This study directly adds to the evidence provided by these two papers. The current research examines the case in which users have access to a highly accurate aid, and a change in environment takes place, after which the aid is of low accuracy. The aim of this study is to look for evidence of reliance and then over-reliance in an ongoing task, and then to determine the factors that affect the strength of these reliance behaviors.

With the availability of decision support increasing at a high rate, research must carefully consider both the positive and negative effects of different types of information, including decision models, on the decision-making process. The findings of the current study can help to underscore the importance of monitoring the use of decision aids. The ability of decision makers to perceive and react to a change in decision model accuracy is an unexplored issue. This topic is as important as is the promotion of reliance, because real business environments are dynamic in nature.

The remainder of this paper is organized as follows: Chapter 2 provides a review of the literature and Chapter 3 describes development of the study's propositions and hypotheses. Chapter 4 gives an overview of the experimental methodology, and Chapter 5 presents the experimental results. Chapter 6 concludes the paper, with a discussion of the contributions and limitations of the study.

Chapter 2

REVIEW OF LITERATURE

2.1 Overview

Decision support comes in many varieties. Decision aids can be designed to assist users with simple or complex tasks, and with tasks of varying structure (Gorry and Scott-Morton, 1971). These aids can vary in complexity from simple checklists, to statistical models, to sophisticated expert systems. They can offer guidance without giving a solution, yield a single solution, or provide an overview of possible solutions, even with probabilities assigned to outcomes. The scope of the present study is limited to intelligent decision aids: those that perform computations based on inputs.

In accounting research, decision aids are most extensively studied in the auditing realm. Auditors have statistical models available to use in going concern prediction ((Altman and McGough, 1974), (Hopwood, McKeown, and Mutchler, 1994)), fraud detection (Bell and Carcello, 2000), selection of a sample size (Kachelmeier and Messier, 1990), and other areas. ExperTAX was an early proprietary system developed by Coopers and Lybrand used by auditors to determine the reasonableness of the income tax accrual presented in financial statements (Shpilberg et al., 1986). In industry, accountants are responsible for budgeting and controlling costs, and use models for forecasting and planning. O'Leary and Lin (1989) present an expert system used to perform cash flow projections based on ratio analysis.

A long line of research has evaluated the ability of human decision-makers relative to machines, and further has attempted to demonstrate how humans and machines best work in a synergistic fashion. Meehl (1957) was one of the first to ask "When shall we use our heads instead of the formula ?" Among his conclusions is that mechanical *combination* of data is quite often superior to subjective or intuitive combination. Einhorn (1972) and Libby and Libby (1989) provide further evidence that humans exhibit many inaccuracies and biases in the combination of data. Libby and Libby perform an experiment with two groups of senior auditors. The first group gave a global assessment of internal controls for a transaction cycle, while the second group gave evaluations for each process within the transaction cycle. The evaluations of the second group were then combined using a mechanical model. The resulting global evaluation of the second group showed 61% less deviation (higher consensus) than that of the first group. Additionally, Peterson and Pitz (1986) show that decision makers underperform a bootstrapping model of their own decision processes. Through a first series of trials, subjects' responses in a task were recorded, and a model of their own judgment strategy (weights) was instantaneously produced. The subjects then performed an additional set of trials, and had the resulting prediction from their own bootstrapped decision model available as an additional item of information. Despite the presence of this prediction, subjects' accuracy in phase two was significantly lower than the model's accuracy.

The evidence of human limitations in information processing is clear. Nevertheless, computerized models do not always provide the best answer to a problem. Blattberg and Hoch (1990) maintain that intuition, particularly expert intuition, cannot be adequately quantified. They provide a thorough analysis of the strengths of models versus the strengths of humans, a summary of which follows. Models are consistent, take base-rates into account, and are immune to social pressures. Models likewise do not get tired or emotional. And models optimally weigh data. Despite these features of models, humans carry their own advantages. Humans know what questions to ask and can identify new variables. Humans are proficient at providing subjective evaluations of variables that are difficult to measure objectively (Einhorn, 1974). Humans are more flexible than models when conditions change, and can often recognize abnormal cases that a model cannot anticipate.

2.2 Reliance Literature

Ultimately, the choice of whether to utilize a model in decision-making is up to the individual. In following with the logic of Blattberg and Hoch, many researchers have attempted to understand to what extent people follow the advice of a model, and to what extent they trust their own evaluative strategies.

Studies consistently show that humans under-rely on decision models in a variety of settings, and due to this under-reliance are often outperformed by models. Arkes, Dawes, and Christensen (1986) performed two seminal experiments on decision rule reliance. Subjects significantly under-performed decision rules because they did not rely on them. Even when warned that unaided decision-makers "very seldom" do better than the rule, under-reliance persisted. These findings are generally robust for both novice and experienced decision makers. Boatsman, Moeckel, and Pei (1997) found evidence that auditors presented with a decision aid to predict management fraud continually rejected the aid's recommendation, and even changed their own initial recommendation away from the aid's advice when the aid agreed with them.

Why are decision-makers unwilling to completely trust the advice of a model ? Dawes (1979) describes the notion of "cognitive conceit". He states that it is "the illusion that the environment is more predictable than it really is and that greater cognitive effort will lead to better predictions than those afforded linear models known to be imperfect." In practical terms, if a model does not provide perfect accuracy, it seems that people will attempt to improve upon the model's algorithm or incorporate information known not to be in the model.

Given that decision-makers have an innate tendency not to be completely trusting of models, research has attempted to identify factors that cause users' reliance on models to vary. Prior literature indicates that models with higher reported accuracy are relied upon more. Powell (1991) conducted a study in which subjects had knowledge of a decision rule that had a 70% success rate if followed exclusively. Three treatment groups were told that the rule's predictive accuracy was 50%, 70%, or 90%. Although subjects in all three groups were only moderately accurate (via less than complete reliance on a 70% rule), higher reported accuracies increased reliance (and thus accuracy) by a significant amount. In addition, Kaplan, Reneau, and Whitecotton (2001) provide evidence that decision makers are more likely to rely on an aid when its predictive ability is *not* disclosed. Thus, a decision aid's rate of failure may be more influential than its rate of success.

A commonly suggested means of motivating decision aid reliance is to provide increased levels of feedback to the user about the (accurate) aid's success. Presumably, knowledge that the aid contributes something to the decision-maker should induce higher reliance. There has been mixed support for this proposition. Ashton (1990) found that feedback helped unaided decision makers improve their judgments. But when feedback was given in the presence of a decision aid, performance actually declined relative to the no-feedback case. Ashton uses a pressure-arousal-performance theory to explain these findings. In this theory, feedback (along with justification and financial incentives) causes increased pressure on the subject (by knowing that performance is sub-optimal), which leads to increased attempts, including strategy-shifting, to beat the aid. Similar results were found by Arkes et al. (1986).

Davis and Kotteman (1995) changed the nature of the feedback manipulation. In a production planning task, subjects' objective was to minimize costs over a long series of trials. Subjects that were given a cumulative, running total of how well they would have done by exclusively following the decision aid's advice eventually increased their reliance. This finding suggests that feedback of sufficient quality and quantity can mitigate some of the prevalent biases in decision-making. On the other hand, what level of feedback is typical of most business decision settings ? Einhorn and Hogarth (1978, p. 395) note that "... in real-world situations, judgments are made for the purpose of choosing between actions. This means that outcome information, which is available only after actions are taken, is frequently the only source of feedback with which to compare judgment." Davis and Kotteman acknowledge this limitation in their own study, noting that if people do not perceive a decision aid as useful, they may stop using it, and thus fail to "generate the feedback needed to disconfirm this erroneous belief" (Davis and Kotteman, p.146).

Some understanding of the factors that go into a model's creation is commonly thought to make it easier for users to trust decision aids. Brown and Jones (1998) propose that knowing an aid's algorithm should raise confidence in the aid, and that this understanding should better help a decision-maker understand the nature of his disagreement with the aid, leading to improved decision quality. For example, a general knowledge of how an aid's answer is derived can lead to a policy of selective disagreement with the aid. Of course, the more complicated this algorithm, the less meaningful knowledge of the algorithm becomes. Davis (1998) finds that auditors prefer a decision aid that involves the user step-by-step to a logit model that incorporates five financial ratios in a fairly complex arrangement.

Whitecotton and Butler (1998) also consider the case in which decision-makers have a role in a decision aid's development. They find that participation in development significantly increases decision aid reliance. This is true even through the simple manipulation of "being able to choose". Subjects selected three of five financial ratios to be included in their decision aid (in predicting bankruptcy). Because these were relatively low experience subjects, it is unlikely that they had compelling reasons for their selections. Nevertheless, participation in model creation was effective at increasing reliance (even though overall performance was lower than the *optimal* model's performance). Becker (1997) provides a model of this phenomenon, in which the ability to choose increases self-determination, which in turn leads to higher intrinsic motivation. Self-determination, a part of Cognitive Evaluation Theory (Deci and Ryan, 1985), is the extent of control one feels over a task, independent of perceived confidence. The combined implications of these studies on participation are valuable to modern organizations, in that they suggest bridging the gap between developers and users.

Decision aid reliance is also related to individual differences. Overconfidence, for example, is a plausible reason for aid under-utilization. Previous findings (Whitecotton, 1996) have generally shown that confidence (in one's own ability) has a negative impact on overall accuracy in the presence of an accurate decision model. The effect of confidence on reliance is apparently greater than the relationship between confidence and ability. Kaplan, Reneau, and Whitecotton (2001) study an innate personality characteristic, locus of control, and find as hypothesized that individuals with an external locus of control are more willing to place their reliance in an aid than are those with an internal locus.

The relationship between experience and reliance is more complex. As a general rule, more experienced decision-makers are less likely to require assistance from an aid. However, more experienced decision-makers are superior at determining the conditions in which an aid is useful and when it is not. They can more readily identify when an aid's logic is invalid. Thus, they are also superior at knowing when to take advantage of an aid's ability. Arnold and Sutton (1998) provide the most comprehensive model to date of novice versus expert reliance differences, in their Theory of Technology Dominance. The first two propositions of the theory are that task experience is negatively related to decision aid reliance, and that task complexity is positively associated with reliance. Technology dominance occurs when the expertise of the decision aid is significantly greater than that of the decision maker: under such conditions, the risk of poor decision-making is high. In contrast, when the expertise fit between user and tool is high, reliance on the tool should result in improved decision making.

2.3 Over-reliance Literature

While the previously discussed studies have focused on how reliance might be improved, other work has taken the opposite perspective, and examined the factors that cause individuals to rely too heavily on the advice of a decision aid. This research can be grouped into two classifications that are similar but not identical: those that describe over-reliance as an instance of the anchoring and adjustment heuristic, and those that describe over-reliance as an effort minimization strategy.

2.3.1 Anchoring and Adjustment

Lichtenstein and Slovic (1971) and Tversky and Kahneman (1974) were the first to describe the anchoring and adjustment heuristic, in which an individual begins a decision process by selecting a certain piece of information as an initial solution, and then adjusts away from it to reach a final decision. This paradigm has been shown to explain judgment behavior in a wide variety of settings (e.g., Northcraft and Neale (1987), Bromiley (1987)). The anchor a person selects may range in utility from highly valuable to completely irrelevant. There are often multiple anchors to choose from, and less relevant anchors can exert sizeable effects on judgment, even when more relevant anchors are present (Whyte and Sebenius, 1997). In many situations, a bias has been found to accompany this strategy, as the decision maker fails to fully adjust away from the anchor. As an example of this phenomenon, Joyce and Biddle (1981) perform an experiment in which auditors are asked to predict the incidence (out of 1000 firms audited by Big 8 accounting firms) of significant executive-level management fraud. One group of auditors was first asked whether the number is greater or less than 10, but the other group was asked whether the number is greater than or less than 200. Then, each subject made a prediction. Subjects in the former group had a mean estimate of 16.52, and those in the latter group estimated the number to be 43.11. In this case, an arbitrary starting point served as a significant anchor in the decision process. Butler (1986) finds that experienced auditors approach assessments of risk by starting with an internally generated anchor of five to ten percent. In an experiment, auditors appeared to conservatively adjust from this starting point.

Can decision aids serve as anchors ? Kowalczyk and Wolfe (1998) create an experiment in which subjects were first given a decision aid's recommendation about the likelihood of a firm's continuation as a going concern. They employ a "curse of knowledge" methodology to test for the presence of anchoring. After seeing the aid's recommendation, subjects were asked how they thought *others* would evaluate the firm assuming they had *not* seen the recommendation. They found that subjects were unable to ignore what they had seen from the decision aid in responding to this question. As a separate test, subjects were given subsequent information that contradicted the advice of the aid. After this refuting evidence was given, subjects had to make an assessment about whether substantial doubt existed (regarding the firm's continuation as a going concern). They found that subjects were unable to dismiss the aid's recommendation even when the subsequent information was conclusive.

2.3.2 Effort Minimization

Decision aid recommendations, then, have the capacity to serve as anchors that are overvalued in the final decision. They also have the capacity to provide a solution quickly and easily. Whether or not this solution is optimal, it is usually complete. The effort minimization role of decision aids has been extensively studied by Todd and Benbasat (1991, 1992, 1994). They declare that the dominant advantage of decision aids is the role they play in minimizing effort. Their experiments concern a preferential choice task, and the aid the subjects are given does not give a recommended solution; instead, it is a tool with which to manipulate data. They find that subjects given this decision aid do not necessarily seek better solutions than do unaided subjects. Instead, their information search and utilization strategies suggest that effort minimization is of higher utility. In other words, the authors show that decision makers place a higher premium on using DSS for efficient decision making than for effective decision making. This cost-benefit framework of cognition (Payne, 1982) is common to all decision making. But the relative trade-off between costs and benefits in the presence of DSS is still an open area of exploration.

Glover, Prawitt, and Spilker (1997) argue that inexperienced decision-makers, when provided with a decision aid, may tend to use the aid mechanistically, as opposed to combining the aid's advice with their own judgment. Two negative ramifications of this behavior are that the aid is used in situations where it should not be (inappropriate reliance), and that task-related learning is impaired. Glover et al. provide an experimental group with a decision aid to compute capital gains taxes, and use a control group without the aid. By design, the aid only worked for tax clients whose marginal tax rate was 28% or higher: when an occasional instance of a 15% marginal tax rate was introduced, the aid continued to compute capital gains taxes at 28%. Aided subjects failed to notice that the aid was not designed for all cases. They also spent less time reviewing the task than did unaided subjects, and were subsequently less able to explain how the computations are made. They conclude that decision aid designers must pay careful attention to the manner in which the aid involves the user in the task. Otherwise, the aid's user may have a poor understanding of its limitations.

Considering the evidence provided by multiple studies, it appears that the recommendations of Blattberg and Hoch (1990) are supported. Because of the relative merits of human cognition and computerized processing, decision aid reliance should be dependent on the situation. If decision aids are flawless, or if they combine information cues in ways that human decision-makers cannot, then they should be relied upon exclusively. But when decision aids do not fit the task, or when they fail to incorporate all information or the right information, they should not be exclusively relied upon. There is, then, an optimal reliance decision that is situation-specific. Little cohesive theory exists towards describing this optimal level of reliance. Brown and Jones (1998) and Arnold and Sutton (1998) provide the most comprehensive models to date. These general models of decision aid reliance describe reliance as contingent on a number of factors such as task difficulty, familiarity with the aid, and expertise. These models do not, however, look at the dynamic process of reliance, as it grows and diminishes over time. Such a perspective is an open area for research, and this paper attempts to take an initial step in that direction.

Chapter 3

MODEL DEVELOPMENT AND HYPOTHESES

Existing models of decision support reliance focus exclusively on the set of static conditions that cause reliance to vary. But by its very nature, reliance can be considered an attribute of behavior that is developed or acquired. This strongly suggests a need for understanding the processes that influence not only the development of reliance, but the persistence of reliance. This study attempts to address this need, by seeking to understand the factors that affect the strength of reliance, through examination of reliance in a changing environment. Earlier in this paper, it is argued that all types of decision support are subject to a limited lifespan, as environmental changes will inevitably take place. Unless a DSS is constantly updated, it is apt to lose an amount of usefulness. Also, empirical evidence provides extensive support that individuals can be directed towards different reliance levels by certain controllable factors. The objectives of this study are to develop a model of decision aid reliance that treats reliance as a dynamic entity, and to empirically test this model in a laboratory experiment.

3.1 Reliance in a recurring task

An initial objective of this study is to validate that decision-makers use decision aids to a significant degree when they are available. Although many of the studies discussed so far report evidence of under-reliance, reliance is usually present to a substantial degree. The first proposition of the current study is, then:

Proposition 1: *Decision-makers will rely on an accurate decision aid.*

3.2 Over-reliance in a recurring task

Studies have shown that people are susceptible to over-reliance on a decision aid. Kowalcyk and Wolfe (1998) show that an aid's recommendation can serve as an anchor in a one-trial setting, in which subjects made going-concern assessments. Subjects who were first given an aid's recommendation, and then were given evidence that contradicted the aid, revised their estimates incompletely in response to the new information.

The remainder of existing studies focus on user behavior when presented with extreme cases not accounted for by the model. Meehl (1957) refers to these pertinent, non-modeled information items as "broken-leg" cues. As an example of this research, Glover et al. (1997) demonstrate that decision-makers can apply a decision aid to incorrect situations, likely due to its perfect performance in other cases. The focus of their study is primarily on passive, or "mechanistic" use of a decision aid. "Broken-leg" cues are often dealt with poorly, because by definition, they are rare enough not to be included in the decision aid. And experts are clearly superior at knowing what constitutes valuable non-modeled information. A useful extension of the above findings is to examine whether anchoring on an aid's advice takes place in a dynamic environment. No study to date has assumed a shift in a decision aid's functionality that persists for multiple trials. The experiment associated with this study includes a one-time shift in aid accuracy, in order to clearly monitor changes in subjects' behavior. The next objective is to look for the presence of over-reliance in a recurring task.

Proposition 2: *Decision-makers will rely (over-rely) on a decision aid after the aid loses accuracy.*

Another important contrast between the current study and that of previous research can be pointed out. In this study, the aid's recommendation is not a perfect solution: the aid does not yield "correct" and "incorrect" answers. Therefore, passive (and complete) acceptance of the aid's solution is less likely, because subjects can always look for evidence to modify the aid's solution. Glover et al. describe the passive reliance found in their study as merely "pushing the buttons", and accepting the model's output as correct.

3.3 Feedback Environment: Summarized Outcome Feedback

Although over-reliance via an anchoring and adjustment heuristic or an alternative mechanism has been demonstrated, little is known about people's ability to recognize and respond to changes in aid accuracy. By implementing a shift in aid usefulness, we can look not only for evidence of over-reliance, but also for the conditions that influence it. Specifically, it will be determined whether certain factors influence reliance initially, and whether these same factors affect the persistence of reliance. Three such factors are investigated, and discussion of each follows below.

The role of feedback in decision-making has been studied from different perspectives. Waller and Felix (1984) study an environment in which outcome feedback is incomplete. Outcome feedback is simply the result (including direction only or both direction and magnitude) of a previous task or trial. Auditors were provided feedback about the success of their decisions only when one of two courses of action (not to perform substantive tests) was taken. The authors find that incomplete outcome feedback can cause extraneous factors (namely, accuracy rate when feedback was given and base-rates) to affect self-perceived judgment ability. The scenario studied in this paper is a good example of how perfect feedback may not be available in all settings faced by accounting professionals.

The role of feedback in the presence of a decision aid is an additional issue. As discussed previously, several academic studies have tried to determine whether variation in levels and types of feedback can induce task participants to accept the usefulness of a decision model. In general, outcome feedback alone (Arkes et al., 1986; Ashton, 1990), as compared to the absence of outcome feedback, does not achieve this goal. In Ashton's study, subjects knew a priori that a decision model would correctly predict 8 of 16 bond ratings based on financial ratios. Subjects that completed the sixteen trials one at a time (with outcome feedback) relied on the aid less than those that completed all sixteen trials with no feedback. As stated earlier, Ashton explains these findings by suggesting that the aid induces pressure, by providing a high benchmark level of performance. This pressure is harmful to performance. A competing, but somewhat similar explanation is that subjects use the feedback to implement trial and error strategies in order to learn and improve at the task, which comes at the expense of aid reliance. Regardless, Ashton's results demonstrate that increased feedback frequency (given that subjects knew base-rate information) can negatively impact reliance.

Only a few studies successfully provide evidence of the beneficial effects of feedback on reliance. Davis and Kotteman (1995) show that a cumulative form of feedback is persuasive. Dzindolet et al. (2000) similarly report that a combination of several types of feedback is effective in inducing reliance. In Davis and Kotteman (1995), subjects given a running total of the aid's performance versus their own performance eventually increased their reliance. The study employed a production planning task, in which the aid recommended actions (the production level and the workforce level) to be taken. Unless the aid's advice was followed, outcome feedback did not indicate how well the aid's suggestion performed. Therefore, cumulative feedback was not directly compared to trial-by-trial feedback information regarding the aid's validity. As such, a question remains unanswered: can feedback regarding the aggregate performance of a decision aid undo the negative effects found by Ashton (1990) ?

Providing feedback to decision-makers about more than just the previous decision is not a simple undertaking. Ideally, outcomes from decisions are entered into the same information system, and are available for comparison when the next decision is to be made. But depending on the costs to the organization, this may not be possible. It may be true that feedback of any sort is only available through additional effort on the part of the person performing the task. But efforts by organizations to "close the loop", and improve the attention paid to outcomes, might be valuable. This study again tests the effects of varying feedback, in a manner somewhat different than in previous work. A treatment group of subjects is presented with a running summary of the past four outcomes, including their own accuracy levels and the accuracy levels of the decision aid. It is expected that offering this more explicit review of performances will have an incremental effect on the degree to which the model's recommendation is used.

Proposition 3A: *The presence of summarized feedback will increase reliance on a decision aid.*

In addition, it is expected that the presence of this summarized feedback will increase decision-makers' sensitivity to changes in decision aid usefulness. Just as aggregate feedback should strengthen the belief in a valid decision model's usefulness, this feedback condition should also strengthen the understanding of when an aid's usefulness is declining.

Proposition 4A: *The presence of summarized feedback will reduce decision aid reliance (over-reliance) when aid accuracy is low.*

3.4 Timing of Aid Availability

Differing interpretations of the anchoring-and-adjustment heuristic suggest that an anchor can be selected in differing ways. One assumes that all information is available simultaneously, and the decision-maker selects an anchor based on its perceived value, before considering other information (Biggs and Wild, 1985). A second model involves information that is presented at different times. Information that is available early in a decision process is used to formulate a tentative, and perhaps partial, solution. When additional information becomes available, it is given an inappropriately low amount of importance. This is often categorized as a type of primacy effect (Hendrick and Constantini,1970; Anderson and Maletta, 1999).

Boatsman et al. (1997) introduced an audit planning aid to subjects only after they had made initial assessments of the likelihood of fraud. Boatsman et al. choose this methodology arguing that failure to create a tentative solution before consulting the decision aid enhances the possibility of an anchoring-and-adjustment effect. If available from the beginning of the decision process, the aid's recommended solution can potentially direct the manner in which other information is searched and utilized. In a different vein, Lim and O'Connor (1996) show that, in a forecasting task, there can be improvements to accuracy when an initial forecast, based on only partial information, is required. Thus, anchors can sometimes have value.

It is proposed that a decision aid that is available throughout the decision process will be more influential than if selectively viewed.

Proposition 3B: *When the aid's recommendation is explicitly presented from the beginning of the task, reliance will be increased.*

Additionally, this study looks at possible over-reliance effects that are a function of the timing of the aid's availability. It is expected that those who have access to the aid's advice at an earlier point in the decision process will more slowly adjust away from the aid after the aid's accuracy falls. This is based on the premise that a two-stage approach (analyze the task, then consider the aid) will force decision-makers to concentrate on their own perceptions of the task as separate from the decision aid's recommendations, and to better distinguish the differences between the two.

Proposition 4B: *When the aid's recommendation is explicitly presented from the beginning of the task, reliance (over-reliance) will be increased when aid accuracy is low.*

3.5 Task Confidence

Reliance studies have examined several decision-maker characteristics, such as confidence, experience, motivation, and locus of control. This study continues the investigation of the effects of a decision-maker's confidence when deciding whether or not to use a decision aid. The literature suggests that those with higher levels of task confidence will demonstrate reduced decision aid reliance levels. Arkes et al. (1986) and Ashton (1990) both provide evidence of the potentially harmful nature of inappropriate confidence. In both studies, those with high expressed confidence were less likely to use the output of a decision aid, and consequently were outperformed by those with lower confidence levels. Whitecotton (1996) finds that this is true when confidence is measured ex ante (instead of ex post). She refines the measurement of confidence by separating confidence in one's own ability from confidence in the decision aid. Citing Pincus (1991), Whitecotton notes that an outcome view of confidence implies a positive relationship between confidence and performance.

But a process view of confidence relates to confidence in one's abilities to perform the task, and not to confidence in solutions that have already been submitted. The first proposition regarding confidence is a replication of the Whitecotton (1996) study, but in a new environment. Following Whitecotton, confidence is defined here as task confidence (a process view), a construct separate from the overall confidence generated from the combination of self and aid (an outcome view).

Proposition 3C: *Task confidence is negatively related to reliance.*

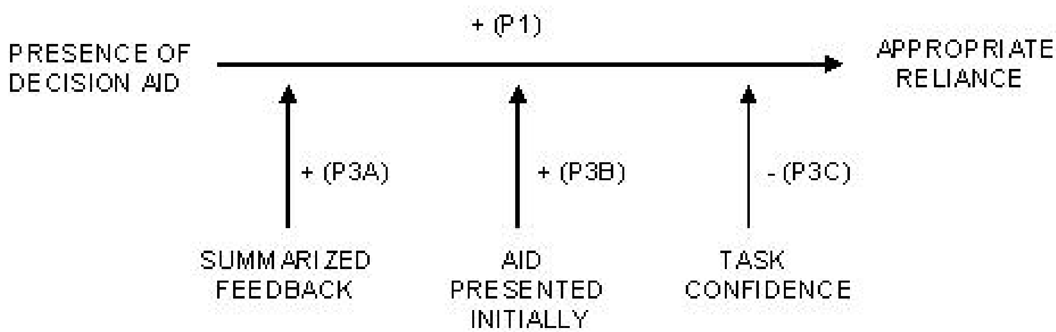
To take the research of Whitecotton a step further, this study seeks to determine whether those with higher confidence are better able to identify and respond to an aid's declining accuracy. Since these individuals seemingly exhibit a higher degree of skepticism towards the value of an aid, it should be the case that they are better attuned to the aid's limitations, and are more sensitive to changes in the aid's usefulness.

Proposition 4C: *Task confidence is negatively related to reliance (over-reliance) when aid accuracy is low.*

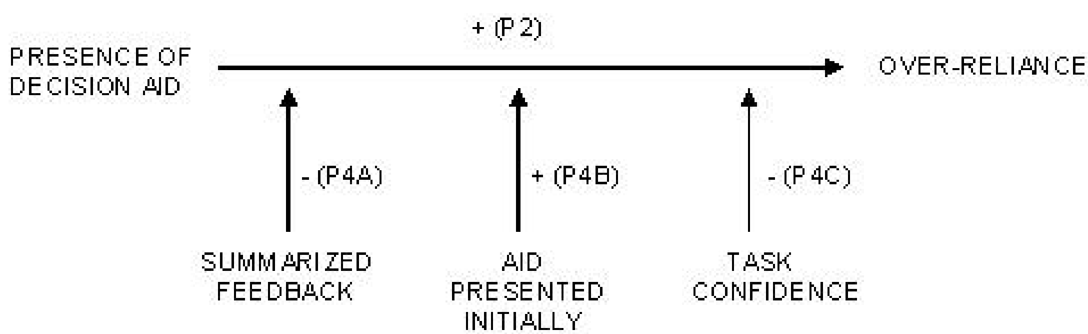
3.6 Experimental Model

A summary of the general models used in this experiment follows.

FACTORS INFLUENCING DECISION AID RELIANCE (HIGH AID ACCURACY):



FACTORS INFLUENCING DECISION AID RELIANCE (LOW AID ACCURACY):



3.7 Hypotheses

From the above propositions comes a set of specific hypotheses that will be tested in a laboratory experiment. Hypotheses 3A, 3B, and 3C are tested over the same time interval, and hypotheses 4A, 4B, and 4C are tested together using a subsequent interval.

As mentioned before, aid reliance in the experimental task can be complete or partial within a single trial, as the aid provides a point estimate. Therefore, reliance in this study is defined as proximity of the decision-maker's estimate to the aid's solution. Similarly, task performance is operationally defined as the proximity of the decision-maker's estimate to the true value.

H1: Aided decision makers will provide solutions that are closer to the decision aid's recommended solution and are more accurate than those of unaided decision makers.

H2: When aid accuracy is low, aided decision makers will provide solutions that are closer to the decision aid's recommended solution and are less accurate than those of unaided decision makers.

H3A: Aided decision makers who receive summarized feedback will provide solutions that are closer to the decision aid's recommended solution and are more accurate than those of decision makers without this feedback.

H3B: Aided decision makers who are presented with the decision aid's recommendation at the beginning of the decision process will provide solutions that are closer to the decision aid's recommended solution and are more accurate than those of decision makers who are not.

H3C: Aided decision makers who have higher levels of task confidence will provide solutions that are further from the decision aid's recommended solution and are less accurate than those of decision makers with lower confidence.

H4A: When aid accuracy is low, aided decision makers who receive summarized feedback will provide solutions that are further away from the decision aid's recommended solution and are more accurate than those of decision makers without this feedback.

H4B: When aid accuracy is low, aided decision makers who are presented with the decision aid's recommendation at the beginning of the decision process will provide solutions that are closer to the decision aid's recommended solution and are less accurate than those of decision makers who are not.

H4C: When aid accuracy is low, aided decision makers who have higher levels of task confidence will provide solutions that are further away from the decision aid's recommended solution and are more accurate than those of decision makers with lower confidence.

Chapter 4

EXPERIMENTAL METHODOLOGY

4.1 Task Overview

The experimental task was a cost accounting task, in which subjects predicted monthly overhead costs for a hypothetical manufacturing company based on measurable factory information (cost drivers).[1] Understanding the relationship between measurable (direct) variables and indirect costs is an important goal for manufacturing entities, as well as for any entity with sizable indirect costs. There were three direct variables available to participants: 1) number of production set-ups, 2) number of machine hours, and 3) number of product tests. Treatment groups also received a model that computed an estimate of overhead. However, after a change in environmental conditions, the relationship between these three factory activity measures and overhead costs changed, causing the static computational model to become significantly less accurate.

4.2 Description of Experiment

There were two factors employed in the design, with task confidence treated as a covariate. Subjects were randomly assigned to one of four groups (figure 1). A control group of subjects (group 1) performed the task without any decision support, and had basic outcome feedback each month. The default aided group (group 2) had the decision aid available by choice, and also received basic

[1] A number of reliance studies use financial accounting tasks, in which subjects have access to financial ratios, and must predict the likelihood that a given company will go bankrupt, or predict the bond rating for that company. It is true that a bankruptcy prediction model can have changing accuracy over time. However, in the current study, an important objective was to start subjects off with equal levels of knowledge. The choice of a cost accounting task eliminates this problem, provides a realistic experimental context, and allows for manipulation of experimental groups.

outcome feedback for each trial. The high feedback treatment group (group 3) additionally had a summary of the past four trials (summarized feedback) at all times. The aid initial group (group 4) had the decision aid's recommendation visible on the screen at all times. Due to sample considerations, there was not a group that received both feedback and aid timing treatments. Figure 1 (in the appendix) provides an overview of the four groups.

The task was run on computers, using a program created with Visual Basic. The protocol for the experiment consisted of the following steps. The researcher met with students during a regular class period, and described the experiment. The classroom script that was used is presented in appendix A. Subjects performed the task on computers at a time they chose. The program was mailed to them electronically. The program gave them thorough instructions. There was practice session of four trials. This was followed by the sequence of thirty trials. Following the trials, subjects returned a data file to the researcher.

Figures A8-A11 show the screen layout. In each trial, subjects were given a hypothetical date (a month), and the relevant information for that month. For each trial, subjects were given the monthly values of each of the factory parameters to use in estimating overhead for the month. The range of possible values for each variable was always presented on the screen. For some of the subjects, a final item of information available on the screen was the recommended solution of the decision aid. Subjects in the "aid initial" condition had the aid's recommendation available at all times. Subjects in the other two aided conditions could only access the aid's recommendation by an explicit request (moving the mouse cursor over the appropriate area). These subjects could view the model's suggestion multiple times if desired. The screen template had an area in which to enter the subject's estimate, in dollars. Subjects had (up to) four minutes to complete each trial, a duration intended to produce only a minimal amount of time pressure. Time remaining was present on the screen, and changed after every ten-second interval. In their attempts to predict overhead costs, subjects could choose to analyze the raw factory production data (three variables), analyze the decision aid's cost estimate, or combine these inputs.

In their effort to work through a given trial, subjects also could use a simple spreadsheet application which is built into the program. This "spreadsheet tool" was provided to lessen the computational requirements of subjects, in order to compress between-subject differences in mathematical ability, and secondly to shorten the duration of the experiment. Despite having this computational tool, subjects still had to iteratively attempt various solutions and recognize relationships in order to improve at the task. The spreadsheet tool consisted of three text boxes, one for each factory cost driver, and a “compute” button (see figure A3). After entering an estimated dollar cost per each unit of each cost driver, subjects could press the compute button to compute a tentative solution. Ultimately, their final solution had to be manually typed in the appropriate field on the screen.

After a subject’s estimate was entered and validated, the correct solution was presented on the screen. The participant then had up to 60 seconds to review the trial, and could proceed to the next trial when ready. In this review process, the original parameter (cue) values, the subject’s estimate, the correct solution, and the decision model’s estimate were all still available.

The three pieces of information (information cues) about the factory’s production for each month, as well as their range of values, are described below. The ranges were explained to come from recent historical evidence from a similar factory:

Production Set-Ups (SETUPS) = number of times the manufacturing equipment must be prepared for use, or converted from one type of use to another: [80 (low) to 250 (high)]

Machine Hours (MH) = hours of machine time: [2400 (low) to 4200 (high)]

Product Tests (TESTS) = number of times that a quality test is performed on inventory: [500 (low) to 1300 (high)]

The instructions explained to subjects that a higher level of each variable is always associated with more overhead costs than a lower level. The distribution of values across the ranges cited above was made to be approximately uniform. Partially due to findings from pilot testing, subjects were also given a limited range of possible values of the cost (weight) of one increment of that variable:

Production Set-Ups (SETUPS): the overhead cost associated with one production set-up will always be between $300 and $700.

Machine Hours (MH): the overhead cost associated with one machine hour will always be between $10 and $40.

Product Tests (TESTS) : the overhead cost associated with one product test will always be between $40 and $100.

Following the practice trials, the actual experiment was given in three sections (hereafter, “section 1”, “section 2”, and “section 3”). In each section, subjects predicted overhead costs for a hypothetical factory (hereafter, “factory 1”, “factory 2”, and “factory 3”, corresponding to each section). Section 1 consisted of 6 trials, section 2 consisted of 10 trials, and section 3 consisted of 14 trials, resulting in 30 total trials. Subjects did not know that the experiment had multiple sections. They also did not know exactly how many trials they would face, but instead were told that the experiment would last about 40-50 minutes. They were not given a count of which trial they are working on.

In section 1 of the experiment, the true overhead amount was equal to the following formula:

OVERHEAD = (SETUPS x $676.71) + (MH x $24.49) + (TESTS x $73.44)

Although correlation between predictor variables would likely be present in a real task of this nature, the variables were made to be uncorrelated for the purpose of generating higher variability in the trials, given the time constraints of the experiment.

Subjects in groups 2, 3, and 4 had access to a decision aid, which provided a recommended solution. The aid was described as follows (in the task instructions):

You have been provided with a decision aid that was developed by the company, based on information from some of its other factories. The aid makes a recommendation that you may refer to when predicting monthly overhead costs. This recommendation is based on a statistical model. You are not required to use the aid’s recommendation in making your prediction.

The decision aid's accuracy was determined based on the results of a pilot group of unaided subjects. This intentional manipulation was carried out by making the weight of one variable too low, and the weight of the other two variables too high. Using this approach, the aid was fairly accurate in its recommendations, but not perfect.

Section 1 (factory 1) consisted of 6 trials, after which subjects were told that they would now estimate overhead for a different factory owned by the same company. Section 2 (factory 2) consisted of 10 trials (7-16). The second factory had cost structure that was nearly the same as that of factory 1, but slightly different, in order to simulate that the factories were, in fact, different.

In this second factory (section 2), the decision aid's model remained unchanged from section 1. The specific weights were selected such that the model's mean accuracy (error) was nearly the same as in section 1. Therefore, in section 2, the task changed slightly, but the model's usefulness in solving the task did not. For both factories 1 and 2, the aid's usefulness was high.

For section 3, the same algorithm was followed. Subjects were told that they would now evaluate a third factory. For this third factory, the new variable weights were selected such that the decision aid had a much higher error size. Therefore, the "fit" of decision aid to environment fell considerably. Exclusive reliance on the aid in section 3 would yield an accuracy rate worse than that of unaided pilot subjects. On the other hand, complete reliance upon the model in section 3 was still clearly superior to any naive strategy.

In real business settings, a decline in decision aid usefulness might take place gradually. Also, in the real world there might not be an obvious signal as to *when* the environment has changed. Here, the decline was one-time and substantial in order to optimally examine aided decision-makers' behavior.

Once a potential series of 30 trials was created, the trials (in each section) were sorted so that aid accuracy was not notably higher at the beginning, middle, or end of a section.

4.3 Independent Variables

4.3.1 Feedback Manipulation

Low feedback subjects received only trial-by-trial outcome feedback, consisting of the solution to the trial. High feedback subjects received both outcome feedback and summarized feedback, consisting of an overview of the outcomes of the past four trials. This overview was shown as a separate item on the screen, present at all times (figure A10).

4.3.2 Timing of Aid Availability Manipulation

Subjects in the "aid initial" condition had the aid's recommendation presented at the beginning of each trial and present throughout the trial (figure A11). Subjects in the "aid by choice" condition, which is treated as the default condition, had to position the mouse cursor over the decision aid's recommendation in order to view it (figure A9).

4.3.3 Measurement of Task Confidence

Task confidence was measured after the practice session and before section 1. The practice session consisted of four trials, and during this session all subjects were unaided. Then, subjects were asked on a 7-point Likert-type scale how confident they felt that they could succeed at such a task. The practice trials were similar to the actual task, but had different information cues (factory variables) and different cost weights attached to them. This approach is similar to that of Whitecotton (1996), who also evaluates task confidence before introducing the decision aid. The shortcoming of this ex ante approach is that task confidence might increase or decrease over the course of the experiment. However, this method completely prevents the presence of the decision aid from affecting the measurement of confidence.

4.4 Dependent Variables

Reliance was measured as the absolute difference between a subject's estimate and the decision aid's recommendation. Thus, a low difference signifies high reliance. The performance measure was the difference between a subject's estimate and the actual outcome. Thus, a low difference signifies high performance. Both of these variables were measured in terms of dollars. The dependent variables were measured across groups of trials: trials 9-16, and trials 23-30 (the last 8 trials of factory two and factory three). In order to attempt to capture the timing of the effect, the final four trials of each section (13-16 and 27-30) were used as a second measurement window. The mean values across these groups of trials were used for statistical testing. The purpose of combining the results from multiple trials was to provide consistent measures of reliance and performance captured in a single variable. The software also measured time spent per trial.

4.5 TESTS OF HYPOTHESES

4.5.1 Hypothesis 1

In order to test hypothesis 1, the unaided control group (group 1) was compared to the default aided group (group 2). Their reliance levels were compared over the final eight (and final four) trials in factory 2. ANOVA was used for this comparison. Naturally, the unaided group's responses are unrelated to the decision aid's recommendation, and so the default aided group's proximity to the decision aid beyond this baseline is a useful measure of actual reliance. Next, the performance of each group was compared to verify that reliance differences had consequences pertaining to task accuracy.

4.5.2 Hypothesis 2

To test hypothesis 2, the unaided control group (group 1) was again compared to the default aided group (group 2). Glover et al. (1998) demonstrate that decision-makers are subject to over-reliance on a decision aid for a single trial, but no research to date has examined over-reliance in an ongoing task. Using ANOVA, the reliance and performance of the default aided group were compared to those of the unaided group in the final trials of factory 3.

4.5.3 Hypotheses 3A, 3B, and 3C

Hypotheses 3A, 3B, and 3C were tested in a manner similar to that of prior research. Trials 9-16 were used as the measurement window. Reliance on the model across this measurement period is modeled as a function of experimental factors. The model that was tested is as follows:

reliance = feedback[a] + aid availability[b] + confidence[c]

a: categorical, manipulated across two levels
b: categorical, manipulated across two levels
c: continuous

This model was tested using analysis of covariance (as a general linear model). The model was also tested (to insure the meaning of results) using performance as the dependent variable. It was expected that higher reliance levels are associated with increased performance (given that aid accuracy is high).

4.5.4 Hypotheses 4A,4B, and 4C

Hypotheses 4A, 4B, and 4C examine reliance levels in a low aid accuracy environment. To test these hypotheses, trials 23-30 (from section 3) served as the measurement period. The complete statistical model is expressed as follows:

reliance = feedback[a] + aid availability[b] + confidence[c]

a: categorical, manipulated across two levels
b: categorical, manipulated across two levels
c: continuous

This model was tested using analysis of covariance (as a general linear model). Again, performance was used as an additional dependent measure to add validity to any reliance findings.

4.5.5 Pilot Test

As described earlier, a pilot test was run prior to the experiment. A small initial pilot test was used to ensure that the objectives of the task were clear, the computer program ran effectively, and that the experimental manipulations were apparent. A larger pilot test was used to help set the task parameters. This procedure is described in the results section.

4.6 Subjects

Subjects were students recruited from upper-level undergraduate and graduate accounting and information systems courses at two universities. They were required to have completed a year of accounting, and virtually all were currently enrolled in an upper-level accounting course. It was expected that subjects had minimal task-related experience. The skill level of subjects was expected to be homogenous, provided that the educational requirements were satisfied. Student subjects have frequently been used before in decision aid studies as reasonable surrogates for entry-level business professionals.

4.6.1 Subject Compensation

Sufficient motivation of subjects was important towards finding results in the study. First, through arrangement with course instructors, subjects were given course credit roughly equal to one percent of their course grade. Secondly, a cash incentive scheme based upon performance was used. Subjects that were in the top 20% of performers were paid $20, and all other subjects were paid $5. Performance was evaluated based on median error size across all trials. Subjects were told the specifics of this incentive scheme, and reminded of these details in the e-mail with which the experimental program was sent. Median performance was selected in order to reward consistent performance, and to avoid unduly penalizing those who used diverse trial-and-error strategies. Using the median also greatly reduced any potential impact of keyboarding or mouse errors on the part of the subjects.

Chapter 5

RESEARCH RESULTS

This chapter describes the research results, and is comprised of four sections. The first section contains preliminary analyses. The second section presents the testing of each research hypothesis. The third section describes analysis of within-groups measures. The fourth section presents supplemental analyses.

5.1 Preliminary Analyses

5.1.1 Pilot Testing

The experimental task was first presented to several pilot subjects, who were undergraduate and graduate business students. The preliminary version of the experimental task was found to be too difficult, based on subjects' performance. It used four factory variables, and did not suggest a minimum and maximum value for the costs (weights) of each of these variables. The task was simplified by lowering the number of factory parameters to three, and by providing the possible range of values that the weights on these parameters could take. Based on the comments of these pilot subjects, some additional changes were made. The screen layout was modified in order to make the experiment easier to read and understand. The experimental directions were revised to more clearly explain the task. And some minor flaws in the calculations performed by the computer program were corrected.

5.1.2 Experimental-Set up

A group of 28 subjects were used to calibrate the experimental task. These subjects were upper division business majors enrolled in accounting courses. They undertook a four trial practice session and then completed 16 trials which were similar to the trials in the experiment. They did not use any decision aid to complete their trials. Their median accuracy levels across trials 9-16 are presented in Table 1.

This data was used as a guide in designing the experimental task. With the help of an iterative statistical optimization tool (the solver add-in for Microsoft Excel), the decision aid formula was defined. It was created to have a median cost error of about $6,900 for factory 1/factory 2. Thus, the decision aid performs as well as the best performers in the calibration group. Appropriately, this represents a very high level of accuracy, but not one that is completely unattainable by unaided subjects. That is, given the variability surrounding the aid's accuracy, individuals could outperform the aid on at least a few trials.

The underlying cost formula for factory 3 was created to be considerably different than the formula for the first 2 factories. The aid's formula was set to have a median error of about $16,000[2]. The aid's error size increases more than two-fold for factory 3. Also, note that in this "low accuracy" phase, the aid was set to be at about the 20th percentile of subjects' performance in the calibration group. As such, it was of low accuracy, but still not worse than a portion of the (unaided pilot) subjects. Ultimately, the shift in aid accuracy was designed to be significant, but not so large as to eliminate reliance completely.

Table 2 shows the specific 30 trials (6->10->16) that constituted the task, along with the underlying cost formula for each factory, the decision aid's formula, and the decision aid's accuracy level. Judging from the results of pilot testing, it was not necessary to introduce any random error into the environment, as it was very difficult to converge on the actual cost formula (for any of the factories) within the given number of trials.

[2] As described in the previous chapter, the actual weights of each cost for factory 3 were chosen to give the aid its pre-determined accuracy level. Hence, the aid's formula was constant throughout the experiment.

5.1.3 Subjects

A total of 172 subjects participated in the experiment: 45 unaided, 44 (default) aided, 41 aided with high feedback, and 42 aided with aid presented initially. Subjects had all completed at least one course in managerial accounting, and the majority indicated that they had completed at least one additional upper-level cost accounting course. As expected, there were no significant differences in age and experience among the subjects.

Forty-two subjects were eliminated from the analysis for varying reasons. One subject was eliminated for not using the spreadsheet tool at all. Nine subjects did not make any changes to their weights for factory 3, and were eliminated for not putting forth appropriate effort. Thirty-two subjects were judged to have misunderstood the directions.

These thirty-two subjects had high variability in their parameter weights: from trial to trial, throughout the experiment, they changed weights significantly. This alone was not a problem, but upon inspection of individual trials, the subjects' parameter coefficients clearly varied inversely with the value of the parameters (# machine set-ups, # machine hours, and # product tests) for that month, implying that they thought that the (cost) weights changed each month, when in reality the values of the factory data changed each month, while the underlying cost structure remained constant[3].

Table 3 overviews the remaining 130 subjects that were used to test the hypotheses.

5.1.4 Manipulation and Reasonableness checks

Feedback Condition

Subjects were given a feedback manipulation check question at the end of the experiment, along with their other questions. The question was stated as "I was aware of my performance over the

[3] Using their weight for machine set-ups as an objective guideline, all of these subjects weighed machine set-ups at least $200 higher when the value for set-ups was high (top quartile of trials) than when it was low (bottom quartile). Interviews with some of these subjects indicated that they failed to realize that the underlying cost formula was consistent across trials for a given factory. They may have assumed an inverse relationship between the size of a measure and the "unit cost" for that measure. Also, performance data show that this sub-sample had inferior performance to the remaining subjects. Fortunately, the relative usefulness of the decision aid (high usefulness then low usefulness) was still intact for the remaining sample.

previous four months", and had a scale of 5 (strongly agree) to 1 (strongly disagree). For the high feedback group, the average score on the question was 4.25 out of 5.0. Those subjects in the default aided group had an average answer of 3.03 out of 5.0. A t-test indicated that these means were significantly different ($p<0.001$).

Timing of Aid Availability Condition

Subjects were also asked to rate their agreement with the statement "I had to use the mouse in order to view the recommended (decision model) solution". All subjects in the "aid initial" treatment appropriately answered this question negatively.

Task Difficulty

Because subjects were able to use the “spreadsheet tool”, the coefficients they used in each trial were observable. From these coefficients, a general examination of task difficulty could be made. When subjects are improving at the task, the percentage deviation between the coefficients provided by the subjects and the true coefficients becomes smaller (i.e., subjects are getting closer to the true formula). Table 4 shows the average percentage by which the median *unaided* (group 1) subjects' coefficients differed from the actual coefficients. Because the cost formulas for factory 1 and factory 2 are nearly the same, the two are presented together. The data in this table verify that two important assumptions are met. First, subjects could improve their performance over the duration of the experiment, implying that they understood the task and had the necessary ability to do it. Second, the task was difficult enough that subjects could not actually "solve" the underlying formula, although some came closer than others.

For factories 1 and 2, there was a negative correlation between trial number and average percentage coefficient error, suggesting that subjects improved at understanding each factory and across factories. Inspection of table 4 shows that this learning was gradual, and that the learning rate slowed down over time.

Decision Aid Accuracy

At the end of the experiment, subjects were asked to rate the decision aid's accuracy from 5 (high) to 1 (low). The mean rating across all aided subjects was 3.2. This result is reasonable in that the decision aid's suggested answer had perceived usefulness to some, but not to others. Also, this mean rating reflects the fact that the aid's accuracy changes during the experiment. Asking this question after factory 2 (during the experiment) was not possible as it would have yielded information about the underlying nature of the experiment.

5.1.5 Tests of Statistical Assumptions

The assumptions underlying the statistical methods were assessed. Anova and Ancova were used to test the proposed hypotheses. These methods require normality and homogeneity of variance.

Graphical representations of the data were initially used to assess the assumption of normality. Although inspection led to the conclusion that the data were approximately normal, the Kolmogorov-Smirnov Normality Test was performed to sustain this conclusion. The assumption of normality was supported (p-values greater than 0.15 in all cases).

Levene's test for Homogeneity of Variance was computed as each statistical test was performed, and showed that the treatment cells had homogeneous variances in each case (all p-values greater than 0.05).

Because these assumptions were met, the statistical models were deemed to be validly calculated.

5.2 Tests of Hypotheses

This section discusses the results of the testing of each hypothesis. For each hypothesis, the dependent measure was captured over a range of trials. The primary measurement window was the final eight trials of each period. A secondary measurement window was the final four trials of each period, and the results of these tests are also presented. Hypotheses 1 and 2 are supported. Hypothesis 3A is not supported. Hypothesis 3B is supported. Hypothesis 3C is not supported. Hypotheses 4A, 4B, and 4C are not supported. This information is summarized in Figure 3. The specific findings for each hypothesis are described in detail in the remainder of this chapter.

5.2.1 Hypothesis 1

For examination of hypothesis 1, reliance and performance were examined when the aid was accurate (Tables 5 and 6). Reliance was measured by mean absolute difference of the subject's response from the decision aid's solution. The degree of reliance can be gauged by comparing the reliance levels of default aided subjects to those without any aid. One-way ANOVA was utilized for all comparisons of the unaided group to the default aided group. As expected, aided subjects were influenced by the decision aid's recommended solution (Table 5). Reliance was present in trials 9-16 (the final 8 trials of factory 2). From Table 5B, unaided subjects had a mean absolute difference of 15,813, while aided subjects' was 9,198.

Additionally, this reliance led to improved decision-making performance, as subjects' mean errors across trials 9-16 were significantly lower in the presence of the decision aid (Table 6). From Table 6B, mean error sizes averaged 14,442 for unaided subjects and 10,942 for aided subjects. It seems that aided subjects performed significantly better than unaided subjects in the presence of a statistically valid decision aid.

Examining a smaller window, the last four trials alone, yields weaker results: the difference in reliance is still significant ($P<.001$), but the corresponding effect on performance is marginal. The performance difference in the last four trials was 1,803 ($p=.105$). One possible explanation for this finding is that the unaided subjects' lower error in the latter trials reflects continuing task improvement with additional time.

5.2.2 Hypothesis 2

The next hypothesis states that users will over-rely on a decision aid when the aid's accuracy declines to a substantially lower level. In the experimental design, this decline is implemented by the aid's error size increasing to a level of approximately 230% of the original level, changing from the equivalent of the best performers to the bottom 20% of performers in the unaided calibration task.

.

Hypothesis 2 was tested by comparing unaided subjects to the default aided group during factory 3. The presence of the decision aid was significantly related to reliance on the aid in trials 23-30 (the final 8 trials of factory 3), meaning that reliance was still significant after a substantial decline in aid accuracy (Table 7B). Unaided subjects had a reliance measure (mean absolute difference) of 19,885 and aided subjects' reliance was 13,389 ($p<.001$).

To determine if this reliance had a material impact on performance, the mean errors of unaided and aided subjects during factory 3 were compared (Table 8). Table 8 shows that the default aided group had lower performance in factory 3 than did unaided (control) subjects. This difference is significant at the $\alpha=.05$ level. Thus, hypothesis 2 is supported. Decision-makers continued to use the aid after its accuracy decreased, even though it caused them to perform worse than unaided subjects.

Examination of the last four trials alone yields similar results: the difference in reliance is similar. The corresponding effect on performance is significant at the $p=.092$ level.

5.2.3 Hypothesis 3A

Hypotheses 3A, 3B, and 3C were tested together using ANCOVA. This was specifically implemented by using process GLM (general linear model) in SPSS. Presence of additional feedback (FEEDBACK) was modeled as a categorical variable. The aid presented at the beginning of each trial (INITIAL) was also modeled as a categorical variable. Task confidence (CONFIDENCE) was included as a continuous variable.

Hypothesis 3A states that differing feedback environments can influence a decision-maker's willingness to rely on a decision aid's advice. Namely, an environment in which a history of both one's own accuracy and aid accuracy is available ("high feedback") will lead subjects to rely on the aid to a greater degree. The results of the testing of this hypothesis are presented in Table 9.

Reliance was greater for those in the higher feedback condition (Table 9). The coefficient on feedback is -$976, which is in the expected direction. However, the t-test on the feedback variable was not significant. Similarly, the effect of feedback on performance was not significant (Table 10).

5.2.4 Hypothesis 3B

Hypothesis 3B pertains to the timing of the decision aid's entrance into the decision process. It states that a decision aid's recommendation that is presented from the beginning of the trial will be more influential than such a recommendation which is available to be accessed by the user at any chosen point during the trial.

Hypothesis 3B is supported. Subjects in the "AID INITIAL" group had significantly lower deviations from the aid's recommendation (Table 9). Their performance (Table 10) was also better (p=.057). Over the last four trials alone, these results are not repeated. The significance level for reliance is marginal (p=.080), and significance for performance is not found (p=.133).

5.2.5 Hypothesis 3C

Hypothesis 3C predicts that subjects' task confidence levels will be inversely related to their reliance on a decision model. Confidence was indicated by subjects after the practice session, and before the decision aid was described and introduced. Table 13 details the frequencies for confidence scores. The grand mean for confidence was 4.35 out of 7, and average confidence scores were approximately equal in all three experimental groups.

From Tables 9 and 10, there is no support for the hypothesis that confidence is associated with lower reliance. Patterns of relying on the decision aid do not appear to be affected by subjects' confidence levels.

5.2.6 Hypothesis 4A

Hypotheses 4A, 4B, and 4C concern the influence of the above three factors on decision aid reliance after a change in environment. The dependent measure in these three hypotheses is the reliance level in factory 3, where the decision aid's model is poorly matched to the environment. All three hypotheses were tested together using ANCOVA (implemented as a general linear model). Presence of additional feedback (FEEDBACK) was modeled as a categorical variable. The aid presented at the beginning of each trial (INITIAL) was also modeled as a categorical variable. Task confidence (CONFIDENCE) was included as a continuous variable.

When the environment changed, aided subjects did reduce their reliance levels. For each group (default aided, feedback, and aid initial) and for the entire sample, reliance was lower in factory 3 than in factory 2 ($P<.001$ in all cases). Subjects detected the change in aid accuracy.

Hypothesis 4A states that in the high feedback condition, subjects will have significantly lower reliance levels after a change in environment (that is accompanied by a decline in aid accuracy) than those without feedback. The predicted results for feedback condition were not found. Post-change reliance level was not affected by the feedback treatment (Table 11).

5.2.7 Hypothesis 4B

Hypothesis 4B states that those in the AID INITIAL treatment group should have higher reliance levels than will those in the default aided control group (as is also expected prior to the environmental change). Hypothesis 4B is not supported (Table 11). The post-change reliance level was not significantly related to the timing of decision aid availability. The aid initial subjects did make estimates that averaged $1117 closer to the recommended solution. But the results of statistical testing do not support the claim that the decision aid serves as a stronger anchor when it is available from the beginning of the trial.

5.2.8 Hypothesis 4C

Hypothesis 4C suggests that subjects with higher task confidence will have lower reliance levels than those with low confidence following a reduction in aid accuracy. Hypothesis 4C is not supported (Table 11). Inherent task confidence levels do not appear to relate to reliance on a decision aid in either environment.

Table 12 adds additional evidence that performance in factory 3, like reliance, did not vary with experimental conditions.

5.3 Analyses of within-groups data

5.3.1 Overview of Within-groups Findings

Table 14 summarizes within-group data across periods. Several conclusions are worth noting. All four groups had sizable reductions in reliance in factory 3. T-tests for each group show that these reductions are significantly different from zero. This is driven partially by a true reduction in reliance, and partially by accurate estimates becoming, by nature, further from the less accurate aid. Unaided subjects had approximately the same performance in factory 3 as in factory 2. A t-test comparing the two samples is not significant. However, aided subjects had substantially reduced performance, statistically significant using a t-test. These results are the same when comparing either the last eight or last four trials of each section.

Although there was not support for H3A nor H3B, table 14 (8 trials) reveals that the feedback group reduced their reliance by $1721 more than the default aided group ($5912 versus $4191). This "overall" difference proved to be significant (p=.095). Therefore, there is evidence for the expected effects of feedback when both experimental sections are viewed in combination.

5.3.2 Trial-by-trial Within-groups Findings

In Table 15, a continuation of within-groups analysis is presented. The table presents information about correlations between data and trial number, and also shows the relationship between the first half of trials and second half of trials for factories 2 and 3.

In factory 2, all three aided groups had negative correlations between (distance from aid) and trial number, indicating that their reliance grew stronger over time. Likewise, their estimates were closer to the aid in the second half of trials than in the first. All four groups gradually improved at the task, and made better estimates in the second half of the segment.

In factory 3, all three aided groups had positive correlations between (distance from aid) and trial number, indicating that their reliance grew weaker over time. Their estimates were further from the aid in the second half of trials, relative to the first half. All four groups gradually improved at the task. But there was less improvement than the improvement shown in factory 2.

5.4 Supplemental Analyses

5.4.1 Examination of Time Spent on Task

The presence of the decision aid was associated with greater time spent on each trial. Default aided subjects had an average of a median 44 seconds spent per trial in factory 2, while unaided subjects used an average of 34 seconds per trial in factory 2. A t-test showed this difference to be significant. Time for both groups declined slightly during factory 3, but the difference between the two groups was still significant.

5.4.2 Time Spent associated with Reliance Level

The relationship between task time and reliance level was also apparent. The correlation between trial time and reliance was about 0.50 in both measurement periods (factory 2 and factory 3). While the exact sequence of cause and effect remains unknown, it is generally likely that those who spend more time contemplating the aid's recommendation are more likely to use it. Previous research (Wedell and Senter, 1997; Pennington, 2002) has found this to be true for information cues in general; it follows that the same expectations apply to decision model recommendations.

5.4.3 Demographic Data and Questionnaires

Subjects indicated that they were motivated to outperform others. The mean value for motivation across all subjects for motivation was 3.8 (on a scale of 1-5). Likewise, subjects indicated that they put forth fairly high effort levels (3.8). Very few subjects indicated effort levels of less than 3.

Confidence levels after the task were not significantly related to performance. The correlation between post-experiment confidence level and error size during the experiment was -0.11, which is in the desired direction but not significant. The correlation was somewhat higher for unaided subjects, suggesting that some aided subjects benefited from relying on the aid while simultaneously not developing confidence about their task ability.

5.4.4 Examination of Selective Aid Access

An additional analysis can be performed by looking at the processes that take place within a single decision (a single trial). It was found that, in cases where the decision aid's recommendation is accessed, viewing the aid's solution earlier in the trial has a greater impact on the decision than viewing the solution later in the trial.

The examination was done using all trials (for all default aided subjects) in which the aid was accessed during the factory 2 segment. There were 32 default aided subjects, who completed 10 trials in factory 2, resulting in 320 trials for potential examination. The aid was not accessed in 41 trials, leaving 279 potential observations. In 143 of the cases, the aid was accessed before any other actions were taken. In 136 of the cases, the aid was accessed later in the trial. Results show that the AID-FIRST subject-trials had associated estimates that were closer to the aid's recommendation than the AID-LATER subject-trials (p=0.01). The results suggest that when the decision aid is *selectively* viewed at the beginning of the task, it carries more influence than if it is *selectively* viewed later in the task.

These findings are difficult to interpret in light of variability in subject ability levels. In other words, subjects that access the aid earlier may be more likely to *need* the aid's recommendation to formulate a high quality estimate. Therefore, it cannot be shown that for a given individual, viewing the aid's recommendation earlier causes inherently different behavior than viewing the aid's recommendation later. Instead, there is an overall association between viewing the aid earlier and reliance. Most individuals had a fairly consistent pattern of aid access across the ten trials.

Chapter 6

CONCLUSIONS, LIMITATIONS AND CONTRIBUTIONS

This research adds evidence to the existing body of literature regarding the effects of using statistical decision aids on individual decision-making processes. The study's main contribution is to extend findings on decision aid reliance by placing decision-makers in an environment in which change occurs that impacts the accuracy of the decision aid. The study presents evidence of both reliance and over-reliance, and also of the impact of feedback environment and aid timing on reliance. The remainder of this chapter discusses these findings.

The sample of aided subjects outperformed the sample of unaided subjects. This shows that the decision aid's recommendation was incorporated into the decision process. These findings of reliance are expected, and are typical of the results of prior literature, such as Brown and Jones (1998) who show that aided subjects have agreement scores that are closer to an aid's selection in a choice task than those scores of an unaided control group. At the same time, the results of this research reflect the fact that humans are subject to under-performing statistical models. No single group of subjects had performance superior to the decision aid's performance. These findings show that the procedures of the experiment are capable of revealing differences expected from previous decision aid reliance research.

Subjects also demonstrated continued reliance in the environment where the decision aid's recommendation became inaccurate. This extends the over-reliance findings of Glover, Prawitt, and Spilker (1997) and Kowalczyk and Wolfe (1998). Glover et. al showed evidence of inappropriate application of an aid used to compute capital gains taxes. Kowalczyk and Wolfe showed that individuals can conservatively anchor around a going concern prediction model's conclusion. In the current study, there was a clear shift in environment in a cost accounting task.

The study also tested the impact of feedback for moderating over-reliance. There is support for the proposition that higher feedback levels, implemented as a summarization of past performance, are incorporated into the decision-making process and reduce the potential for over-reliance. Differences in feedback environment had some bearing on subjects' willingness to trust the decision aid. Subjects were given a history of recent outcomes, which included the error size of the decision aid. After the accuracy of the decision aid was substantively reduced, significant differences existed on the change in reliance between the feedback and default aided groups. The change in reliance for the feedback group was $1721 higher (when comparing the last eight trials) than for the default aided group. The results suggest that DSS users do not optimally utilize past information. Arkes et al. (1986) and Ashton (1990) both show that feedback can even lead to errant behavior. But, like Davis and Kotteman (1995), who provide subjects with ongoing statistics concerning the aid's accuracy, this study shows that appropriate and well-designed feedback schemes can make a difference in reliance on a decision aid.

The study also finds that the timing of decision aid availability can influence the user's reliance behavior. When the decision aid was initially available on the screen, it was more influential than if viewed only by choice. This finding implies that merely viewing a model's recommendation affects users: if using a decision aid is required in an accounting task, individuals are subject to be influenced by it more than if they have the option to consult the decision aid. Thus, this finding may suggest that initially viewing the decision aid's recommendation may initiate fundamentally different cognitive processes, such as seeking to confirm or disconfirm the recommendation versus attempting to solve the problem. But this study provides an initial look at decision aid timing.

Another finding is that confidence had no bearing on decision aid reliance. Whitecotton (1996) found that those with high inherent task confidence used the output of a statistical model to a lesser degree. It was expected that individuals with higher confidence would have lower reliance in both before and after changes in the environment, meaning that confidence would have both positive and negative effects. But there was no support for these hypotheses. High confidence individuals acted in a manner similar to that of low confidence individuals. A possible counter-argument to the Whitecotton (1996) proposition is that individuals with higher confidence may be superior in some other trait, giving them increased ability to comprehend and evaluate the accuracy of a decision model, thereby clouding the result. Another alternative explanation has to do with the self-reporting of task confidence, and subject characteristics. In the present study, subjects made a confidence assessment based on a relatively short practice session. Whitecotton uses a mixture of financial analysts and students. Therefore, her main effect for confidence, although reportedly unrelated to task experience, may include secondary effects of experience, such as familiarity with similar types of analytical judgments.

A final result of the study is that the presence of the decision aid results in increased time spent on each decision. This result was found for all aided groups when the aid was accurate or inaccurate. This finding contradicts an effort minimization explanation (Todd and Benbasat, 1991, 1992, 1994) as minimization implies that having the recommendation would lead to no additional thinking and almost perfect reliance. When time pressure is significant, it may be advantageous for decision-makers to reduce their information search or information processing. But when time pressure is less intense, the decision aid recommendation provides an additional input cue, and therefore can increase information processing time. Chu and Spires (2002) argue that users make effort and quality trade-offs when using computerized decision aids, and that they follow a pure cost/benefit approach. Finally, the results indicate that higher time spent completing a task is associated with higher reliance levels. Among other possibilities, this may imply that motivation or effort could be positively associated with reliance.

Implications for Practice

The results of this paper suggest that organizations should attempt to keep decision aids up to date with changing environments. This might entail recalibration, addition of new variables, removal of variables, or abandonment of the aid entirely. Attempts to provide feedback to decision-makers are warranted. Feedback mechanisms may provide a partial remedy when aids cannot be updated.

Also, managers should be concerned with the procedures in place regarding the manner and timing in which a statistical model fits into the decision process. Decision-makers are more likely to incorporate the advice of a model if the model is made explicit to them. That is, the process within which an aid is to be used should be designed to fit the decision being made.

Limitations of the Research

The limitations of the study presented here must be considered. As is always the case, it is difficult to ensure the external validity of a laboratory experiment. Most specific to the current study, the environmental changes induced in this experiment may be more subtle, more gradual, and take a longer period of time to unfold outside of the laboratory.

Another potential limitation is the use of student subjects. To address this issue, this study used a relatively straightforward cost accounting task which students can readily understand, using predictor variables whose relationship to the dependent value is logical, but whose weights in this relationship could vary widely. Also, the relationship between the subject and the task was made to reasonably approximate the setting faced by a relatively inexperienced business professional. The task was of a difficulty level that permitted improvement over the course of the experiment.

Lastly, this study raises questions about the means of administering behavioral experiments. Many current studies use a web-based interface, and often the experimental participant operates in a different-time, different-place relationship to the researcher. The aim of this approach is to improve convenience, with a likely goal of increasing participation rates. But this study found evidence that on-line subjects are not likely to take full advantage of asking for help, a situation that may be ameliorated in a traditional laboratory setting. This is an open area of questioning for future experimental protocols.

Contributions of the Research

The results of this study make an important contribution to our understanding of judgment behavior in the presence of computerized assistance. Decision support offers the ability to raise the limits of an individual's cognitive processing capabilities, and thereby contribute to an organization's success. But due to situations in which decision aids cannot be continually updated, we must be able to understand how individuals react when decision aids fail. The current research provides evidence towards this goal, through an examination of decision-making assisted by a decision aid that comes to lose its utility due to an environmental change. This study provides additional evidence on the ways that the presence of a decision aid can influence user behavior and strategy selection. The study contributes to existing literature by extending research on reliance into a dynamic realm much more consistent with organizational environments by applying a methodology which includes multiple trials in an ongoing task.

Opportunities for Future Research

Several opportunities for additional research are available. While this study considers the most significant factors contributing to reliance and over-reliance on a model, many other factors, including other behavioral characteristics, might be explored.

Further research could also examine the differences in how experts and novices change their reliance behavior in response to environmental changes. And other task environments might be studied to see if the results, as expected, hold in different situations.

Decision aid studies have generally not considered the frequency or the importance of the decision in question. Decisions made infrequently are undoubtedly different than routine decisions, and the corresponding decision aid reliance issues are likely different as well. Similarly, decisions of greater importance may yield different behavior. Future research must attempt to separate routine, recurring decisions from those with more severe consequences.

Decision-makers are more likely to incorporate the advice of a model if the model is made explicit to them. Simulating decision aid timing in a setting that includes communications with colleagues, examination of historically similar cases, consideration of recent trends, etc., also has potential as a challenge for researchers.

Lastly, as the quantities of data available to business decision-makers continues to grow, and the techniques used to aggregate that data into decision tools become more defined, future research will be able to add more precision when describing and studying the environments faced by users of decision-aids.

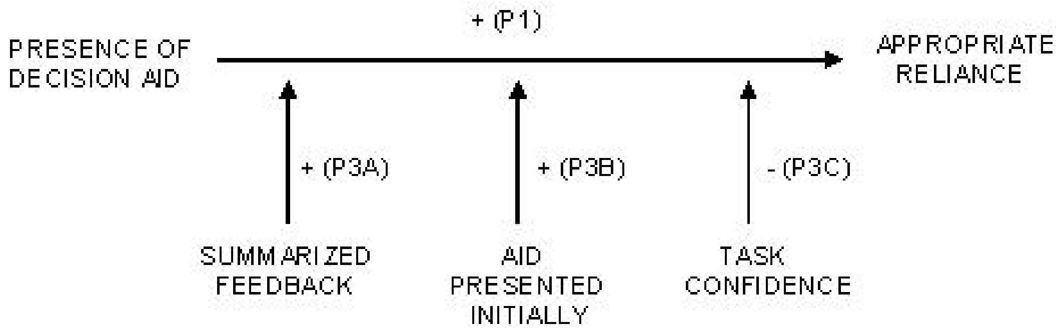
FIGURES

FIGURE 1: OVERVIEW OF EXPERIMENTAL GROUPS

	low feedback (outcome)	**high feedback (outcome and summary)**
unaided	group 1	
aid available by choice	group 2 (default aided)	group 3
aid presented initially	group 4	

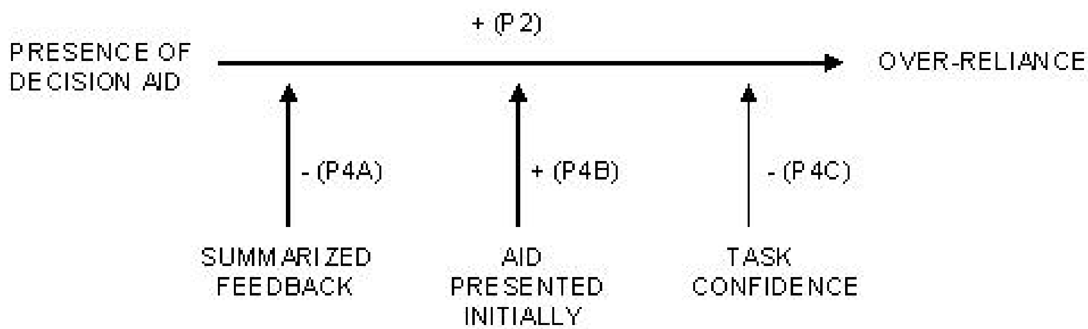
FIGURE 2: OVERVIEW OF EXPERIMENTAL MODELS

FACTORS INFLUENCING DECISION AID RELIANCE (HIGH AID ACCURACY):



:

FACTORS INFLUENCING DECISION AID RELIANCE (LOW AID ACCURACY):



.

FIGURE 3: OVERVIEW OF RESULTS OF HYPOTHESIS TESTING

H1	Aided decision makers will provide solutions that are closer to the decision aid's recommended solution and are more accurate than those of unaided decision makers.	supported
H2	When aid accuracy is low, aided decision makers will provide solutions that are closer to the decision aid's recommended solution and are less accurate than those of unaided decision makers.	supported
H3A	Aided decision makers who receive summarized feedback will provide solutions that are closer to the decision aid's recommended solution and are more accurate than those of decision makers without this feedback.	not supported *
H3B	Aided decision makers who are presented with the decision aid's recommendation at the beginning of the decision process will provide solutions that are closer to the decision aid's recommended solution and are more accurate than those of decision makers who are not.	supported
H3C	Aided decision makers who have higher levels of task confidence will provide solutions that are further from the decision aid's recommended solution and are less accurate than those of decision makers with lower confidence.	not supported
H4A	When aid accuracy is low, aided decision makers who receive summarized feedback will provide solutions that are further away from the decision aid's recommended solution and are more accurate than those of decision makers without this feedback.	not supported *
H4B	When aid accuracy is low, aided decision makers who are presented with the decision aid's recommendation at the beginning of the decision process will provide solutions that are closer to the decision aid's recommended solution and are less accurate than those of decision makers who are not.	not supported
H4C	When aid accuracy is low, aided decision makers who have higher levels of task confidence will provide solutions that are further away from the decision aid's recommended solution and are more accurate than those of decision makers with lower confidence.	not supported

* There is support for the effect of summarized feedback when both experimental sections are considered together.

TABLES

.

Table 1: Median Error Sizes of Pilot Group

	median error
100th percentile (best performance)	$6,912
75th percentile	$10,387
50th percentile	$13,212
25th percentile	$15,465
n=28	

Table 2: Overview of Experimental Trials

		P1	P2	P3		
	aid formula:	**$660**	**$34.50**	**$44**		
factory 1	**cost formula:**	**$676.71**	**$24.49**	**$73.44**		**aid error$**
1		180	3530	710		11433
2		140	4090	1130		5342
3		120	3160	630		11086
4		200	2790	1220		-11327
5		220	2610	790		-803
6		110	3720	960		7144
factory 2	**cost formula:**	**$674**	**$26**	**$71**		
7		160	3350	540		11655
8		230	2980	1050		-6240
9		250	2420	880		-6690
10		90	3900	1300		-3210
11		220	2880	1260		-12620
12		80	3620	1170		-1940
13		240	3070	500		9235
14		110	2700	1010		-5860
15		100	3810	580		15325
16		130	3250	920		965
median absolute aid error (1-16):						6917
factory 3	**cost formula:**	**$363.82**	**$37.93**	**$81.91**		
17		180	3530	710		14276
18		140	4090	1130		-15415
19		90	3900	1300		-36015
20		110	3720	960		-16585
21		220	2610	790		26251
22		200	2790	1220		3410
23		160	3350	540		15415
24		230	2980	1050		18086
25		250	2420	880		32377
26		120	3160	630		808
27		220	2880	1260		7508
28		80	3620	1170		-33087
39		240	3070	500		41587
30		110	2700	1010		-14977
median absolute aid error (17-30):						16000

Table 3: Overview of Subjects

Group:	A (unaided)	B (aided)	C (aided with feedback)	D (aided with aid initial)
initial sample:	45	44	41	42
didn't use spreadsheet		1		
didn't attempt factory 3	2	3	3	1
misinterpreted directions	8	8	6	10
final sample:	35	32	32	31

Table 4: Coefficient Error Rates by Trial (Factories 1 and 2): Unaided Subjects (group A)

trial number	median coefficient percentage error
factory 1	
1	31%
2	34
3	25
4	25
5	27
6	23
factory 2	
7	21
8	21
9	22
10	23
11	21
12	21
13	18
14	20
15	20
16	19

Table 5: ANOVA: Effect of Aid Presence on Reliance (Factory 2)

Panel A. ANOVA Table: Aid Presence (8 trials)					
	df	SS	MS	F	P-value
Between Groups	1	7.32E+08	731587409	44.568	0.000
Within Groups	65	1.07E+09	16414924		
Total	66	1.80E+09			

Panel B. Descriptive Statistics (8 trials)			
	mean	stdev	N
UNAIDED	15813	4402	35
AIDED	9198	3628	32
Total	12654	5220	67

Panel C. ANOVA Table: Aid Presence (4 trials)					
	df	SS	MS	F	P-value
Between Groups	1	8.43E+08	843130572	34.149	0.000
Within Groups	65	1.60E+09	24689476		
Total	66	2.45E+09			

Panel D. Descriptive Statistics (4 trials)			
	mean	stdev	N
UNAIDED	15661	5426	35
AIDED	8559	4413	32
Total	12269	6090	67

Definition of Variables:
AIDED: categorical variable indicating whether subjects had access to decision aid.

Reliance was measured as the mean absolute difference from the decision aid's recommendation across trials 9-16 or 13-16 (factory 2).

Table 6: ANOVA: Effect of Aid Presence on Error Size (Factory 2)

Panel A. ANOVA Table: Aid Presence (8 trials)					
	df	SS	MS	F	P-value
Between Groups	1	2.05E+08	204729107	15.442	0.000
Within Groups	65	8.62E+08	13258153		
Total	66	1.07E+09			

Panel B. Descriptive Statistics (8 trials)			
	mean	stdev	N
UNAIDED	14442	3907	35
AIDED	10942	3324	32
Total	12770	4019	67

Panel C. ANOVA Table: Aid Presence (4 trials)					
	df	SS	MS	F	P-value
Between Groups	1	54366696	54366696	2.695	0.105
Within Groups	65	1.31E+09	20173515		
Total	66	1.37E+09			

Panel D. Descriptive Statistics (4 trials)			
	mean	stdev	N
UNAIDED	12945	4421	35
AIDED	11142	4567	32
Total	12084	4548	67

Definition of Variables:
AIDED: categorical variable indicating whether subjects had access to decision aid.

Error size was measured as the mean absolute difference from the correct solution across trials 9-16 or 13-16 (factory 2).

Table 7: ANOVA: Effect of Aid Presence on Reliance (Factory 3)

Panel A. ANOVA Table: Aid Presence (8 trials)					
	df	SS	MS	F	P-value
Between Groups	1	7.05E+08	705481309	32.640	0.000
Within Groups	65	1.40E+09	21613797		
Total	66	2.11E+09			

Panel B. Descriptive Statistics (8 trials)			
	mean	stdev	N
UNAIDED	19885	5228	35
AIDED	13389	3916	32
Total	16782	5654	67

Panel C. ANOVA Table: Aid Presence (4 trials)					
	df	SS	MS	F	P-value
Between Groups	1	8.52E+08	851597922	19.228	0.000
Within Groups	65	2.88E+09	44289813		
Total	66	3.73E+09			

Panel D. Descriptive Statistics (4 trials)			
	mean	stdev	N
UNAIDED	22828	7757	35
AIDED	15691	5183	32
Total	19419	7518	67

Definition of Variables:
AIDED: categorical variable indicating whether subjects had access to decision aid.

Reliance was measured as the mean absolute difference from the decision aid's recommendation across trials 23-30 or 27-30 (factory 3).

Table 8: ANOVA: Effect of Aid Presence on Error Size (Factory 3)

Panel A. ANOVA Table: Aid Presence (8 trials)					
	df	SS	MS	F	P-value
Between Groups	1	42180934	42180934	4.293	0.042
Within Groups	65	6.39E+08	9826630		
Total	66	6.81E+08			

Panel B. Descriptive Statistics (8 trials)			
	mean	stdev	N
UNAIDED	13908	2932	35
AIDED	15497	3342	32
Total	14667	3211	67

Panel C. ANOVA Table: Aid Presence (4 trials)					
	df	SS	MS	F	P-value
Between Groups	1	47838145	47838145	2.917	0.092
Within Groups	65	1.07E+09	16401547		
Total	66	1.11E+09			

Panel D. Descriptive Statistics (4 trials)			
	mean	stdev	N
UNAIDED	14015	4342	35
AIDED	15706	3702	32
Total	14823	4108	67

Definition of Variables:
AIDED: categorical variable indicating whether subjects had access to decision aid.

Error size was measured as the mean absolute difference from the correct solution across trials 23-30 or 27-30 (factory 3).

Table 9: ANCOVA: Effect of Feedback, Aid Timing, and Confidence on Reliance (Factory 2)

Panel A. Ancova Results (8 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	68938100	3	22979366	2.185	.095
Intercept	1112552111	1	1112552111	105.771	.000
CONFIDENCE	20792962	1	20792962	1.977	.163
FEEDBACK	15249025	1	15249025	1.450	.232
AID INITIAL	49090717	1	49090717	4.667	.033
Error	957185312	91	10518519		
Total	7571785661	95			
Corrected Total	1026123413	94			

Panel B. Descriptive Statistics (8 trials)

	Mean	Std. Deviation	N
Default Aided	9198	3628	32
FEEDBACK	8222	2986	32
AID INITIAL	7455	3126	31
Total	8300	3303	95

Panel C. Ancova Results (4 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	60992488	3	20330829	1.240	.300
Intercept	717102249	1	717102249	43.724	.000
CONFIDENCE	1143246	1	1143246	.070	.792
FEEDBACK	1166940	1	1166940	.071	.790
AID INITIAL	51334329	1	51334329	3.130	.080
Error	1492459533	91	16400654	.	.
Total	7448780023	95			
Corrected Total	1553452021	94			

Panel D. Descriptive Statistics (4 trials)

	Mean	Std. Deviation	N
Default Aided	8559	4413	32
FEEDBACK	8289	4067	32
AID INITIAL	6748	3545	31
Total	7877	4065	95

Definition of Variables:
FEEDBACK: presence of summarized feedback.
INITIAL: decision aid presented initially and visible throughout trial
CONFIDENCE: subject's task confidence

Reliance was measured as the mean absolute difference from the decision aid's recommendation across trials 9-16 or 13-16 (factory 2).

Table 10: ANCOVA: Effect of Feedback, Aid Timing, and Confidence on Error Size (Factory 2)

Panel A. Ancova Results (8 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	49551111	3	16517037	1.763	.160
Intercept	1527174595	1	1527174595	162.995	.000
CONFIDENCE	13380419	1	13380419	1.428	.235
FEEDBACK	17709368	1	17709368	1.890	.173
AID INITIAL	34689762	1	34689762	3.702	.057
Error	852622304	91	9369475	.	.
Total	10612039838	95			
Corrected Total	902173415	94			

Panel B. Descriptive Statistics (8 trials)

	Mean	Std. Deviation	N
Default Aided	10942	3324	32
FEEDBACK	9890	2711	32
AID INITIAL	9476	3138	31
Total	10109	3097	95

Panel C. Ancova Results (4 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	38410412	3	12803470	.887	.451
Intercept	1217877979	1	1217877979	84.327	.000
CONFIDENCE	2907746	1	2907746	.201	.655
FEEDBACK	16698460	1	16698460	1.156	.285
AID INITIAL	33145235	1	33145235	2.295	.133
Error	1314248173	91	14442287	.	.
Total	11474494363	95			
Corrected Total	1352658586	94			

Panel D. Descriptive Statistics (4 trials)

	Mean	Std. Deviation	N
Default Aided	11142	4567	32
FEEDBACK	10120	3522	32
AID INITIAL	9682	3086	31
Total	10322	3793	95

Definition of Variables:
FEEDBACK: presence of summarized feedback.
INITIAL: decision aid presented initially and visible throughout trial
CONFIDENCE: subject's task confidence

Error size was measured as the mean absolute difference from the correct solution across trials 9-16 or 13-16 (factory 2).

Table 11: ANCOVA: Effect of Feedback, Aid Timing, and Confidence on Reliance (Factory 3)

Panel A. Ancova Results (8 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	55295447	3	18431815	1.385	.253
Intercept	2234529631	1	2234529631	167.867	.000
CONFIDENCE	110317	1	110317	.008	.928
FEEDBACK	8871462	1	8871462	.666	.416
AID INITIAL	19575036	1	19575036	1.471	.228
Error	1211328157	91	13311298		
Total	18009704121	95			
Corrected Total	1266623605	94			

Panel B. Descriptive Statistics (8 trials)

	Mean	Std. Deviation	N
Default Aided	13389	3916	32
FEEDBACK	14133	3417	32
AID INITIAL	12272	3530	31
Total	13275	3670	95

Panel C. Ancova Results (4 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	137999098	3	45999699	1.735	.165
Intercept	2517495838	1	2517495838	94.976	.000
CONFIDENCE	46985768	1	46985768	1.773	.186
FEEDBACK	24730729	1	24730729	.933	.337
AID INITIAL	19939702	1	19939702	.752	.388
Error	2412095638	91	26506545	.	.
Total	26060813622	95			
Corrected Total	2550094737	94			

Panel D. Descriptive Statistics (4 trials)

	Mean	Std. Deviation	N
Default Aided	15691	5183	32
FEEDBACK	16934	5411	32
AID INITIAL	14531	4893	31
Total	15731	5208	95

Definition of Variables:
FEEDBACK: presence of summarized feedback.
INITIAL: decision aid presented initially and visible throughout trial
CONFIDENCE: subject's task confidence

Reliance was measured as the mean absolute difference from the decision aid's recommendation across trials 23-30 or 27-30 (factory 3).

Table 12: ANCOVA: Effect of Feedback, Aid Timing, and Confidence on Error Size (Factory 3)

Panel A. Ancova Results (8 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	50734124.570	3	16911374.857	1.412	.244
Intercept	3091562101.391	1	3091562101.391	258.216	.000
CONFIDENCE	1779158.256	1	1779158.256	.149	.701
FEEDBACK	3303306.250	1	3303306.250	.276	.601
AID INITIAL	24908904.418	1	24908904.418	2.080	.153
Error	1089521497.851	91	11972763.713	.	.
Total	24714383834.000	95			
Corrected Total	1140255622.421	94			

Panel B. Descriptive Statistics (8 trials)

	Mean	Std. Deviation	N
Default Aided	15497	3342	32
FEEDBACK	15043	3208	32
AID INITIAL	16748	3767	31
Total	15752	3482	95

Panel C. Ancova Results (4 trials)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	65180271	3	21726757	.846	.472
Intercept	3661419459	1	3661419459	142.574	.000
CONFIDENCE	13930598	1	13930598	.542	.463
FEEDBACK	7347487	1	7347487	.286	.594
AID INITIAL	18781188	1	18781188	.731	.395
Error	2336949633	91	25680765		
Total	26241488421	95			
Corrected Total	2402129904	94			

Panel D. Descriptive Statistics (4 trials)

	Mean	Std. Deviation	N
Default Aided	15706	3702	32
FEEDBACK	15029	5470	32
AID INITIAL	16817	5767	31
Total	15841	5055	95

Definition of Variables:
FEEDBACK: presence of summarized feedback.
INITIAL: decision aid presented initially and visible throughout trial
CONFIDENCE: subject's task confidence

Error size was measured as the mean absolute difference from the correct solution across trials 23-30 or 27-30 (factory 3).

Table 13: Confidence Scores by group

confidence	N	default aided	feedback	aid initial
7	11	3	4	4
6	17	6	6	5
5	20	7	7	6
4	16	6	5	5
3	13	5	3	5
2	13	4	5	4
1	5	1	2	2
	95	32	32	31
	mean	4.38	4.38	4.29

Table 14: Group Data summarized by Period

last 8 trials	factory 2 reliance	factory 3 reliance	factory 2 performance	factory 3 performance
1 unaided	15814	19886	14442	13909
2 aided	9198	13389	10943	15498
3 aided: feedback	8222	14134	9891	15043
4 aided: aid initial	7455	12272	9476	16749

last 4 trials	factory 2 reliance	factory 3 reliance	factory 2 performance	factory 3 performance
1 unaided	15661	22829	12946	14015
2 aided	8560	15691	11143	15707
3 aided: feedback	8290	16934	10121	15029
4 aided: aid initial	6748	14531	9683	16818

Table 15: Group Data Within each Period

	factory 2 reliance correlation	2nd half / 1st half	factory 2 performance correlation	2nd half / 1st half
unaided	-	-	-0.09	1.17
aided	-0.15	1.25	-0.08	1.11
feedback	-0.13	1.27	-0.12	1.22
aid initial	-0.10	1.30	-0.09	1.13

	factory 3 reliance correlation	2nd half / 1st half	factory 3 performance correlation	2nd half / 1st half
unaided	-	-	-0.06	1.07
aided	0.20	0.75	-0.06	1.07
feedback	0.19	0.82	-0.10	1.13
aid initial	0.18	0.74	-0.04	1.02

Explanation of Terms:
correlation: the correlation between the average value and trial number (7-16 or 17-30).
2nd half / 1st half: the average value for the 2nd half of trials relative to the 1st half of trials.

REFERENCES

Altman, E.I. and T.P. McGough. 1974. Evaluation of a company as a going concern. *The Journal of Accountancy* (December): 50-57.

Anderson, B.H. and M.J. Maletta. 1999. Primacy effects and the role of risk in auditor belief-revision processes. Auditing (18): 75-89.

Arkes, H.R., R.M. Dawes, and C.Christiansen. 1986. Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes* 37: 93-110.

Arnold, V. and S.G. Sutton. 1998. The theory of technology dominance: understanding the impact of intelligent decision aids on decision makers' judgments. *Advances in Accounting Behaviora Research*, 1: 175-194

Ashton, R. 1990. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research* 28 (Supplement): 148-180.

Becker, D.A. 1997. The effects of choice on auditors' intrinsic motivation and performance. *BehavioralResearch in Accounting* (9): 1-19.

Bell, T. B., and J. V. Carcello. 2000. Development of a Decision Aid for Assessing the Likelihood of Fraudulent Financial Reporting. *Auditing: A Journal of Practice & Theory* (Spring): 167-182.

Biggs, S. and J. Wild. 1985. An investigation of auditor judgment in analytical review. *The Accounting Review* (October): 607-633.

Blattberg, R.C., and S.J. Hoch. 1990. Database models and managerial intuition: 50% model + 50% manager. *Management Science* 36 (August): 887-899.

Boatsman, J.R., C. Moeckel, and B.K.W. Pei. 1997. The effects of decision consequences on auditors' reliance on decision aids in audit planning. *Organizational Behavior and Human Decision Processes* (August): 211-247.

Bromiley, P. 1987. Do forecasts produced by organizations reflect anchoring and adjustment ? *Journal of Forecasting*, 6: 201-210.

Brown, D. and M.M. Eining. 1997. Judgment and decision making, Part IV: Information technology and decision aids. In V.Arnold and S.G.Sutton (Eds.), Behavioral Accounting Research: Foundations and Frontiers (pp.164-187). Sarasota, FL: American Accounting Association.

Brown, D.L., and D.R. Jones. 1998. Factors that influence reliance on decision aids: A model and an experiment. *Journal of Information Systems* (Fall): 75-94.

Butler, S. 1986. Anchoring in the judgmental evaluation of audit samples. *The Accounting Review* (January): p.101

Chu, P.C. and E.E. Spires. 2000. The joint effects of effort and quality on decision strategy choice with computerized decision aids. *Decision Sciences* 31 (2): 259-292.

Davis, E.B. 1998. Decision Aids for Going-Concern Evaluation: Expectations of Partial Reliance. *Advances in Accounting Behavioral Research* 1: 33-59.

Davis, F.D. and J.E. Kotteman. 1995. Determinants of decision rule use in a production planning task. *Organizational Behavior and Human Decision Processes* 63(2): 145-157..

Dawes, R.M. 1979. The robust beauty of improper linear models in decision making. *American Psychologist* 34: 571-582.

Deci, E.L. and R.M. Ryan. 1985. *Intrinsic Motivation and Self Determination in Human Behavior*. New York: Plenum Press.

Dzindolet, M.T., H.P. Beck and L.G. Pierce. 2000. Encouraging human operators to appropriately rely on automated decision aids. Working paper, Cameron University.

Einhorn, H.J. 1972. Expert measurement and mechanical combination. *Organizational Behavior and Human Performance* 7: 86-106.

Einhorn, H.J. 1974. Cue definition and residual judgment. *Organizational Behavior and Human Performance* 12: 30-49.

Einhorn, H.J. and R.M. Hogarth. 1978, Confidence in judgment: Persistence of the illusion of validity. *Psychological Review* (85): 395-416.

Glover, S., D. Prawitt, and B. Spilker. 1997. Decision aids and user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes* 72 (2): 232-255.

Gorry, G.A. and Scott-Morton, M. 1971. A framework for management information systems. Sloan Management Review (13): pp.55-70.

Hendrick, C. and A.F. Constantini. 1970. Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. *Journal of Personality and Social Psychology*: 58-164.

Hopwood, W., J.C. McKeown and J.F. Mutchler. A reexamination of auditor versus model accuracy within the context of the going concern opinion decision. *Contemporary Accounting Research* (Spring): 409-431.

Joyce, E. and G. Biddle. 1981. Anchoring and adjustment in probabilistic inference in auditing. Journal of Accounting Research (spring): p. 120

Kaplan, S.E., J.H. Reneau and S.Whitecotton. 2001. The effects of predictive ability information, locus of control, and decision maker involvement on decision aid reliance. *Journal of Behavioral Decision Making* (14): 35-50.

Kachelmeier, S. and W.F.Messier. 1990. An investigation of the influence of a non-statistical decision aid on auditor sample size decisions. *The Accounting Review* 65: 209-226.

Kowalczyk, T.K., and C.J. Wolfe. 1998. Anchoring effects associated with recommendations from expert decision aids: An experimental analysis. *Behavioral Research in Accounting* 10 (Supplement): 147-169

Libby, R., and P.A. Libby. 1989. Expert measurement and mechanical combination in control reliance decisions. *The Accounting Review* (October): 729-747.

Lichtenstein, S., and P. Slovic. 1971. Reversal of preferences between bids and choices in gambling decisions. Journal of Experimental Psychology, 89: 46-55.

Lim, J.S., and M. O'Connor. 1996. Judgmental forecasting with interactive forecasting support systems. Decision Support Systems 16: 339-357.

Meehl, P.E. 1957. When shall we use our heads instead of the formula ? *Journal of Counseling Psychology*4: 268-273.

McClish, D.K. and S.H. Powell. 1989. How well can physicians estimate mortality in a medical intensive care unit ? *Medical Decision Making*, 9: 125-132

Northcraft , G.B. and M.A. Neale. 1987. Experts, amateurs, and real estate: an anchoring and adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, 39: 84-97.

O'Leary, D., and T. Lin. 1989. An expert system for cash flow analysis. In *Artificial Intelligence in Accounting and Auditing*, edited by M. Vasarhelyi, (New York: Markus Weiner Publishing, Inc.)

Payne, J.W. 1982. Contingent Decision Behavior. *Psychological Bulletin*, 92: 382-402.

Pennington, R. 2002. The effects of information overload on information processing and judgment in an information systems audit task. Ph.D. dissertation. University of South Carolina.

Peterson, D.K. and G.F.Pitz. 1986. Effect of input from a mechanical model on clinical judgment. *Journal of Applied Psychology*, 71: 163-167.

Pincus, K.V. 1991. Audit judgment confidence. *Behavioral Research in Accounting* 3:39-65.

Powell, J.L. 1991. An attempt at increasing decision rule use in a judgment task. *Organizational Behavior and Human Decision Processes* 48: 89-99.

Prawitt, D. 1998. Commentary on: Anchoring effects associated with recommendations from expert decision aids: An experimental analysis. *Behavioral Research in Accounting* 10 (Supplement): 170-181.

Remus, W. and J.Kottemann. 1995. Anchor-and-adjustment behaviour in a dynamic decision environment. *Decision Support Systems*, 15: 63-74.

Shpilberg, D., L. Graham, and H. Shatz. 1986. ExperTAX^{sm}: An expert system for corporate tax planning. *Expert Systems* (July).

Todd, P., and I. Benbasat. 1991. The impact of computer based decision aids on the decision making process. *Information Systems Research* 2: 87-115.

Todd, P., and I. Benbasat. 1992. An experimental investigation of the impact of computer based decision aids on processing effort. *MIS Quarterly* 16: 373-393.

Todd, P., and I. Benbasat. 1994. The influence of decision aids on choice strategies: An empirical analysis of the role of cognitive effort. *Organizational Behavior and Human Decision Processes* 60: 36-74.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* (September): 1124-1131.

Waller, W.S. and W.L. Felix. 1984. The effects of incomplete outcome feedback on auditors' self-perceptions of judgment ability. *The Accounting Review* 59:4 (October): 637-46.

Wedell, D.H. and S.M. Senter. 1997. Looking and weighting in judgment and choice. *Organizational Behavior and Human Decision Processes, 70: 41-64.*

Whitecotton, S.M. 1996. The effects of experience and confidence on decision aid reliance: A causal model. *Behavioral Research in Accounting* 8: 194-216.

Whitecotton, S.M. and S.A.Butler. 1998. Influencing decision aid reliance through involvement in information choice. Behavioral Research in Accounting 10(supplement): 182-200.

Whyte, G. and J. Sebenius. 1997. The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Processes,* 69(1): 75-85.

APPENDIX A: EXPERIMENTAL MATERIALS

.

EXPERIMENTAL SCRIPT: CLASSROOM PRESENTATION

Hello.

My name is John Briggs, and I'm a graduate student in the department of Accounting and Information Systems. As your instructor has told you, I am conducting a research project and am here to ask you if you would like to participate.

Participation in this study will be worth (x) points extra credit in this course. Also, you will have an opportunity to earn money based upon your performance in the study. So there are two personal benefits from participation.

The purpose of the research is to study decision-making behavior in a complex environment. You will act as a managerial accountant who has the responsibility of predicting manufacturing overhead costs on a month-to-month basis.

As you are aware, manufacturing overhead is the group of indirect costs associated with production operations. Companies usually apply these indirect costs to products based on some directly measurable characteristic. Your goal will not be to apply overhead to products, but instead to predict the total amount of overhead costs for each month for a factory. This prediction will be based on various pieces of information about monthly operations, such as the number of machine hours incurred for the month.

The experiment involves a computer program written in Visual Basic. A requirement is that you have a PC. You will e-mail me to request the program. Once I hear from you, I will e-mail you the program. The file is an .exe file, and so you just need to click on the file, and it will run automatically. If you do not have access to a PC, you can contact me and we will make other arrangements.

I won't go into too much detail today, because the program itself contains detailed instructions. If you read the instructions slowly and carefully, you should not have any problems. There is also a practice session before the actual experiment begins. In addition, if you feel you do not understand the instructions sufficiently, there is a place to quit and start over. You can contact me if you do this, and would like additional explanation.

When you are finished with the experiment, a file will be written to the C:\ root drive of your computer. The program will tell you the name of this file, and you will e-mail the file back to me.

To earn the course credit offered, you must complete the experiment and demonstrate that you put forth a reasonable level of effort.

Here's how you can earn money. Just for completing the experiment, you will be paid $5. However, the top 20% of performers will be paid $20 instead of $5. Your performance will be judged based on how close you are in your monthly estimates. But specifically, your score will be based on your MEDIAN error size. So, if it takes you a while to make good monthly estimates, or you happen to have very large errors for a couple of months, you can still be a top performer, because it is your final MEDIAN monthly error size that will be evaluated.

Once you finish with the monthly estimates, there will be a few questions for you to answer. All of your responses will be completely confidential, and never shared with anyone else.

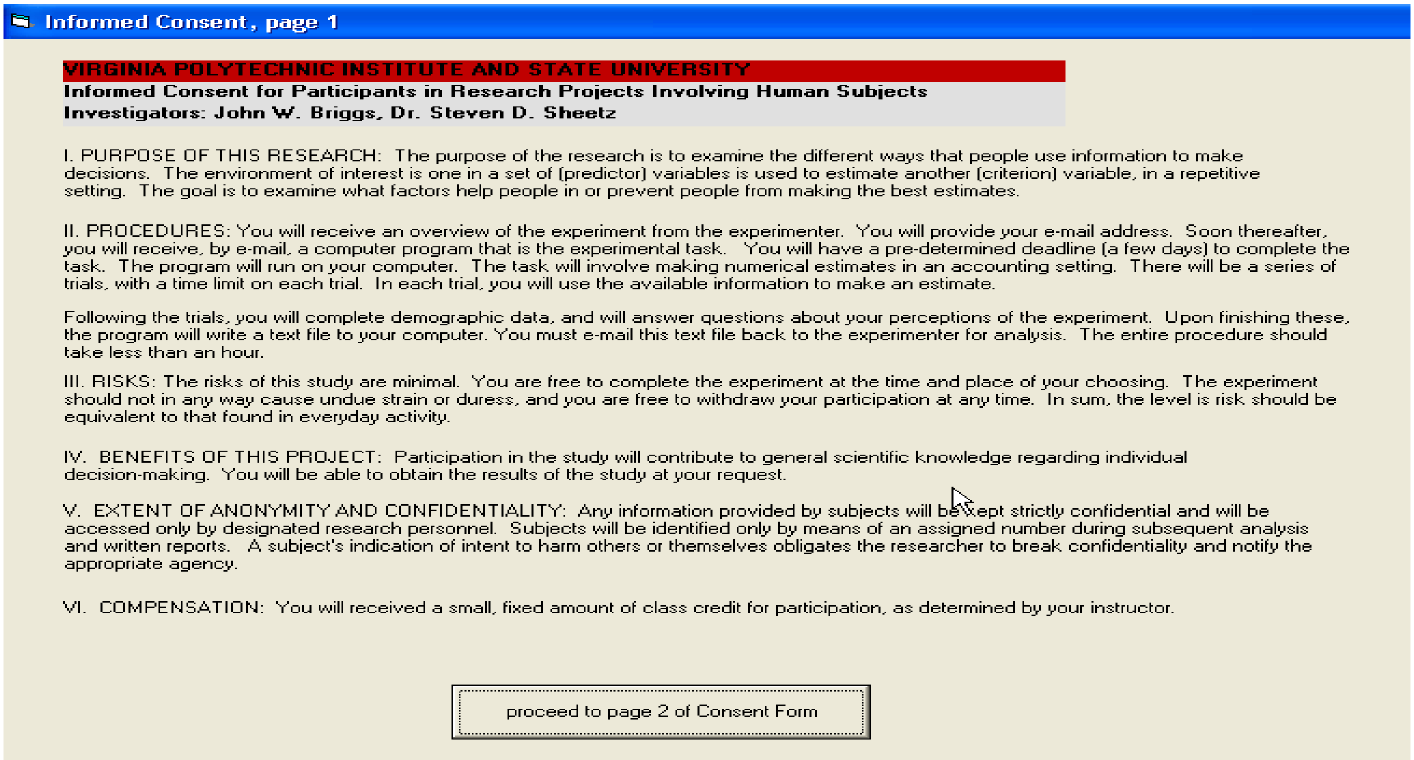
I will return to your class to give you a brief explanation of the purpose of the experiment. If you would like a detailed explanation or an overview of the results, you are free to contact me.

You may not discuss this experiment with anyone else. Also, do not receive the experiment from another student. You must acquire the program directly from me.

This experiment has been approved by the institutional review board for research involving human subjects.

My e-mail address is __________ when you are ready to receive the computer program.
I look forward to hearing from you.

Figure A1: Informed Consent



Informed Consent, page 1

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
Informed Consent for Participants in Research Projects Involving Human Subjects
Investigators: John W. Briggs, Dr. Steven D. Sheetz

I. PURPOSE OF THIS RESEARCH: The purpose of the research is to examine the different ways that people use information to make decisions. The environment of interest is one in a set of (predictor) variables is used to estimate another (criterion) variable, in a repetitive setting. The goal is to examine what factors help people in or prevent people from making the best estimates.

II. PROCEDURES: You will receive an overview of the experiment from the experimenter. You will provide your e-mail address. Soon thereafter, you will receive, by e-mail, a computer program that is the experimental task. You will have a pre-determined deadline (a few days) to complete the task. The program will run on your computer. The task will involve making numerical estimates in an accounting setting. There will be a series of trials, with a time limit on each trial. In each trial, you will use the available information to make an estimate.

Following the trials, you will complete demographic data, and will answer questions about your perceptions of the experiment. Upon finishing these, the program will write a text file to your computer. You must e-mail this text file back to the experimenter for analysis. The entire procedure should take less than an hour.

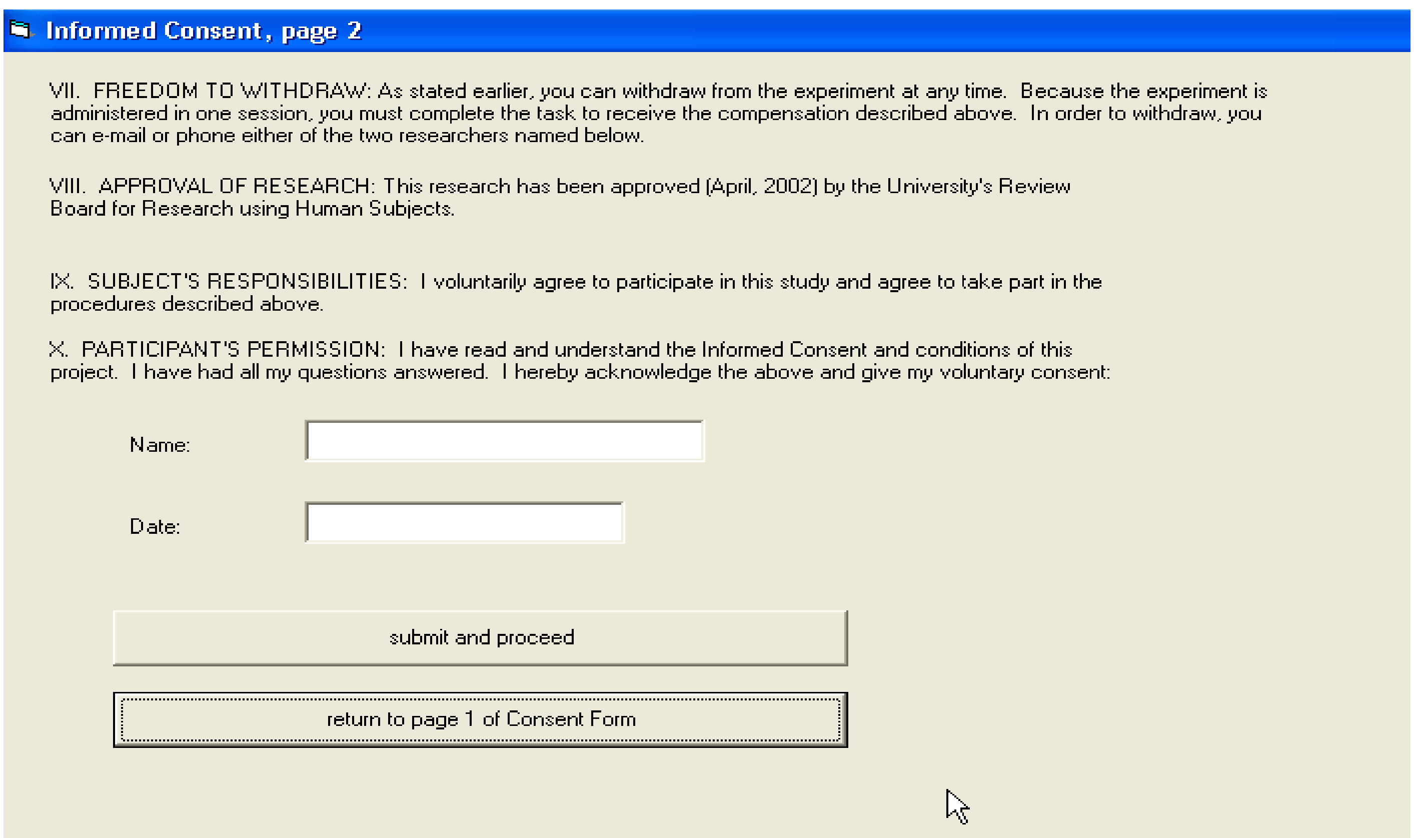
III. RISKS: The risks of this study are minimal. You are free to complete the experiment at the time and place of your choosing. The experiment should not in any way cause undue strain or duress, and you are free to withdraw your participation at any time. In sum, the level is risk should be equivalent to that found in everyday activity.

IV. BENEFITS OF THIS PROJECT: Participation in the study will contribute to general scientific knowledge regarding individual decision-making. You will be able to obtain the results of the study at your request.

V. EXTENT OF ANONYMITY AND CONFIDENTIALITY: Any information provided by subjects will be kept strictly confidential and will be accessed only by designated research personnel. Subjects will be identified only by means of an assigned number during subsequent analysis and written reports. A subject's indication of intent to harm others or themselves obligates the researcher to break confidentiality and notify the appropriate agency.

VI. COMPENSATION: You will received a small, fixed amount of class credit for participation, as determined by your instructor.

proceed to page 2 of Consent Form



Informed Consent, page 2

VII. FREEDOM TO WITHDRAW: As stated earlier, you can withdraw from the experiment at any time. Because the experiment is administered in one session, you must complete the task to receive the compensation described above. In order to withdraw, you can e-mail or phone either of the two researchers named below.

VIII. APPROVAL OF RESEARCH: This research has been approved (April, 2002) by the University's Review Board for Research using Human Subjects.

IX. SUBJECT'S RESPONSIBILITIES: I voluntarily agree to participate in this study and agree to take part in the procedures described above.

X. PARTICIPANT'S PERMISSION: I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

Name:

Date:

submit and proceed

return to page 1 of Consent Form

FIGURE A2: Introduction 1

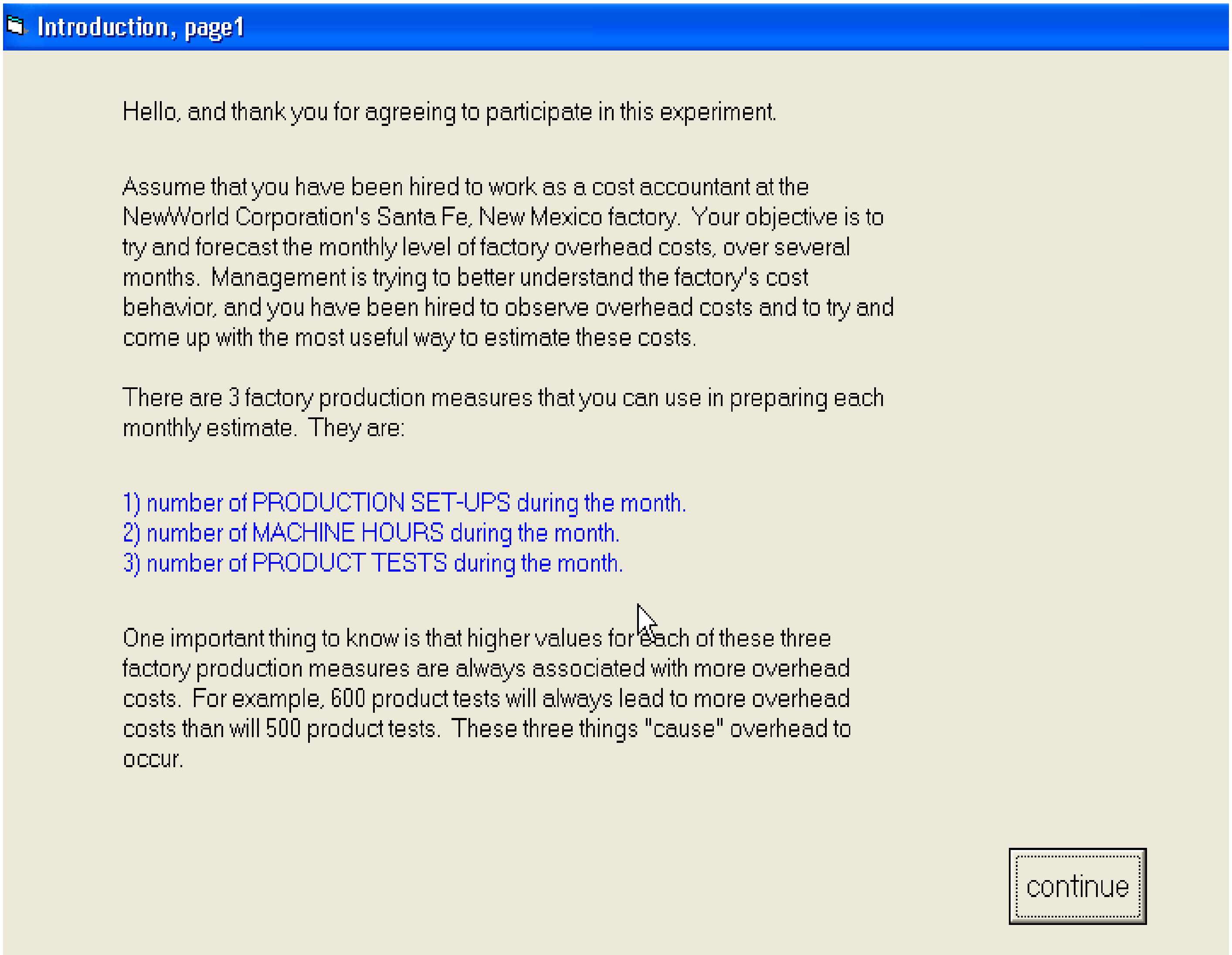


FIGURE A3: Introduction 2

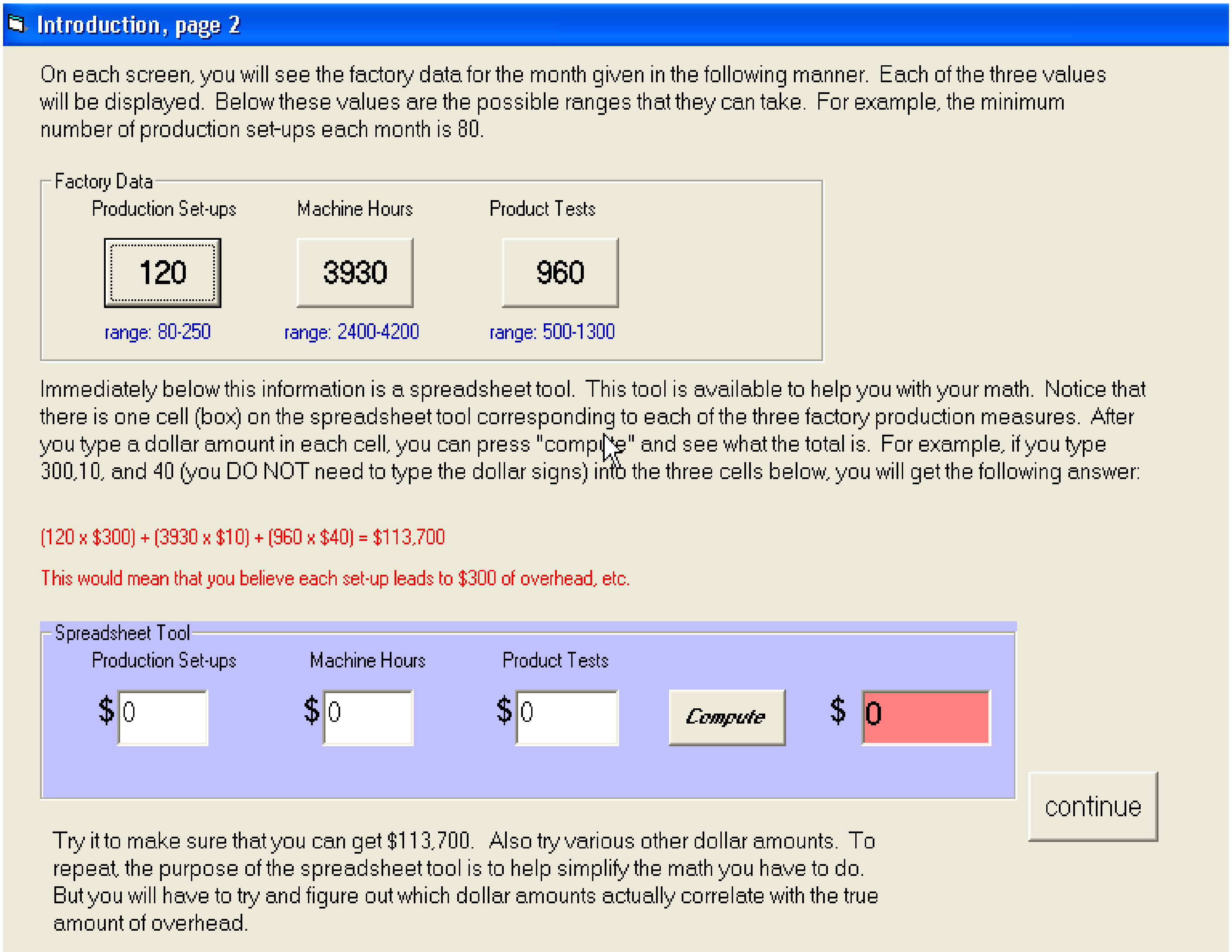


FIGURE A4: Introduction 3

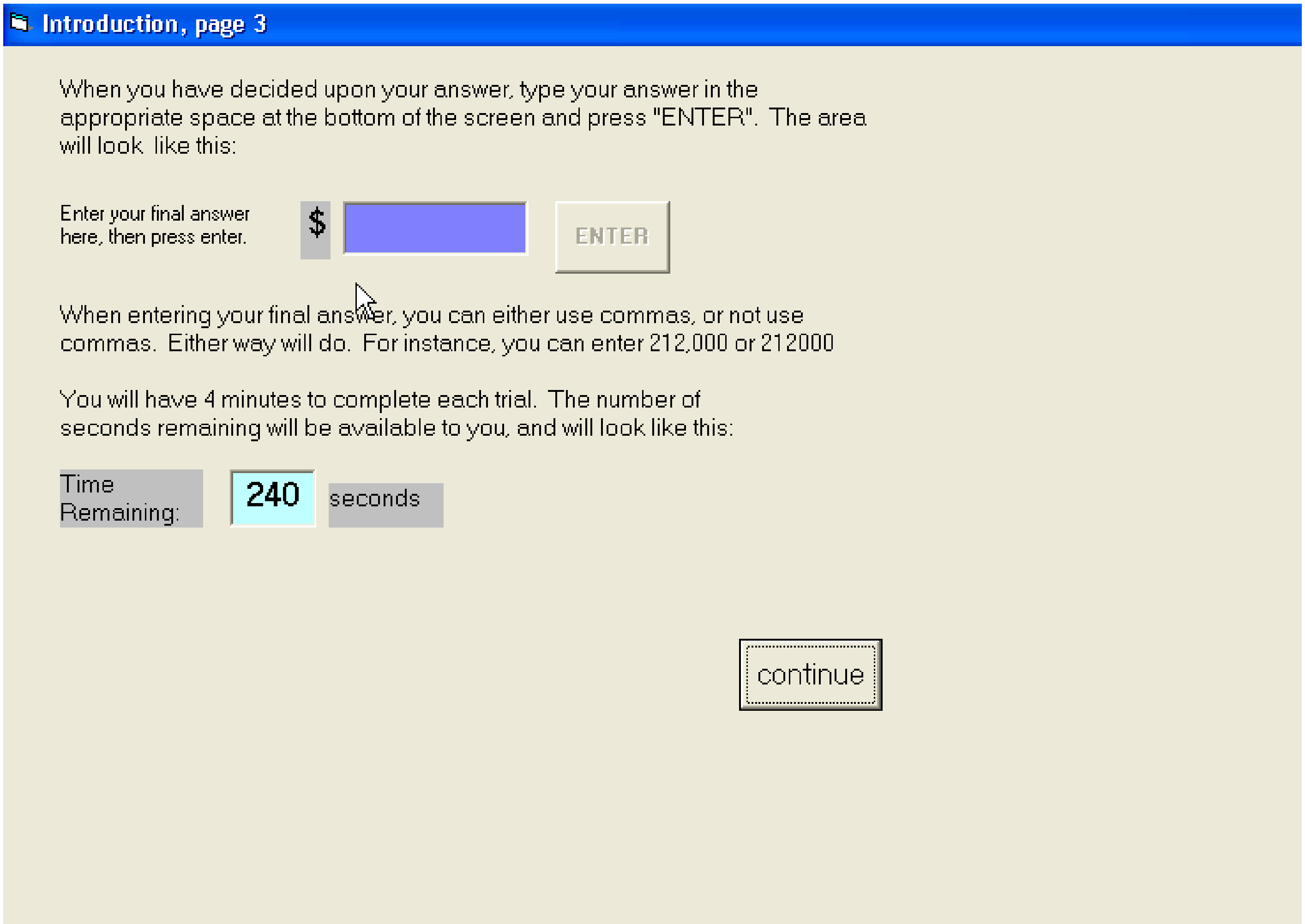


FIGURE A5: Introduction 4

Introduction, page 4

Now, here's some very important information. Your goal is to predict factory overhead costs as accurately as possible. Obviously, your first few tries will involve some guesswork, but with practice, you will improve. Strive for improvement; your goal should be to make the most accurate estimates that you can. You will be told how accurate you are after each month. Learn from experience.

Please do not use a calculator, and do not use pencil and paper to assist you. Only use the spreadsheet tool to help you with your computations.

So your objective is to predict the amount of overhead each month, and the way to go about achieving this objective is to try and uncover the relationship between each of 3 factory measures and overhead. There is an underlying relationship (each variable has a consistent weight) which you can probably get closer to as the experiment progresses.

The true values of the weight (in dollars) of each of the three factory variables fall between these ranges:

SET-UPS: the amount of overhead resulting from one machine set-up is between $300 and $700
MACHINE HOURS: the amount of overhead resulting from one machine hour is between $10 and $40
PRODUCT TESTS: the amount of overhead resulting from one test is between $40 and $100

Now, let's try some practice sessions to make sure you get the hang of it. Once you get to the next screen, press "BEGIN" to start.

continue

FIGURE A6: Introduction 5 (for group 3)

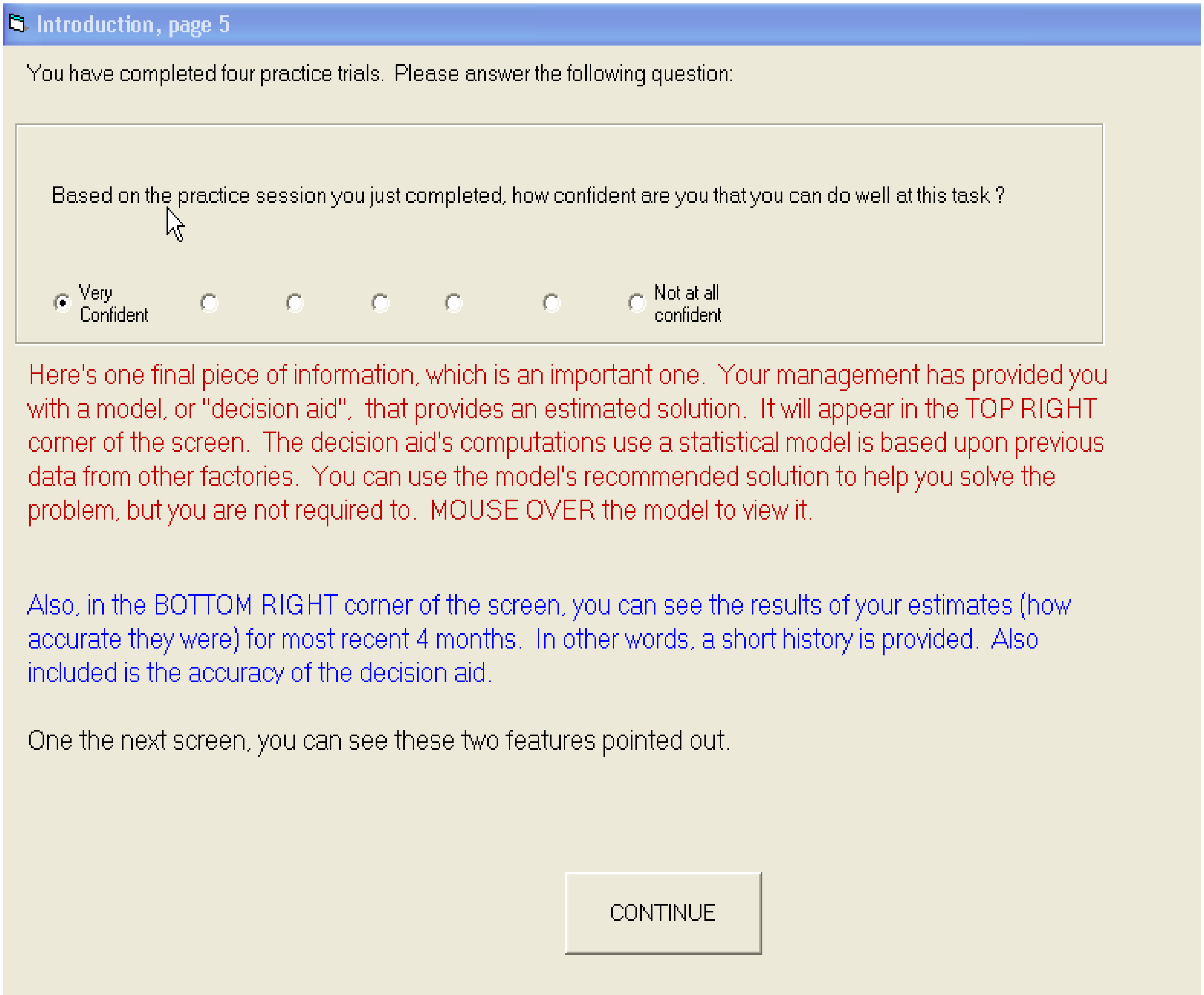


FIGURE A7: Introduction 6 (for group 3)

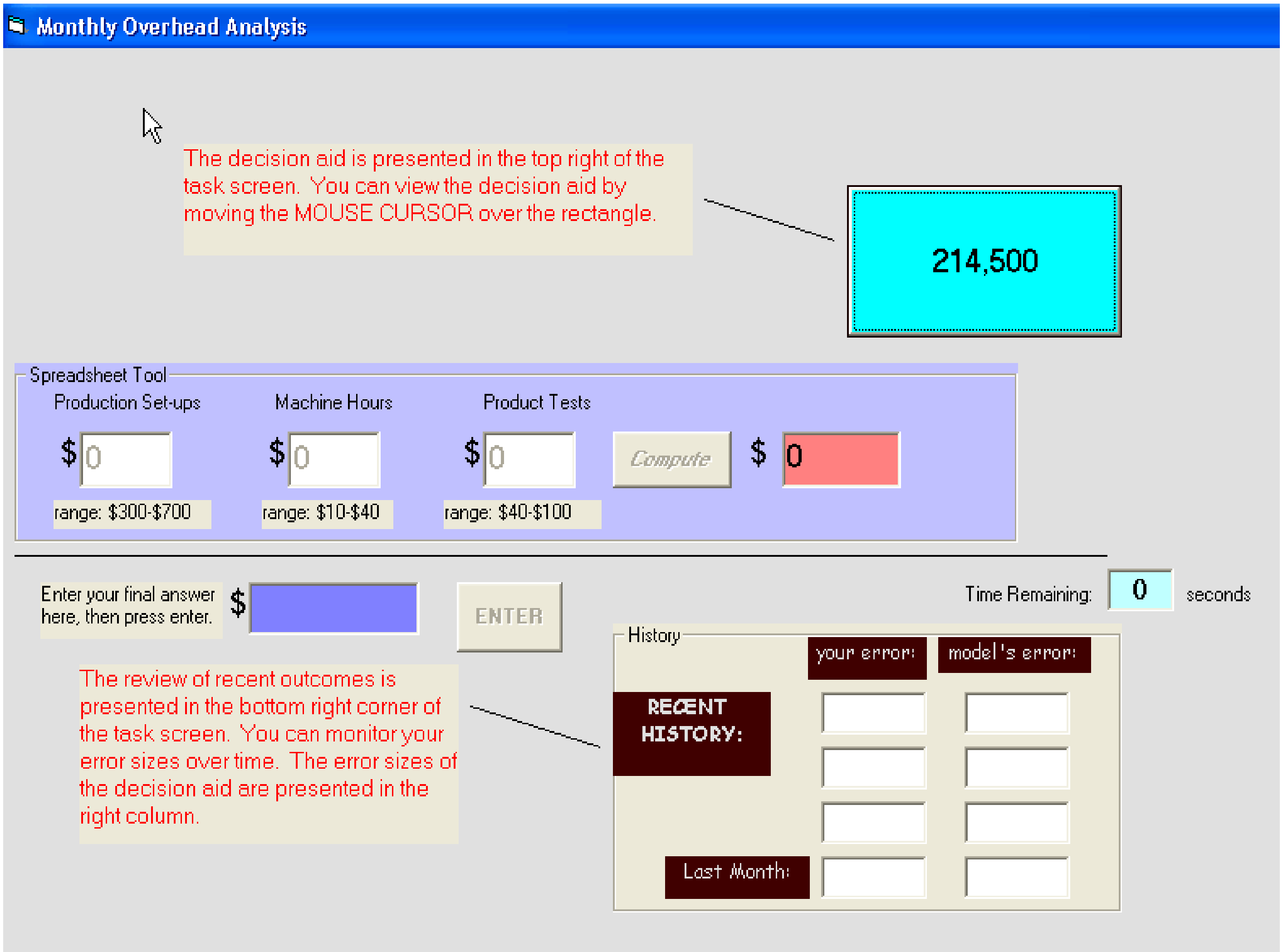


FIGURE A8: Unaided Group (group 1) Task Screen

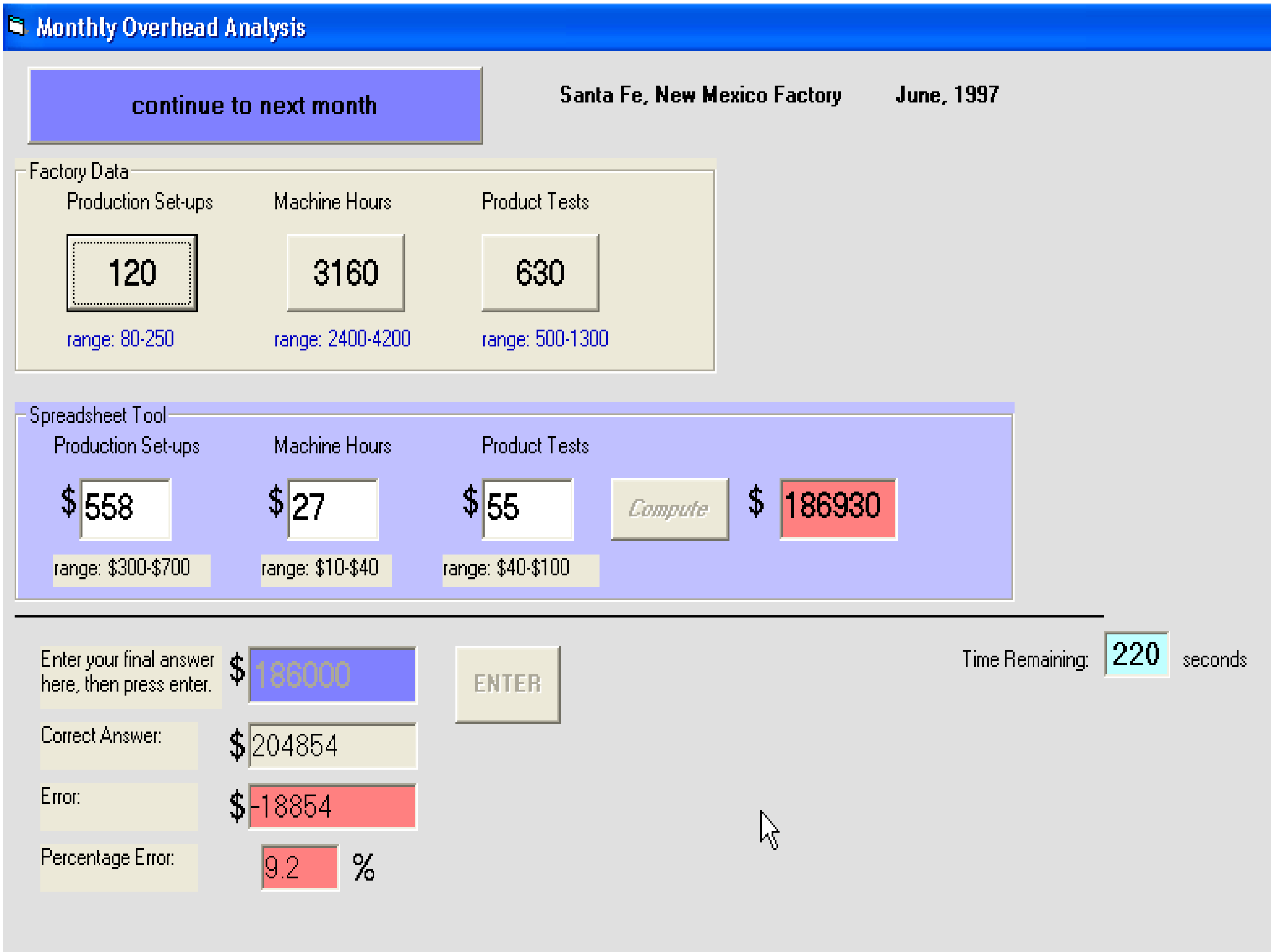


FIGURE A9: Default Aided Group (group 2) Task Screen

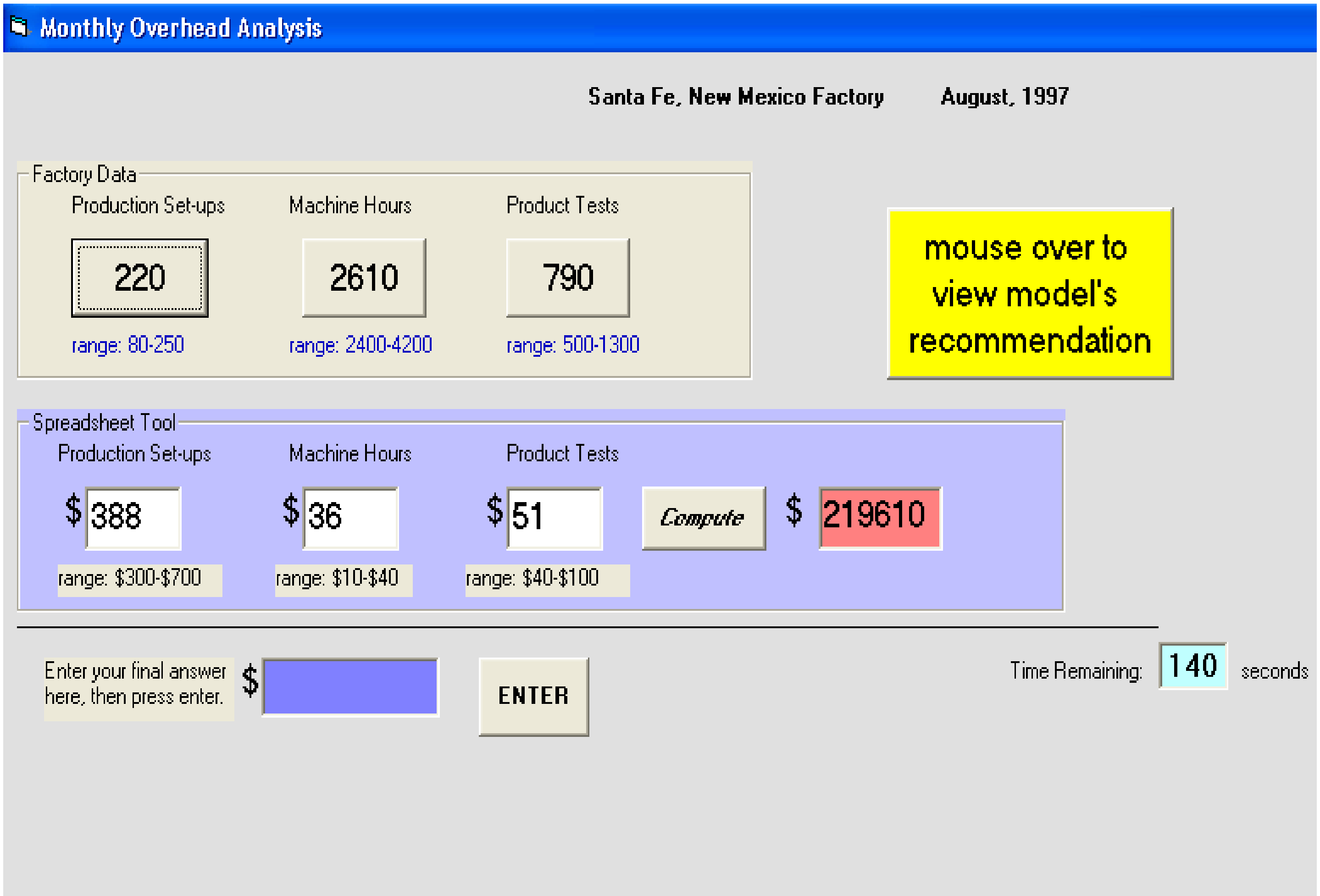


FIGURE A10: High Feedback Group (Group 3) Task Screen

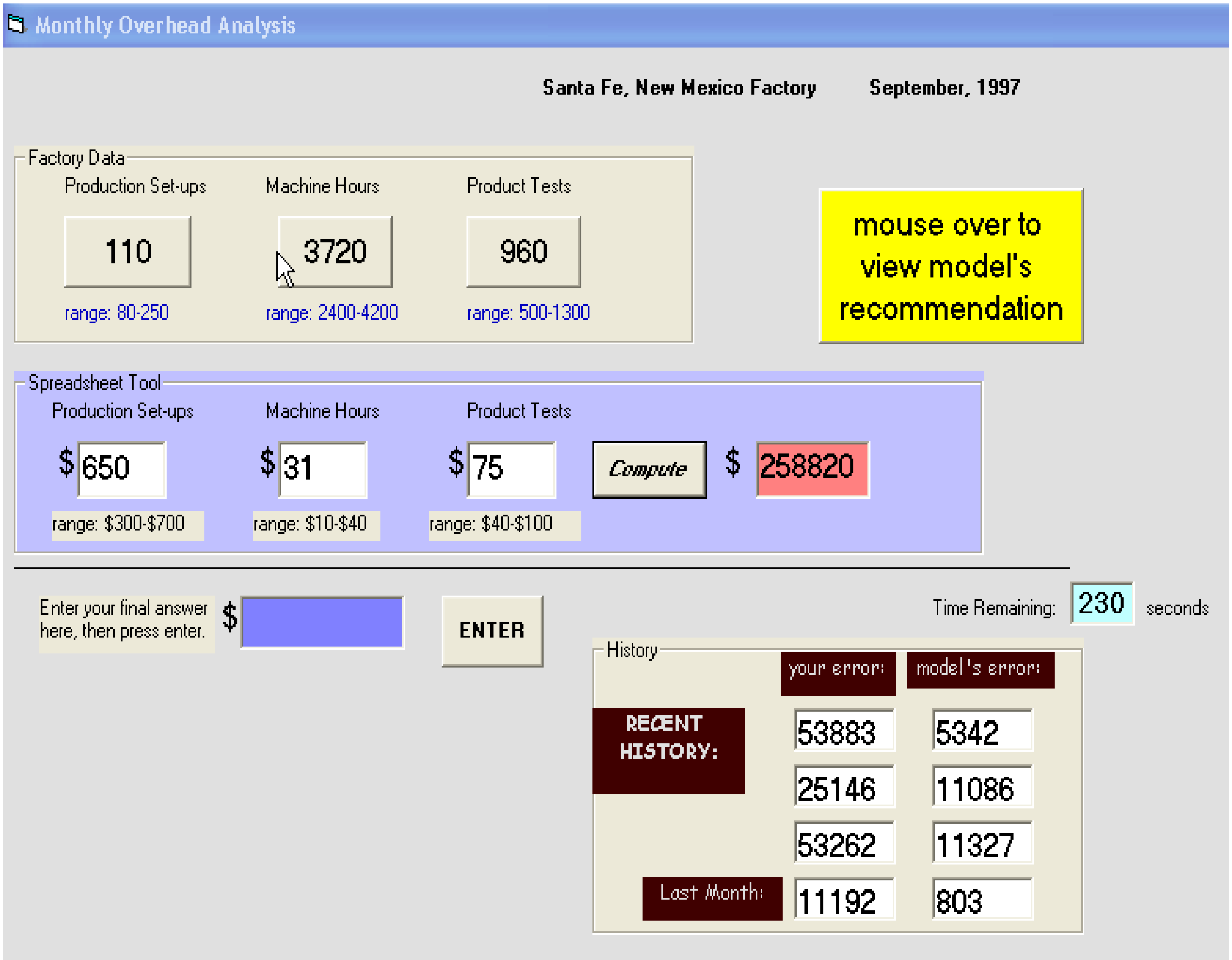
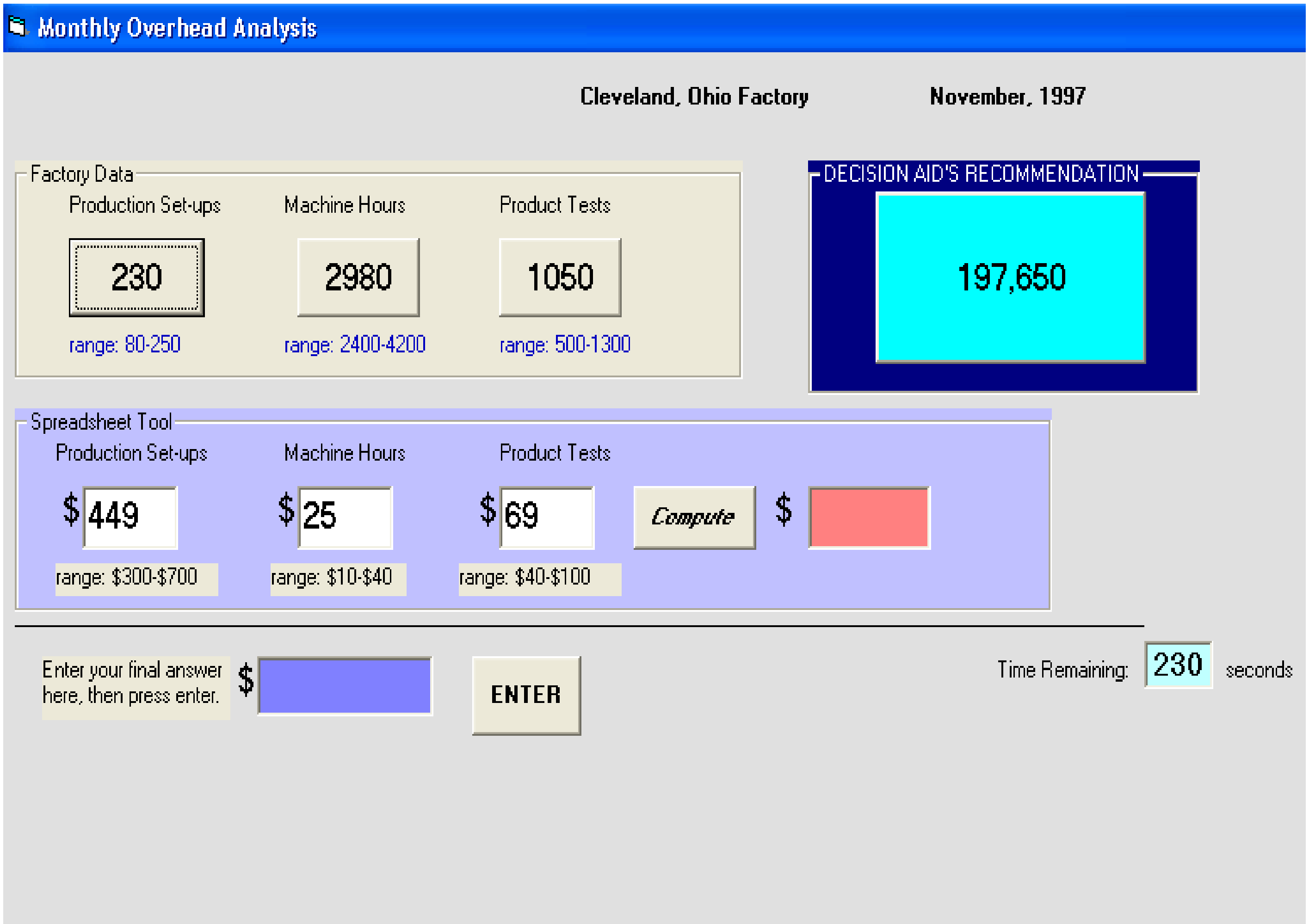


FIGURE A11: Aid Initial (group 4) Task Screen



Vita

John W. Briggs grew up in Roanoke, Virginia, where he attended Cave Spring High School. He earned a Bachelor of Science degree in Biology from the College of William and Mary. He received a Master of Accountancy degree from Virginia Polytechnic Institute and State University. He is currently an Assistant Professor in the accounting department at James Madison University.