

**THE EFFECTS OF DECISION AID RELIABILITY AND
PRESSURES TO PERFORM ON
DECISION AID RELIANCE**

Mohamed I. Gomaa
Suffolk University

James E. Hunton
Bentley College / Universiteit Maastricht

Eddy H. J. Vaassen
Universiteit Maastricht

Martin A. Carree
Universiteit Maastricht

January 2008

Draft. Please do not quote without permission of the authors.

THE EFFECTS OF DECISION AID RELIABILITY AND PRESSURES TO PERFORM ON DECISION AID RELIANCE

Abstract

We investigate the effects of decision aid reliability and pressure to perform on decision aid reliance. A total of 403 students took part in a four (pressures to perform: one through four) by five (decision aid reliability: 50%, 60%, 70%, 80% and 90%) between-participants experiment. We compared, contrasted and tested three competing models of decision aid reliance behavior: rationality, pressure-induced rationality and pressure arousal theory. We find that pressure arousal theory predominantly explains decision aid reliance behavior at all but the highest level of decision aid reliability tested in this study (90%). Our results indicate that there are reliance peaks across the reliability levels as more pressure to perform is applied, and that continually increasing pressures can eventually lead to decreased, rather than increased, reliance.

Key Words: decision aid reliance, decision aid reliability, pressure to perform

I. INTRODUCTION

Decision aids appear in many forms and levels of sophistication, including expert systems, neural networks and decision support systems (Dungan 1983; Steinbart 1987; Vasarhelyi & Harper, 1990), and less advanced tools such as algorithms, procedures, decision rules, and standard forms (Brown 1962; Mock & Turner 1981; Kachelmeier & Messier 1990; Powell 1991; McDaniel & Kinney 1995). It is generally argued that proper reliance on sufficiently reliable decision aids can improve decision accuracy (Benbasat and Nault 1990; Kleinmuntz 1990; Ashton 1992).

At the one end of the decision aid reliance continuum, decision makers will totally rely on the advice of the decision aid, and at the other end, they will completely ignore the aid's advice and instead rely on their own judgment. Prior research indicates that users generally under-rely on decision aids (i.e., they overweigh their judgment relative to the decision aids' advice), although reliance findings are inconsistent across studies (Ashton 1990, Powell 1991, Ashton 1992; Brown & Jones 1998). It is difficult to unravel precisely why under-reliance and contradictory reliance results occur, as there are numerous influential decision aid factors to consider and many research idiosyncrasies that complicate cross-study comparisons (Rose 2002).

Brown and Jones (1998) offer a model of decision aid reliance that includes decision aid features, strategy evaluation factors, decision-maker characteristics and decision task factors. In the current study, we examine the singular and joint impacts of one decision aid feature (decision aid reliability) and one strategy evaluation factor (pressures to perform) on decision aid reliance, as these two variables have been found to be influential over a wide array of decision tasks and across a diverse set of decision makers (Rose 2002). By experimentally focusing on decision aid reliability and pressures to perform, we can rule out other potential confounds, thus better understand the effects of these two key variables on decision aid reliance.

In a between-participants experiment, we manipulated decision aid reliability at five levels (50%, 60%, 70%, 80%, and 90%) and number of pressures to perform at four levels (one through four). All participants were exposed to the decision aid's recommendation, which represented a constant pressure across all conditions.¹ For participants in conditions with two, three and four pressures, we randomized litigation risk, time pressure, and decision justification. Participants first read about a situation where they were required to estimate an allowance for uncollectible accounts for an audit client without the advice of a decision aid. Next, they were exposed to the decision aid's advice (estimate), after which they could revise their initial estimate if they chose. The dependent variable, decision aid reliance, reflects the extent to which the participants' revised estimate moved from their initial anchor toward the aid's advice.

A total of 403 undergraduate and graduate business students participated in the study. Results indicate that at 50%, 60% and 70% reliability levels, decision aid reliance first rises as pressure to perform increases, reaches a maximum point, and then decreases as more pressure is exerted (inverted U-shape). At the 80% reliability level, reliance increases through three pressures and then reaches an inflection point, after which reliance increases at a diminishing rate. Interestingly, at each of the reliability levels 50% through 80%, the level of reliance peaks at less than the decision aid's reliability level, suggesting continued under-reliance in violation of economic rationality. At the highest reliability level (90%), reliance is high, stable and at least equivalent to reliability across the four pressures.

We tested three models to describe reliance on a decision aid: a rationality model where reliance is equal to reliability across all pressure levels; a pressure induced rationality model where reliance continually increases along with more pressure; and a quadratic model where

¹ Heiman (1990), in a discussion of Ashton (1990), argues that the mere availability of a decision aid can be considered one form of pressure to use and rely on the aid.

reliance increases, reaches a maximum and then decreases as pressure intensifies. We furthermore introduce a general model of reliance on a decision aid. The quadratic model is the best fitting model of decision aid reliance behavior, which is consistent with pressure arousal theory.

The remainder of this paper is organized as follows. Section II reviews prior research in the area of decision aid use and develops the hypotheses. Section III provides an overview of our research method. Section IV presents study results and develops a general model of decision aid reliance under various reliability and pressure conditions. The final section discusses the implications of our findings.

II. BACKGROUND AND HYPOTHESES

Decision aids are developed to help decision-makers enhance decision-making processes and judgments (Kleinmuntz 1990; Rose 2002). Existing decision aid literature documents that decision aids not only improve the accuracy of decisions, but also enhance the efficiency of decision making processes (Kumar 1997; Smith et al. 1998; Chu and Spires 2003; and Arnold et al. 2000). Despite potential improvements afforded by the use of decision aids, prior research has demonstrated that decision aid users, particularly experienced professionals, typically hold higher confidence in their own judgments relative to decision aids' recommendations, which typically leads them to discount and even ignore the advice of reliable decisions aids (e.g., Messier 1995; Sutton et al. 1995). This results in situations where the potential benefits of decision aids often go unrealized (Rose 2002).

One objective of the current research is to compare and contrast three competing perspectives of decision aid reliance: rational behavior, pressure-induced rational behavior and

pressure-arousal theory (Ashton 1990). We also attempt to develop a general model of decision aid reliance. We begin with rational behavior.

Model 1: Rational behavior

Economic rationality would suggest that perceived decision aid reliability and decision aid reliance should be equal, *ceteris paribus*. In the case where the probability of the decision aid being correct is p , and the subject's initial judgment and decision aid outcomes are x and y , respectively, it is optimal to choose the expected value $(1-p)x+py$, as this minimizes expected, absolute or quadratic error in judgement. Thus, the level of decision aid reliance should be equal to the aid's level of reliability, which can be expressed as: $([(1-p)x+py]-x) / (y-x) = p$.

In one study (Arkes et al. 1987), participants were provided with a decision aid that was reported to be 70% accurate and asked to make 40 predictions. If the participants strictly followed the decision aid's rule, they would be expected to make 28 correct predictions (70%). While decision aid reliability was held constant at 70%, other confounding treatments were manipulated such that only the control group is useful for analyzing rational behavior. Control group participants, who were only exposed to the decision aid's reported reliability, recorded a mean of around 28 correct predictions, which is the outcome of rational behavior. Hence, we begin our study of decision aid reliance by positing rational behavior:

H1: The level of reliance on a decision aid will equal the level of the decision aid's reliability, *ceteris paribus*.

Model 2: Pressure-induced rational behavior

While the notion of rational behavior provides an economic explanation for reliance on a decision aid, extant literature typically does not support rationality. For instance, Powell (1991) conducted the only study of which we are aware where decision aid reliability was manipulated

with no other confounding treatments. In his study, he examined the effect of varying a decision rule's reported accuracy from 50% to 70% to 90% on utilization of the decision rule. His results showed that the reported accuracy of the decision aid significantly influenced rule utilization. At 50% decision rule accuracy, the rule was utilized in 48.8% of the time, at 70% accuracy, utilization was 53.4%, and at 90% accuracy, utilization was 55.3%. The rule utilization percentages were significant from one another ($p < 0.05$). These results indicate a positive relationship between decision aid reliance and perceived reliability of the aid; however, as perceived reliability increased, the distance between rule utilization and reported accuracy became wider. These findings suggest that psychological factors other than rationality appear to be dampening decision aid reliance, particularly as decision aid reliability moves to higher levels. Powell (1991) documented a phenomenon that other decision aid researchers had observed prior to his study; that is, decision aid users often under-rely on reliable decision aids.

The extent to which individuals exhibit rational behavior may be partially dependent on the level of motivational pressures to render high quality decisions. We term such motivations as *pressures to perform*. If reliance without any pressure to perform is less than what would be expected under a strictly rational behavior assumption, perhaps the addition of such pressures will lead to more rational behavior, where reliance increases until rationality is eventually reached. Here, a notion of reliance behavior emerges which we refer to as *pressure-induced rational behavior* (PIRB).

Extant decision aid literature provides conflicting evidence related to the effects of various pressures to perform, as increased pressures of this nature have been known to dampen decision aid reliance in some instances and enhance reliance in other instances (Rose 2002). One reason for the equivocal results in this regard is that reliance behaviors associated with pressures

to perform could be partially confounded by a host of other factors (Brown & Jones 1998), including the perceived reliability of the decision aid (a decision aid feature), the expected causality between decision aid use and judgment quality (a decision task factor), and the decision maker's judgment confidence or risk attitude (a decision maker characteristic). Below, we briefly overview several studies where various pressures to perform have been examined.

Arkes et al. (1986) conducted a seminal study wherein they examined decision aid reliance. In their experiment, they examined conditions under which a decision aid was used or ignored under different incentive schemes. Holding the perceived accuracy of the decision aid constant across treatment conditions at 70%, the authors manipulated three incentive levels: none, a dime for each correct answer (piece-rate), or a tournament-style award. Their results indicate that individuals in both the piece-rate and tournament-style incentive groups deviated more from the decision aid's advice (i.e., relied less on the aid) than the no-incentive group. The result in the tournament-style group was expected, as participants in this group would likely deviate from the advice given to all participants in an attempt to outperform the others, thus win the tournament; however, deviation of the piece-rate group was not expected. One explanation for exhibited behavior in the piece-rate group is that their motivation to receive the incentive could have caused them to be less tolerant of probabilistic errors in the task, as reflected by the 30% error rate if they totally relied on the decision aid. The authors also manipulated the consequences of ignoring the decision aid. One of the groups was warned of the negative consequences of ignoring the decision aid while the other was encouraged to try to beat the decision aid. The results show that the manipulation of negative decision consequences increased the level of reliance significantly more than either in a control group or the group that was

encouraged to beat the aid. The authors speculated that negative consequences might have increased the users' tolerance for probabilistic errors.

It is difficult to determine precisely whether decision aid reliance was equal to decision aid reliability (70%) in the Arkes et al (1986) study, as the data were presented in graphical not tabular form. The participants made a total of 40 predictions, and had they followed decision rule consistently, they likely would have made 28 correct predictions (70%). As reported earlier in this paper, the results indicate that the only treatment condition to display rational behavior (mean number of correct answers around 28), was the 'no feedback x no incentive group' control group. In all other conditions, where incentives and consequences were manipulated, the mean number of correct answers fell below 28. These findings suggest that increased pressures to perform reduced reliance behavior. Ashton (1990) and Boatsman et al. (1997) found a similar relationship between increased pressures to perform and decreased decision aid reliance.

On the other hand, Gomaa (2005) and Gomaa et al. (2008) reported a positive relationship between pressures to perform and decision aid reliance. The Gomaa et al. (2008) study involved 118 audit partners who were asked to make an initial judgment, review a decision aid's recommendation related to the judgment, and then revise their judgment. The pressures to perform were operationalized as litigation risk (low or high) and internal control risk (low or high). Decision aid reliability was held constant at 80%. The main effect of litigation risk brought reliance from a mean of 17% (low) to 67% (high), the main effect of internal control risk raised reliance from a mean of 28% (low) to 57% (high), and the interactive effect of high litigation and high internal control risk increased reliance to the highest level of 89%. As indicated, reliance only approached reliability with both risks high, at which point some degree of apparent over-reliance was obtained.

Below, we present a second perspective of decision reliance behavior, PIRB, in which marginal reliance will eventually diminish to zero as pressures to perform continue to increase, up to the point where reliance equals reliability:

H2: The level of reliance on a decision aid will start below the decision aid's level of reliability and will eventually equal the decision aid's reliability level as pressures to perform increase.

Model 3: Pressure arousal theory

Pressure arousal theory (PAT), as suggested by Yerkes & Dodson (1908) and introduced in the accounting literature by Ashton (1990), maintains that pressures to improve decision-making performance can initially yield the intended effect, as a result of increased attention and effort; however, too much pressure can heighten an individual's level of psychological arousal to a point where performance begins to deteriorate. As indicated earlier, Arkes et al. (1986) observed lower reliance on a decision aid and worse performance in the presence, relative to absence, of incentives and decision consequences—findings that are consistent with PAT

Ashton (1990) examined the effects of four pressures to perform on decision-making—incentives, feedback, justification, and the mere presence of a decision aid. He conducted several experiments using professional auditors to predict bond ratings based on financial ratios. In his study, he adopted the pressure-arousal-framework to explain the effects of the pressures generated by incentives, feedback and justification in the decision making process. Using this framework, he suggested that each of these three factors adds pressure on the decision maker that increases attention and effort, and in-turn improves performance, up to a point after which too much pressure will lead to anxiety-induced, worse performance. In his experiments, Ashton used a tournament-style incentive scheme (i.e. based on performance relative to other contestants) and

used a decision aid with an accuracy rating (statistical validity) of 50%. The task was to predict the correct bond rating for 16 different companies. Mean prediction accuracy ranged from 4.64 correct predictions (without the aid, incentives, feedback or justification) to 6.38 correct predictions (with the aid, but without incentives, feedback or justification). When the decision aid's advice was accompanied by incentives, feedback or justification, mean prediction accuracy deteriorated from the high of 6.38, thereby indicating too much pressure arousal. Two additional experiments within the Ashton (1990) study also supported the inverted-U nature of PAT; that is, some pressure can lead to improved reliance and performance, but there is a tipping point over which too much pressure can yield the opposite effects. Consistent with PAT, our third model of decision aid reliance is as follows:

H3: The level of reliance on a decision aid will start below the decision aid's level of reliability, display increases that will initially lead to a maximum point as pressures to perform increase, after which it will display decreases as pressures to perform continue to increase.

Comparing the rationality, PIRB and PAT models of decision aid reliance

If there is an optimum in any of the curves that model the relationship between reliance on a decision aid and the number of pressures to perform for various levels of decision aid reliability, then a question arises regarding how these optimum levels are related. One could argue that at very low levels of decision aid reliability, say 50%, decision makers might rely more on their judgments than the decision aid's advice without much pressure to perform, as the accuracy odds of the relying on the aid would not any better than flipping a coin in the long run. Hence, a relatively low level of pressure to perform might drive the decision maker away from relying on the aid. However, as the decision aid's level of reliance increases, decision-makers

will rely more on the aid and inflection points of excessive pressure arousal (where decision makers decide to rely more on their judgment relative to the aid's advice) will occur at higher levels of pressure. To test the expected interaction between decision aid reliance and pressures to perform, we will model a value function of decision aid reliance. Hopefully, our unique approach at modeling a value function will help to reconcile the three competing hypotheses presented herein. Hence our research question (RQ) is presented as:

RQ: Can decision aid reliance be modeled such that the assumptions of rational behavior, pressure-induced rational behavior and pressure arousal theory are simultaneously met?

III. METHODOLOGY

Design

The experiment involved a five (decision aid reliability: 50%, 60%, 70%, 80% and 90%) by four (number of pressures: one, two, three and four) between-participants design, for a total of 20 treatment conditions. All participants were exposed to the decision aid's recommendation, which represented one pressure. In addition, participants in the two, three or four pressure conditions randomly received one or more of the following pressures: litigation risk, time pressure and justification. The dependent variable reflected the level of reliance on the decision aid.

Procedure

The experiment was computerized and installed on a password-controlled server in an experimental laboratory of a large state university. The study was available for a period of six weeks during which participants could take part on any day at any time from 8 am until 10 pm.

Access to the laboratory was restricted and participants were monitored at all times. Students signed-up for the experiment during classroom hours and were awarded extra credit (5% of their overall course grade) points for participating. In addition, students could earn up to \$10.00 for their participation, depending on the accuracy of their estimate of the allowance for uncollectible accounts.² Only students listed on sign-up sheets who displayed proper identification were issued a password to gain entry to the computerized experiment. Each computer was placed in a separate room to ensure privacy and independence among participants.

The software randomly assigned each participant to one of the 20 treatment conditions at the beginning of the experiment. The experiment incorporated various internal controls, such as validating input, ensuring completeness on one screen before moving to the next screen, preventing changes to earlier responses and randomizing response items where practical. Participant data were automatically recorded into a database during the experiment.

After agreeing to informed consent, the general instructions emphasized the confidentiality of participation and explained the course credit and financial incentive. Afterward, the participants began reading the case materials.

Task

Participants read about a case company, called Amalgamated Manufacturing, Inc. (AMI) that produced and sold high-tech equipment. They were presented with key ratios for the current year and prior year, and a summary balance sheet and income statement for the current year. They were informed that management had estimated the allowance for uncollectible accounts as

² Participants were informed that the quality/accuracy of their judgment would be assessed by an expert who has considerable experience in making such judgments, and that they would be paid by the laboratory at the end of the semester. After all data were collected, a senior audit manager from one of the Big-4 CPA firms reviewed all estimates. She placed the estimates into five categories: excellent, very good, average, below average and poor. Her categories were based on her subjective judgment of the reasonableness of the estimates based on the historical financial information of the case firm. The monetary payments ranged from \$10 (excellent), \$8 (very good), \$6 (average), \$4 (below average) and \$2 (poor). The median payout was \$8 per individual.

of 12/31 at \$600,000, the audit committee of the board of directors agreed with management's estimate and both parties were reluctant to change the estimate. The participants were provided with the past five-year accounts receivable and write-off history. Participants were asked to role play as an auditor for a CPA firm and informed that their task was to independently review the accounts receivable history, arrive at their own estimate of uncollectible accounts, and propose a year-end adjustment (if any) to the \$600,000 estimate provided by management.

The case materials were designed such that a relatively straightforward linear extrapolation of historical write-off activity would lead to an allowance amount of approximately \$700,000. Participants next offered their initial estimate of what they believed the allowance for uncollectible accounts should be as of 12/31, and proposed a year-end adjustment.

Participants next read about a decision aid, provided by their firm, which used historical information as well as contemporary economic conditions to predict the allowance for uncollectible accounts. They were told that the decision aid recommended an allowance of \$800,000 and had given correct estimates of allowances for uncollectible accounts in every X out of 10 of cases (manipulated at 5 out of 10 cases [50% reliability], 6 out of 10 cases [60% reliability], 7 out of 10 cases [70% reliability], 8 out of 10 cases [80% reliability] and 9 out of 10 cases [90% reliability]). After reading this information, participants were afforded an opportunity to change their initial estimate and proposed year-end adjustment, if desired. They were made fully aware that they did not have to change their initial estimate and proposed adjustment or they could change their estimate and adjustment in any direction and magnitude.

Independent Variables ³

Reported Reliability of the Decision Aid

The decision aid manipulation read as follows:

Your firm has provided you with a computerized decision aid to assist you during the audit of AMI. The output of the decision aid is intended to be advisory only, as the ultimate decision regarding the estimate of the allowance for uncollectible accounts is your judgment call.

In addition to analyzing AMI's historical financial records, the decision aid incorporates current economic conditions into its estimate of uncollectible accounts.

The decision aid has been used many times to estimate the allowance for uncollectible accounts for various clients. Experience indicates that it yields an acceptably accurate estimate every x out of 10 times; thus, the decision aid is considered to be X% reliable.

The decision aid has just estimated AMI's allowance for uncollectible accounts as of December 31st, 2004 at \$800,000.

The variable X reflected 50%, 60%, 70%, 80%, or 90% reliability.

Pressures to Perform

All participants were exposed to the decision aid, at varying degrees of reliability, which represented one source of pressure. The remaining three pressures were randomized between participants. For instance, all participants in the one-pressure condition had the decision aid as the source of pressure. If a participant was randomized into the two-pressure condition, the decision aid always represented one pressure, and one of three remaining pressures was randomized for that participant. For a participant in the three-pressure treatment, the decision aid reflected one pressure, and two of the three remaining pressures were randomized for that individual. In the four-pressure condition, all pressures were present. Aside from the decision aid itself, the three other pressures are described next.

³ We used the results of four pilot tests over a 6-month period to refine the case materials, independent variable manipulations and response item wordings. Pilot participants included a total of 16 senior, 11 manager and 7 partner auditors from the Big-4 CPA firms. The participants helped us to word the manipulations so that they would be realistic and effective.

Litigation Risk

Litigation risk was manipulated as follows:

Your legal department has indicated to you that the risk of your CPA firm being sued as a result of this engagement is very high should the financial statements be materially misstated. Specifically, your legal department has evaluated the likelihood of being sued at 95%. If your firm is sued, the likelihood of the plaintiffs winning the suit is 90%. If the suit is successful, the fine and penalties imposed on your firm would be extremely high.

Time Pressure

Participants who were exposed to the time pressure manipulation read that they would have 13 minutes to complete the upcoming task.⁴ They then began reading the case materials. To further reinforce the time pressure manipulation, a count-down clock was placed in the upper-right hand corner of the participants' computer screens. The clock remained active and visible until the participants made their final estimate of the allowance for uncollectible accounts.

Written Justification

Immediately prior to reading the case materials, participants in the justification treatment read the following:

You will be asked to make an estimate of the company's allowance for uncollectible accounts. To help the judge assess the accuracy of your estimate, after you make your estimate, you will be asked to justify in writing how and why you arrived at your estimate.

To reinforce the manipulation, while the participants were making their final estimates, they read:

The judge who will be evaluating the accuracy of your final estimate needs to understand how and why you arrived at your final estimate of the allowance for uncollectible accounts. In the response box below, please justify your final estimate.

⁴ A pilot test revealed that 13 minutes offered sufficient time to complete the task under all pressure conditions, but, participants had to work quickly and efficiently, as they perceived a relatively high level of time pressure.

Dependent Variable

The dependent variable examined in this study is the level of reliance on the decision aid. As with Hayes (2002), the reliance measure determines the extent to which participants change their initial prediction to agree with the decision aid's advice. It is calculated as follows (range from 0.00 to 1.00):

$$RELIANCE = \frac{Final\ Proposed\ Adjustment - Initial\ Proposed\ Adjustment}{(Decision\ Aid's\ Recommendation - Management's\ Estimate) - Initial\ Proposed\ Adjustment}$$

Since we held *Decision Aid's Recommendation – Management's Estimate* constant at \$200,000, this algorithm can be simplified as follows:

$$RELIANCE = \frac{Final\ Proposed\ Adjustment - Initial\ Proposed\ Adjustment}{200,000 - Initial\ Proposed\ Adjustment}$$

The decision aid's recommendation can influence the participants' final estimates of the allowance for uncollectible accounts and their final proposed adjustments. To incorporate the total effect of the decision aid on the participants' ultimate audit judgments, the metric used in the current study incorporates their final proposed adjustment.

IV. RESULTS

Participants

Participants were business students attending a large state university. There were 403 participants in total, comprised of 193 male and 210 female participants. The mean (standard deviation) age was 23.76 (2.54). The breakdown by college major was: 225 accounting, 150

finance, 1 management and 27 economics. Their current college levels were: 130 junior, 195 senior, and 78 masters. All students had successfully passed Accounting Principles I and Finance I, 257 passed Accounting principles II, and 151 passed Financial Auditing. The mean (standard deviation) number of total accounting and finance courses successfully passed were 4.74 (2.90) and 3.27 (1.31), respectively. The mean (standard deviation) total years work experience was 0.33 (0.74) and no students had ever worked as an auditor. A combination of chi-square and regression analyses indicated no significant difference ($p > .15$) across treatment conditions on any of the demographic variables (above); thus, the randomization procedure was deemed successful.

Manipulation Checks

Wording of manipulation check items and statistical results are shown in Appendix A. The first three items indicate that all participants understood the task, felt that the task was moderately difficult ($m = 5.38$), and were moderately motivated to participate in the study ($m = 5.15$).

Item four suggests that all participants who were told that they would be asked to justify their estimates in writing understood the manipulation. Item five asked the participants how concerned they were about justifying their estimates. This question was asked because participants who were not told that they would have to justify in writing nevertheless might have been concerned about some form of *ex post* justification. The results indicate that participants in the justification treatment were relatively concerned ($m = 7.88$), while those who were not asked to justify were relatively unconcerned about having to justify to someone else how and why they arrived at their final estimates ($m = 1.22$).

Item six indicates that participants in the litigation risk treatment recalled being advised of the client's litigation risk profile. Item seven asks about perceived litigation risk. This item was asked because participants in the non-litigation risk treatment might have presumed a relatively high level of risk despite receiving any explicit information of this nature. Based on the results, participants in the litigation risk treatment perceived a relatively high level of risk ($m = 8.14$), while those who were not told about the client's litigation risk profile perceived a relatively low level of risk ($m = 1.66$).

Item eight asked how many minutes the participants had to make their initial and final estimates. Participants in the time pressure treatment correctly indicated 13 minutes and those not in the time pressure condition indicated that there was "no stated time limit". What is important with time pressure is how much pressure the participants "felt" internally, as addressed by item nine. Those who were exposed to the time pressure treatment felt a relatively high level of pressure ($m = 8.46$) while the remaining participants felt relatively little pressure ($m = 1.15$). A related finding emerges from analyzing results of the computer-recorded time it took for participants to complete the initial and final estimates. Participants in the time pressure condition took 11.47 (0.84) minutes and the other participants took 11.59 (1.82) minutes, and the two means are not significantly different ($t = 0.87$, $p = .39$). Interestingly, the mere mention of a time limitation (reinforced by a visible count-down clock) excited "felt" time pressure despite the equivalence of task time use across all participants.

Item ten asked about the reported reliability of the decision aid and participants gave the correct answer based on their treatment conditions. Item eleven asked participants how they felt about the reliability of the decision aid. The responses are ordered from relatively low to relatively high across the increasing levels of reported decision aid reliability, and all means

were significantly different from each other. Items twelve through fourteen suggest that the participants understood management's estimate, the decision aid's recommendation, and the difference between the two.

The next set of four items (fifteen through eighteen) address the extent to which each of the pressures motivated the participants to provide accurate estimates. These are key questions in this study, as the experiment presumes that each of the pressures, including the decision aid's recommendation itself, provided motivation or incentive for the participants to be accurate. The results indicate that participants in the presence of litigation risk, written justification and time pressure were relatively highly motivated to provide accurate estimates, and those who were not exposed to such manipulations reported a relatively low level of such motivation. All participants felt that the decision aid's recommendation provided a relatively high level of motivation ($m = 8.10$) to provide accurate estimates. These findings suggest that the four decision aid reliance pressures activated motivations to be accurate.

One final consideration was examined. We analyzed whether the decision weight of each of the pressures were equivalent. If the weights were not equivalent, then a numeric count of pressures without considering discrepant weights of each pressure could have differential effects on RELIANCE. Using an OLS regression with RELIANCE as the dependent variable and the three manipulated pressures as independent variables, we found no significant differences among the regression coefficients.⁵ Taken as a whole, the set of manipulation check items indicates that the experimental treatments were successful. The next section provides the results of preliminary testing of the study data.

Value function

⁵ The regression coefficients of litigation risk, justification and time pressure were 0.251, 0.272 and 0.194 respectively, with no significant differences (p -values > 0.50).

The extent to which an auditor relies upon the decision aid is determined by two factors: decision aid reliability (DAR) and the number of pressures to perform (PTP). The extent to which an auditor relies upon the decision aid (RELIANCE) can range from 0 (not at all) to 1 (completely).⁶ In our experiment the value for DAR ranged from 0.5 (50%) to 0.9 (90%). The higher reliability, the higher we expect decision aid reliance will be. The number of pressures to perform ranges from 1 to 4. The presence of the decision aid is considered to be a pressure in all experimental conditions. The general relation between reliance, decision aid reliability and pressures to perform is as follows:

$$\text{RELIANCE} = g(\text{DAR}, \text{PTP}) + e$$

where $g(\cdot)$ is a general unknown function and e is the error term. The error term is included to accommodate for differences in individual perceptions and weighting of different decision criteria. For a given number of pressures, DAR is expected to have a positive effect on RELIANCE (Powell 1991). For given levels of DAR, the effect of PTP on RELIANCE is less straightforward (Gomaa 2005; Boatsman et al. 1997; Ashton 1990; Arkes 1986).

Covariates. A preliminary ANCOVA model (independent variables - decision aid reliability, number of pressures and the interaction term: dependent variable - decision aid reliance) included demographic factors (i.e., gender, age, college major, total accounting courses, total finance courses and college level) as potential covariates. The model also included as covariates the 'week' (1 through 6), 'day of week' (Monday through Sunday) and 'time of day' (8 am through 10 pm) during which the participants took part in the experiment. While the main effect and interaction term variables were all significant ($p < .01$), only one of the covariates, age

⁶ Theoretically, an auditor might over-rely, i.e. have a RELIANCE score above 1. However, this did not occur in our experiment.

($F = 6.816$, $p = .009$), was significant. Accordingly, age was included as a covariate in hypotheses testing.

One more potential covariate, risk propensity, was considered. Risk propensity was measured using a standard two-outcome lottery scenario (Young 1985; Hunton et al. 2005). Specifically, participants were asked to respond to the following item (post-experiment):

Given the choice between receiving a guaranteed \$5 or a gamble of receiving \$10, there would have to be at least a _____% chance that I would win the gamble before I would accept the gamble.

The mean (standard deviation) response across all conditions was 63.50% (21.11%). A preliminary ANCOVA model, with decision aid reliance as the dependent variable, decision aid reliability, number of pressures and the interaction term as independent variables, and risk attitude as the covariate, indicated that the covariate was non-significant (p -value = .61). Therefore, it was not included in the upcoming statistical analyses.

Assumptions. The Shapiro-Wilkes test for normality was significant ($p < .01$), indicating that the dependent variable (decision aid reliance) was not normally distributed across treatment conditions. Inspection of frequency distributions suggested that the non-normal distributions were due to skewness. ANOVA is robust to violations of the normality assumption where skewness is the underlying cause; thus the data were not transformed.

Levene's test was also significant ($< .01$), indicating unequal variance of the response variable (decision aid reliance) across treatment conditions. Data transformation was not performed, as ANOVA is robust to moderate departures from homogeneity of variance if sample sizes within each condition are fairly equal, as in the current study.

As a result of the assumption testing, ANOVA is used to test the hypotheses. In addition, inferences are made based on relative, not absolute, differences among treatment conditions, thereby treating the data as ordinal rather than interval.

Hypotheses testing

Figure 1 depicts the reliance means at each level of DAR and PTP. There appears to be a tendency toward inverted U-shaped curves at reliability levels of 50-70%, an asymptotic curve at a reliability level of 80%, and a flat curve at a reliability level of 90%. Table 1 presents the results of the ANCOVA model with AGE as a covariate.⁷ As indicated, the main effects for decision aid reliability and number of performance pressures are significant ($p < .01$), as is the interaction term ($p < .01$).

[Insert Table 1 and Figure 1 about here]

Table 2 presents the reliance means and tests of differences between subsequent reliance levels from one through four PTP, across the different DAR levels. The p-values comparing mean reliance to the DAR level are reported in the rows. The p-values to test for the PTP increases are reported in the columns between the means.

[Insert Table 2 about here]

Hypothesis H1 posits a rational model of decision aid reliance, such that RELIANCE should equal DAR, regardless of the number of PTP. Results indicate that reliance does not equal reliability for any number of pressures to perform on the decision aid (all p-values < 0.05). At DAR levels of 50% through 80%, RELIANCE is significantly below DAR, indicating under-reliance. Interestingly, at 90% reliability, RELIANCE exceeds DAR, suggesting a over-reliance. Hence, H1 must be rejected, which suggests that the decision aid users in this study did not display rational behavior within or across the ranges of PTP and DAR examined.

⁷ An ANCOVA was also performed using ranked data, where the decision aid reliance metric was ranked within each treatment. Additionally, the Kruskal-Wallis test was run to determine if one or more of the treatment means are significantly different from the others, and Mann-Whitney tests were conducted to compare each pair of means. The significance and direction of all test results were substantially the same, in both direction and significance, as the ANCOVA results shown in Table 1.

Hypothesis H2 posits a pressure-induced model of rational behavior, where RELIANCE begins below DAR, but after enough pressure is exerted, RELIANCE will rise to the level of DAR and sustain thereafter. Indeed, at the 80% reliability level, this pattern seems to emerge, as RELIANCE rises from one through three PTP, and approaches (but does not equal) DAR at four pressures. Hence, H2 must be rejected because within and across the range of PTP and DAR we tested in our experiment, the PIRB pattern does not obtain.

Hypothesis H3 posits that RELIANCE will resemble the inverted-U shape predicted by pressure-arousal theory. At DAR of 50%, reliance peaks for two pressures at 0.36 and then drops. At DAR of 60%, reliance peaks for two and three pressures at 0.46, and then decreases. At DAR of 70%, reliance peaks for three pressures at 0.65, and then declines. However, at DAR of 80% there is no indication of a peak, nor is there at DAR of 90%. Hence, H3 must also be rejected across the range of PTP and DAR tested in this experiment.

General Model of the Reliance Value function

Next we try to model reliance behavior from the data obtained in this study. The model to test for rational behavior is:

$$\text{RELIANCE} = a(\text{DAR}) + e$$

Where $a(\cdot)$ is an unknown positive increasing function and the rationality hypothesis is: $a(\text{DAR}) = \text{DAR}$. Estimates are in the first column of Table 3.

The model to test for pressure-induced rational behavior is:

$$\text{RELIANCE} = a(\text{DAR}) - b(\text{DAR})/\text{PTP} + e$$

Where $a(\cdot)$ and $b(\cdot)$ are unknown positive functions of decision aid reliability. Estimates are in the second column of Table 3.

The concave relation between PTP and RELIANCE, as suggested by pressure-arousal theory, can be tested using the following quadratic model:

$$\text{RELIANCE} = a(\text{DAR}) + c(\text{DAR}) * \text{PTP} + d(\text{DAR}) * \text{PTP}^2 + e$$

Where $a(\cdot)$, $c(\cdot)$ and $d(\cdot)$ are unknown functions of decision aid reliability. Pressure arousal theory predicts that $c(\cdot)$ is positive while $d(\cdot)$ is negative. The functions $a(\cdot)$, $c(\cdot)$ and $d(\cdot)$ must fulfil some conditions. A first condition is that for DAR equal to 0 there should also be a predicted reliance equal to zero, independent of the number of pressures. That is, if the decision aid is completely unreliable, then an auditor should not use the information given by the decision aid, irrespective of the value of PTP. Hence, $a(0) = c(0) = d(0) = 0$. A second condition is that for DAR equal to 1 there should be complete reliance, again independent of the number of pressures. A predicted RELIANCE of one, for each value of PTP, is achieved through $a(1) = 1$ and $c(1) = d(1) = 0$. One approach to handle the unknown functions $a(\cdot)$, $c(\cdot)$ and $d(\cdot)$ is to run separate regressions for each value of DAR. In this way we estimate no functions, but only specific values: $a(0.5)$, $c(0.5)$ and $d(0.5)$ when DAR is 50%, $a(0.6)$, $c(0.6)$ and $d(0.6)$ when DAR is 60%, etc. Since we have five different values for DAR we then get 15 parameters to be estimated. These are reported in the third column of Table 3.

[Insert Table 3 about here]

A general description

A disadvantage of the approach to estimate the quadratic relation for each different level of DAR is that we fail to obtain a general description of the relationship between reliance, reliability and pressures to perform. An approach that may reconcile the three mentioned theories

is to choose flexible functions for $a(\cdot)$, $c(\cdot)$ and $d(\cdot)$ defined on the $[0,1]$ -interval. An adequate choice is then to use the Beta distribution. This function is defined for the $[0,1]$ -interval and can take a wide range of shapes. The function has two parameters that determine the shape (just like the normal distribution). For $a(\cdot)$, a function that should rise from 0 to 1, we use the cumulative Beta distribution function. For $c(\cdot)$ and $d(\cdot)$ we choose a multiple of the Beta density function. For appropriate parameter values (both shape parameters should be in excess of one) we then have that $c(0) = d(0) = c(1) = d(1) = 0$, as required. The resulting model is:

$$\text{RELIANCE} = F(\text{DAR}, a_1, a_2) + c_0 * f(\text{DAR}, c_1, c_2) * \text{PTP} + d_0 * f(\text{DAR}, d_1, d_2) * \text{PTP}^2 + e$$

where $F(\cdot)$ is the cumulative Beta distribution function with the two shape parameters a_1 and a_2 , and $f(\cdot)$ is the Beta density function with shape parameters c_1 and c_2 , and d_1 and d_2 , respectively. The parameters c_0 and d_0 scale the density function and may be positive or negative. In total there are eight parameters to be estimated. The function $F(\cdot)$ has a minimum of 0 when DAR is 0% and a maximum of 1 when DAR is 100%. The function $f(\cdot)$ is always positive and can have shapes ranging from constant (both parameters equal to one), linear (one parameter equal to two, the other equal to one), quadratic (both parameters equal to two), U-shaped (both parameters less than one), etc. The function is bell-shaped when both shape parameters are in excess of one. This may be symmetric (just like the normal distribution) when the two parameters are equal or asymmetric when one shape parameter is larger than the other. See the Appendix B for more details. Table 3 shows the results for the models that test for rationality (5 parameters), pressure-induced rational behavior (10 parameters) and pressure arousal (15 parameters).

The models were estimated using ordinary least square. The results in the first column show that rationality is not supported, R^2 is relatively low at 0.91 and for no value of DAR the hypothesis of $a(\text{DAR})$ unequal to DAR can be rejected. The results in the second column provide some support for PIRB, especially at intermediate values of reliability. The test of the hypothesis of $a(0.7)$ unequal to 0.7 returns a p-value of 0.13. The test of the hypothesis of $a(0.8)$ unequal to 0.8 gives a p-value of 0.02. In both these cases there is a positive value for $b(\cdot)$ meaning that there is convergence towards rational behavior when pressure increases. However, the PIRB model is not a good description for DAR equal to 0.5, 0.6 or 0.9. The PIRB model provides a much better description of the data than the simple rational behavior model ($R^2 = 0.94$). The R^2 of the PAT model is relatively high at 0.98. The results clearly show that there is a concave relationship for DAR equal to 50%, 60%, 70% and 80%, and that $c(\text{DAR})$ is significantly positive, while $d(\text{DAR})$ is significantly negative. For a very high reliability (DAR equal to 90%) there is no significant concave relationship: reliance is high and constant across different levels of pressure. For such high levels of reliability the participants appear to trust the decision aid almost completely, to the extent that they even over-rely on the aid.

The results show that a quadratic relationship between reliance and the number of pressures has a superior fit. Figure 2 depicts this model for each of the five DAR levels. As can be seen, for DAR=50% to DAR=80% the curves' peaks move to the right as pressures are added for values of PTP and RELIANCE of (2.176, 0.296); (2.474, 0.474); (3.057, 0.620) and (4.222, 0.778) respectively. This pattern is disrupted for DAR=90% since this curve peaks for values of PTP and RELIANCE of (2.333, 0.952). The significant over-reliance on the decision aid when DAR=90% at any level of PTP explains this phenomenon, yet the mere fact that there is over-reliance when DAR approaches 100% needs further explication.

[Insert Figure 2 about here]

Table 4 shows the estimation results for the model with the Beta distribution assumed for the unknown functions. It was estimated using non-linear least squares. There are eight parameters estimated, which are all strongly significant. The R^2 of the model is 0.978, slightly less than for the previous model. The log-likelihood is 712.250. There are seven parameters less in this approach and three different tests suggest that the null hypothesis of using the eight parameter model versus the 15 parameter model cannot be rejected. The first test is the likelihood ratio test. Twice the difference between the log-likelihoods is 12.99. The 5% critical value from the chi-square table is 14.07. This indicates that the null hypothesis of reducing the model with seven parameters is not rejected (the procedure suggested by Vuong (1989, p.318) on non-nested models conforms a p-value > 0.05). The second test is the Akaike information criterion, which is lower for the eight parameter model (-3.521 versus -3.519). The third test is the Schwarz information criterion, which is also lower for the eight parameter model (-3.441 versus -3.369).⁸

[Insert Table 4 about here]

The results presented in Table 4 suggest clear bell-shaped Beta distributions for $c(\cdot)$ and $d(\cdot)$: hence, the shape parameters are expected to be both above one. The estimate for c_0 is

⁸ The Akaike information criterion (AIC) equals $-2LL/n + 2k/n$ where k is the number of estimated parameters, n is the number of observations and LL is the value of the log-likelihood function. The Schwarz information criterion (SIC) equals $-2LL/n + k\log(n)/n$. Hence, in both model selection criteria there is a penalty imposed for using more parameters.

positive and for d_0 is negative, as expected. Hence, pressure arousal theory finds support: initially pressure leads to more reliance, and when pressure further increases reliance starts to decrease. The number of pressures at which reliance starts to decrease may however be at levels higher than investigated in the current study. The shape parameters for the three Beta distributions are all clearly in excess of one.⁹ It is interesting to see when the Beta distributions for the linear and quadratic terms reach their maximum, i.e. when the distribution attains its mode.¹⁰ For the $c(\cdot)$ -function this is for DAR equal to $7.244/(7.244+4.430) = 0.62$. For the $d(\cdot)$ -function this is for DAR equal to $7.593/(7.593+5.404) = 0.58$. That is, for values of DAR around 60% the impact of the pressures is strongest. The mode for the Beta distribution related to the $a(\cdot)$ -function is at $18.518/(18.518+3.978) = 0.82$. That is, around 80% reliability, there is the strongest increase of RELIANCE when reliability goes up. In fact, the constant increases from just 0.12 when DAR is 70% to 0.92 when DAR is 90%.

V. DISCUSSION

The current study examines the joint impact of decision aid reliance and pressures to perform on decision aid reliance. After comparing and contrasting three potential decision aid models, rationality, pressure-induced rationality and pressure arousal theory, we find that a quadratic model best explains reliance behavior, which is consistent with pressure arousal theory.

We contribute to extant decision aid literature in several ways. First, we help to explain inconsistent results across decision aid studies by demonstrating that various reliability levels and pressures to perform operant in prior studies make it difficult to compare research findings,

⁹ The parameter estimates for c_1 and d_1 , and for c_2 and d_2 seem quite close to one another. However, a model restricting $c_1 = c_2$ and $d_1 = d_2$ has an R^2 of only 0.9454 and must be rejected using all three tests mentioned.

¹⁰ The mode of a variate distributed as $\beta(\alpha, \beta)$ is $(\alpha-1) / (\alpha+\beta-2)$.

and can interact with other factors (such as decision tasks and decision maker characteristics) to produce seemingly contradictory results. Second, we find that pressure arousal theory predominantly explains decision aid reliance behavior at all but very high levels of decision aid reliability. Finally, we develop a generalized economic model of reliance behavior that can be used in future research endeavors to predict reliance outcomes based on the level of reliability and pressure used in a particular study, and better hypothesize the expected effects of other decision aid reliance factors under examination.

Our findings contribute to pressure-arousal theory, indicating that as a decision aid's reliability increases, the number of pressures to perform will become less relevant to the choice whether or not to rely on the decision aid. Our findings also hold implications for organizations that deploy decision aids to increase decision quality; that is, the mere existence of inflection points in the reliance curves indicate that there are reliance peaks, and that more effort to make decision makers rely on the decision aid may not lead to increased, but rather to decreased, reliance.

However, as in any experiment there are inherent limitations to our study that must be further investigated. We selected litigation risk, time pressure, and required written justification as pressures to rely on the decision aid. Another choice was that we investigated decision makers' behavior within the decision aid's reliability range of 50% to 90%. Hence, a strand for further research may be to add other pressures to rely and also investigate decision aids with less than 50% reliability. An interesting avenue for further research may be to investigate the underlying explanations of the over-reliance that we observed at reliability levels of 90%. In addition to adding more pressures to rely, delving deeper into the chosen pressures may reveal that each of these pressures has a different weight, hence causes the number of pressures to be a

variable that does not accurately measure the real pressure that is exerted on the decision maker because it assumes equal weights. Further, we picked the estimation of uncollectible accounts as the task to perform. This is a task with a high performance standard and may be considered a complex task. In addition, we did not measure the respondents' cognitive styles. Since, decision making requires cognitive effort, variations in reliance behavior may occur that can be explained by differences in cognitive styles. Manipulating task complexity and measuring and manipulating participants' cognitive styles may lead to new insights into the factors that explain the use of decision aids. Finally, we did not investigate the effects of financial pressures to rely on decision aids. There is an extensive body of literature in the area of performance-based rewards as incentives to enhance performance. Combining this study with insights from extant decision aid literature may reveal interesting new insights.

REFERENCES

- Arkes, H. R., R. M. Dawes, and C. Christensen (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processing*, 37: 93-110
- Arnold, V., S. Sutton, S. Hayne, and C. Smith (2000). Group Decision Making: The Impact of Opportunity-Cost Time Pressure and Group Support Systems. *Behavioral Research in Accounting*, 12: 70-96
- Ashton, R.H. (1974). An Experimental Study of Internal Control Judgments. *Journal of Accounting Research*, Spring
- Ashton, R.H. (1990). Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback and Justification. *Journal of Accounting Research*, 28, Supplement
- Ashton, R. H. (1992). Effects of Justification and a Mechanical Aid on Judgment Performance. *Organizational Behavior and Human Decision Processing*, 52: 292-306
- Bedard, J.C., & Lynford, G.L. (2002). The Effects of Decision Aid Orientation on Risk Factor Identification and Audit Test Planning. *Auditing: a Journal of Practice and Theory*, 21(2), 39-57
- Benbasat, I., B. R. Nault (1990). An Evaluation of Empirical Research in Managerial Support Systems. *Decision Support Systems*, 6(3):203-226
- Boatsman, J.R., C. Moeckel, and B.K.W. Pei. (1997). The Effects of Decision Consequences on Auditor's Reliance on Decision Aids in Audit Planning. *Organizational Behavior and Human Decision Processing*, 71: 211-247
- Brown, R.G. (1962). Objective Internal Control Evaluation. *Journal of Accountancy*, November
- Brown, D.L., and D.R. Jones (1998). Factors that Influence Reliance on Decision Aids: a Model and an Experiment. *Journal of Information Systems*, 12(2), 75-94
- Chu, P.C., and E.E. Spire. (2003). Does Time Pressure on Users Negate the Efficacy of Decision Support Systems? Working Paper
- Dungan, C.W. (1983). *A Model of an Audit Judgment in the Form of an Expert System*. Dissertation. University of Illinois
- Gomaa, M.I.I. (2005). *The Effect of Time Pressure, Task Complexity and Litigation Risk on Auditors' Reliance on Decision Aids*. Dissertation, Maastricht University.
- Gomaa, M.I.I., J. E. Hunton and J. Rose. (2008) The Effects of Control Risk and Litigation Risk Decision Aid Reliance. *International Journal of Applied Decision Sciences*, forthcoming.
- Hayes, D. (2002). *Factors affecting reliance on a decision aid's advice: the effects of the group decision-making process on decision aid reliance*. Ph.D. Dissertation
- Heiman, V. (1990). Discussion of Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback and Justification. *Journal of Accounting Research*, 28: 181-186.
- Joyce, E.J. (1976). Expert Judgment in Audit Program Planning. *Journal of Accounting*

- Research*, Supplement, 14(3): 29-60.
- Kachelmeier, S.J., and Messier, W.F. (1990). An Investigation of the Influence of a Nonstatistical Decision Aid on Auditor Sample Size Decisions. *The Accounting Review*, 65(1)
- 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
- Kinney, W.R. and W.C. Uecker (1982). Mitigating the Consequences of Anchoring in Audit Judgments. *The Accounting Review*, January
- Kleinmuntz, B. (1990). Why We Still Use our Heads Instead of Formulas: Toward an Integrative Approach. *Psychological Bulletin*. 107 (3). 296-310
- Kumar, L. (1997). DSS Values and Time-Constrained Decision Making. *Proceedings of the Americas Conference on Information Systems*, Association for Information Systems
- Lewis, B.L. (1980). A Decision Theory Approach to the Sampling Problem in Auditing. *Journal of Accounting Research*, Autumn
- McDaniel, L., W. Kinney (1995). Expectation Formation Guidance in the Auditor's Review of Interim Financial Information. *Journal of Accounting Research*, 33: 59-76
- Messier, W. (1995). Research and Development of Audit Decision Aids. *Judgment and Decision Making Research in Accounting and Auditing* (pp. 207-228). Cambridge University Press
- Mock, Th.J., and J.L. Turner (1981). Internal Accounting Control Evaluation and Auditor Judgment. *Auditing Monograph*, 3, New York, NY: AICPA
- Moriarty, S. and F.H. Barron (1976). Modeling the Materiality Judgments of Audit Partners. *Journal of Accounting Research*, Autumn.
- Powell, J.L. (1991). An Attempt at Increasing Decision Rule Use in a Judgment Task. *Organizational Behavior and Human Decision Processes*, 48: 89-99
- Rose, J. (2002). Behavioral Decision Aid Research: Decision Aid Use and Effects. *Researching Accounting as an Information Systems Discipline* (pp. 111-133). American Accounting Association, Sarasota, Florida
- Smith, C., S. Hayne, V. Arnold, and S. Sutton (1998). An investigation of the effect of computerized support systems on managerial group decision making under time pressure. *Advances in Management Accounting*, 6: 17-37
- Steinbart, P.J. (1987) The Construction of a Rule-Based Expert System as a Method for Studying Materiality Judgments. *The Accounting Review*, 13(1)
- Sutton, S., R. Young, and P. McKenzie (1995). An Analysis of Potential Legal Liability Incurred Through Audit Expert Systems. *International Journal of Intelligent Systems in Accounting, Finance, & Management*. 4 (September): 191-204.
- Vasarhelyi, M., & Halper, F.B. (1991). The Continuous Audit of Online Systems. *Auditing; a Journal of Practice and Theory*, 10(1)
- Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses.

Econometrica, 57 (March): 307-333

Ward, B.H. (1976). An Investigation of the Materiality Construct in Auditing. *Journal of Accounting Research*, Spring

Yerkes, R.M. and J.D. Dodson (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18: 459-482

APPENDIX A: Manipulation Check Items

- 1) *Did you understand the task? (Yes, No)*
 - a) All respondents answered 'Yes'
- 2) *How difficult was the task? (1 = not very difficult, 4 = moderately difficult, 9 = very difficult)*
 - a) Overall Mean (standard deviation): 5.38 (1.26)
 - b) Main effects and interaction term were non-significant ($p > .90$)
- 3) *How motivated are you to participate in this study? (1 = not very motivated, 4 = moderately motivated, 9 = highly motivated)*
 - a) Overall Mean (standard deviation): 5.15 (1.30)
 - b) Main effects and interaction term were non-significant ($p > .90$)
- 4) *Were you asked to justify in writing how you arrived at your final estimate for the allowance for uncollectible accounts?*
 - a) All participants in the justification treatment recorded "Yes"
 - b) All participants not in the justification treatment indicated "No"
- 5) *While you were making your initial and final estimates of uncollectible accounts, how concerned were you that you would have to justify to someone else how and why you arrived at your final estimate? (1 = not concerned, 4 = moderately concerned, 9 = very concerned)*
 - a) Mean (standard deviation) in the justification treatment: 7.88 (1.15)
 - b) Mean (standard deviation) in the no justification treatment: 1.22 (0.47)
 - c) t-statistic (p-value): 77.49 ($<.01$)
- 6) *Did the CPA firm advise you of the level of litigation risk of this client?*
 - a) All participants in the litigation risk treatment recorded "Yes"
 - b) All participants not in the litigation risk treatment indicated "No"
- 7) *I felt that the risk of being sued by this client was: (1 = low risk, 4 = moderate risk, 9 = high risk)*
 - a) Mean (standard deviation) in the litigation treatment: 8.14 (0.94)
 - b) Mean (standard deviation) in the no litigation treatment: 1.66 (0.68)
 - c) t-statistic (p-value): 79.91 ($<.01$)

- 8) *How many minutes were you allowed to arrive at your initial and final estimates?*
- All participants in the time pressure treatment recorded “13 Minutes”
 - All participants not in the time pressure treatment indicated “There was no time limit”
- 9) *How do you feel about the time you had to make your initial and final estimates: (1 = I felt a low level of time pressure, 4 = I felt a moderate level of time pressure, 9 = I felt a high level of time pressure)*
- Mean (standard deviation) in the time pressure treatment: 8.46 (0.86)
 - Mean (standard deviation) in the no time pressure treatment: 1.15 (0.67)
 - t-statistic (p-value): 109.55 (<.01)
- 10) *What was the reported reliability of the decision aid?*
- All participants indicated the DA reliability of the treatment to which they were randomized (50%, 60%, 70%, 80% or 90%)
- 11) *I felt that the decision aid was: (1 = very unreliable, 4 = moderately reliable, 9 = very reliable)*
- DA Reliability main effect: F-ratio (p-value): 449.22 (<.013)
 - Reliance Pressures main effect: F-ratio (p-value): 0.02 (= .99)
 - Interaction term: F-ratio (p-value): 0.01 (= .99)
 - Mean (standard deviation):
 - 50% DA Reliability 1.15 (0.72)
 - 60% DA Reliability 2.76 (0.85)
 - 70% DA Reliability 4.26 (1.33)
 - 80% DA Reliability 7.25 (1.71)
 - 90% DA Reliability 8.20 (1.02)
 - Based on Duncan’s multiple pairwise testing ($\alpha = .05$), all means are significantly different from each other.
- 12) *What was the dollar amount of the allowance for uncollectible accounts initially estimated by AMI’s management?*
- All participants correctly indicated \$600,000
- 13) *What was the dollar amount of the allowance for uncollectible accounts indicated by the decision aid?*
- All participants correctly indicated \$800,000
- 14) *What was the difference between management’s estimate of the allowance for uncollectible accounts and the amount indicated by the decision aid?*
- All participants correctly indicated \$200,000

- 15) *To what extent did your perception of the client's litigation risk motivate you to ensure that your final estimate for the allowance for uncollectible accounts was accurate? (1=No Extent, 4=Moderate Extent, 9=Great Extent)*
- a) Mean (standard deviation) in the litigation risk treatment: 7.97 (1.10)
 - b) Mean (standard deviation) in the no litigation risk treatment: 1.25 (0.48)
 - c) t-statistic (p-value): 80.41 (<.01)
- 16) *To what extent did your concern over having to justify your final estimate motivate you to ensure that your final estimate for the allowance for uncollectible accounts was accurate? (1=No Extent, 4=Moderate Extent, 9=Great Extent)*
- a) Mean (standard deviation) in the justification treatment: 7.96 (1.09)
 - b) Mean (standard deviation) in the no justification treatment: 1.31 (0.51)
 - c) t-statistic (p-value): 76.65 (<.01)
- 17) *To what extent did your perception of time pressure to complete screens 1 and 2 motivate you to ensure that your final estimate for the allowance for uncollectible accounts was accurate? (1=No Extent, 4=Moderate Extent, 9=Great Extent)*
- a) Mean (standard deviation) in the time pressure treatment: 8.03 (1.06)
 - b) Mean (standard deviation) in the no time pressure treatment: 1.17 (1.12)
 - c) t-statistic (p-value): 85.00 (<.01)
- 18) *To what extent did the decision aid's recommendation motivate you to ensure that your final estimate for the allowance for uncollectible accounts was accurate? (1=No Extent, 4=Moderate Extent, 9=Great Extent)*
- a) Overall mean (standard deviation): 8.10 (1.10)
 - b) Main effects and interaction term were non-significant ($p > .82$)

Appendix B: Beta distribution

The Beta density function for a variable x and shape parameters $\alpha > 0$ and $\beta > 0$ is as follows:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad 0 \leq x \leq 1$$

where $\Gamma(\cdot)$ is the Gamma function, where for integers: $\Gamma(k) = (k-1)!$ The mean value for x is equal to $\alpha / (\alpha + \beta)$. When α and β are both larger than one, the density function has a maximum. The point at which this is reached is the mode and this equals $(\alpha - 1) / (\alpha + \beta - 2)$.

For $\alpha = \beta = 1$ we get a uniform density function: $f(x; 1, 1) = 1$. For $\alpha = 2$ and $\beta = 1$ we get an increasing line: $f(x; 2, 1) = 2x$. For $\alpha = 1$ and $\beta = 2$ the function is a declining line: $f(x; 1, 2) = 2 - 2x$. For $\alpha = \beta = 2$ there is a quadratic function: $f(x; 2, 2) = 6x - 6x^2$. In general, when both α and β are below 1, there is a U-shape for $f(\cdot)$ and when they are both above 1, $f(\cdot)$ is bell-shaped with $f(0; \alpha, \beta) = f(1; \alpha, \beta) = 0$. The cumulative Beta distribution function is:

$$F(x; \alpha, \beta) = \int_0^x f(z; \alpha, \beta) dz$$

Which has no general closed form. Table B-1 gives values for the cumulative distribution function $F(\cdot)$ and the density function $f(\cdot)$ for the shape parameters as estimated in this paper. The values for the three modes of the Beta distributions (0.8232, 0.6205 and 0.5842) are incorporated next to steps of 0.05 from a DAR of 40% to complete reliability (100%). It allows for easy calculation of the predicted reliance with the general model:

$$\text{RELIANCE} = F(\text{DAR}, a_1, a_2) + c_0 * f(\text{DAR}, c_1, c_2) * \text{PTP} + d_0 * f(\text{DAR}, d_1, d_2) * \text{PTP}^2 + e$$

For example, when DAR is 75% and the number of pressures (PTP) is 3, using the regression results as depicted in Table 4, the expected reliance on the decision aid is $0.261 + 0.127 * 1.85 * 3 - 0.0249 * 1.32 * 9 = 0.670$ or 67%.

Table B-1:

x	$F(x; 19.518, 4.978)$	$f(x; 8.244, 5.430)$	$f(x; 8.593, 6.404)$
0.4000	0.0000	0.9400	1.2600
0.4500	0.0000	1.5100	1.9200
0.5000	0.0010	2.1200	2.5600
0.5500	0.0040	2.6500	2.9900
0.5842	0.0110	2.8900	3.0800

0.6000	0.0160	2.9500	3.0600
0.6205	0.0250	2.9800	2.9700
0.6500	0.0470	2.9100	2.7300
0.7000	0.1210	2.5200	2.0900
0.7500	0.2610	1.8500	1.3200
0.8000	0.4760	1.1000	0.6400
0.8232	0.5920	0.7800	0.4100
0.8500	0.7250	0.4800	0.2200
0.9000	0.9190	0.1200	0.0400
0.9500	0.9940	0.0100	0.0000
1.0000	1.0000	0.0000	0.0000

Figure 1: Effects of Pressures-to-Perform (PTP) & Decision Aid Reliability (DAR) on Decision Aid Reliance

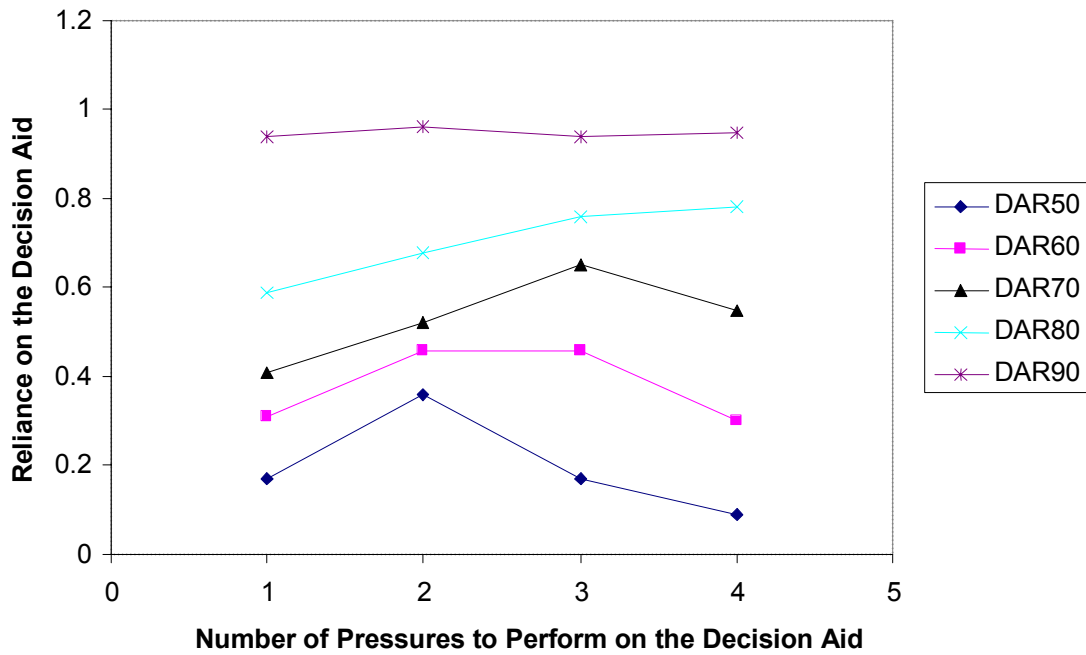


Figure 2: General Model of Decision Aid Reliance—Consistent with Pressure Arousal Theory

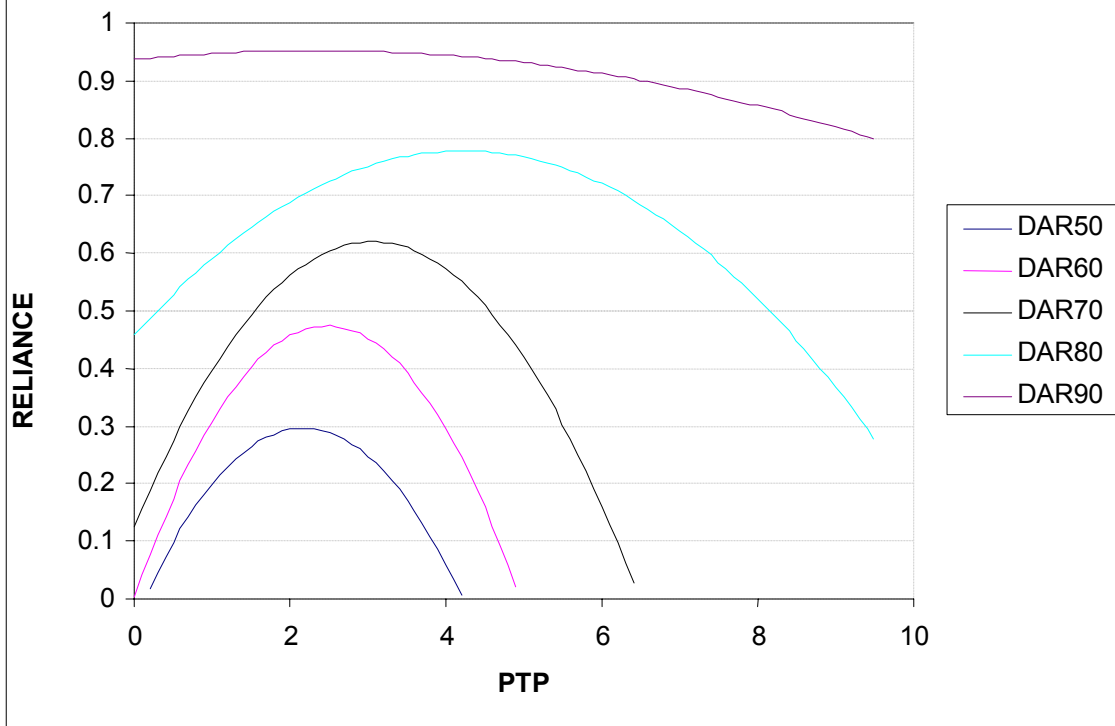


Table 1: Overall ANCOVA Model: Dependent Variable = RELIANCE¹¹

Source	Sum of Squares	df	Mean Square	F	Sig.
AGE	0.007	1	0.007	6.816	.009
Decision Aid Reliability (DAR)	27.305	4	6.826	6,630.954	.000
Pressures to Rely (PTR)	0.855	3	0.285	276.990	.000
DAR x PTR	1.473	12	0.123	119.267	.000
Error	0.39	379	0.001		
Total	153.368	400			

¹¹ The number of observations is 400. Three observations were considered outliers.

Table 2: Test of hypotheses: Mean RELIANCE (Standard Deviation) [Sample Size] and p-values^{12, 13}

Decision Aid Reliability (DAR)	Number of Pressures to Perform on the Decision Aid (PTP)						
	One	p-value	Two	p-value	Three	p-value	Four
50%	0.17 (0.01) [20]	0.000	0.36 (0.02) [21]	0.000	0.17 (0.02) [19]	0.000	0.09 (0.05) [22]
p-value	0.000		0.000		0.000		0.000
60%	0.31 (0.03) [19]	0.000	0.46 (0.03) [20]	0.992	0.46 (0.04) [21]	0.000	0.30 (0.04) [19]
p-value	0.000		0.000		0.000		0.000
70%	0.41 (0.04) [18]	0.000	0.52 (0.04) [19]	0.000	0.65 (0.03) [21]	0.000	0.55 (0.03) [17]
p-value	0.000		0.000		0.000		0.000
80%	0.59 (0.02) [21]	0.000	0.68 (0.02) [22]	0.000	0.76 (0.05) [19]	0.436	0.78 (0.04) [22]
p-value	0.000		0.000		0.004		0.030
90%	0.94 (0.03) [21]	0.199	0.96 (0.03) [20]	0.184	0.94 (0.03) [19]	0.816	0.95 (0.03) [20]
p-value	0.000		0.000		0.000		0.000

¹² P-values are based on Scheffé's multiple pairwise test.

¹³ The RELIANCE metric is calculated as follows: (Final Recommended Adjustment – Initial Recommended Adjustment) / ((Decision Aid's Recommendation – Management's Estimate) – Initial Recommended Adjustment)

Table 3: Estimation results for five levels of DAR

Parameter	H1 Rationality		H2 PIRB		H3 PAT	
	Coeff.	Std.Er.	Coeff.	Std.Er.	Coeff.	Std.Er.
a(0.5)	0.195	(0.009)	0.158	(0.015)	-0.040	(0.025)
a(0.6)	0.383	(0.009)	0.428	(0.016)	0.003	(0.026)
a(0.7)	0.538	(0.010)	0.675	(0.016)	0.125	(0.026)
a(0.8)	0.705	(0.009)	0.835	(0.015)	0.457	(0.025)
a(0.9)	0.951	(0.009)	0.955	(0.016)	0.936	(0.025)
b(0.5)			-0.072	(0.026)		
b(0.6)			0.087	(0.027)		
b(0.7)			0.265	(0.027)		
b(0.8)			0.249	(0.025)		
b(0.9)			0.007	(0.026)		
c(0.5)					0.309	(0.023)
c(0.6)					0.381	(0.023)
c(0.7)					0.324	(0.024)
c(0.8)					0.152	(0.023)
c(0.9)					0.014	(0.023)
d(0.5)					-0.071	(0.005)
d(0.6)					-0.077	(0.005)
d(0.7)					-0.053	(0.005)
d(0.8)					-0.018	(0.004)
d(0.9)					-0.003	(0.005)
R ² adj.	0.908		0.940		0.979	
Log Likelih.	427.640		512.934		718.746	
Akaike info	-2.113		-2.515		-3.519	
Schwarz info	-2.063		-2.415		-3.369	

Table 4: Regression results for the general reliance model using Beta distributions

Parameter	Constant	Parameter	Linear	Parameter	Quadratic
		c_0	0.127 (0.004)	d_0	-0.025 (0.001)
a_1	19.518 (1.056)	c_1	8.244 (0.481)	d_1	8.593 (0.741)
a_2	4.978 (0.260)	c_2	5.430 (0.316)	d_2	6.404 (0.510)

Note: standard errors between brackets.